Clustering

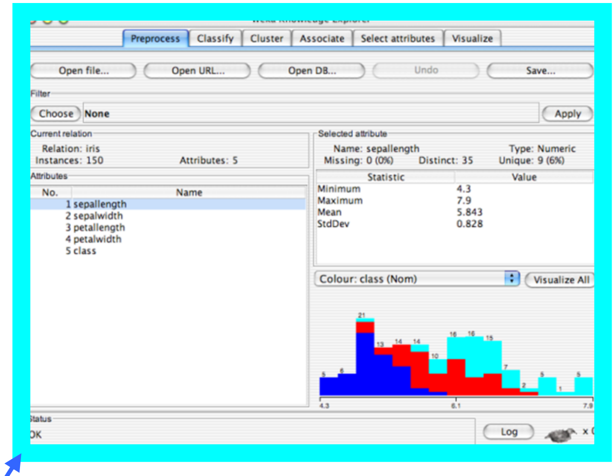
Explorer: clustering data

- 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)

Waikato Environment for
Knowledge Analysis

Version 3.4.5

(c) 1999 - 2005
University of Waikato
New Zealand



Filter
None

Current relation
Relation: iris
Instances: 150
Attributes: 5

Attributes

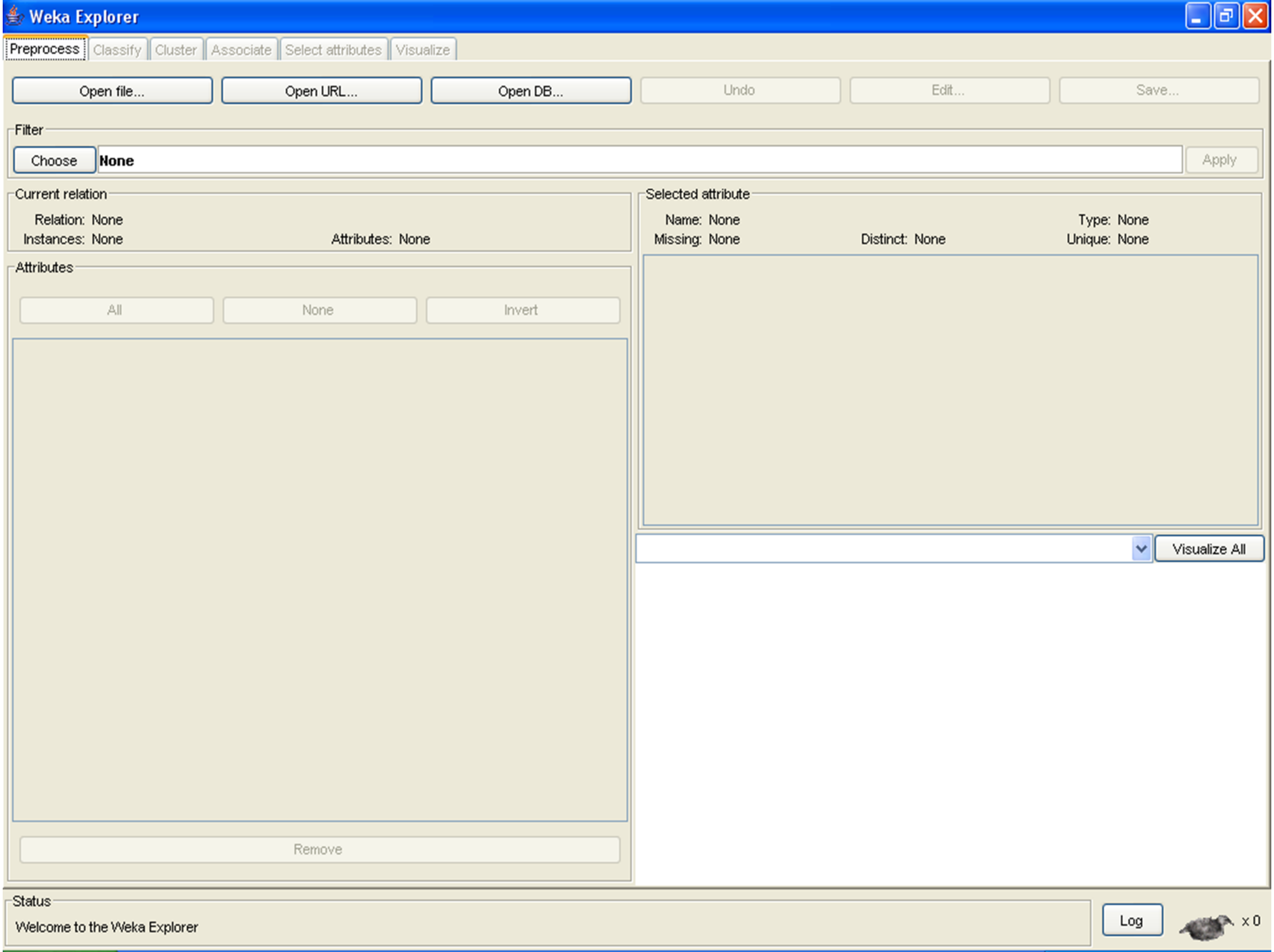
No.	Name
1	sepallength
2	sepalwidth
3	petallength
4	petalwidth
5	class

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)

Status
OK

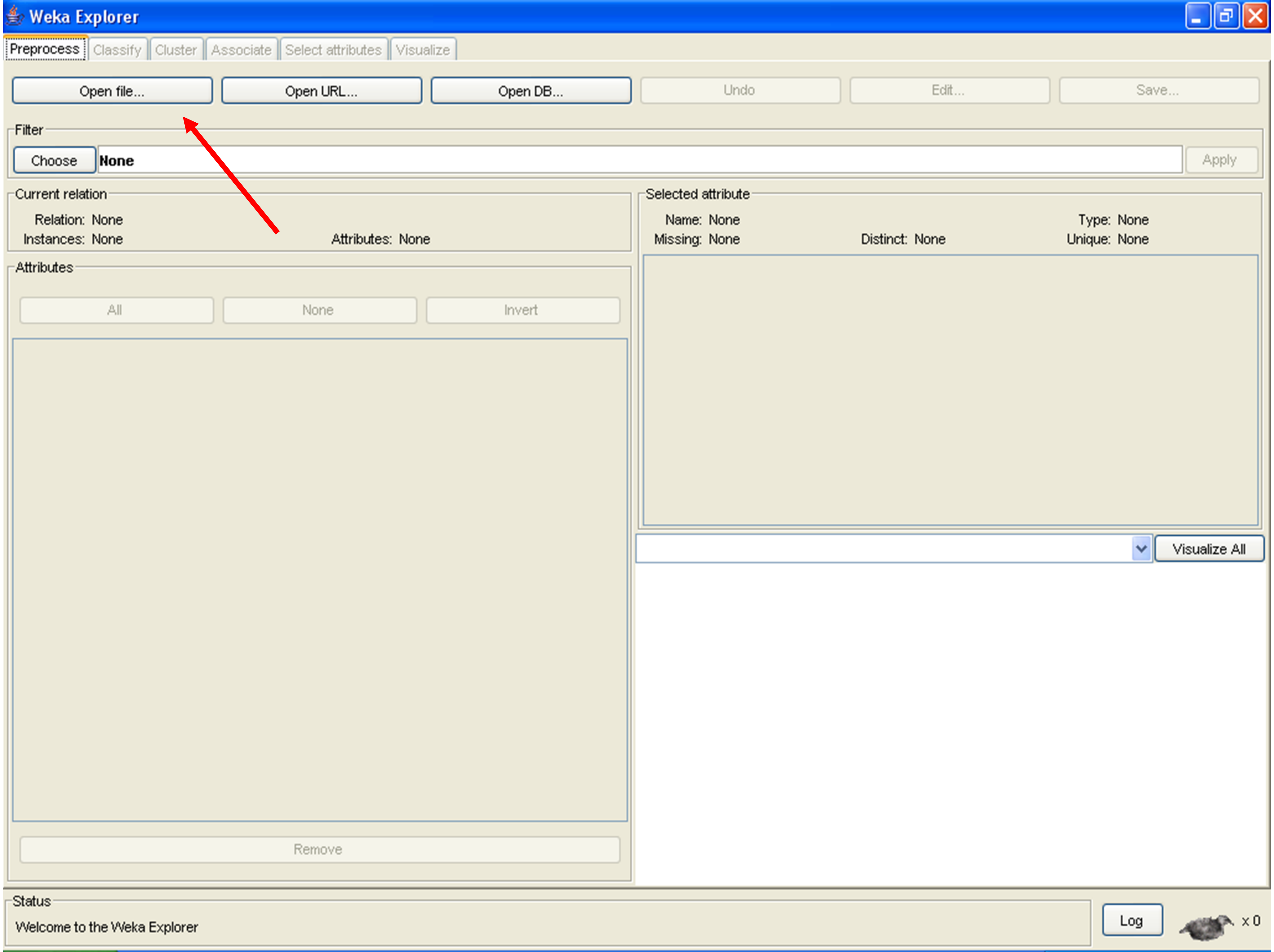


Filter: None

Relation: None
Instances: None
Attributes: None

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

Welcome to the Weka Explorer

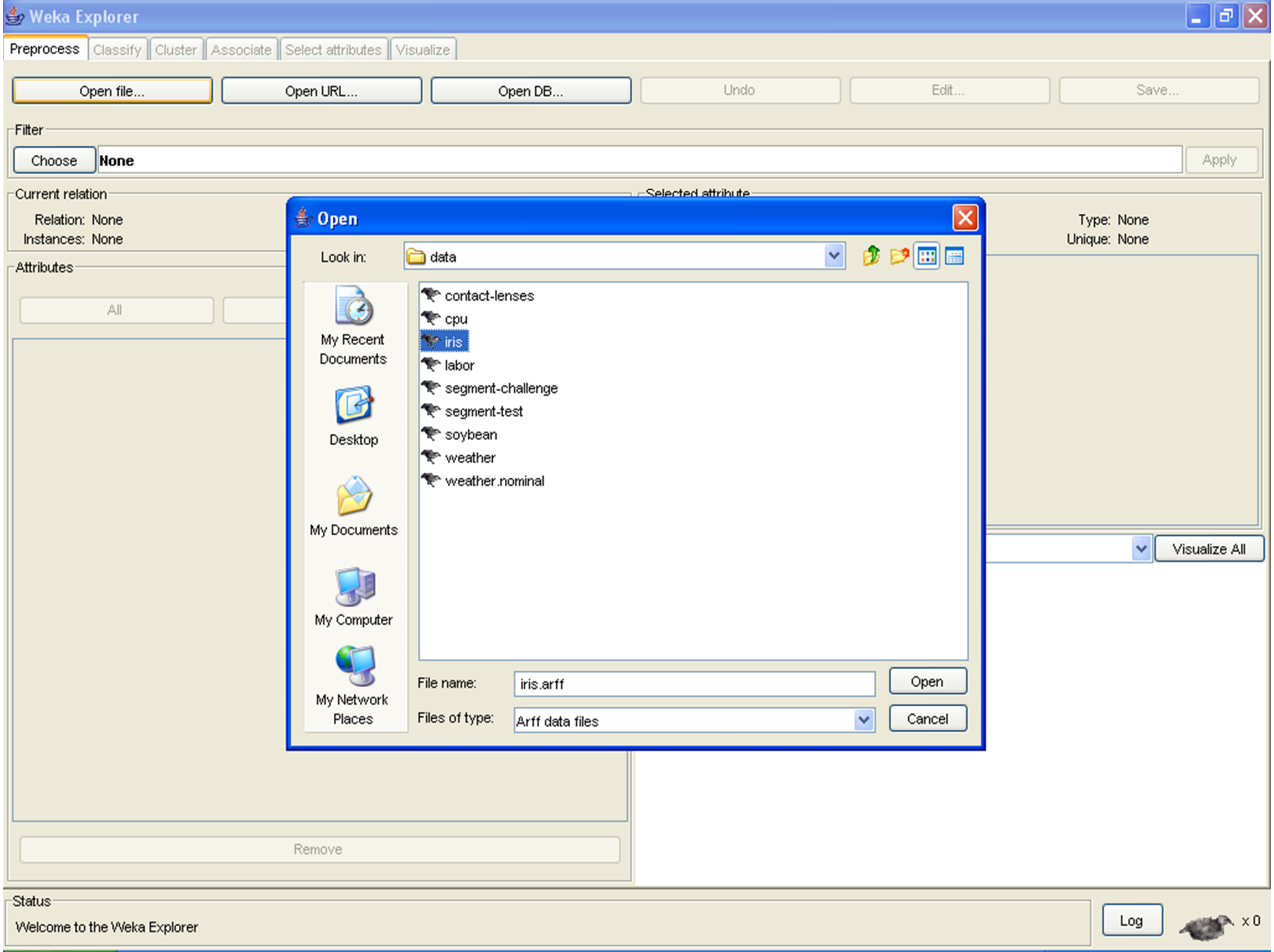


None

Relation: None
Instances: None
Attributes: None

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

Welcome to the Weka Explorer



Filter: None

Relation: None
Instances: None

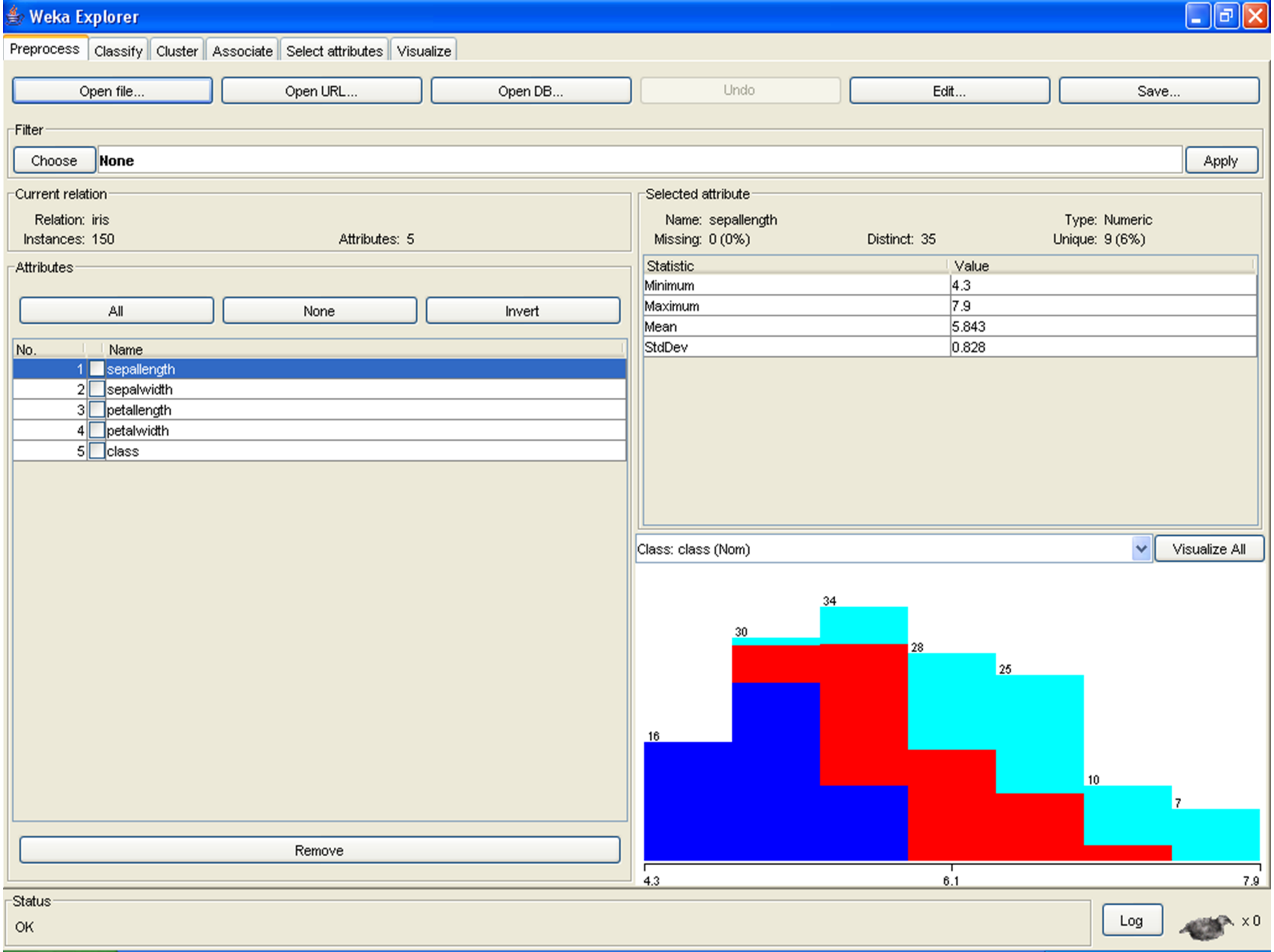
Type: None
Unique: None

Look in: data

File name: iris.arff

Files of type: Arff data files

Welcome to the Weka Explorer



Filter: None

Relation: iris
Instances: 150
Attributes: 5

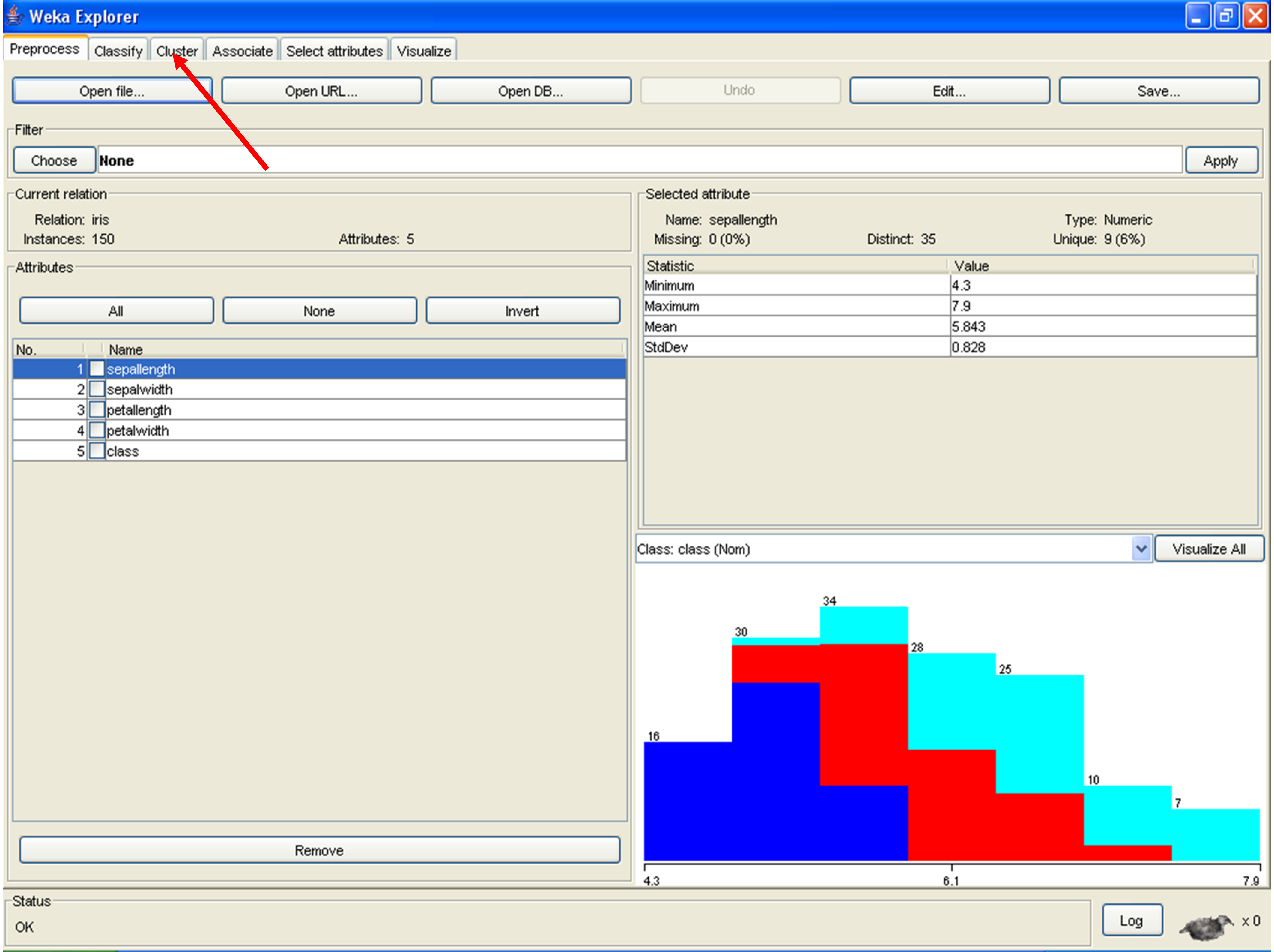
No.		Name
1	☐	sepallength
2	☐	sepalwidth
3	☐	petallength
4	☐	petalwidth
5	☐	class

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

Class: class (Nom)

OK



Filter: None

Current relation

Relation: iris
Instances: 150
Attributes: 5

Attributes

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

Selected attribute

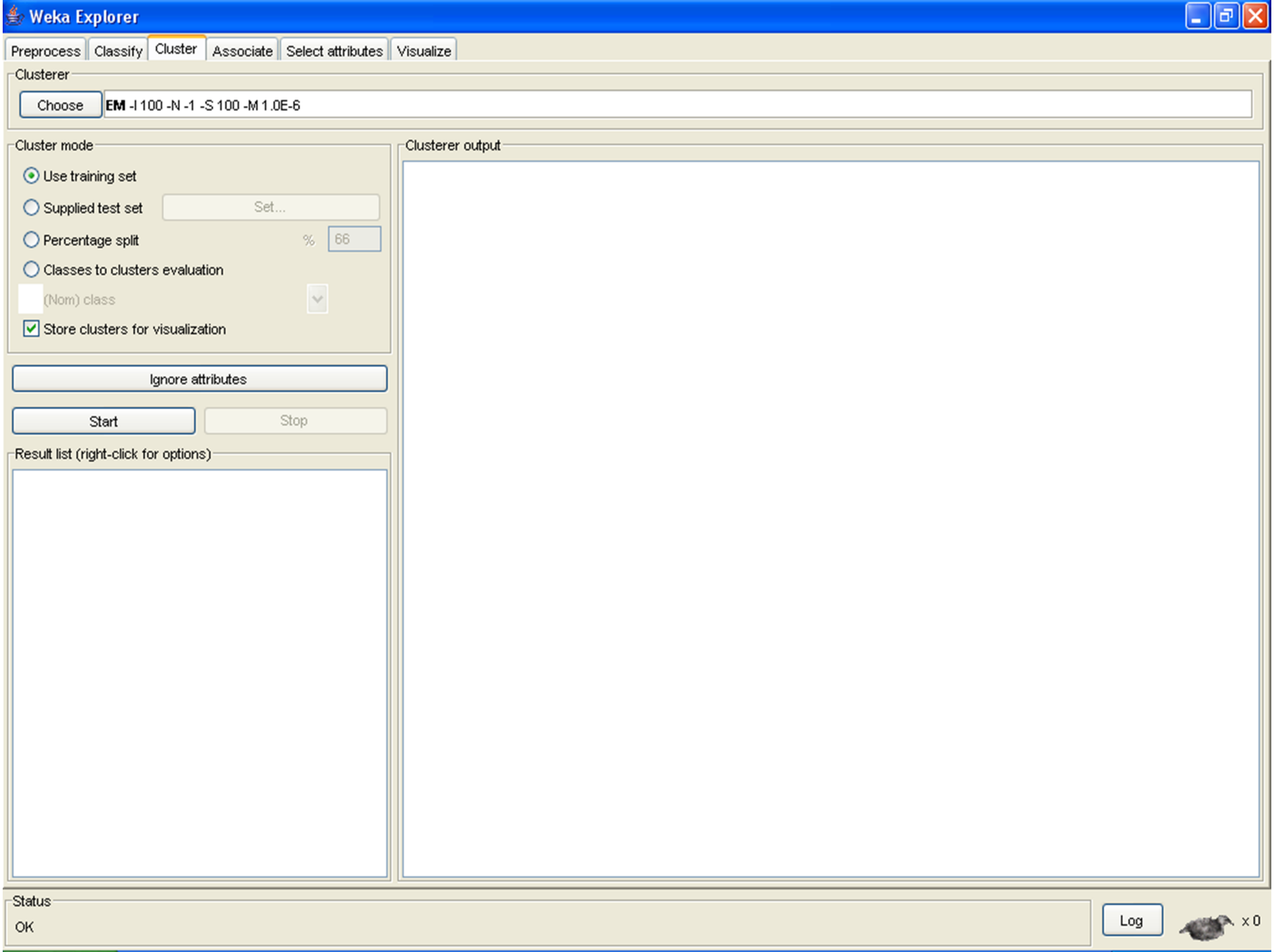
Name: sepallength
Type: Numeric
Missing: 0 (0%)
Distinct: 35
Unique: 9 (6%)

Statistic	Value
Minimum	4.3
Maximum	7.9
Mean	5.843
StdDev	0.828

Class: class (Nom)

Status

OK



Clusterer

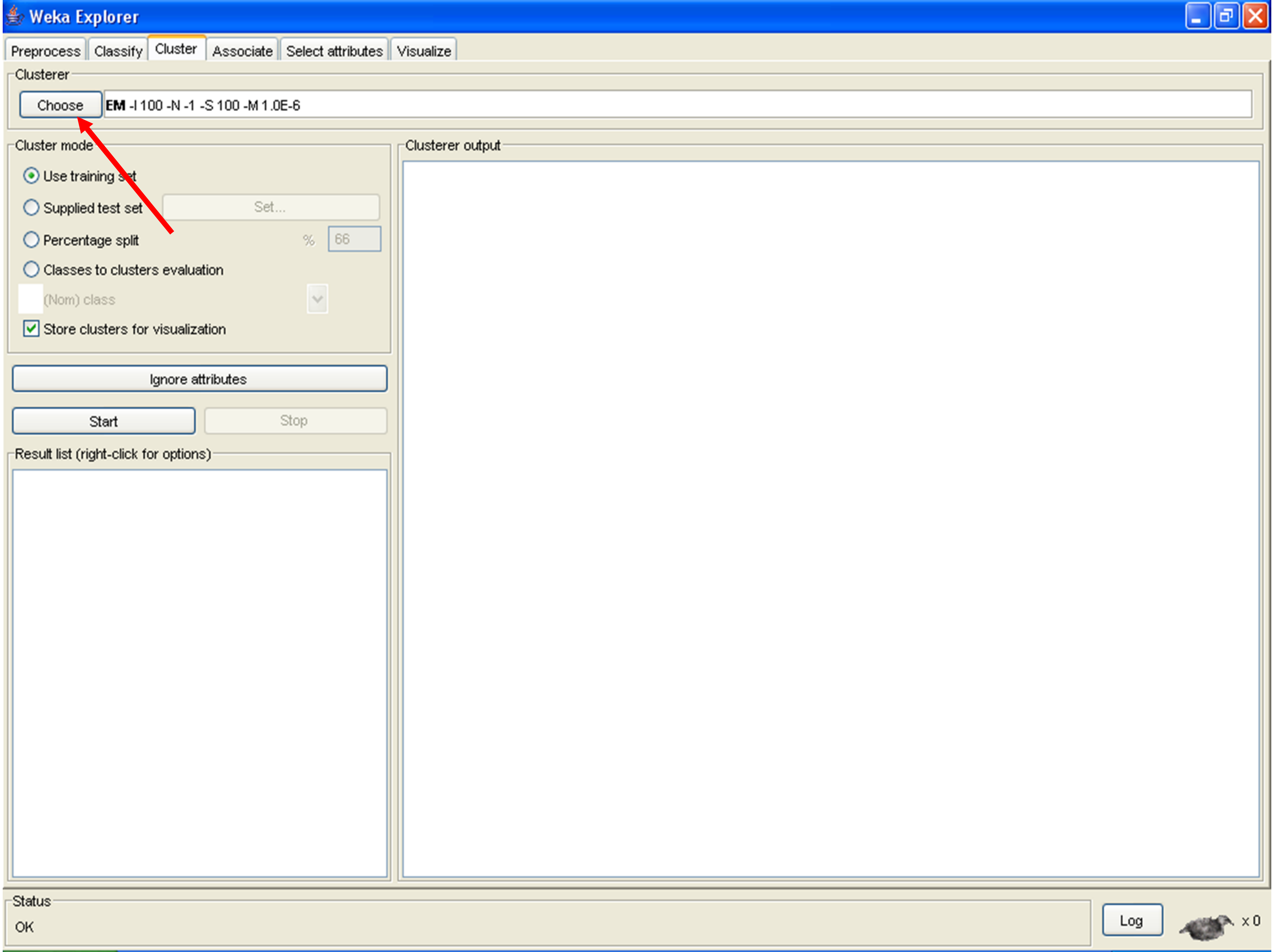
EM -I 100 -N -1 -S 100 -M 1.0E-6

Cluster mode

- ● Use training set
- ○ Supplied test set
- ○ Percentage split % 66
- ○ Classes to clusters evaluation
- (Nom) class
- ☑ Store clusters for visualization

Status

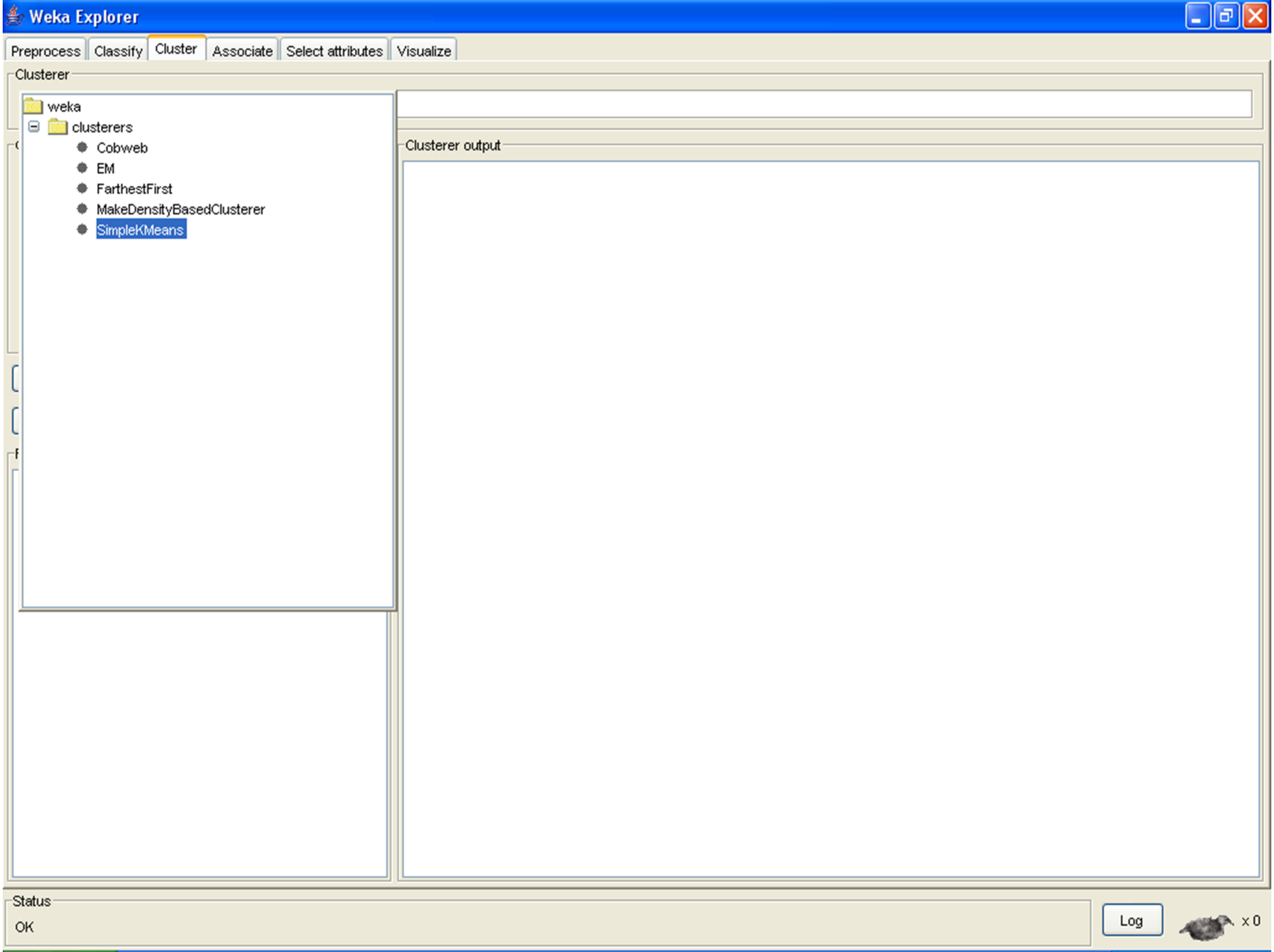
OK

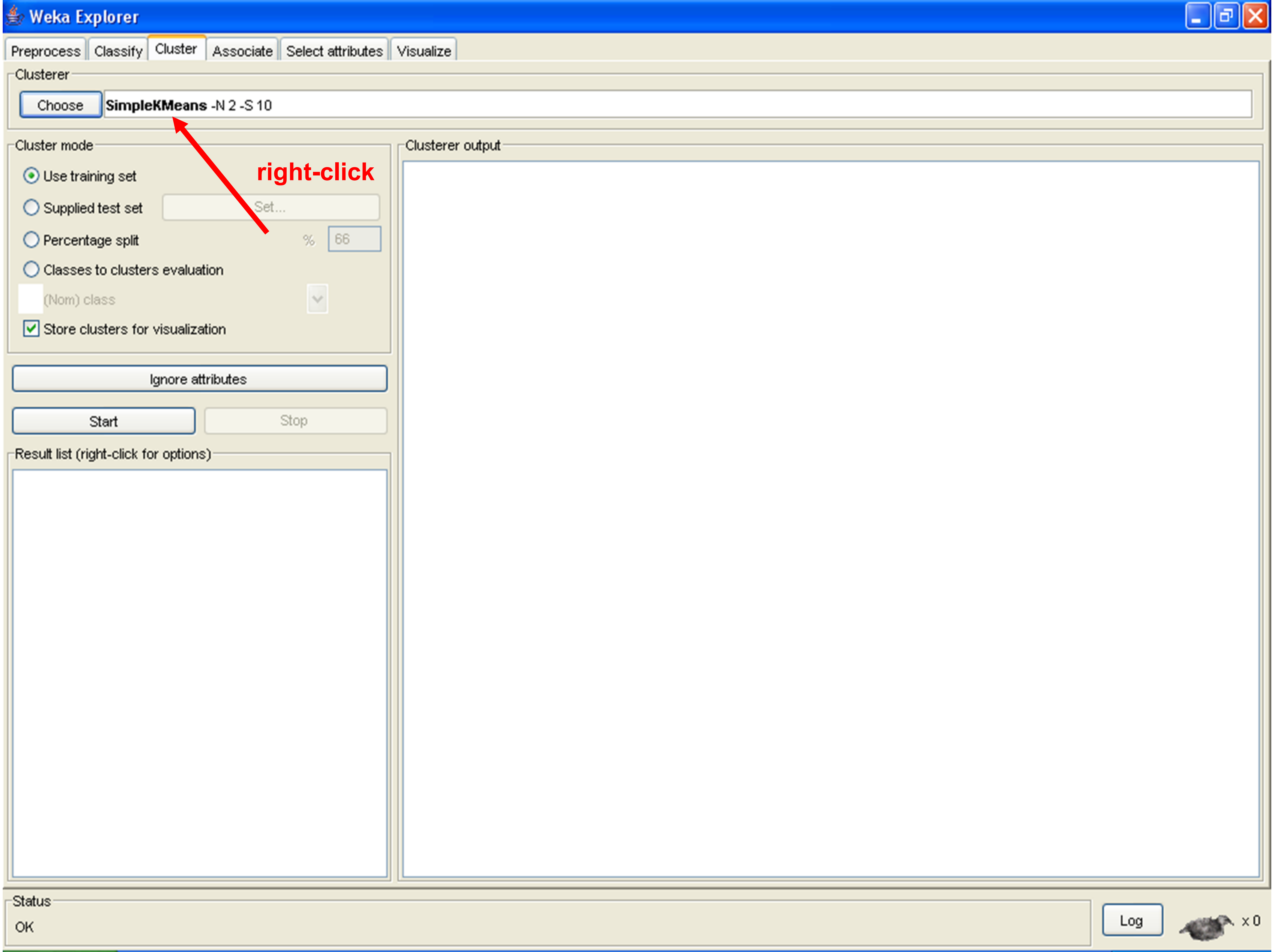


Clusterer: **EM** -I 100 -N -1 -S 100 -M 1.0E-6

Cluster mode:
● Use training set
○ Supplied test set
○ Percentage split % 66
○ Classes to clusters evaluation
☑ Store clusters for visualization

Status: OK

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
weka
clusterers
Cobweb
EM
FarthestFirst
MakeDensityBasedClusterer
SimpleKMeans
Clusterer output
Status
OK
Log
x 0



Clusterer

SimpleKMeans -N 2 -S 10

right-click

Cluster mode

● Use training set

○ Supplied test set

○ Percentage split % 66

○ Classes to clusters evaluation

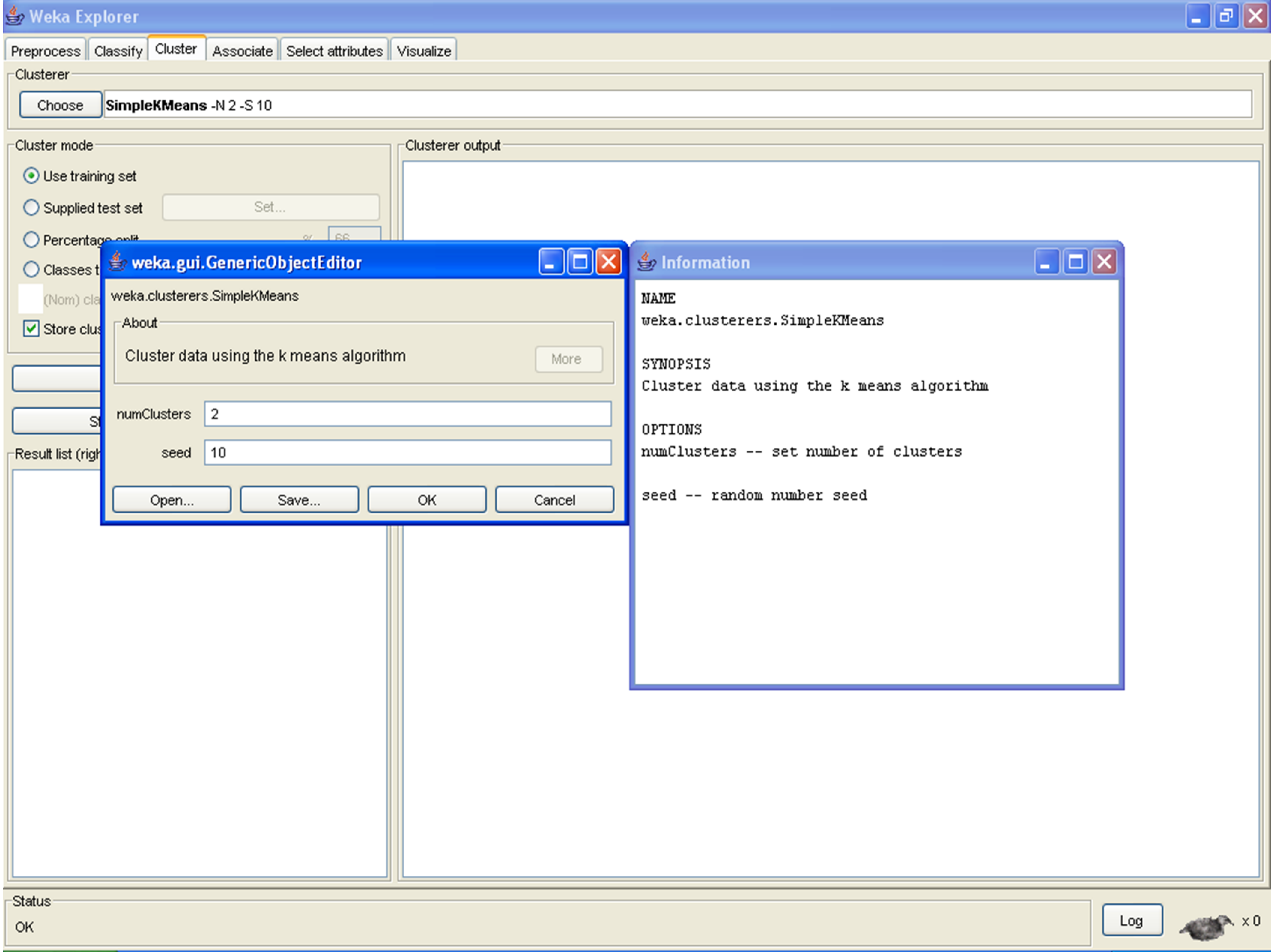
(Nom) class

☑ Store clusters for visualization

Result list (right-click for options)

Status

OK



SimpleKMeans -N 2 -S 10

weka.clusterers.SimpleKMeans

Cluster data using the k means algorithm

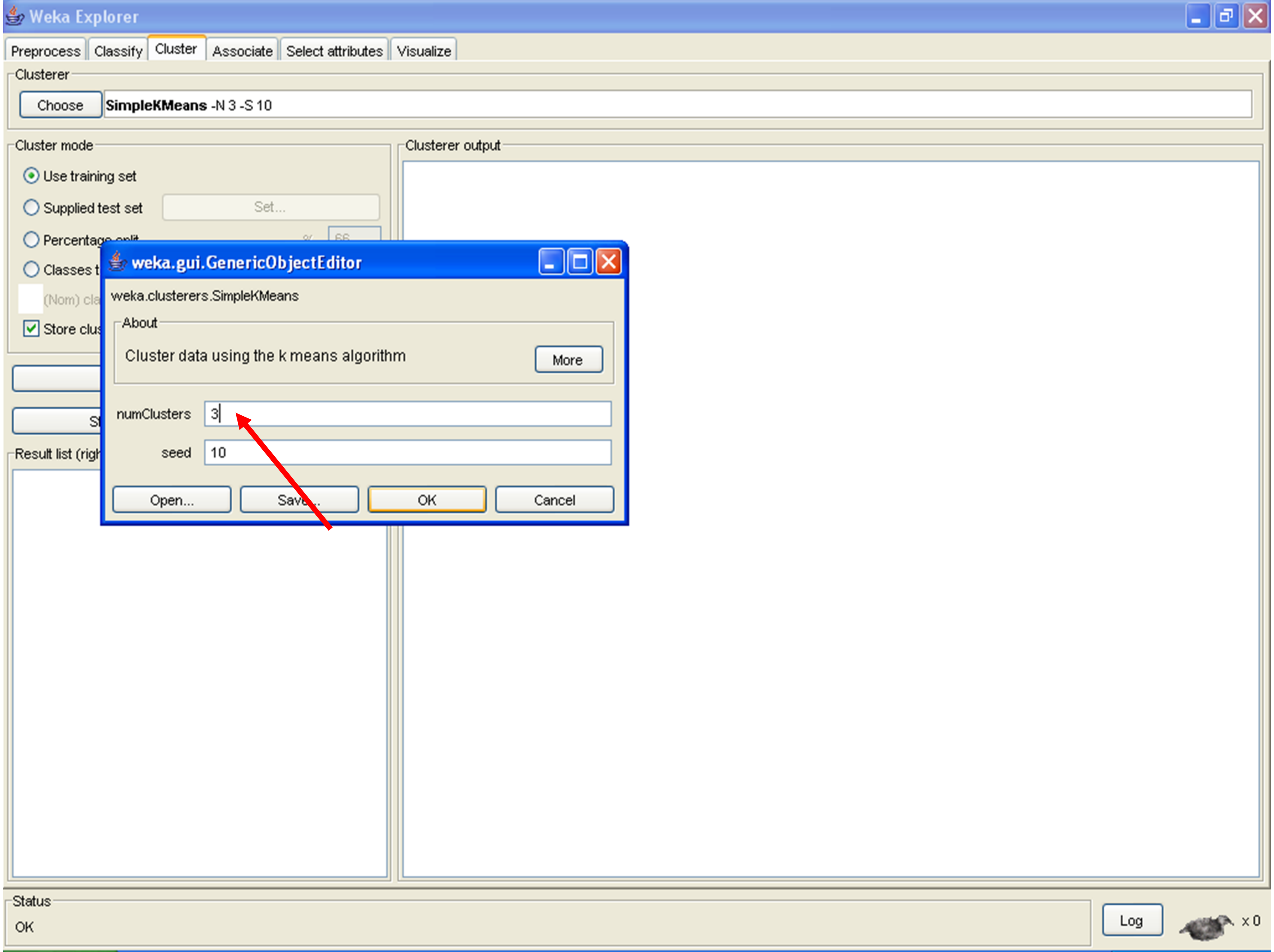
numClusters: 2

seed: 10

```
NAME
weka.clusterers.SimpleKMeans

SYNOPSIS
Cluster data using the k means algorithm

OPTIONS
numClusters -- set number of clusters

seed -- random number seed
```

Status

OK



SimpleKMeans -N 3 -S 10

● Use training set

○ Supplied test set

○ Percentage split

○ Classes t

(Nom) cla

☑ Store clus

weka.gui.GenericObjectEditor

weka.clusterers.SimpleKMeans

About

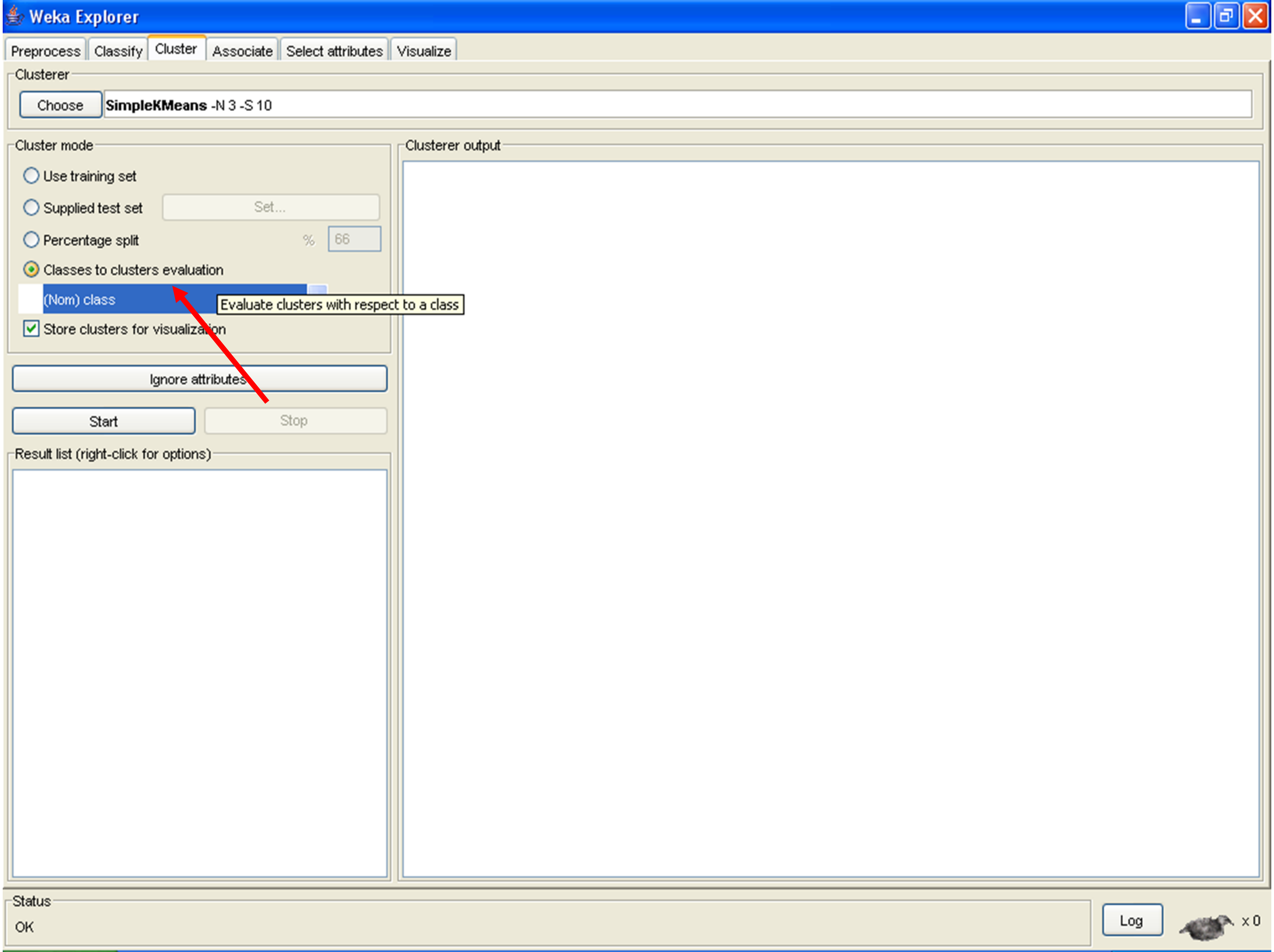
Cluster data using the k means algorithm

numClusters 3

seed 10

Status

OK



Clusterer: **SimpleKMeans** -N 3 -S 10

Cluster mode:

○ Use training set

○ Supplied test set

○ Percentage split

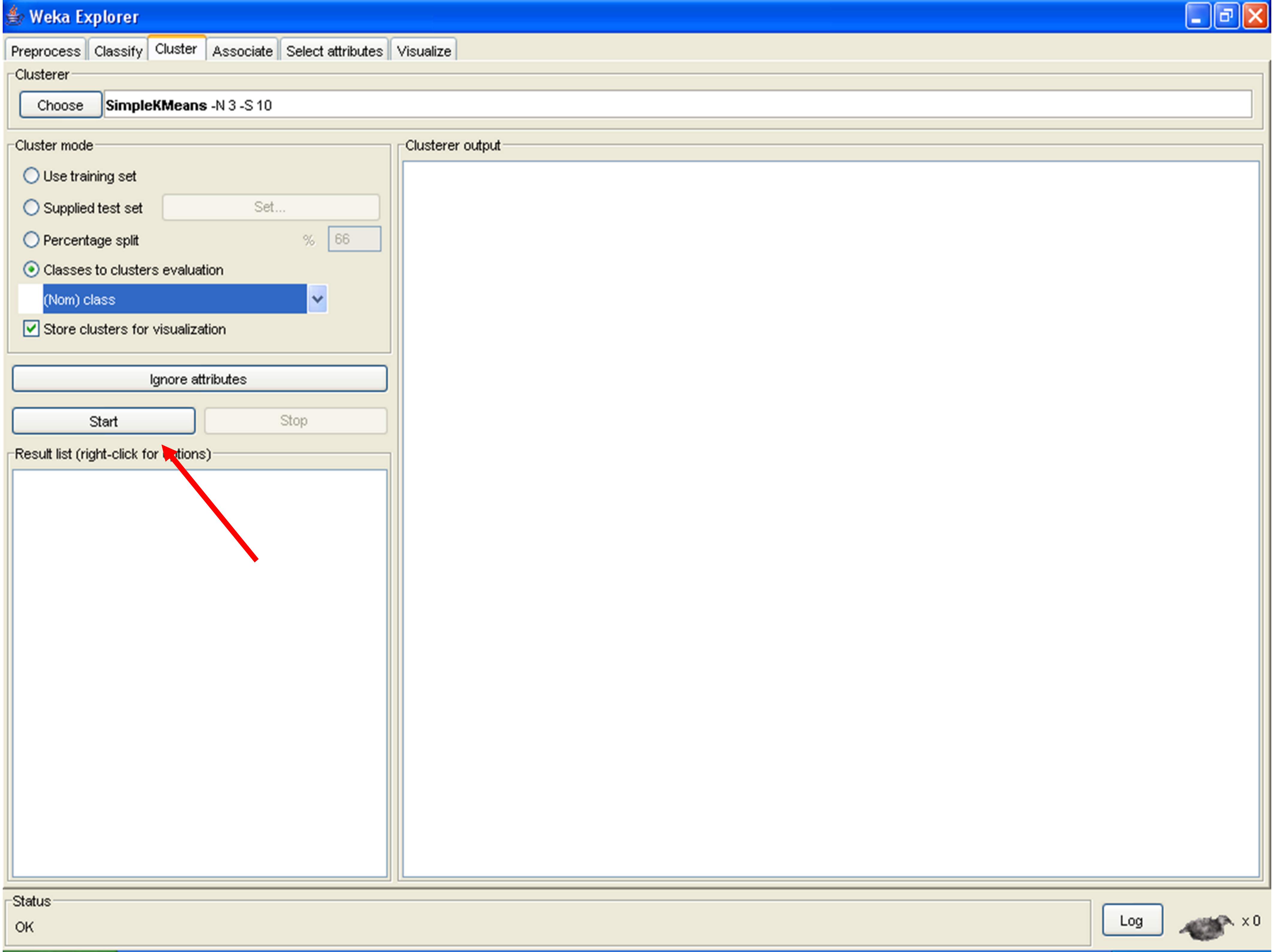
● Classes to clusters evaluation

(Nom) class

Evaluate clusters with respect to a class

☑ Store clusters for visualization

Status: OK

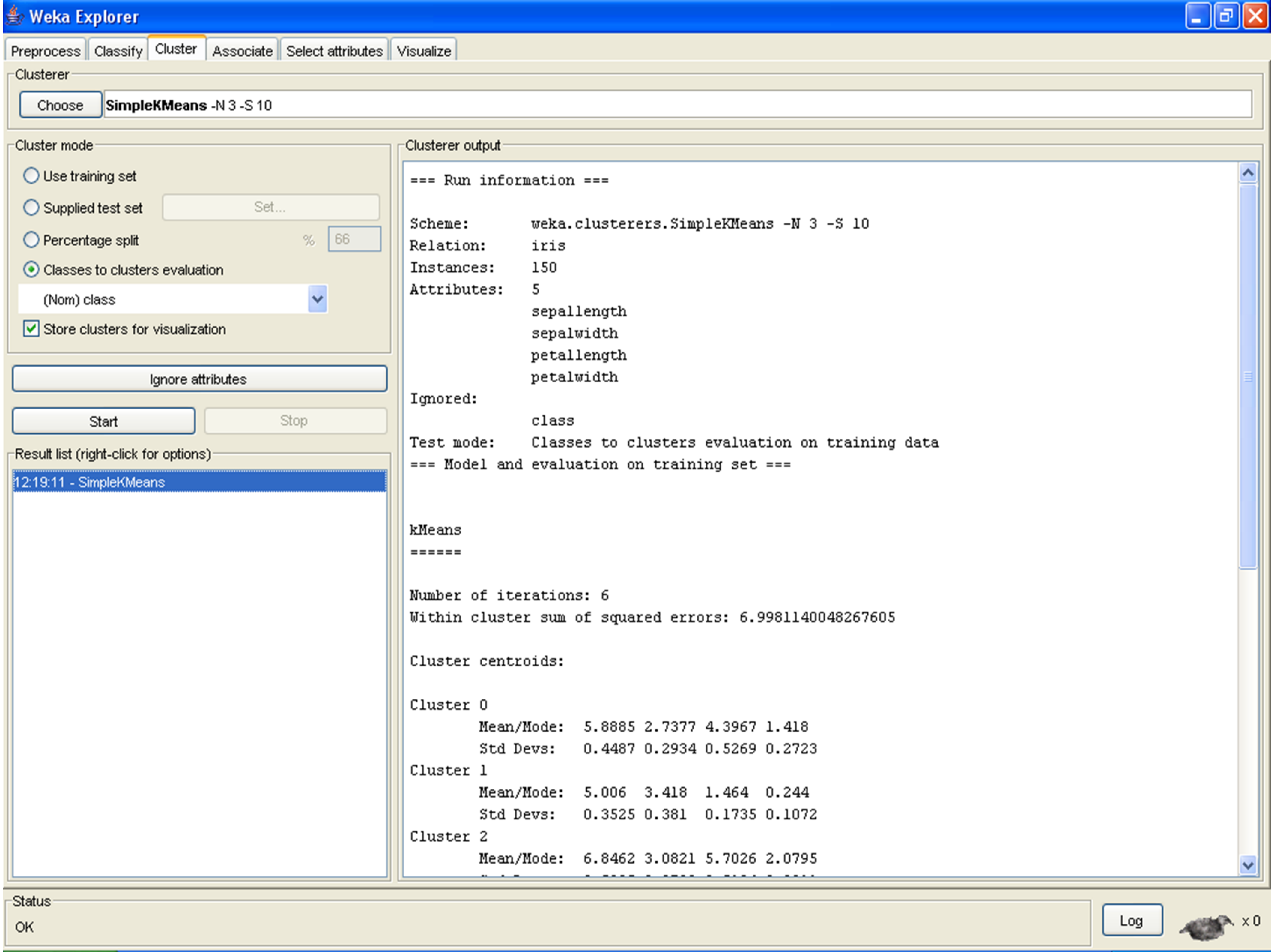


Clusterer: **SimpleKMeans** -N 3 -S 10

Cluster mode:

○ Use training set
○ Supplied test set
○ Percentage split % 66
● Classes to clusters evaluation
(Nom) class
☑ Store clusters for visualization

Status: OK



Clusterer: **SimpleKMeans** -N 3 -S 10

Cluster mode:

○ Use training set

○ Supplied test set

○ Percentage split % 66

● Classes to clusters evaluation

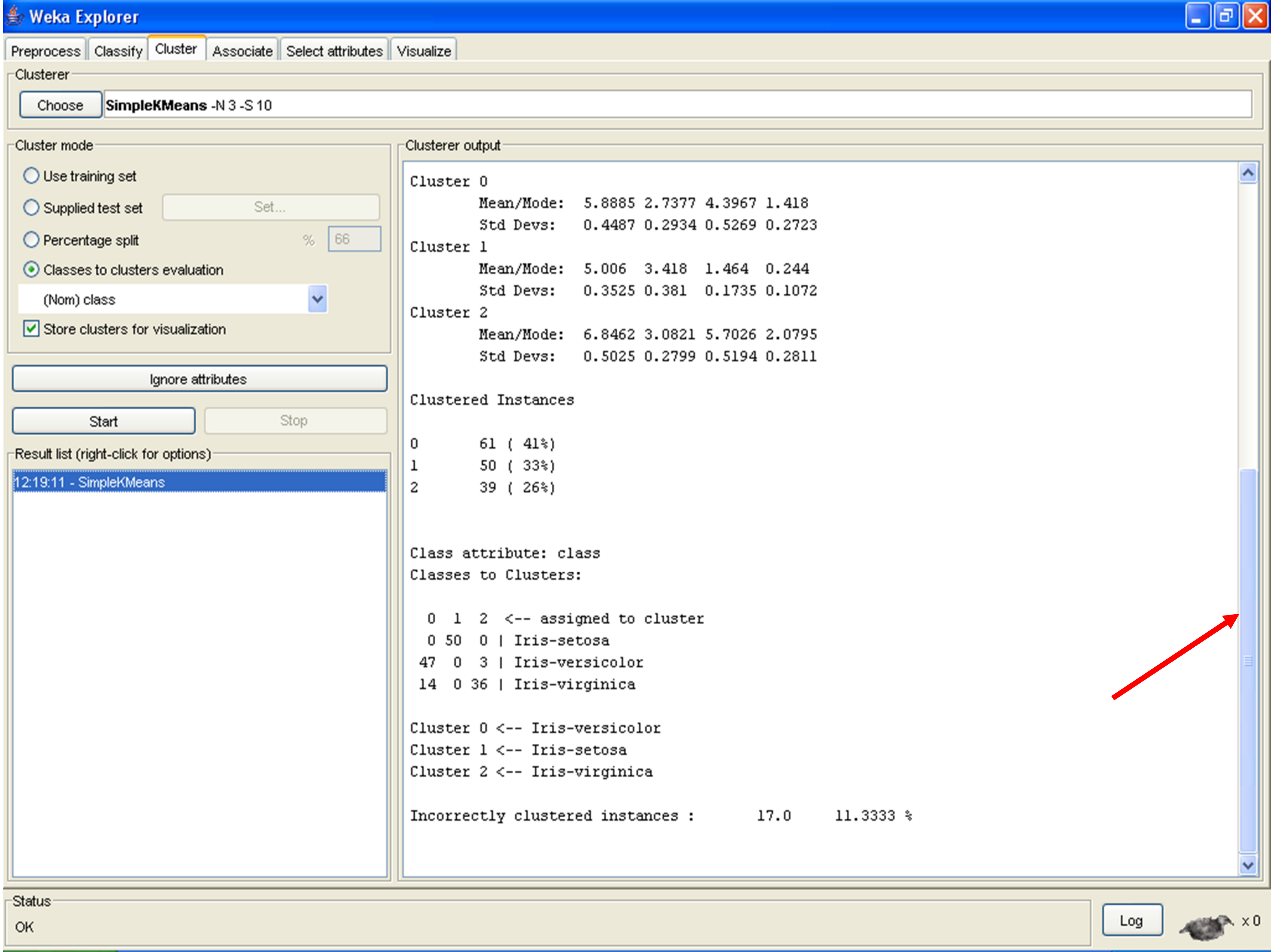
(Nom) class

☑ Store clusters for visualization

Result list: 12:19:11 - SimpleKMeans

Status: OK

```
=== Run information ===

Scheme:       weka.clusterers.SimpleKMeans -N 3 -S 10
Relation:     iris
Instances:    150
Attributes:   5
              sepallength
              sepalwidth
              petallength
              petalwidth
Ignored:
              class
Test mode:    Classes to clusters evaluation on training data
=== Model and evaluation on training set ===


kMeans
======

Number of iterations: 6
Within cluster sum of squared errors: 6.998114004826705

Cluster centroids:

Cluster 0
        Mean/Mode:  5.8885 2.7377 4.3967 1.418
        Std Devs:   0.4487 0.2934 0.5269 0.2723
Cluster 1
        Mean/Mode:  5.006  3.418  1.464  0.244
        Std Devs:   0.3525 0.381  0.1735 0.1072
Cluster 2
        Mean/Mode:  6.8462 3.0821 5.7026 2.0795
```



Clusterer: **SimpleKMeans** -N 3 -S 10

Cluster mode

○ Use training set

○ Supplied test set

○ Percentage split % 66

● Classes to clusters evaluation

(Nom) class

☑ Store clusters for visualization

Result list (right-click for options)

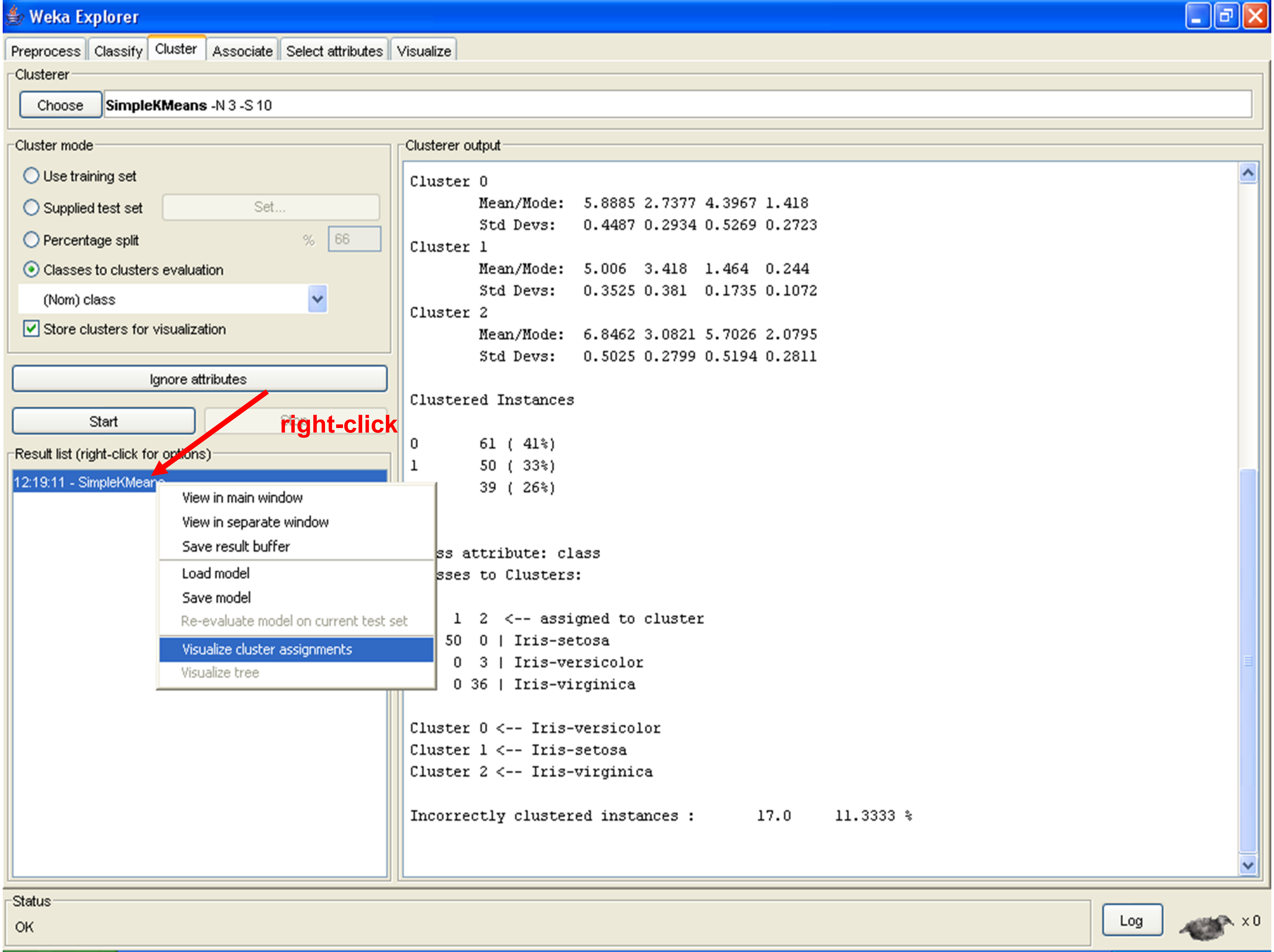
12:19:11 - SimpleKMeans

Clusterer output

```
Cluster 0
        Mean/Mode:   5.8885 2.7377 4.3967 1.418
        Std Devs:    0.4487 0.2934 0.5269 0.2723
Cluster 1
        Mean/Mode:   5.006  3.418  1.464  0.244
        Std Devs:    0.3525 0.381  0.1735 0.1072
Cluster 2
        Mean/Mode:   6.8462 3.0821 5.7026 2.0795
        Std Devs:    0.5025 0.2799 0.5194 0.2811

Clustered Instances

0        61 ( 41%)
1        50 ( 33%)
2        39 ( 26%)


Class attribute: class
Classes to Clusters:

  0  1  2  <-- assigned to cluster
  0 50  0 | Iris-setosa
 47  0  3 | Iris-versicolor
 14  0 36 | Iris-virginica

Cluster 0 <-- Iris-versicolor
Cluster 1 <-- Iris-setosa
Cluster 2 <-- Iris-virginica

Incorrectly clustered instances :       17.0     11.3333 %
```

Status

OK



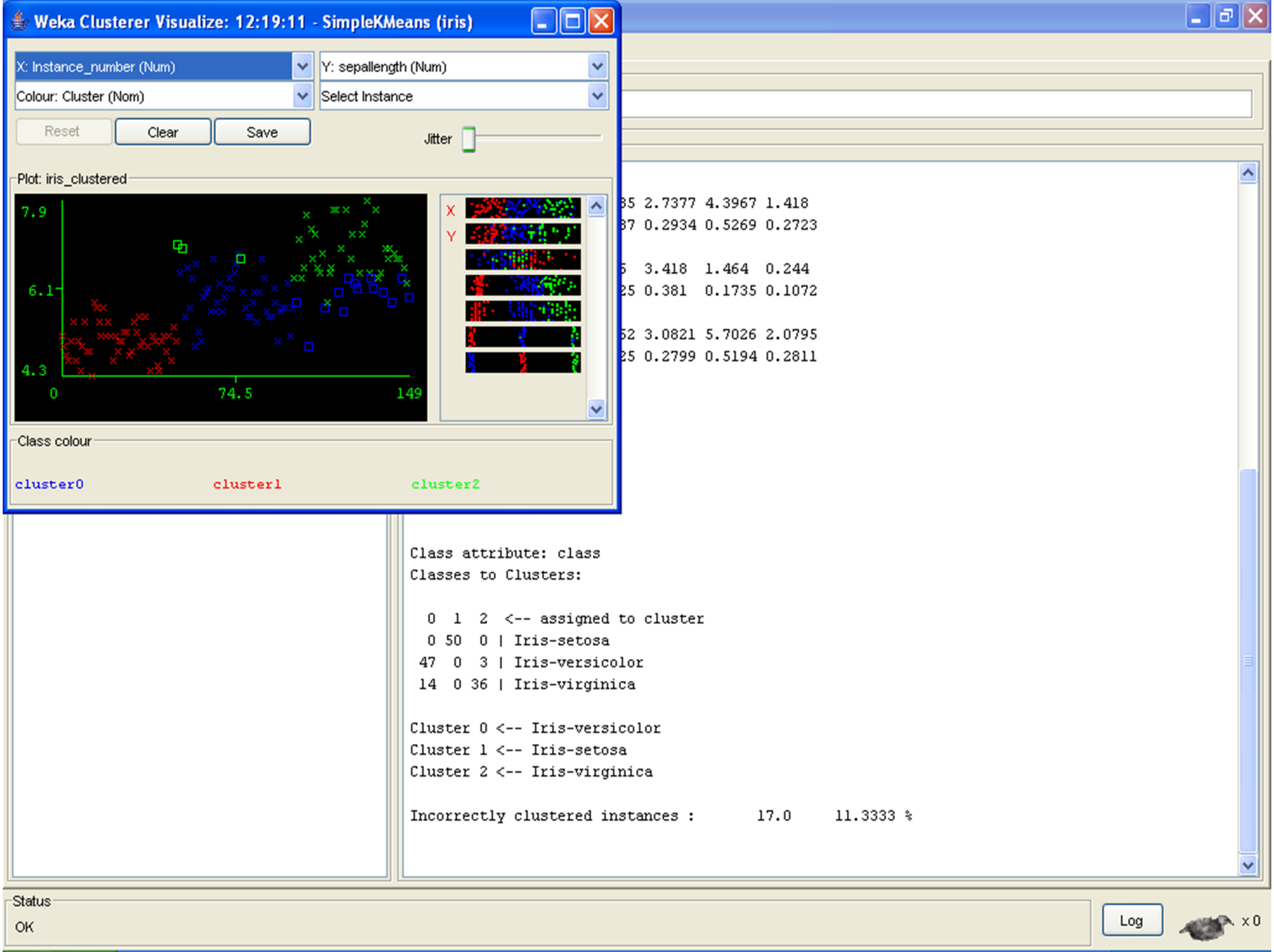
Clusterer: **SimpleKMeans** -N 3 -S 10

Cluster mode:
- ○ Use training set
- ○ Supplied test set
- ○ Percentage split % 66
- ● Classes to clusters evaluation
  - (Nom) class
- ☑ Store clusters for visualization

Result list: 12:19:11 - SimpleKMeans

```
Cluster 0
        Mean/Mode:   5.8885 2.7377 4.3967 1.418
        Std Devs:    0.4487 0.2934 0.5269 0.2723
Cluster 1
        Mean/Mode:   5.006  3.418  1.464  0.244
        Std Devs:    0.3525 0.381  0.1735 0.1072
Cluster 2
        Mean/Mode:   6.8462 3.0821 5.7026 2.0795
        Std Devs:    0.5025 0.2799 0.5194 0.2811

Clustered Instances

0       61 ( 41%)
1       50 ( 33%)
        39 ( 26%)


     ss attribute: class
     sses to Clusters:

      1  2  <-- assigned to cluster
     50  0 | Iris-setosa
      0  3 | Iris-versicolor
      0 36 | Iris-virginica

Cluster 0 <-- Iris-versicolor
Cluster 1 <-- Iris-setosa
Cluster 2 <-- Iris-virginica

Incorrectly clustered instances :       17.0     11.3333 %
```

Status: OK



X: Instance_number (Num)

Y: sepallength (Num)

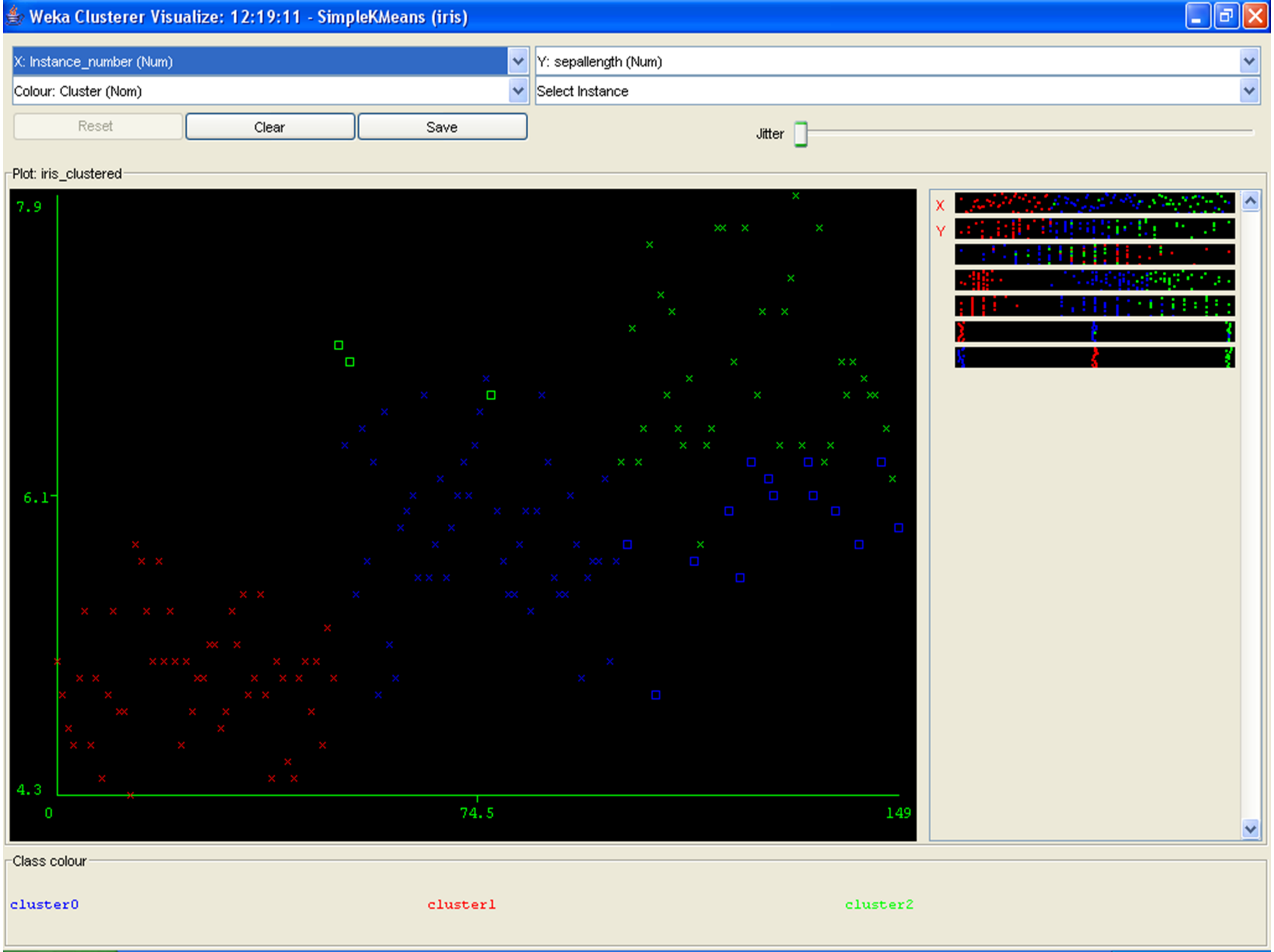
Colour: Cluster (Nom)

Select Instance

```
35 2.7377 4.3967 1.418
37 0.2934 0.5269 0.2723

5  3.418  1.464  0.244
25 0.381  0.1735 0.1072

52 3.0821 5.7026 2.0795
25 0.2799 0.5194 0.2811
```

```
Class attribute: class
Classes to Clusters:

  0  1  2  <-- assigned to cluster
  0 50  0 | Iris-setosa
 47  0  3 | Iris-versicolor
 14  0 36 | Iris-virginica

Cluster 0 <-- Iris-versicolor
Cluster 1 <-- Iris-setosa
Cluster 2 <-- Iris-virginica

Incorrectly clustered instances :       17.0     11.3333 %
```

OK



Weka Clusterer Visualize: 12:19:11 - SimpleKMeans (iris)

X: Instance_number (Num)

Y: sepallength (Num)

Colour: Cluster (Nom)

Select Instance

Plot: iris_clustered

Class colour

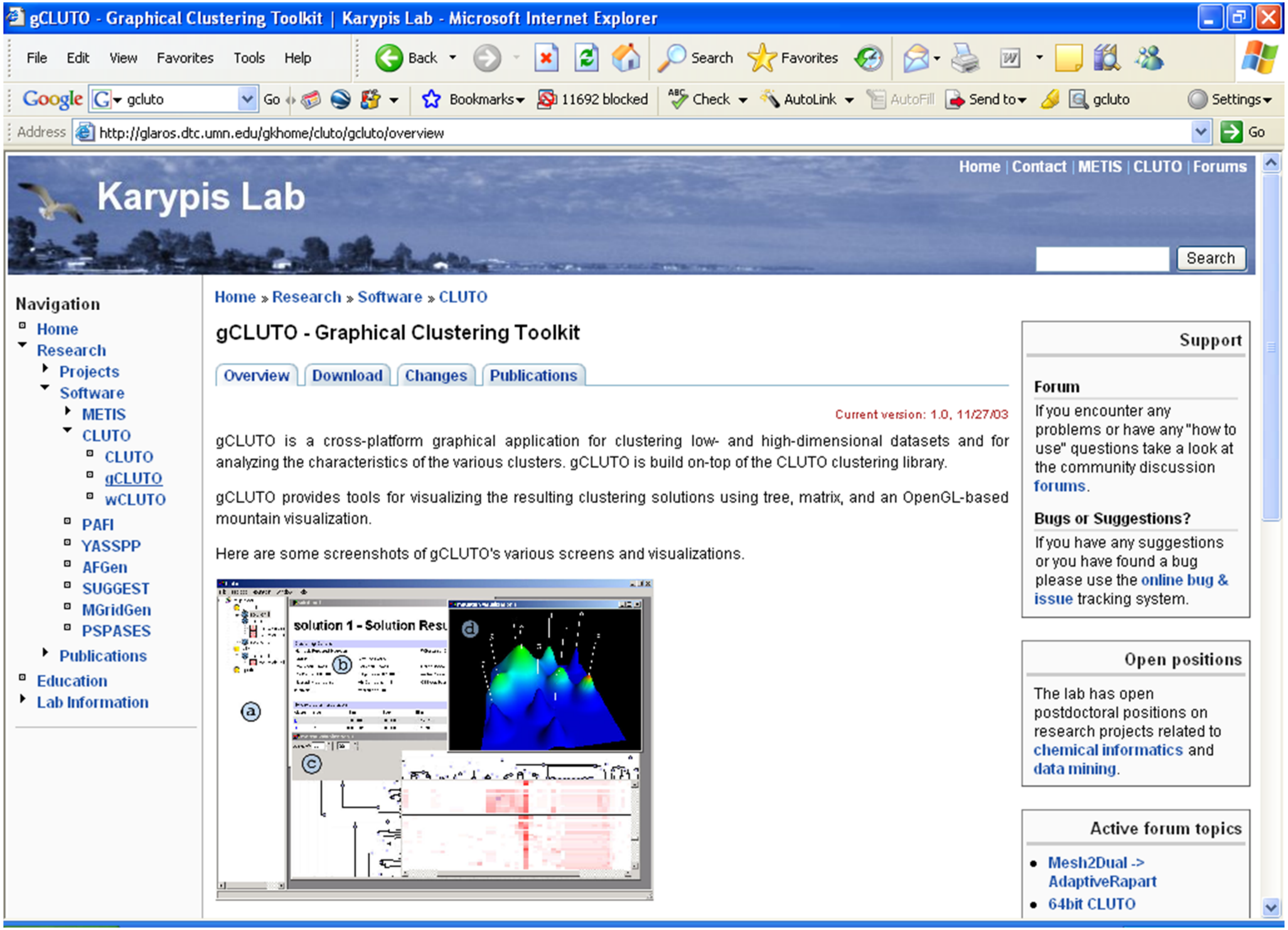
cluster0

cluster1

cluster2

Experiment other visualization options

Experiment different clustering parameters and algorithms



Navigation

- Home
- Research
  - Projects
  - Software
    - METIS
    - CLUTO
      - CLUTO
      - gCLUTO
      - wCLUTO
    - PAFI
    - YASSPP
    - AFGen
    - SUGGEST
    - MGridGen
    - PSPASES
  - Publications
- Education
- Lab Information

Home » Research » Software » CLUTO

gCLUTO - Graphical Clustering Toolkit

Overview | Download | Changes | Publications

Current version: 1.0, 11/27/03

gCLUTO is a cross-platform graphical application for clustering low- and high-dimensional datasets and for analyzing the characteristics of the various clusters. gCLUTO is build on-top of the CLUTO clustering library.

gCLUTO provides tools for visualizing the resulting clustering solutions using tree, matrix, and an OpenGL-based mountain visualization.

Here are some screenshots of gCLUTO's various screens and visualizations.

Support

Forum

If you encounter any problems or have any "how to use" questions take a look at the community discussion forums.

Bugs or Suggestions?

If you have any suggestions or you have found a bug please use the online bug & issue tracking system.

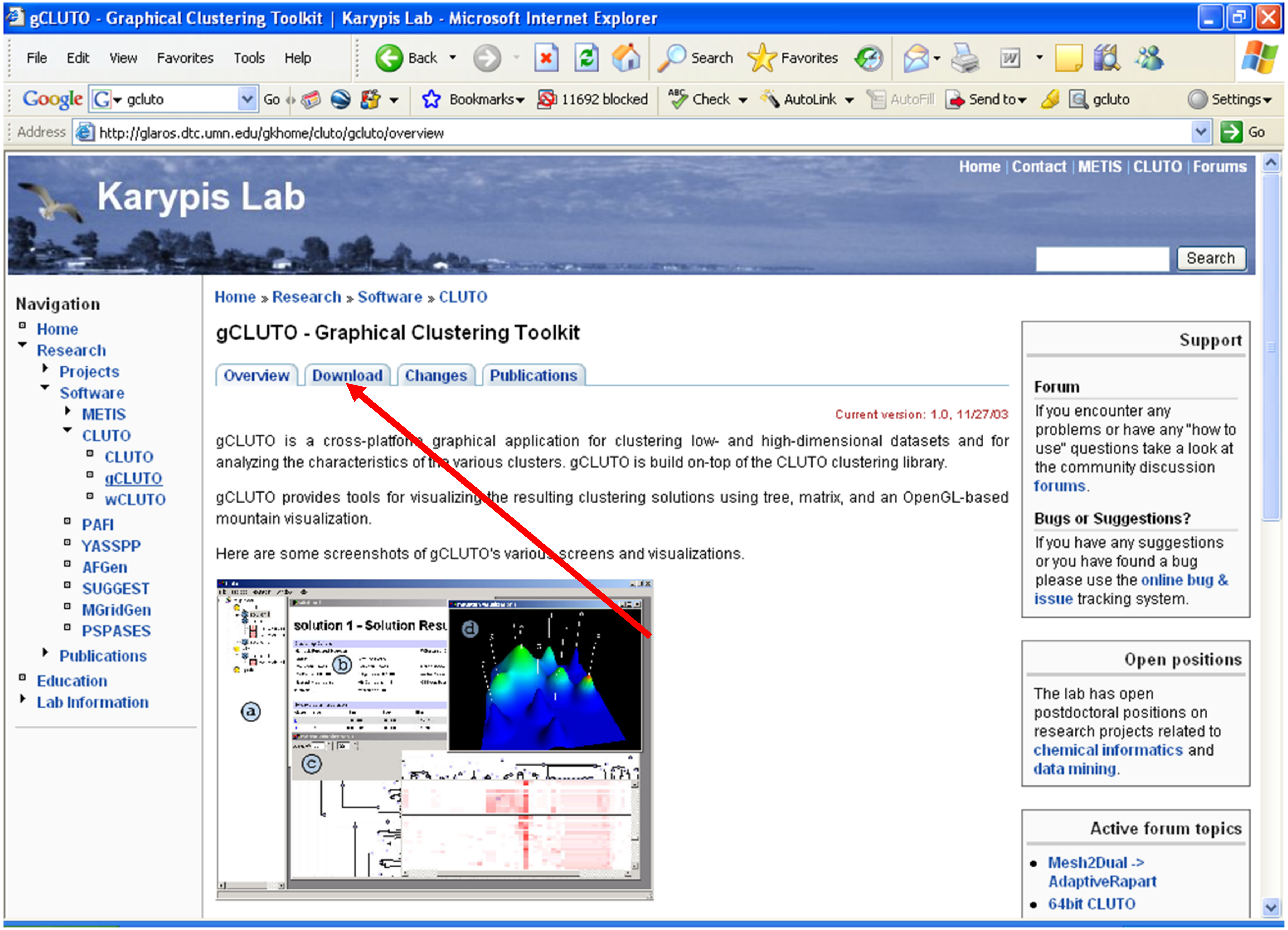
Open positions

The lab has open postdoctoral positions on research projects related to chemical informatics and data mining.

Active forum topics

- Mesh2Dual -> AdaptiveRapart
- 64bit CLUTO

http://glaros.dtc.umn.edu/gkhome/cluto/gcluto/overview



gCLUTO - Graphical Clustering Toolkit

gCLUTO is a cross-platform graphical application for clustering low- and high-dimensional datasets and for analyzing the characteristics of the various clusters. gCLUTO is build on-top of the CLUTO clustering library.

gCLUTO provides tools for visualizing the resulting clustering solutions using tree, matrix, and an OpenGL-based mountain visualization.

Here are some screenshots of gCLUTO's various screens and visualizations.

Forum

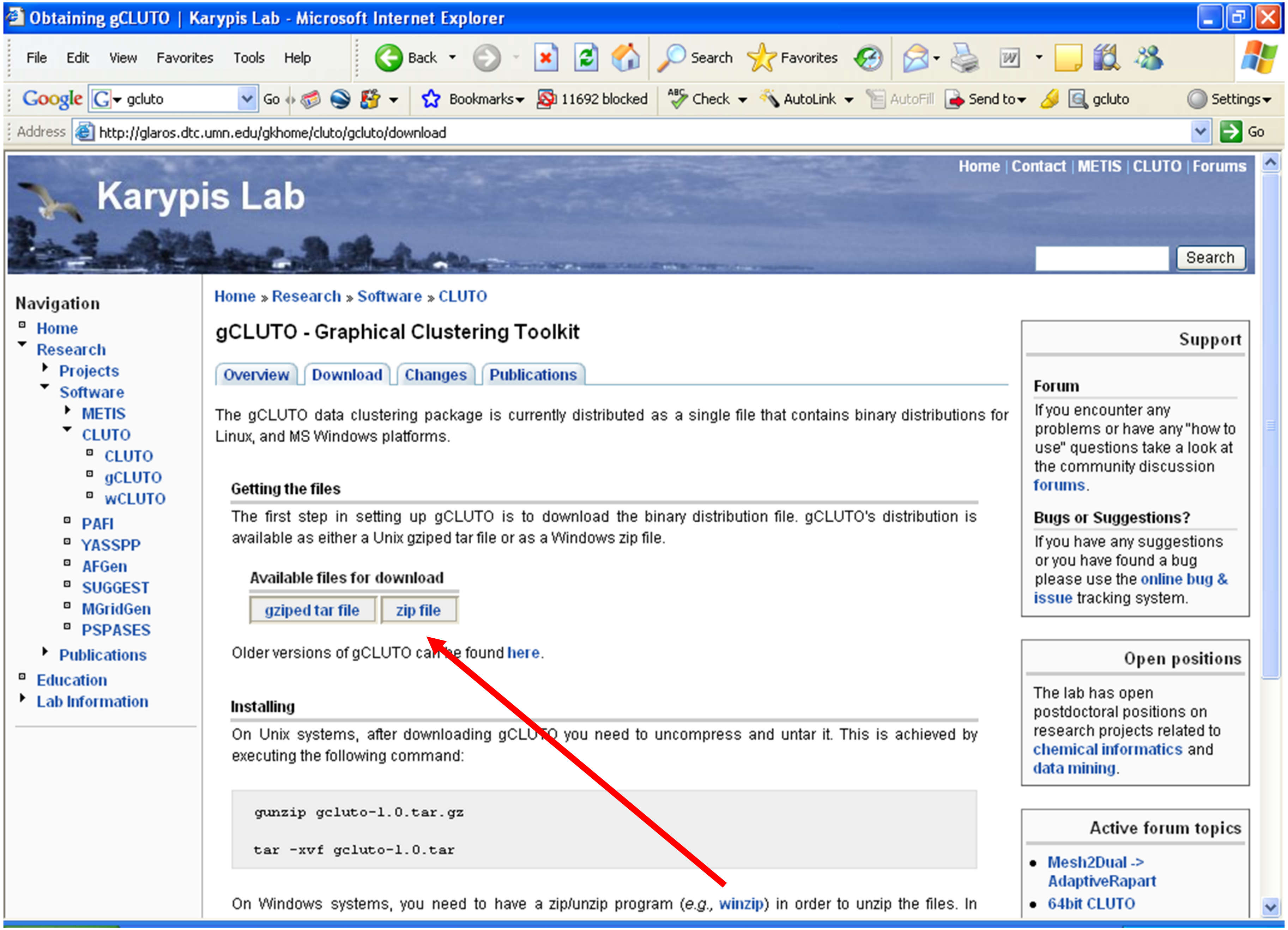
If you encounter any problems or have any "how to use" questions take a look at the community discussion forums.

Bugs or Suggestions?

If you have any suggestions or you have found a bug please use the online bug & issue tracking system.

The lab has open postdoctoral positions on research projects related to chemical informatics and data mining.

http://glaros.dtc.umn.edu/gkhome/cluto/gcluto/overview



gCLUTO - Graphical Clustering Toolkit

The gCLUTO data clustering package is currently distributed as a single file that contains binary distributions for Linux, and MS Windows platforms.

Getting the files

The first step in setting up gCLUTO is to download the binary distribution file. gCLUTO's distribution is available as either a Unix gziped tar file or as a Windows zip file.

Available files for download

Older versions of gCLUTO can be found here.

Installing

On Unix systems, after downloading gCLUTO you need to uncompress and untar it. This is achieved by executing the following command:

```
gunzip gcluto-1.0.tar.gz

tar -xvf gcluto-1.0.tar
```

On Windows systems, you need to have a zip/unzip program (*e.g.*, winzip) in order to unzip the files. In

Forum

If you encounter any problems or have any "how to use" questions take a look at the community discussion forums.

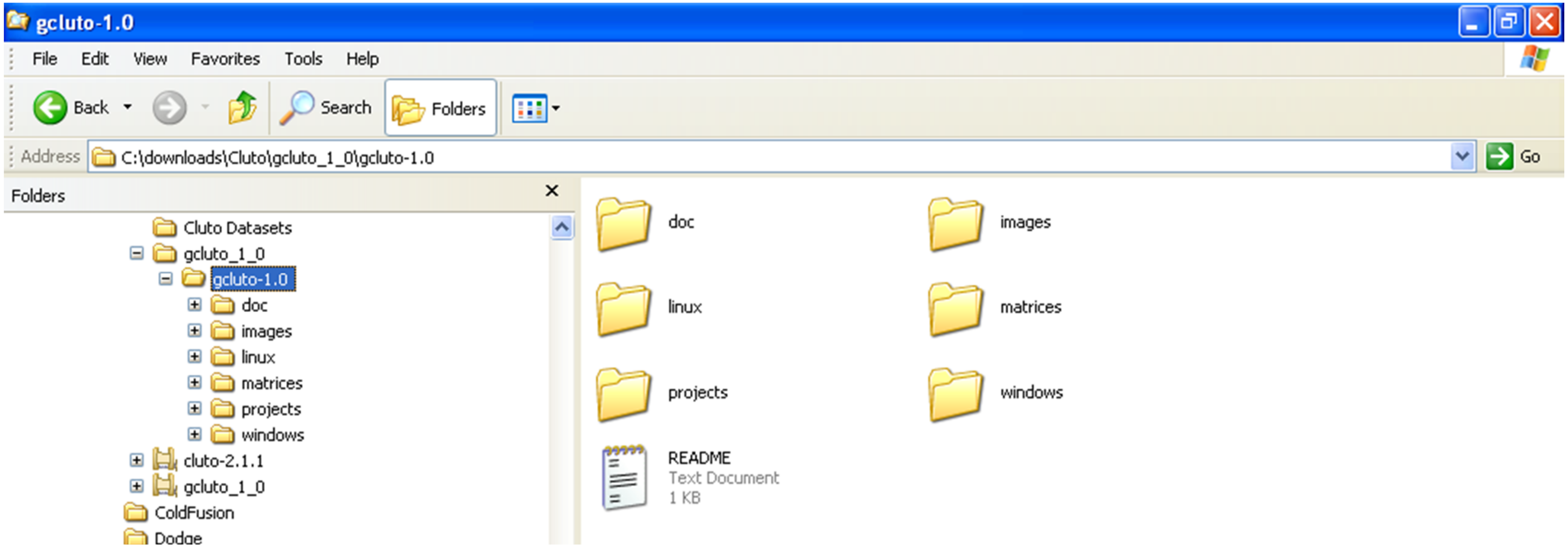
Bugs or Suggestions?

If you have any suggestions or you have found a bug please use the online bug & issue tracking system.

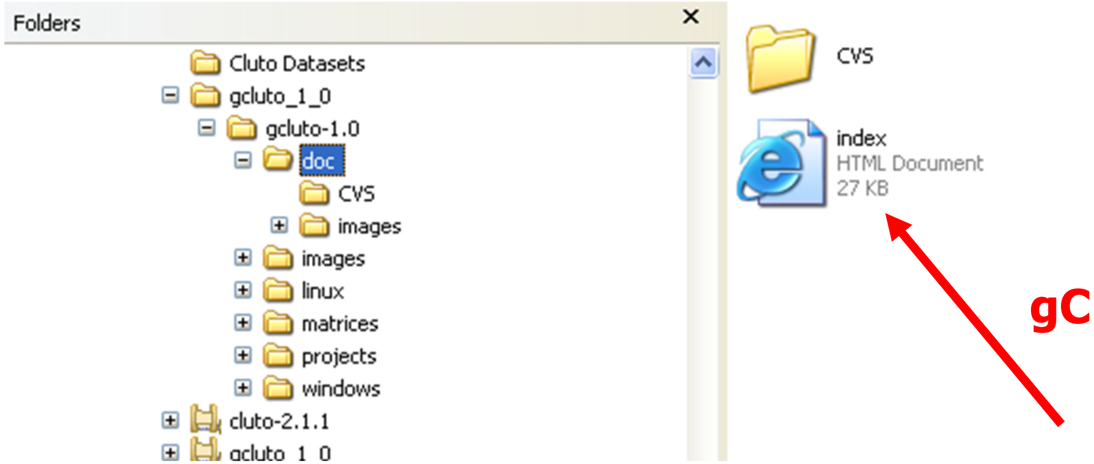
The lab has open postdoctoral positions on research projects related to chemical informatics and data mining.

Download and Unzip “gcluto_1_0.zip”

gCLUTO Folders and Files (extracted)



gCLUTO “doc” Subfolder and Files



gCLUTO Documentation

gCLUTO Documentation

Version 1.0

Matt Rasmussen, Mark Newman, George Karypis
University of Minnesota. Copyright 2003
Last Modified: Wed Nov 19 15:16:53 CST 2003
http://www.cs.umn.edu/~mrasmus/gcluto

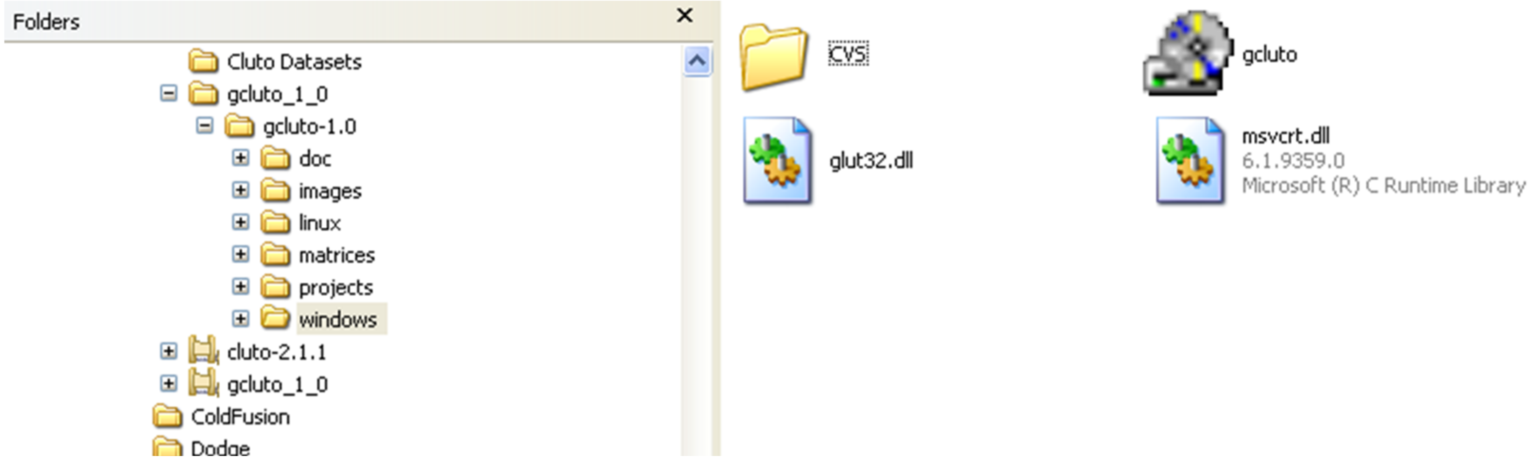
Table of Contents

- 1 Introduction
    - 1.1 What is gCLUTO
    - 1.2 Features

For more information on CLUTO (original non-graphical version), download and Unzip "cluto-2.1.1.zip" at:

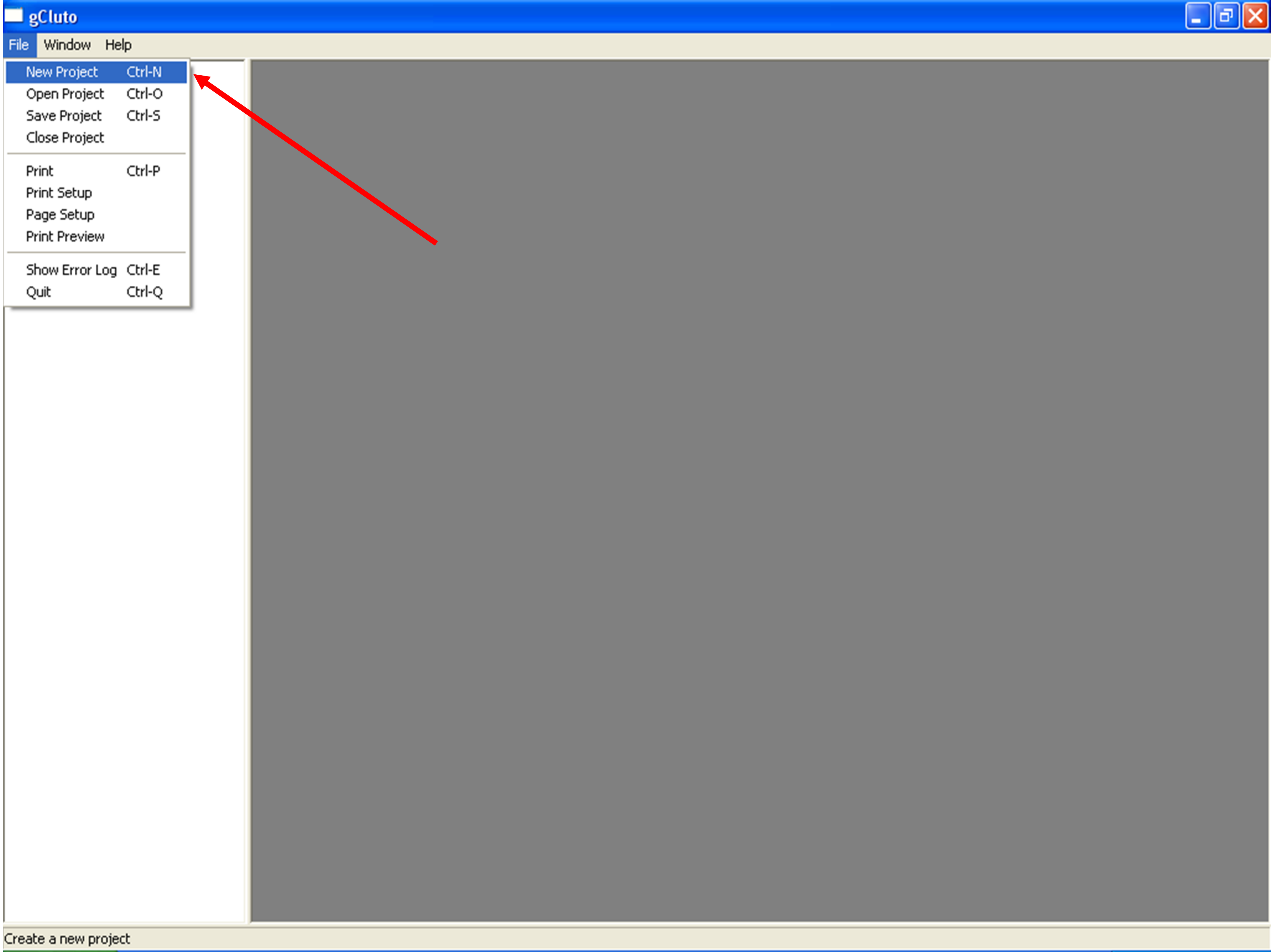
http://glaros.dtc.umn.edu/gkhome/cluto/cluto/download

gCLUTO “windows” Subfolder and Files

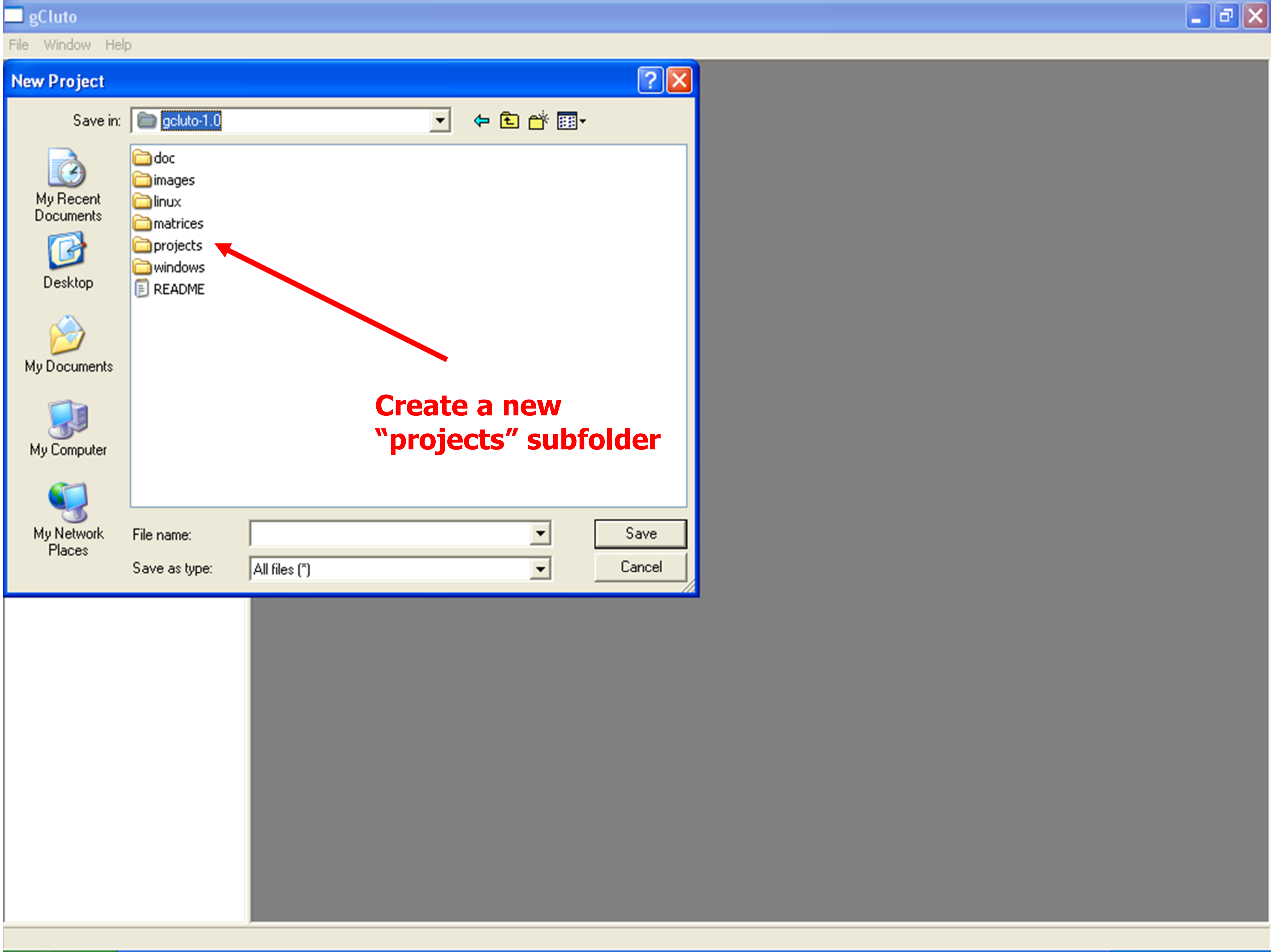


Executable file;
Double-click to run the program

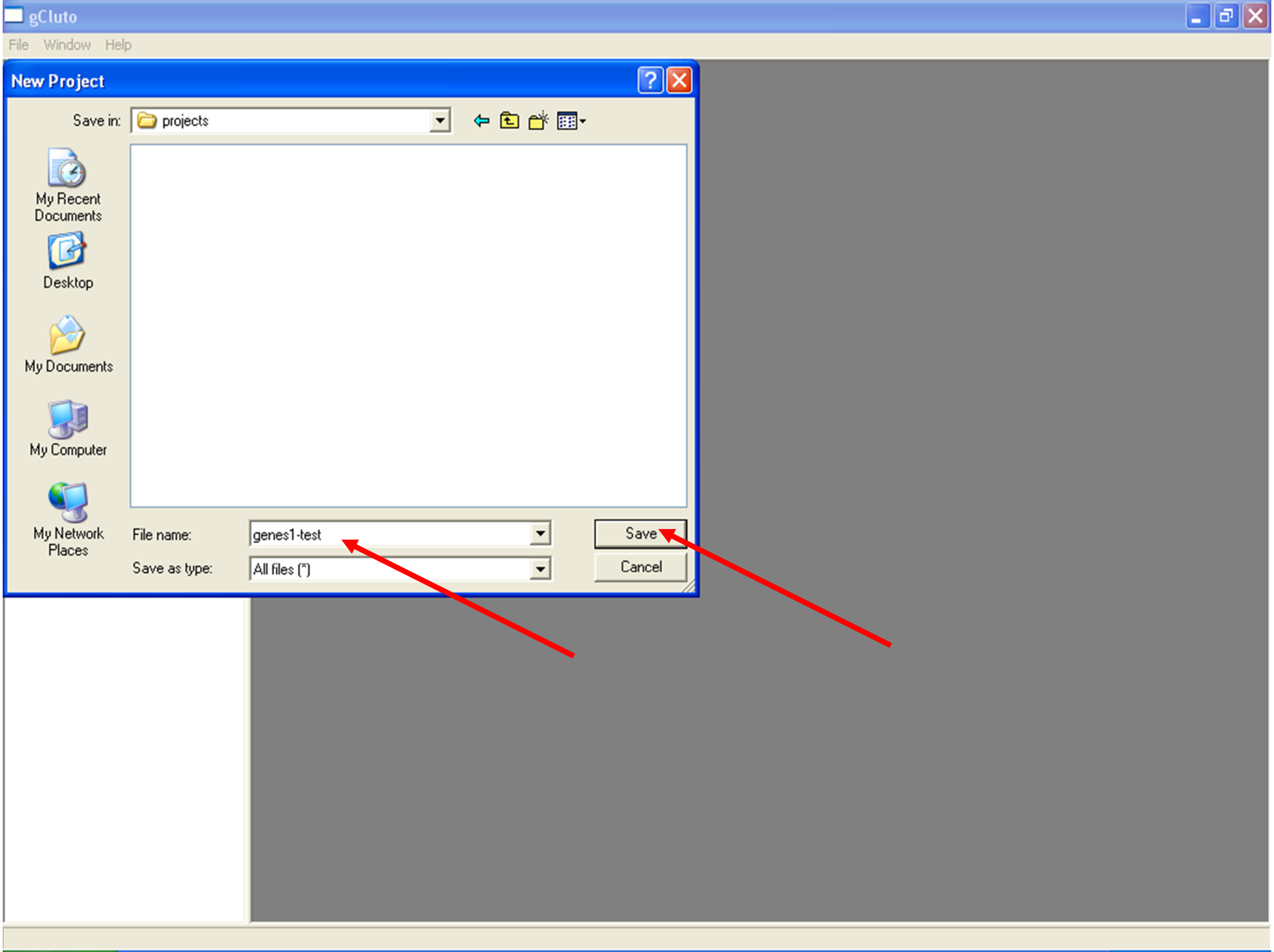

gCluto
File
Window
Help



Create a new project



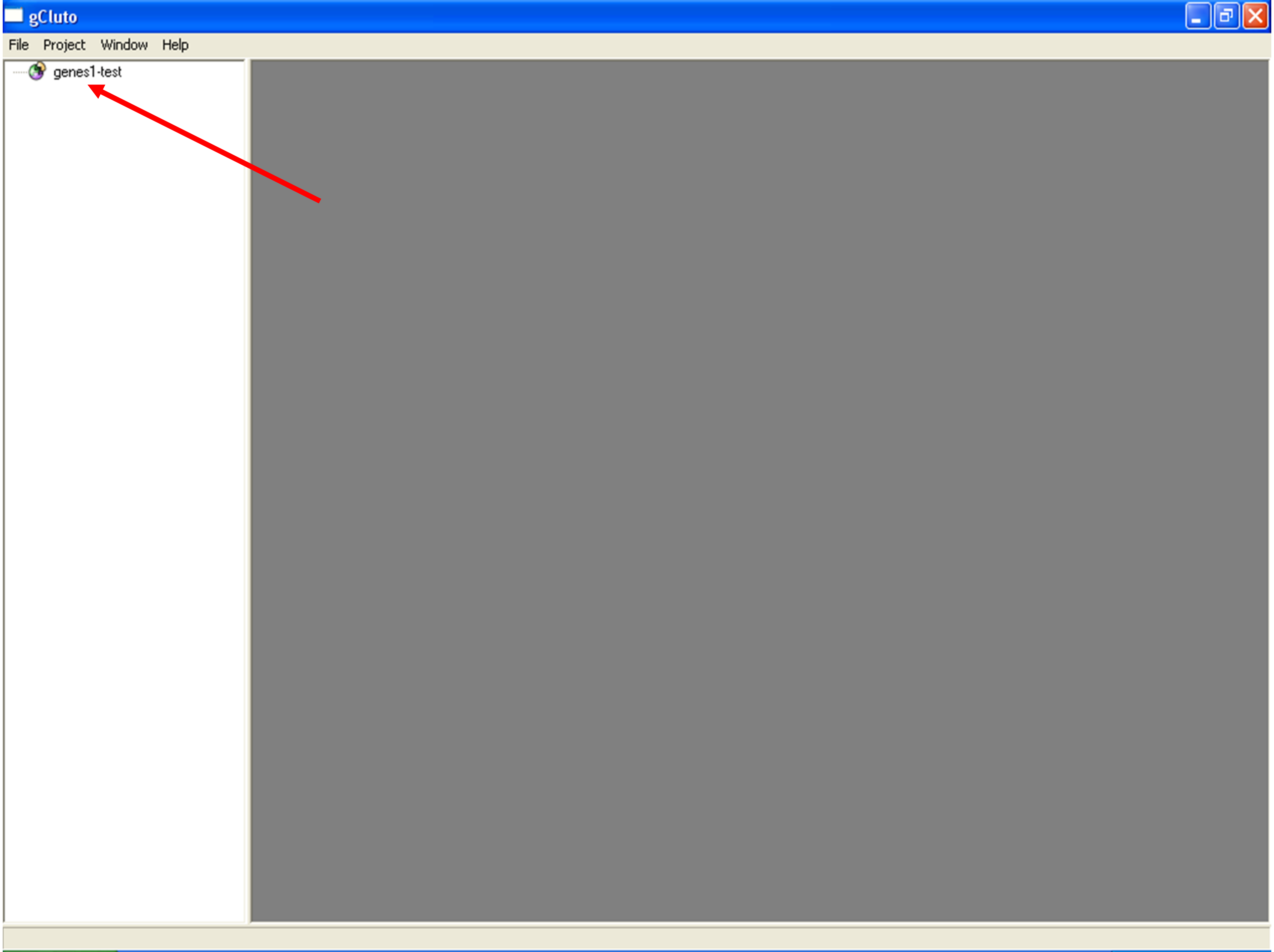
Create a new “projects” subfolder

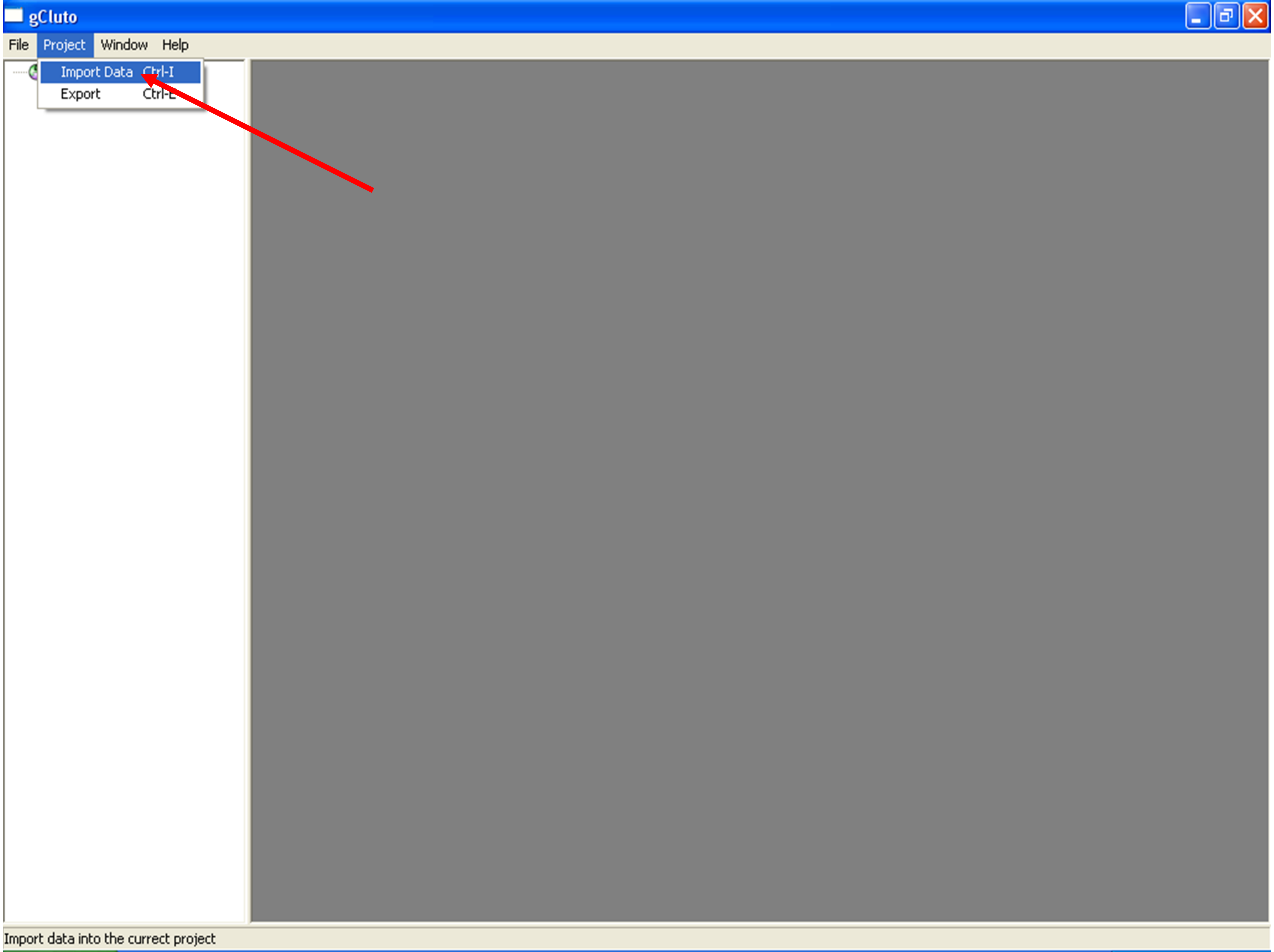


Save in: projects

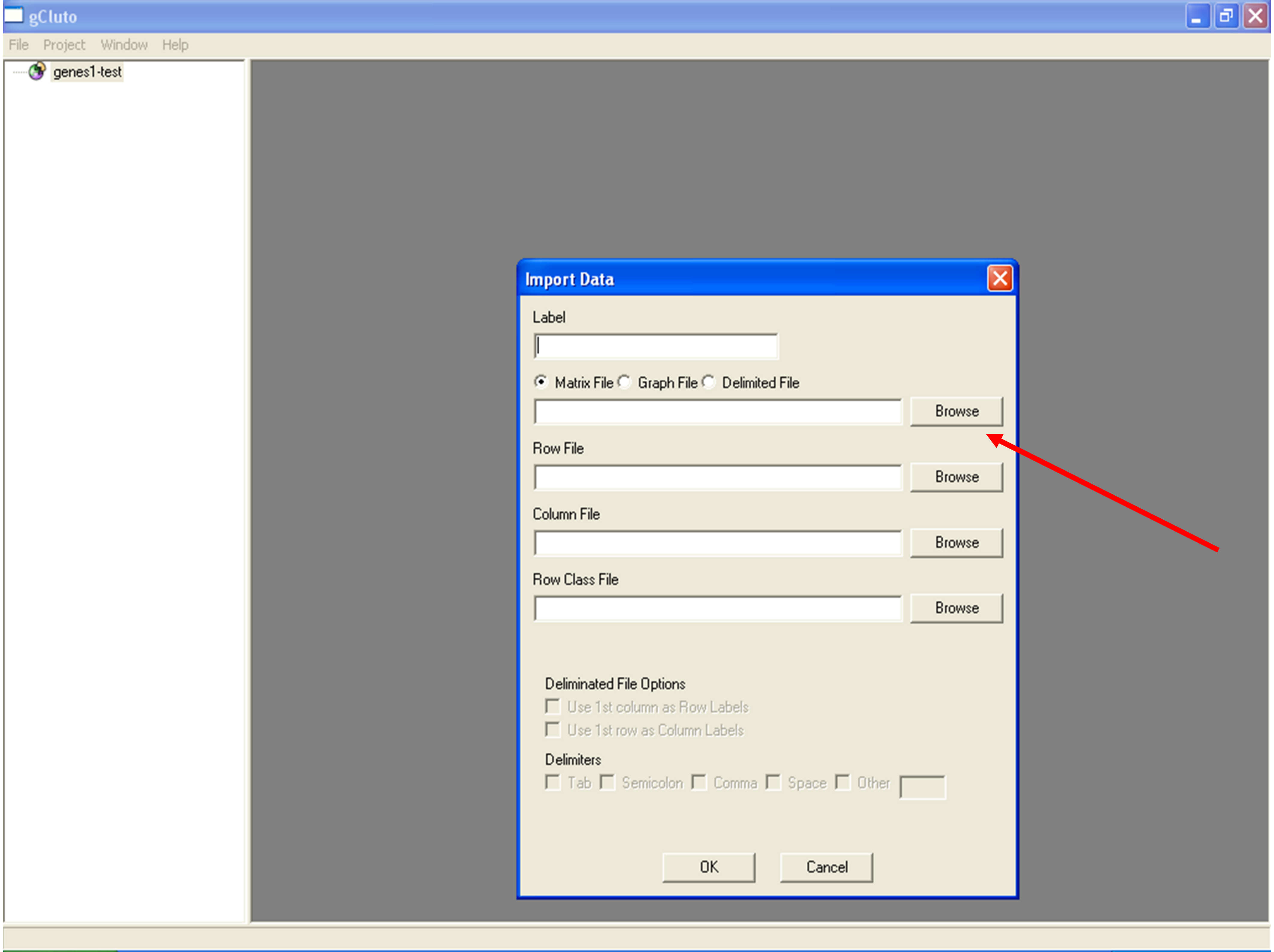
File name: genes1-test

Save as type: All files (*)

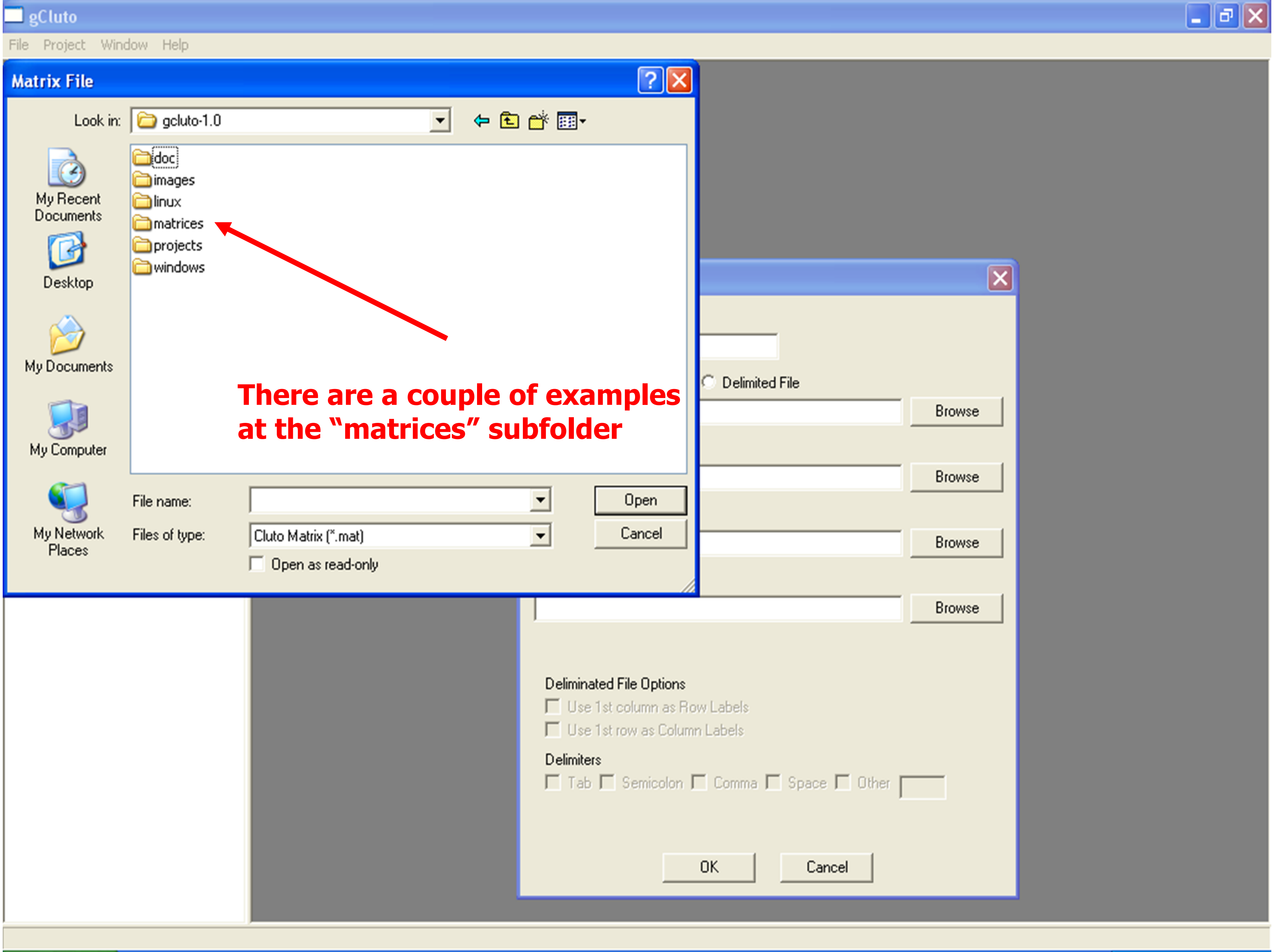

gCluto
File
Project
Window
Help
genes1-test



Import data into the currect project



● Matrix File ○ Graph File ○ Delimited File

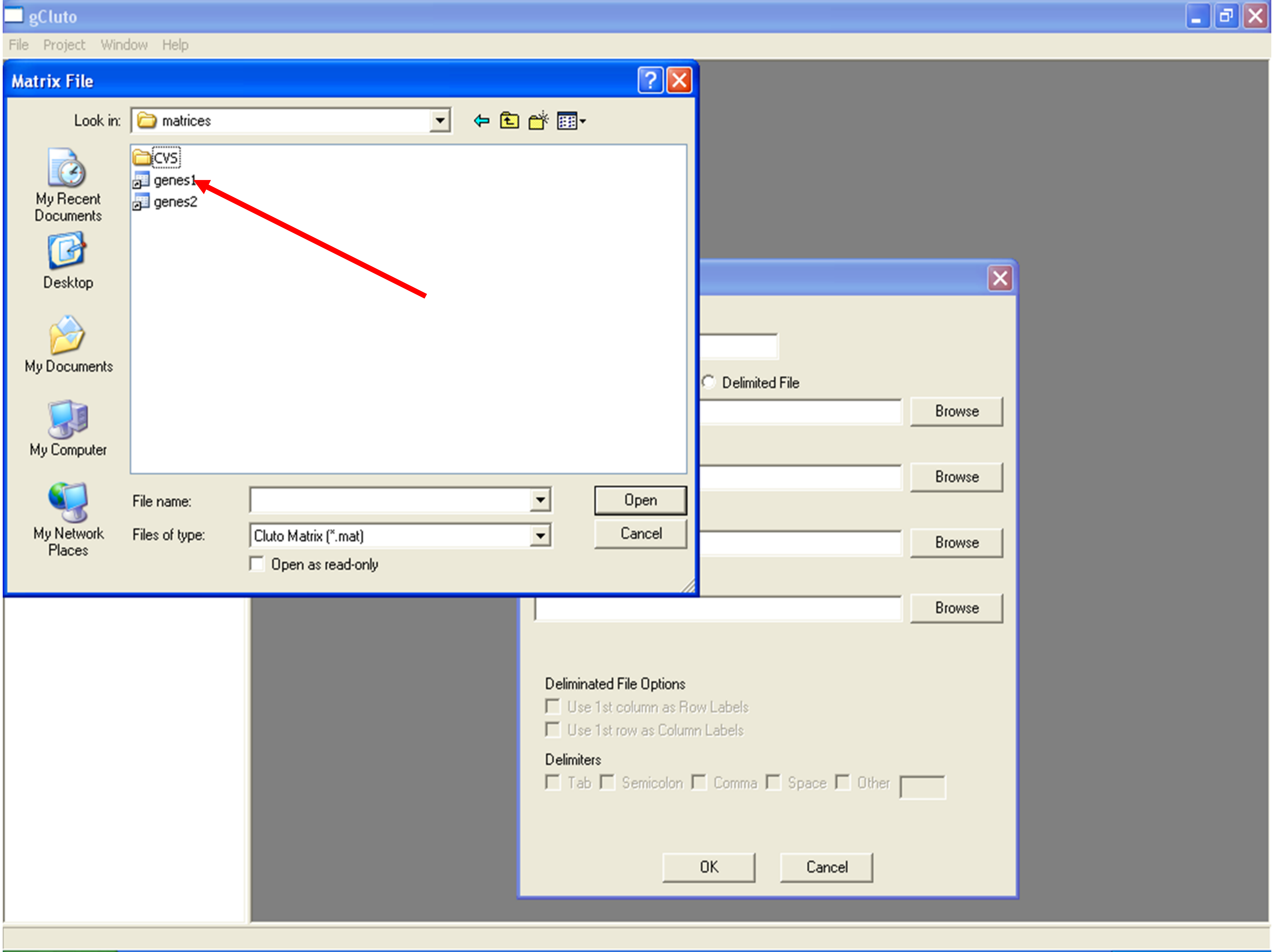


Look in: gcluto-1.0

File name:

Files of type: Cluto Matrix (*.mat)

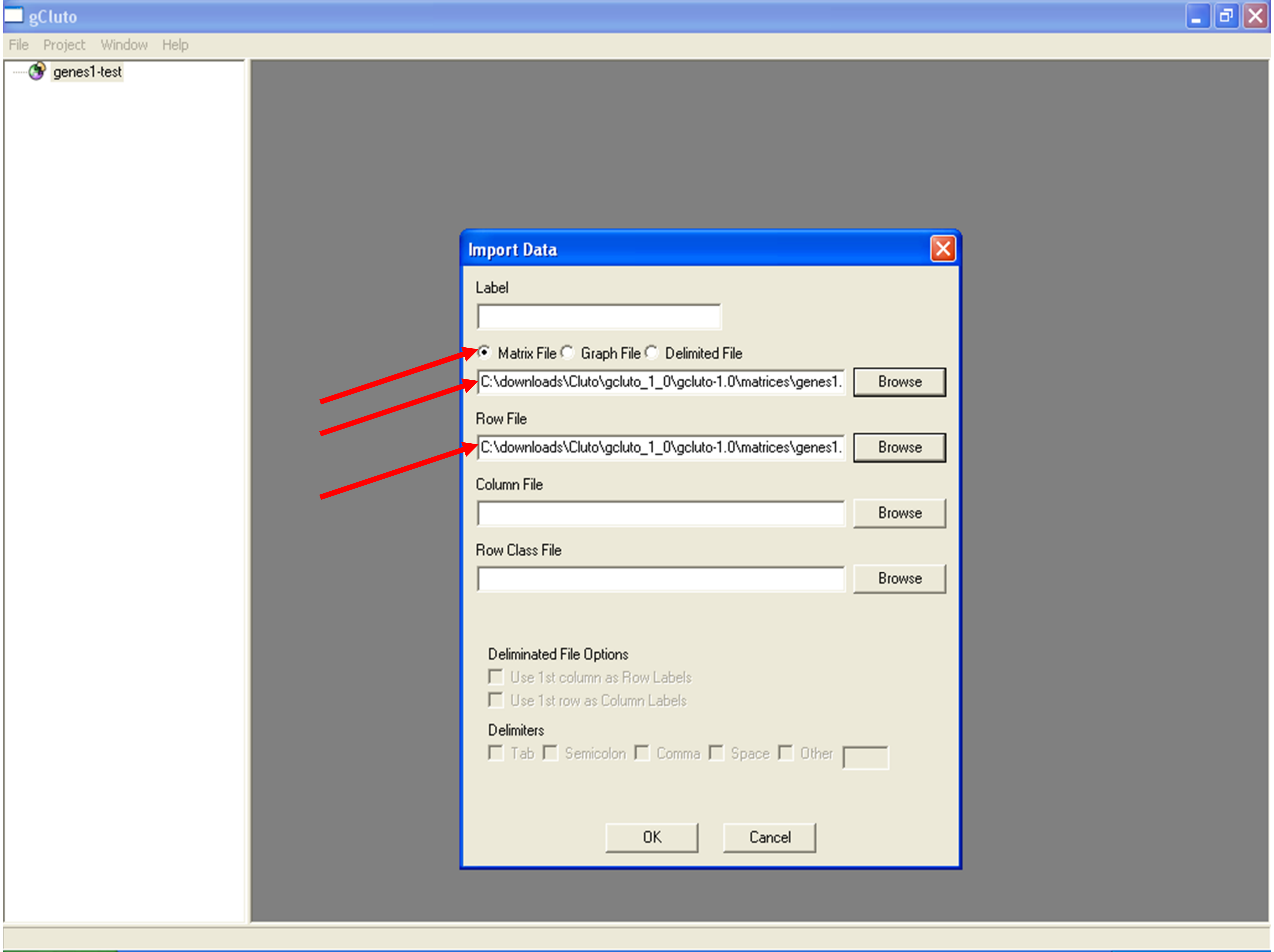
There are a couple of examples at the "matrices" subfolder



Look in: matrices

Files of type: Cluto Matrix (*.mat)

☐ Open as read-only



Label

● Matrix File ○ Graph File ○ Delimited File

C:\downloads\Cluto\gcluto_1_0\gcluto-1.0\matrices\genes1.

Row File

C:\downloads\Cluto\gcluto_1_0\gcluto-1.0\matrices\genes1.

Column File

Row Class File

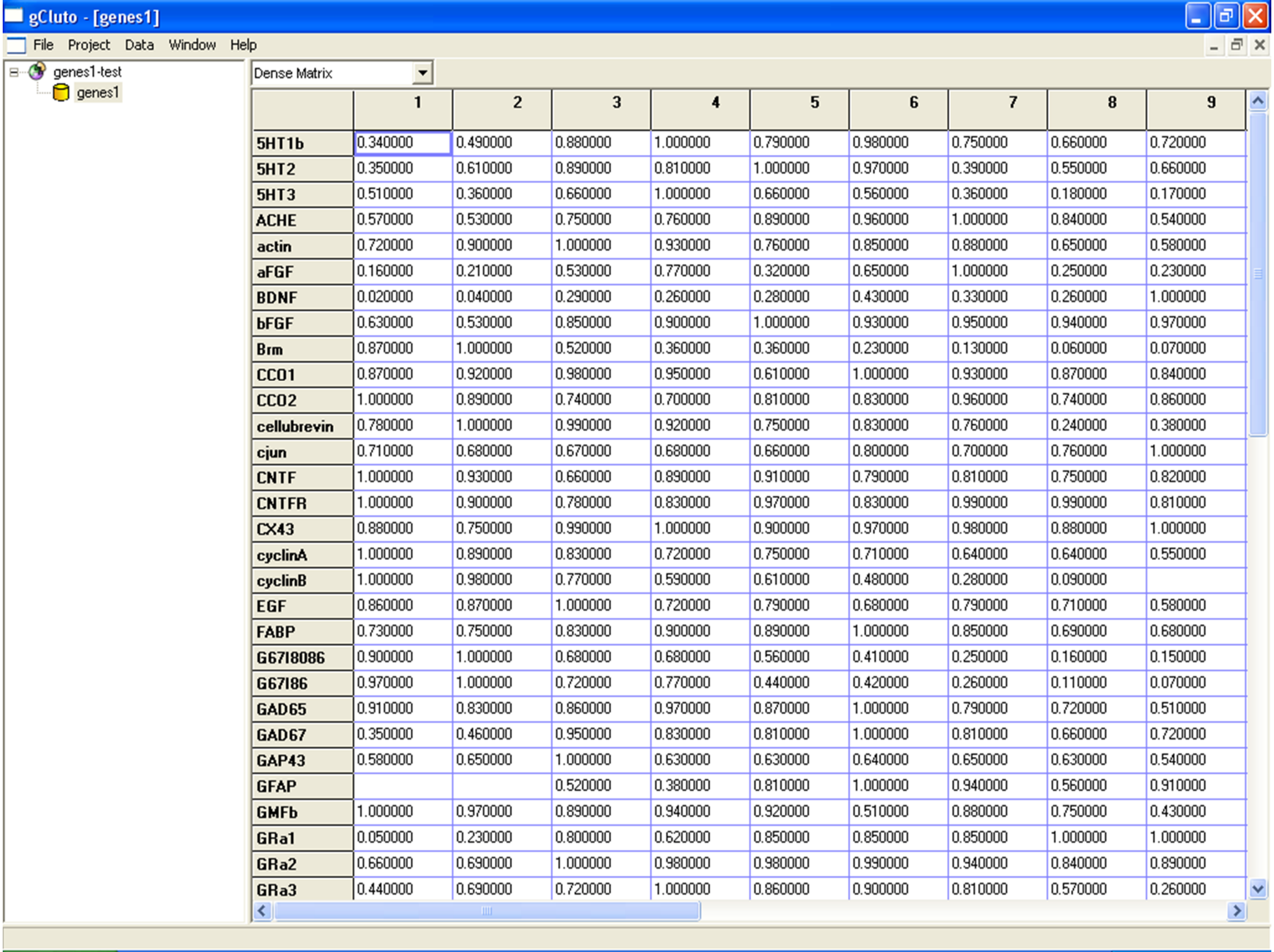
Deliminated File Options

☐ Use 1st column as Row Labels

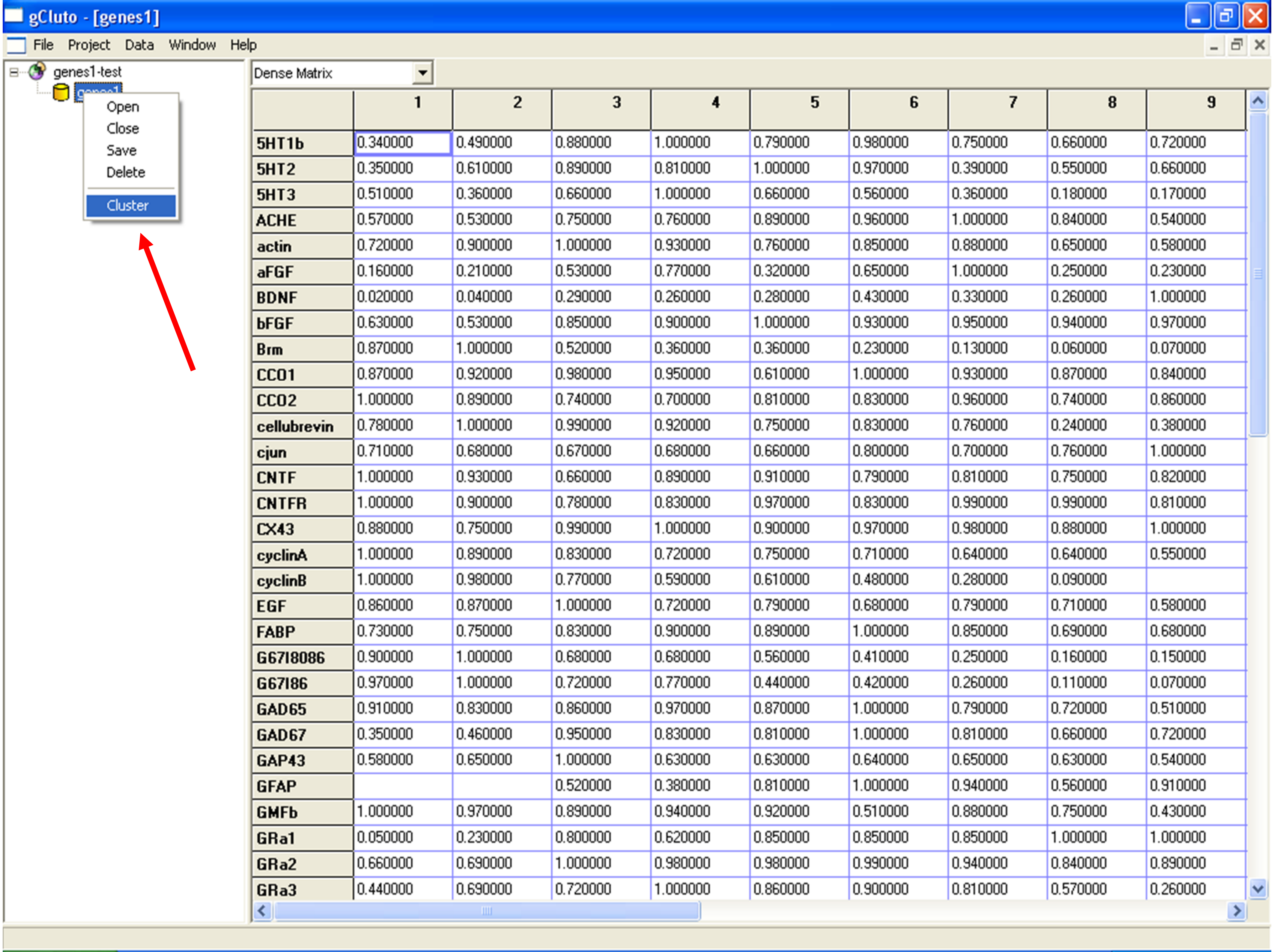
☐ Use 1st row as Column Labels

Delimiters

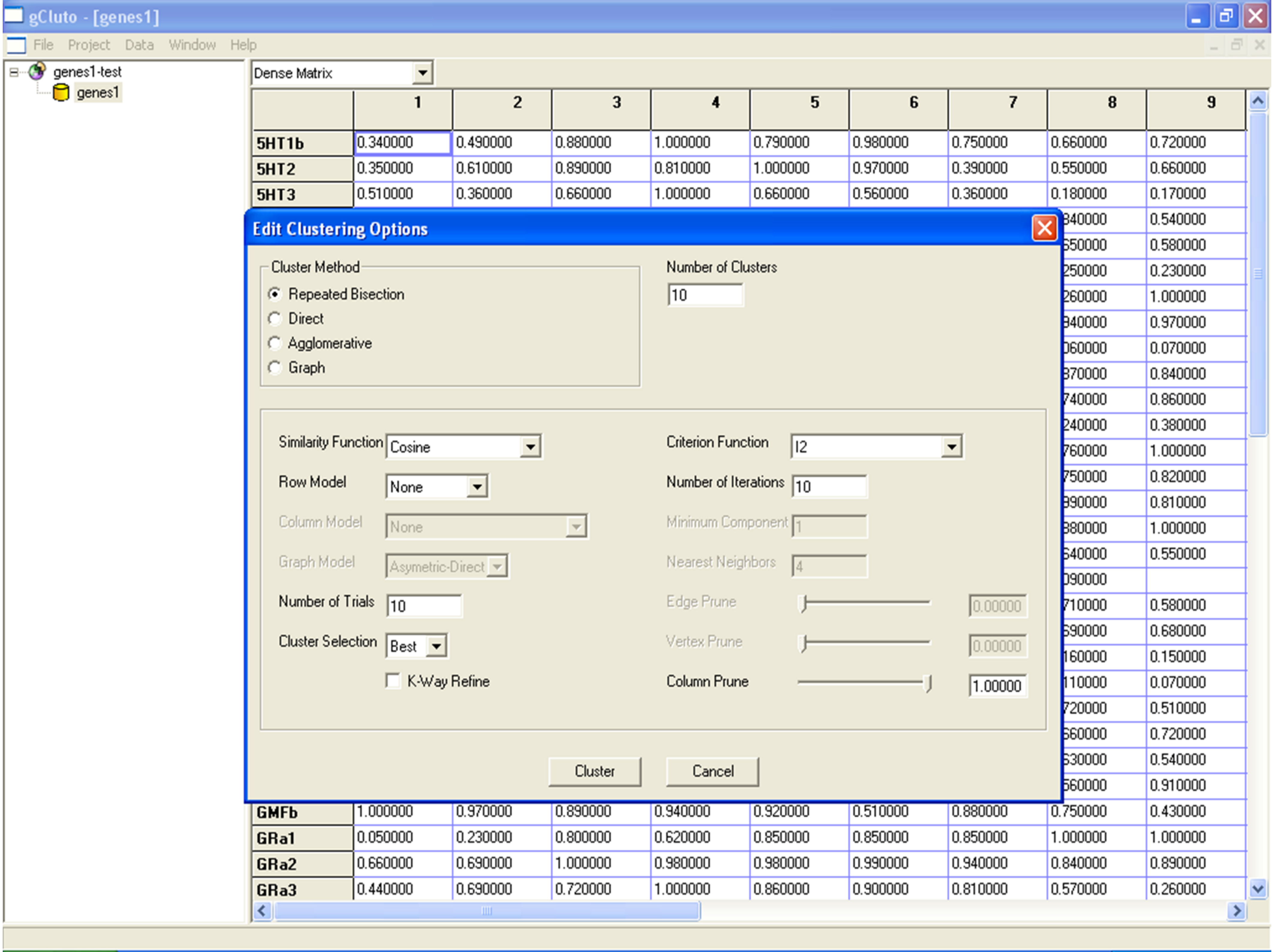
☐ Tab ☐ Semicolon ☐ Comma ☐ Space ☐ Other



	1	2	3	4	5	6	7	8	9
5HT1b	0.340000	0.490000	0.880000	1.000000	0.790000	0.980000	0.750000	0.660000	0.720000
5HT2	0.350000	0.610000	0.890000	0.810000	1.000000	0.970000	0.390000	0.550000	0.660000
5HT3	0.510000	0.360000	0.660000	1.000000	0.660000	0.560000	0.360000	0.180000	0.170000
ACHE	0.570000	0.530000	0.750000	0.760000	0.890000	0.960000	1.000000	0.840000	0.540000
actin	0.720000	0.900000	1.000000	0.930000	0.760000	0.850000	0.880000	0.650000	0.580000
aFGF	0.160000	0.210000	0.530000	0.770000	0.320000	0.650000	1.000000	0.250000	0.230000
BDNF	0.020000	0.040000	0.290000	0.260000	0.280000	0.430000	0.330000	0.260000	1.000000
bFGF	0.630000	0.530000	0.850000	0.900000	1.000000	0.930000	0.950000	0.940000	0.970000
Brm	0.870000	1.000000	0.520000	0.360000	0.360000	0.230000	0.130000	0.060000	0.070000
CCO1	0.870000	0.920000	0.980000	0.950000	0.610000	1.000000	0.930000	0.870000	0.840000
CCO2	1.000000	0.890000	0.740000	0.700000	0.810000	0.830000	0.960000	0.740000	0.860000
cellubrevin	0.780000	1.000000	0.990000	0.920000	0.750000	0.830000	0.760000	0.240000	0.380000
cjun	0.710000	0.680000	0.670000	0.680000	0.660000	0.800000	0.700000	0.760000	1.000000
CNTF	1.000000	0.930000	0.660000	0.890000	0.910000	0.790000	0.810000	0.750000	0.820000
CNTFR	1.000000	0.900000	0.780000	0.830000	0.970000	0.830000	0.990000	0.990000	0.810000
CX43	0.880000	0.750000	0.990000	1.000000	0.900000	0.970000	0.980000	0.880000	1.000000
cyclinA	1.000000	0.890000	0.830000	0.720000	0.750000	0.710000	0.640000	0.640000	0.550000
cyclinB	1.000000	0.980000	0.770000	0.590000	0.610000	0.480000	0.280000	0.090000	
EGF	0.860000	0.870000	1.000000	0.720000	0.790000	0.680000	0.790000	0.710000	0.580000
FABP	0.730000	0.750000	0.830000	0.900000	0.890000	1.000000	0.850000	0.690000	0.680000
G6718086	0.900000	1.000000	0.680000	0.680000	0.560000	0.410000	0.250000	0.160000	0.150000
G67186	0.970000	1.000000	0.720000	0.770000	0.440000	0.420000	0.260000	0.110000	0.070000
GAD65	0.910000	0.830000	0.860000	0.970000	0.870000	1.000000	0.790000	0.720000	0.510000
GAD67	0.350000	0.460000	0.950000	0.830000	0.810000	1.000000	0.810000	0.660000	0.720000
GAP43	0.580000	0.650000	1.000000	0.630000	0.630000	0.640000	0.650000	0.630000	0.540000
GFAP			0.520000	0.380000	0.810000	1.000000	0.940000	0.560000	0.910000
GMFb	1.000000	0.970000	0.890000	0.940000	0.920000	0.510000	0.880000	0.750000	0.430000
GRa1	0.050000	0.230000	0.800000	0.620000	0.850000	0.850000	0.850000	1.000000	1.000000
GRa2	0.660000	0.690000	1.000000	0.980000	0.980000	0.990000	0.940000	0.840000	0.890000
GRa3	0.440000	0.690000	0.720000	1.000000	0.860000	0.900000	0.810000	0.570000	0.260000



	1	2	3	4	5	6	7	8	9
5HT1b	0.340000	0.490000	0.880000	1.000000	0.790000	0.980000	0.750000	0.660000	0.720000
5HT2	0.350000	0.610000	0.890000	0.810000	1.000000	0.970000	0.390000	0.550000	0.660000
5HT3	0.510000	0.360000	0.660000	1.000000	0.660000	0.560000	0.360000	0.180000	0.170000
ACHE	0.570000	0.530000	0.750000	0.760000	0.890000	0.960000	1.000000	0.840000	0.540000
actin	0.720000	0.900000	1.000000	0.930000	0.760000	0.850000	0.880000	0.650000	0.580000
aFGF	0.160000	0.210000	0.530000	0.770000	0.320000	0.650000	1.000000	0.250000	0.230000
BDNF	0.020000	0.040000	0.290000	0.260000	0.280000	0.430000	0.330000	0.260000	1.000000
bFGF	0.630000	0.530000	0.850000	0.900000	1.000000	0.930000	0.950000	0.940000	0.970000
Brm	0.870000	1.000000	0.520000	0.360000	0.360000	0.230000	0.130000	0.060000	0.070000
CCO1	0.870000	0.920000	0.980000	0.950000	0.610000	1.000000	0.930000	0.870000	0.840000
CCO2	1.000000	0.890000	0.740000	0.700000	0.810000	0.830000	0.960000	0.740000	0.860000
cellubrevin	0.780000	1.000000	0.990000	0.920000	0.750000	0.830000	0.760000	0.240000	0.380000
cjun	0.710000	0.680000	0.670000	0.680000	0.660000	0.800000	0.700000	0.760000	1.000000
CNTF	1.000000	0.930000	0.660000	0.890000	0.910000	0.790000	0.810000	0.750000	0.820000
CNTFR	1.000000	0.900000	0.780000	0.830000	0.970000	0.830000	0.990000	0.990000	0.810000
CX43	0.880000	0.750000	0.990000	1.000000	0.900000	0.970000	0.980000	0.880000	1.000000
cyclinA	1.000000	0.890000	0.830000	0.720000	0.750000	0.710000	0.640000	0.640000	0.550000
cyclinB	1.000000	0.980000	0.770000	0.590000	0.610000	0.480000	0.280000	0.090000	
EGF	0.860000	0.870000	1.000000	0.720000	0.790000	0.680000	0.790000	0.710000	0.580000
FABP	0.730000	0.750000	0.830000	0.900000	0.890000	1.000000	0.850000	0.690000	0.680000
G67I8086	0.900000	1.000000	0.680000	0.680000	0.560000	0.410000	0.250000	0.160000	0.150000
G67I86	0.970000	1.000000	0.720000	0.770000	0.440000	0.420000	0.260000	0.110000	0.070000
GAD65	0.910000	0.830000	0.860000	0.970000	0.870000	1.000000	0.790000	0.720000	0.510000
GAD67	0.350000	0.460000	0.950000	0.830000	0.810000	1.000000	0.810000	0.660000	0.720000
GAP43	0.580000	0.650000	1.000000	0.630000	0.630000	0.640000	0.650000	0.630000	0.540000
GFAP			0.520000	0.380000	0.810000	1.000000	0.940000	0.560000	0.910000
GMFb	1.000000	0.970000	0.890000	0.940000	0.920000	0.510000	0.880000	0.750000	0.430000
GRa1	0.050000	0.230000	0.800000	0.620000	0.850000	0.850000	0.850000	1.000000	1.000000
GRa2	0.660000	0.690000	1.000000	0.980000	0.980000	0.990000	0.940000	0.840000	0.890000
GRa3	0.440000	0.690000	0.720000	1.000000	0.860000	0.900000	0.810000	0.570000	0.260000



	1	2	3	4	5	6	7	8	9
5HT1b	0.340000	0.490000	0.880000	1.000000	0.790000	0.980000	0.750000	0.660000	0.720000
5HT2	0.350000	0.610000	0.890000	0.810000	1.000000	0.970000	0.390000	0.550000	0.660000
5HT3	0.510000	0.360000	0.660000	1.000000	0.660000	0.560000	0.360000	0.180000	0.170000
								340000	0.540000
								650000	0.580000
								250000	0.230000
								260000	1.000000
								340000	0.970000
								060000	0.070000
								370000	0.840000
								740000	0.860000
								240000	0.380000
								760000	1.000000
								750000	0.820000
								390000	0.810000
								380000	1.000000
								640000	0.550000
								090000	
								710000	0.580000
								690000	0.680000
								160000	0.150000
								110000	0.070000
								720000	0.510000
								660000	0.720000
								630000	0.540000
								560000	0.910000
GMFb	1.000000	0.970000	0.890000	0.940000	0.920000	0.510000	0.880000	0.750000	0.430000
GRa1	0.050000	0.230000	0.800000	0.620000	0.850000	0.850000	0.850000	1.000000	1.000000
GRa2	0.660000	0.690000	1.000000	0.980000	0.980000	0.990000	0.940000	0.840000	0.890000
GRa3	0.440000	0.690000	0.720000	1.000000	0.860000	0.900000	0.810000	0.570000	0.260000

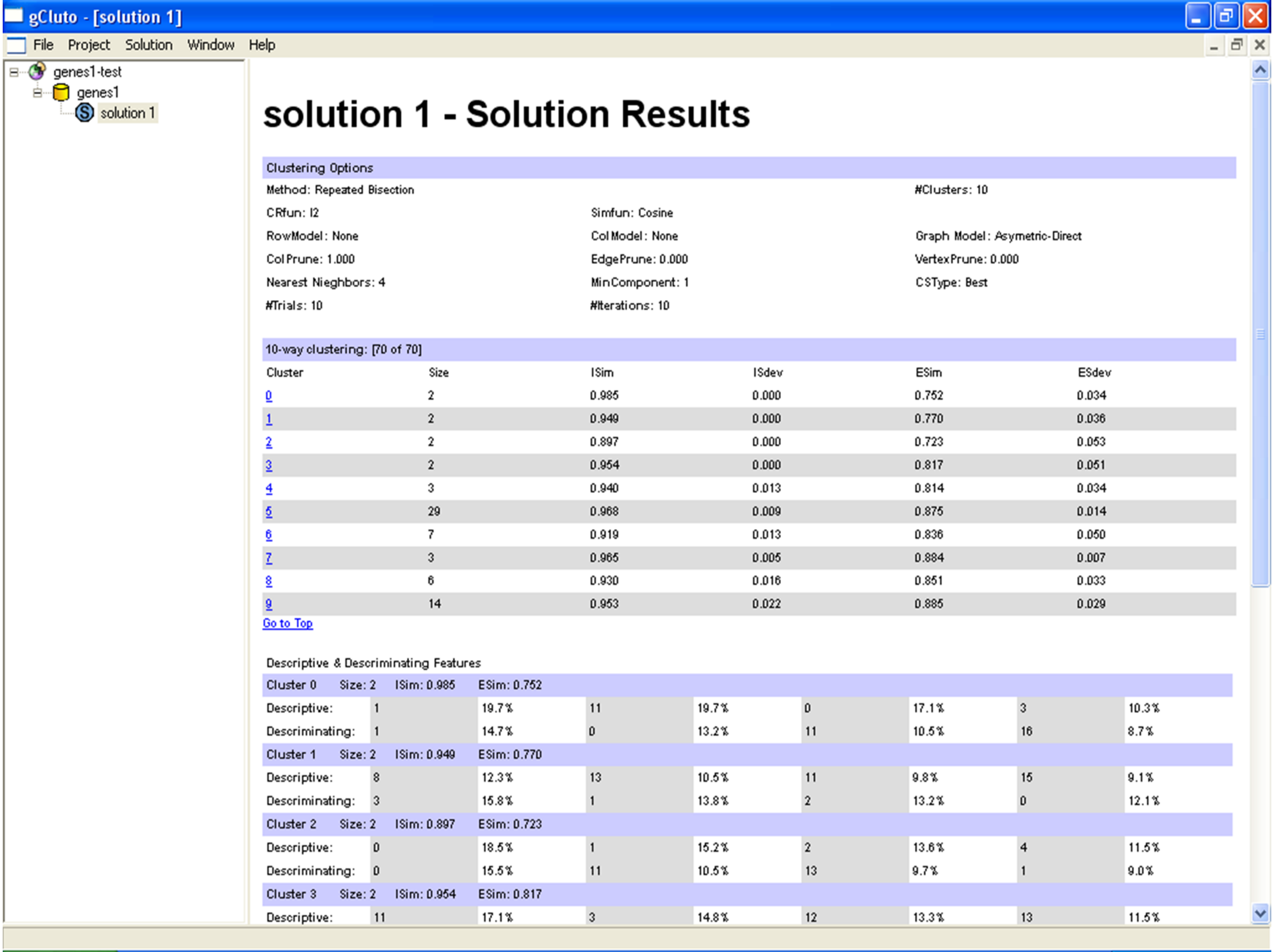
Edit Clustering Options

Cluster Method
- ● Repeated Bisection
- ○ Direct
- ○ Agglomerative
- ○ Graph

Number of Clusters: 10

Similarity Function: Cosine
Row Model: None
Column Model: None
Graph Model: Asymetric-Direct
Number of Trials: 10
Cluster Selection: Best
☐ K-Way Refine

Criterion Function: I2
Number of Iterations: 10
Minimum Component: 1
Nearest Neighbors: 4
Edge Prune: 0.00000
Vertex Prune: 0.00000
Column Prune: 1.00000



solution 1 - Solution Results

Clustering Options

Method: Repeated Bisection
#Clusters: 10
CRfun: I2
Simfun: Cosine
RowModel: None
ColModel: None
Graph Model: Asymetric-Direct
ColPrune: 1.000
EdgePrune: 0.000
VertexPrune: 0.000
Nearest Nieghbors: 4
MinComponent: 1
CSType: Best
#Trials: 10
#Iterations: 10

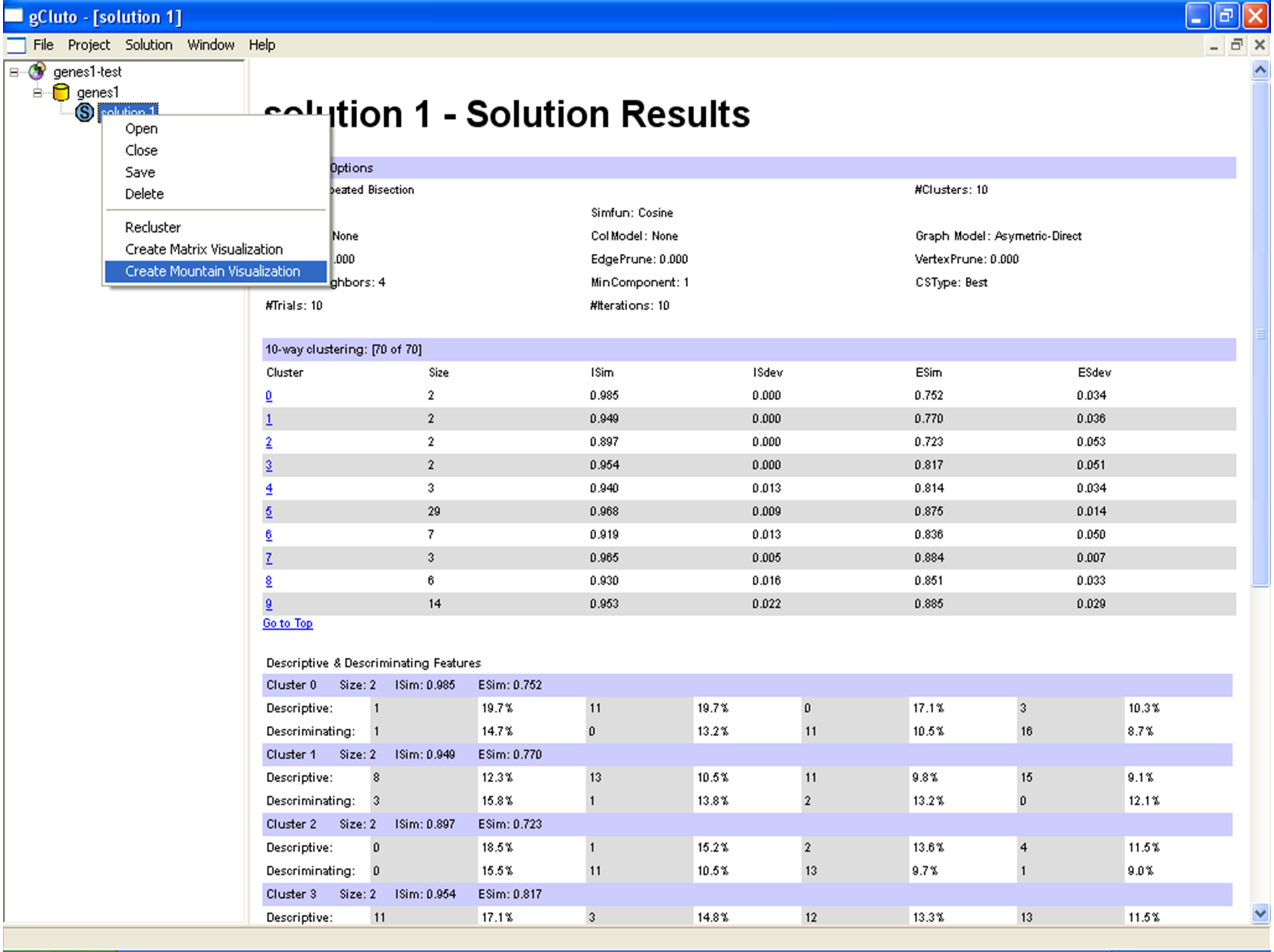
10-way clustering: [70 of 70]

Cluster	Size	ISim	ISdev	ESim	ESdev
0	2	0.985	0.000	0.752	0.034
1	2	0.949	0.000	0.770	0.036
2	2	0.897	0.000	0.723	0.053
3	2	0.954	0.000	0.817	0.051
4	3	0.940	0.013	0.814	0.034
5	29	0.968	0.009	0.875	0.014
6	7	0.919	0.013	0.836	0.050
7	3	0.965	0.005	0.884	0.007
8	6	0.930	0.016	0.851	0.033
9	14	0.953	0.022	0.885	0.029

Go to Top

Descriptive & Descriminating Features

Cluster 0	Size: 2	ISim: 0.985	ESim: 0.752					
Descriptive:	1	19.7%	11	19.7%	0	17.1%	3	10.3%
Descriminating:	1	14.7%	0	13.2%	11	10.5%	16	8.7%
Cluster 1	Size: 2	ISim: 0.949	ESim: 0.770					
Descriptive:	8	12.3%	13	10.5%	11	9.8%	15	9.1%
Descriminating:	3	15.8%	1	13.8%	2	13.2%	0	12.1%
Cluster 2	Size: 2	ISim: 0.897	ESim: 0.723					
Descriptive:	0	18.5%	1	15.2%	2	13.6%	4	11.5%
Descriminating:	0	15.5%	11	10.5%	13	9.7%	1	9.0%
Cluster 3	Size: 2	ISim: 0.954	ESim: 0.817					
Descriptive:	11	17.1%	3	14.8%	12	13.3%	13	11.5%



solution 1 - Solution Results

Options

peated Bisection

#Clusters: 10

Simfun: Cosine

None

Col Model: None

Graph Model: Asymetric-Direct

.000

EdgePrune: 0.000

VertexPrune: 0.000

ghbors: 4

MinComponent: 1

CSType: Best

#Trials: 10

#Iterations: 10

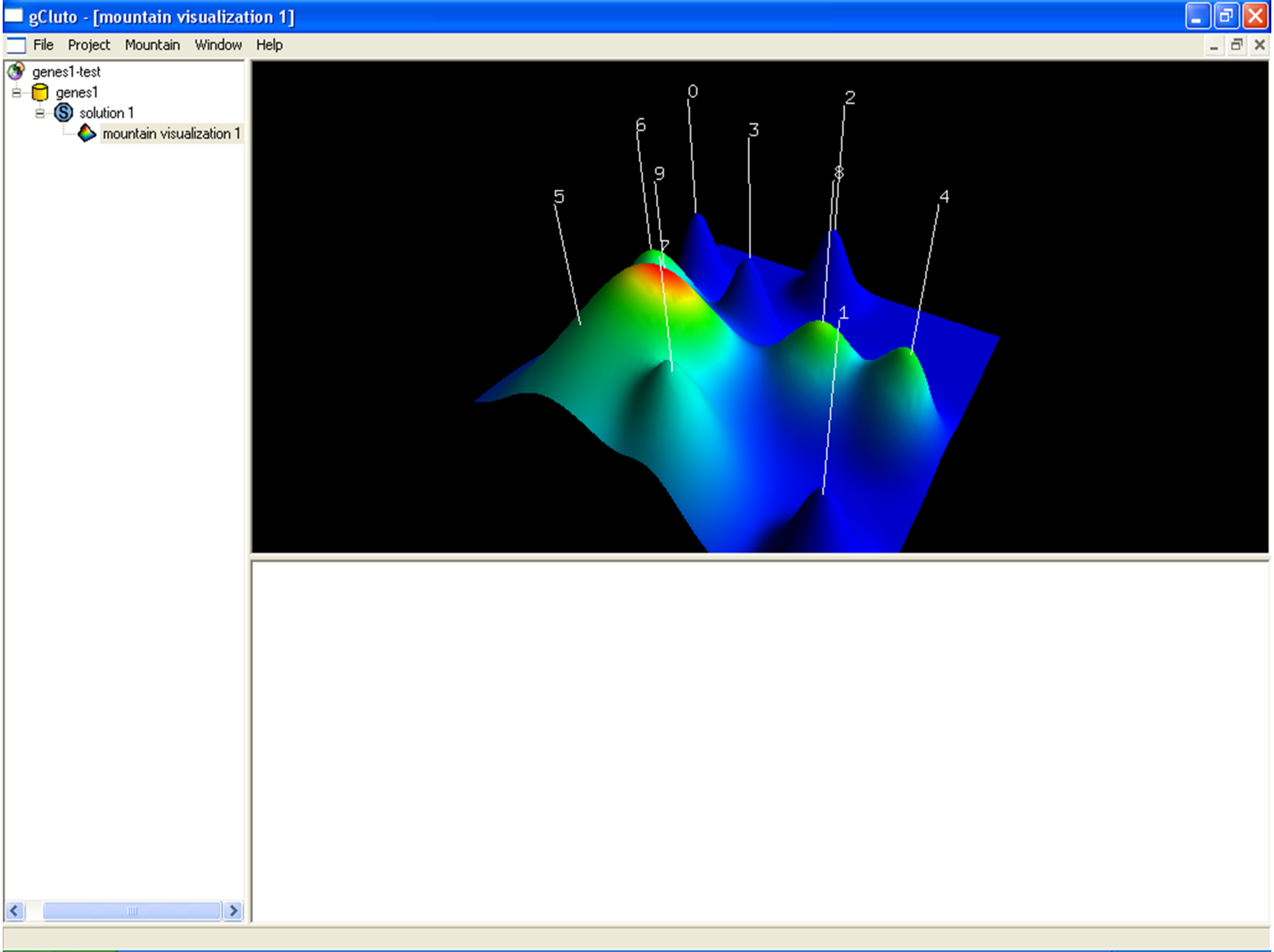
10-way clustering: [70 of 70]

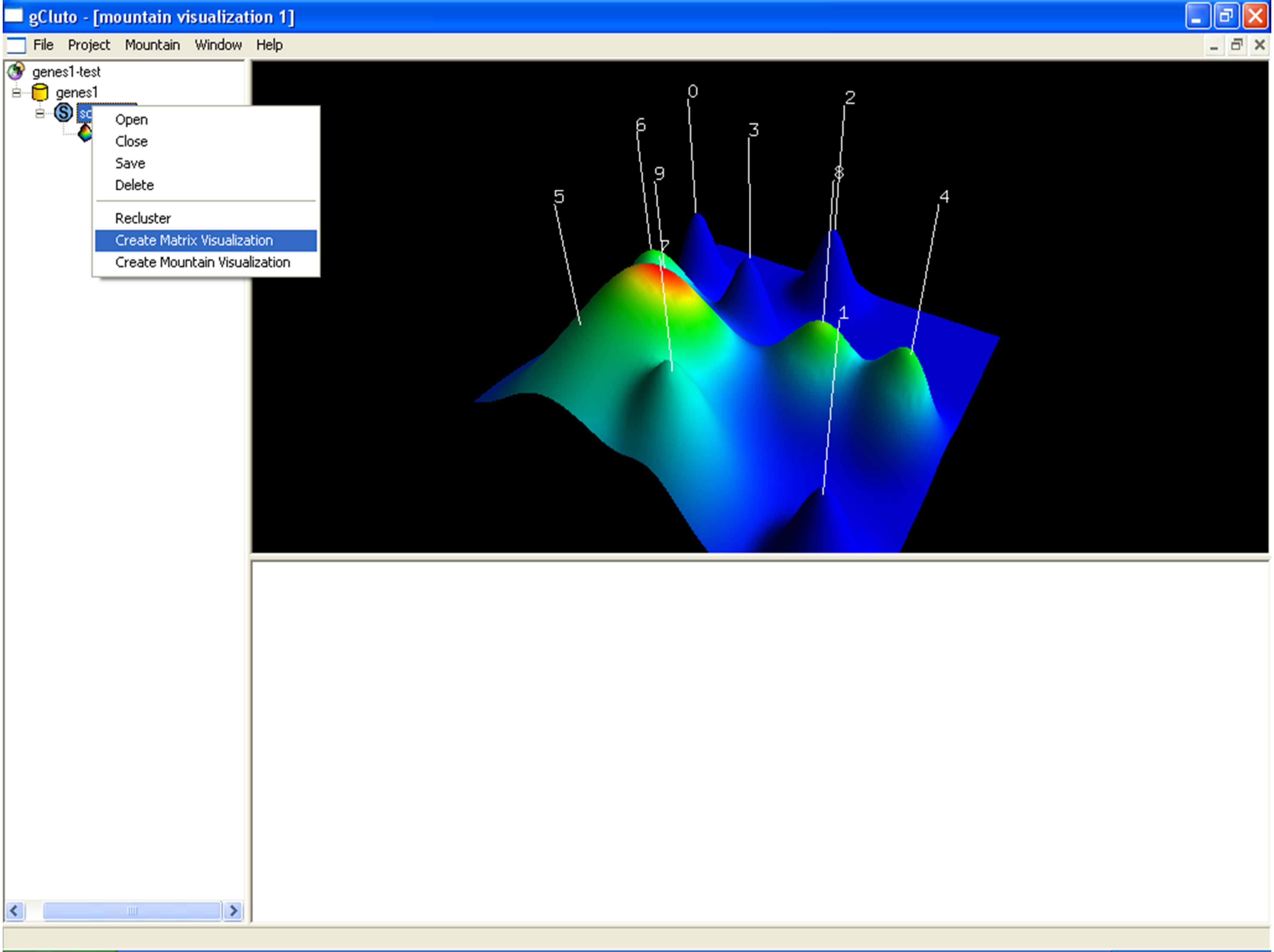
Cluster	Size	ISim	ISdev	ESim	ESdev
0	2	0.985	0.000	0.752	0.034
1	2	0.949	0.000	0.770	0.036
2	2	0.897	0.000	0.723	0.053
3	2	0.954	0.000	0.817	0.051
4	3	0.940	0.013	0.814	0.034
5	29	0.968	0.009	0.875	0.014
6	7	0.919	0.013	0.836	0.050
7	3	0.965	0.005	0.884	0.007
8	6	0.930	0.016	0.851	0.033
9	14	0.953	0.022	0.885	0.029

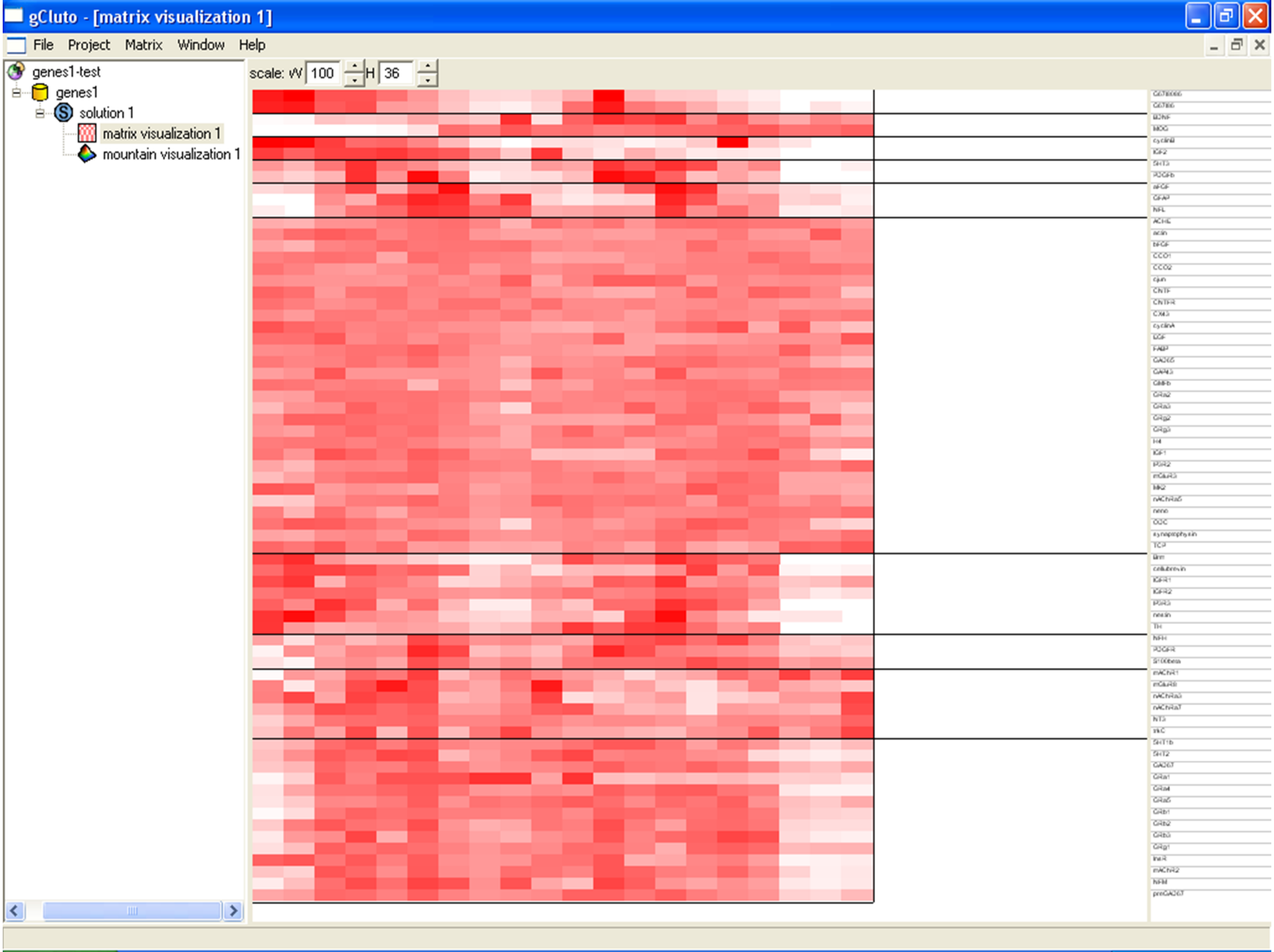
Go to Top

Descriptive & Descriminating Features

Cluster 0 Size: 2 ISim: 0.985 ESim: 0.752								
Descriptive:	1	19.7%	11	19.7%	0	17.1%	3	10.3%
Descriminating:	1	14.7%	0	13.2%	11	10.5%	16	8.7%
Cluster 1 Size: 2 ISim: 0.949 ESim: 0.770								
Descriptive:	8	12.3%	13	10.5%	11	9.8%	15	9.1%
Descriminating:	3	15.8%	1	13.8%	2	13.2%	0	12.1%
Cluster 2 Size: 2 ISim: 0.897 ESim: 0.723								
Descriptive:	0	18.5%	1	15.2%	2	13.6%	4	11.5%
Descriminating:	0	15.5%	11	10.5%	13	9.7%	1	9.0%
Cluster 3 Size: 2 ISim: 0.954 ESim: 0.817								
Descriptive:	11	17.1%	3	14.8%	12	13.3%	13	11.5%


gCluto - [mountain visualization 1]
File
Project
Mountain
Window
Help
genes1-test
genes1
solution 1
mountain visualization 1
0
2
6
3
9
8
5
4
7
1

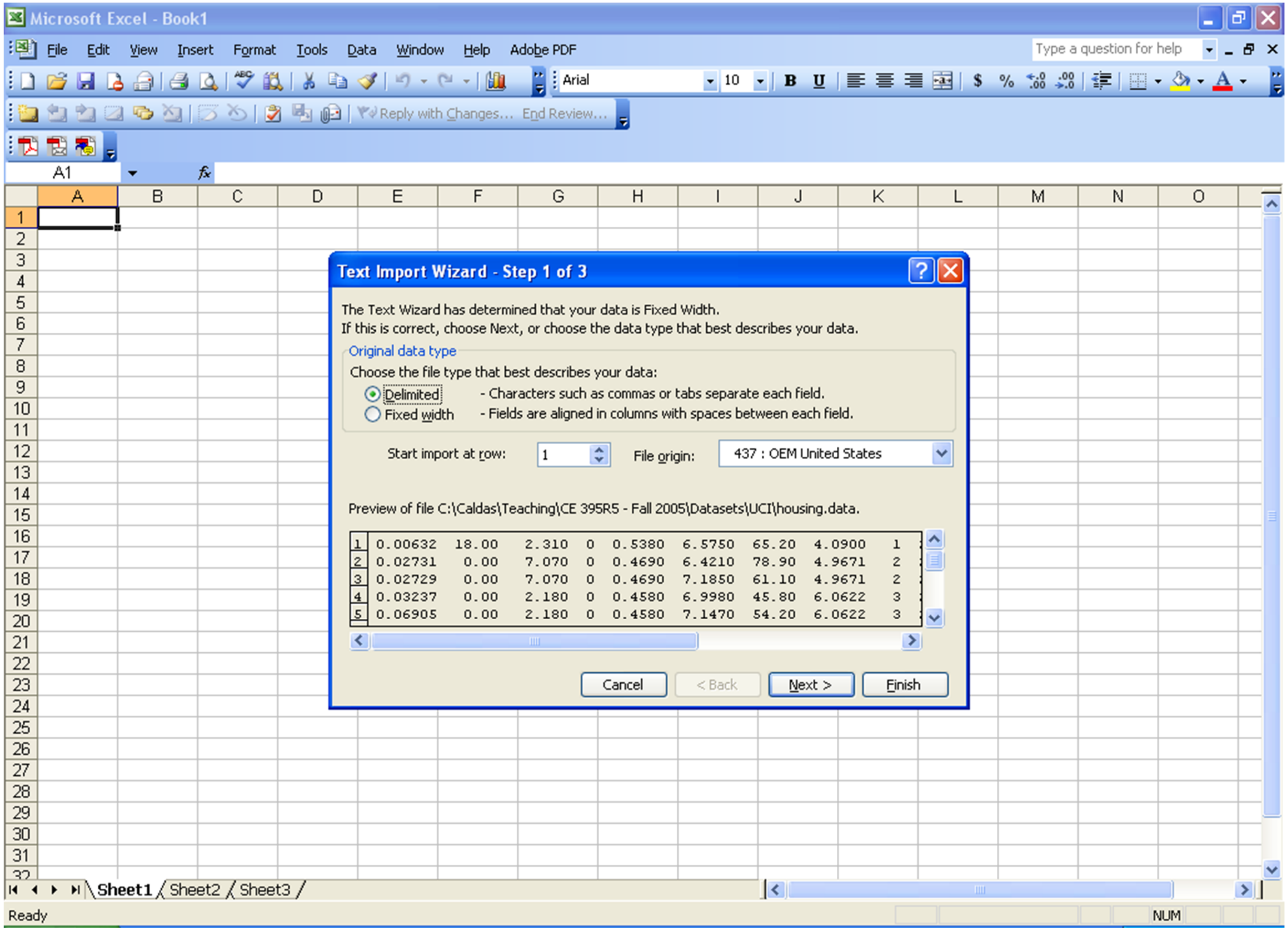

gCluto - [mountain visualization 1]
File
Project
Mountain
Window
Help
genes1-test
genes1
sc
Open
Close
Save
Delete
Recluster
Create Matrix Visualization
Create Mountain Visualization
0
2
6
3
9
8
5
4
7
1


gCluto - [matrix visualization 1]
File
Project
Matrix
Window
Help
genes1-test
genes1
solution 1
matrix visualization 1
mountain visualization 1
scale: W 100
H 36
G67I80/86
G67I86
BDNF
MOG
cyclinB
IGF2
5HT3
PDGFb
aFGF
GFAP
NFL
ACHE
actin
bFGF
CCO1
CCO2
cjun
CNTF
CNTFR
CX43
cyclinA
EGF
FABP
GAD65
GAP43
GMfb
GRa2
GRa3
GRg2
GRg3
H4
IGF1
IP3R2
mGluR3
MK2
nAChRa5
neno
ODC
synaptophysin
TCP
Brm
cellubrevin
IGFR1
IGFR2
IP3R3
nestin
TH
NFH
PDGFR
S100beta
mAChR1
mGluR8
nAChRa3
nAChRa7
NT3
trkC
5HT1b
5HT2
GAD67
GRa1
GRa4
GRa5
GRb1
GRb2
GRb3
GRg1
InsR
mAChR2
NFM
preGAD67

Using other data sets

Preparing the data

Opening “housing.data” in Excel



The Text Wizard has determined that your data is Fixed Width.
If this is correct, choose Next, or choose the data type that best describes your data.

Original data type

Choose the file type that best describes your data:

- ● Delimited - Characters such as commas or tabs separate each field.
- ○ Fixed width - Fields are aligned in columns with spaces between each field.

Start import at row: 1

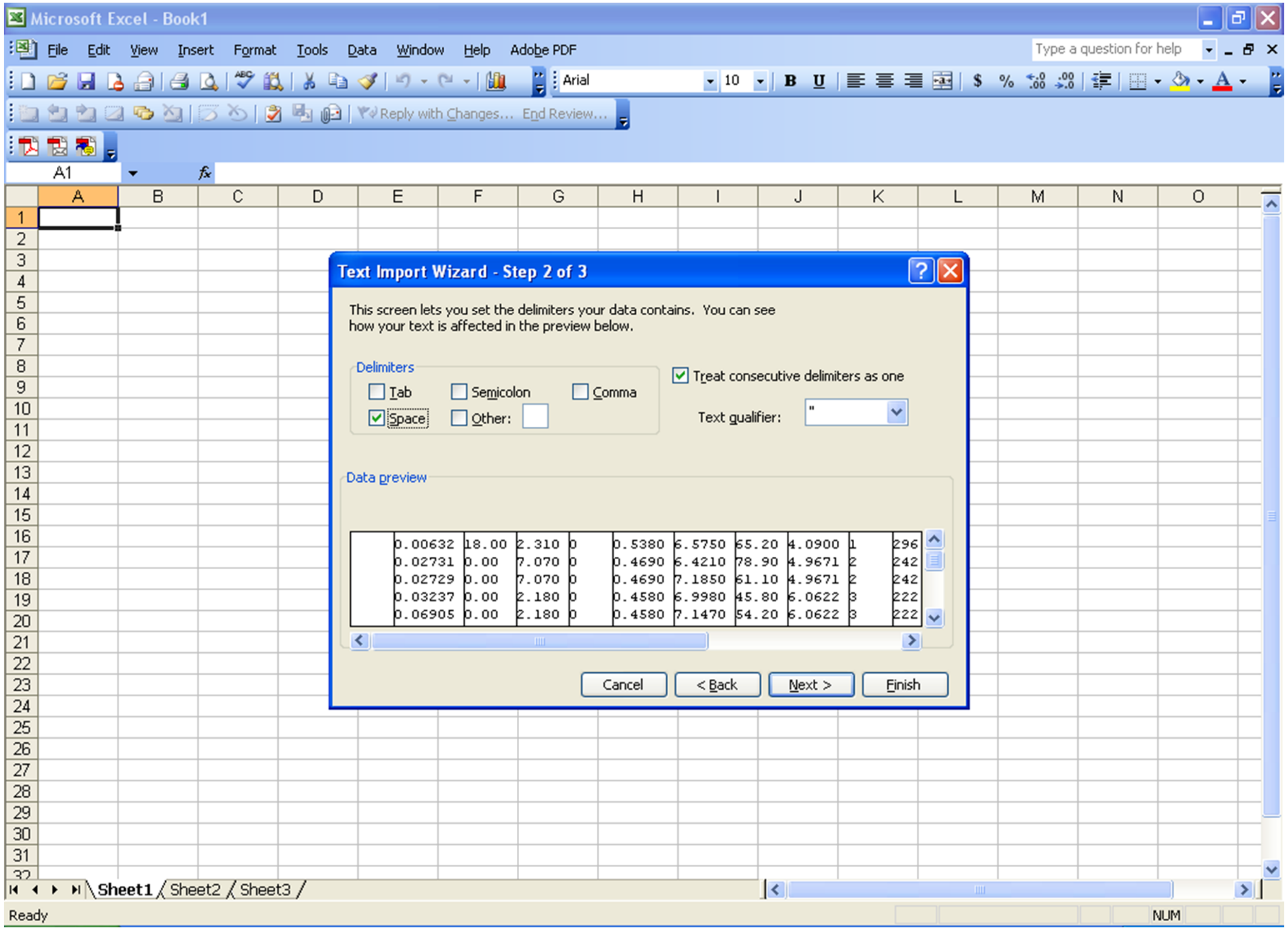
File origin: 437 : OEM United States

Preview of file C:\Caldas\Teaching\CE 395R5 - Fall 2005\Datasets\UCI\housing.data.

```
1  0.00632  18.00   2.310  0  0.5380  6.5750  65.20  4.0900   1
2  0.02731   0.00   7.070  0  0.4690  6.4210  78.90  4.9671   2
3  0.02729   0.00   7.070  0  0.4690  7.1850  61.10  4.9671   2
4  0.03237   0.00   2.180  0  0.4580  6.9980  45.80  6.0622   3
5  0.06905   0.00   2.180  0  0.4580  7.1470  54.20  6.0622   3
```

In this data set “space” is the delimiter

That’s not always the case



Text Import Wizard - Step 2 of 3

This screen lets you set the delimiters your data contains. You can see how your text is affected in the preview below.

Delimiters

☐ Tab ☐ Semicolon ☐ Comma

☑ Space ☐ Other:

☑ Treat consecutive delimiters as one

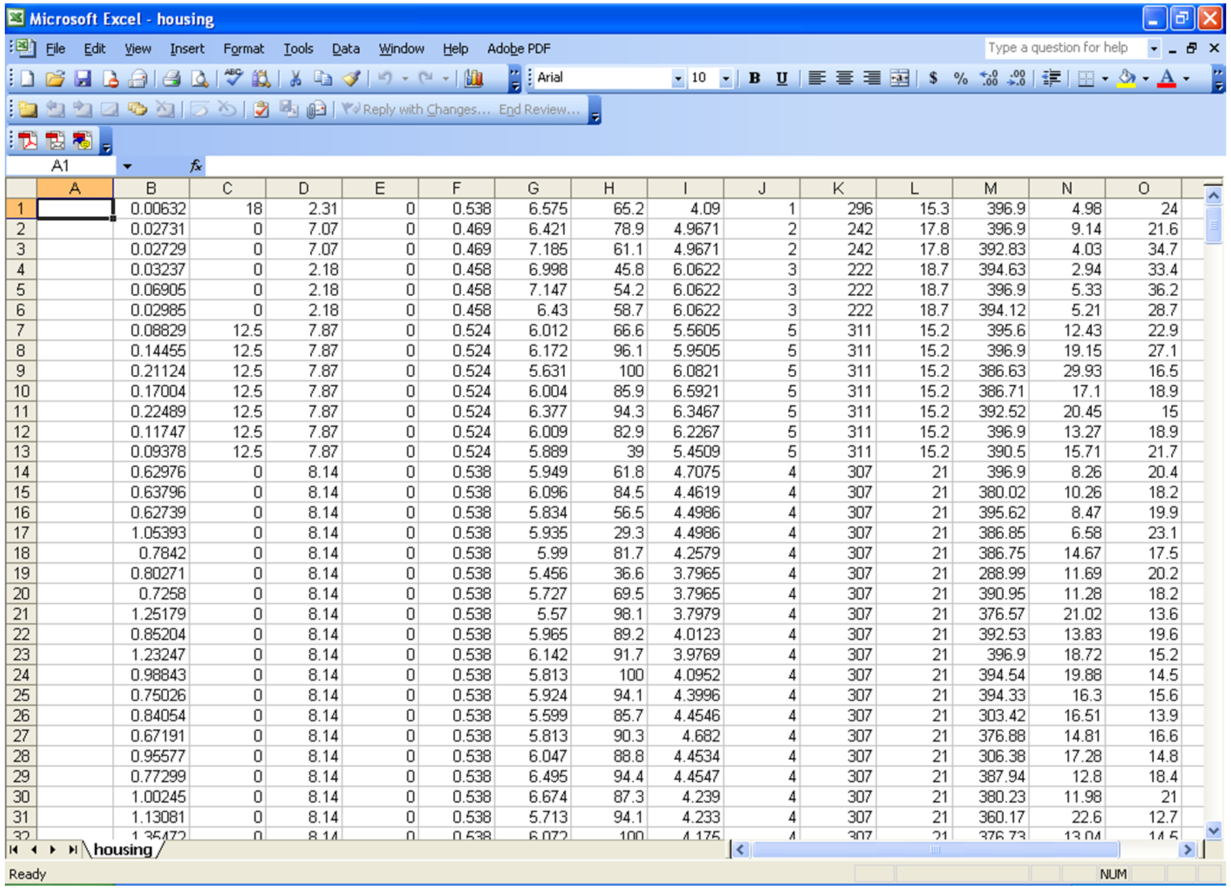
Text qualifier: "

Data preview

	0.00632	18.00	2.310	0	0.5380	6.5750	65.20	4.0900	1	296
	0.02731	0.00	7.070	0	0.4690	6.4210	78.90	4.9671	2	242
	0.02729	0.00	7.070	0	0.4690	7.1850	61.10	4.9671	2	242
	0.03237	0.00	2.180	0	0.4580	6.9980	45.80	6.0622	3	222
	0.06905	0.00	2.180	0	0.4580	7.1470	54.20	6.0622	3	222

Check the data

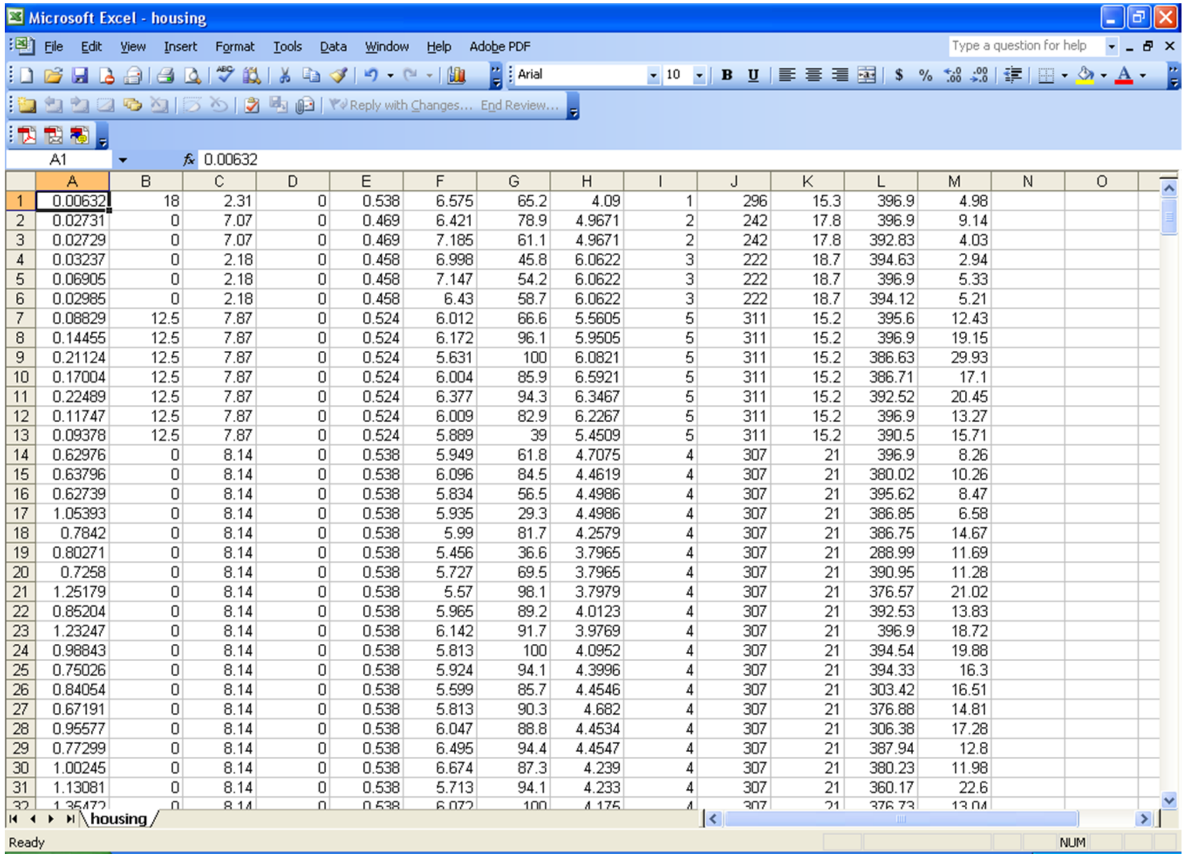
- make sure that all columns are complete
- preprocess the data, if needed
- eliminate unnecessary columns



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		0.00632	18	2.31	0	0.538	6.575	65.2	4.09	1	296	15.3	396.9	4.98	24
2		0.02731	0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.9	9.14	21.6
3		0.02729	0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7
4		0.03237	0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4
5		0.06905	0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.9	5.33	36.2
6		0.02985	0	2.18	0	0.458	6.43	58.7	6.0622	3	222	18.7	394.12	5.21	28.7
7		0.08829	12.5	7.87	0	0.524	6.012	66.6	5.5605	5	311	15.2	395.6	12.43	22.9
8		0.14455	12.5	7.87	0	0.524	6.172	96.1	5.9505	5	311	15.2	396.9	19.15	27.1
9		0.21124	12.5	7.87	0	0.524	5.631	100	6.0821	5	311	15.2	386.63	29.93	16.5
10		0.17004	12.5	7.87	0	0.524	6.004	85.9	6.5921	5	311	15.2	386.71	17.1	18.9
11		0.22489	12.5	7.87	0	0.524	6.377	94.3	6.3467	5	311	15.2	392.52	20.45	15
12		0.11747	12.5	7.87	0	0.524	6.009	82.9	6.2267	5	311	15.2	396.9	13.27	18.9
13		0.09378	12.5	7.87	0	0.524	5.889	39	5.4509	5	311	15.2	390.5	15.71	21.7
14		0.62976	0	8.14	0	0.538	5.949	61.8	4.7075	4	307	21	396.9	8.26	20.4
15		0.63796	0	8.14	0	0.538	6.096	84.5	4.4619	4	307	21	380.02	10.26	18.2
16		0.62739	0	8.14	0	0.538	5.834	56.5	4.4986	4	307	21	395.62	8.47	19.9
17		1.05393	0	8.14	0	0.538	5.935	29.3	4.4986	4	307	21	386.85	6.58	23.1
18		0.7842	0	8.14	0	0.538	5.99	81.7	4.2579	4	307	21	386.75	14.67	17.5
19		0.80271	0	8.14	0	0.538	5.456	36.6	3.7965	4	307	21	288.99	11.69	20.2
20		0.7258	0	8.14	0	0.538	5.727	69.5	3.7965	4	307	21	390.95	11.28	18.2
21		1.25179	0	8.14	0	0.538	5.57	98.1	3.7979	4	307	21	376.57	21.02	13.6
22		0.85204	0	8.14	0	0.538	5.965	89.2	4.0123	4	307	21	392.53	13.83	19.6
23		1.23247	0	8.14	0	0.538	6.142	91.7	3.9769	4	307	21	396.9	18.72	15.2
24		0.98843	0	8.14	0	0.538	5.813	100	4.0952	4	307	21	394.54	19.88	14.5
25		0.75026	0	8.14	0	0.538	5.924	94.1	4.3996	4	307	21	394.33	16.3	15.6
26		0.84054	0	8.14	0	0.538	5.599	85.7	4.4546	4	307	21	303.42	16.51	13.9
27		0.67191	0	8.14	0	0.538	5.813	90.3	4.682	4	307	21	376.88	14.81	16.6
28		0.95577	0	8.14	0	0.538	6.047	88.8	4.4534	4	307	21	306.38	17.28	14.8
29		0.77299	0	8.14	0	0.538	6.495	94.4	4.4547	4	307	21	387.94	12.8	18.4
30		1.00245	0	8.14	0	0.538	6.674	87.3	4.239	4	307	21	380.23	11.98	21
31		1.13081	0	8.14	0	0.538	5.713	94.1	4.233	4	307	21	360.17	22.6	12.7
32		1.35472	0	8.14	0	0.538	6.072	100	4.175	4	307	21	376.73	13.04	14.5

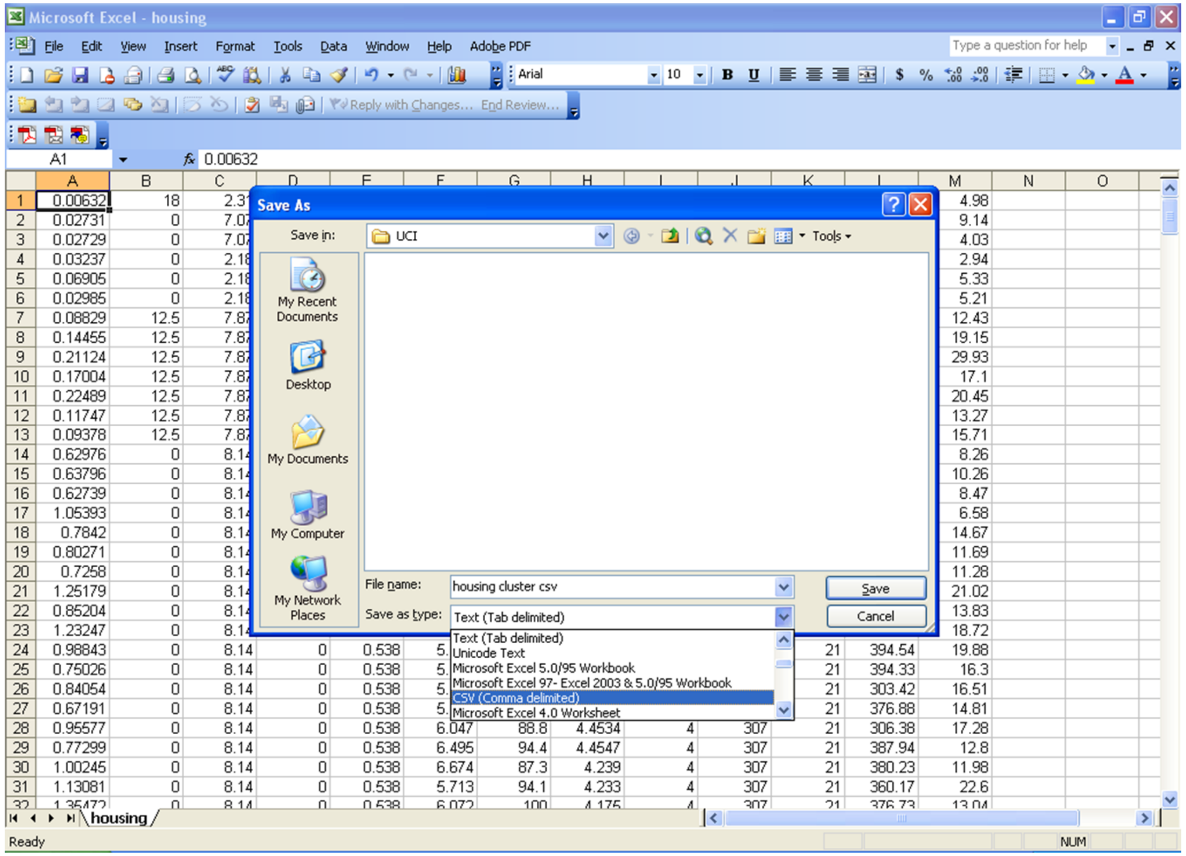
In this example, I will eliminate:

- First column: empty
- Last column: predicted value (don’t need it for clustering



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	0.00632	18	2.31	0	0.538	6.575	65.2	4.09	1	296	15.3	396.9	4.98		
2	0.02731	0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.9	9.14		
3	0.02729	0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03		
4	0.03237	0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94		
5	0.06905	0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.9	5.33		
6	0.02985	0	2.18	0	0.458	6.43	58.7	6.0622	3	222	18.7	394.12	5.21		
7	0.08829	12.5	7.87	0	0.524	6.012	66.6	5.5605	5	311	15.2	395.6	12.43		
8	0.14455	12.5	7.87	0	0.524	6.172	96.1	5.9505	5	311	15.2	396.9	19.15		
9	0.21124	12.5	7.87	0	0.524	5.631	100	6.0821	5	311	15.2	386.63	29.93		
10	0.17004	12.5	7.87	0	0.524	6.004	85.9	6.5921	5	311	15.2	386.71	17.1		
11	0.22489	12.5	7.87	0	0.524	6.377	94.3	6.3467	5	311	15.2	392.52	20.45		
12	0.11747	12.5	7.87	0	0.524	6.009	82.9	6.2267	5	311	15.2	396.9	13.27		
13	0.09378	12.5	7.87	0	0.524	5.889	39	5.4509	5	311	15.2	390.5	15.71		
14	0.62976	0	8.14	0	0.538	5.949	61.8	4.7075	4	307	21	396.9	8.26		
15	0.63796	0	8.14	0	0.538	6.096	84.5	4.4619	4	307	21	380.02	10.26		
16	0.62739	0	8.14	0	0.538	5.834	56.5	4.4986	4	307	21	395.62	8.47		
17	1.05393	0	8.14	0	0.538	5.935	29.3	4.4986	4	307	21	386.85	6.58		
18	0.7842	0	8.14	0	0.538	5.99	81.7	4.2579	4	307	21	386.75	14.67		
19	0.80271	0	8.14	0	0.538	5.456	36.6	3.7965	4	307	21	288.99	11.69		
20	0.7258	0	8.14	0	0.538	5.727	69.5	3.7965	4	307	21	390.95	11.28		
21	1.25179	0	8.14	0	0.538	5.57	98.1	3.7979	4	307	21	376.57	21.02		
22	0.85204	0	8.14	0	0.538	5.965	89.2	4.0123	4	307	21	392.53	13.83		
23	1.23247	0	8.14	0	0.538	6.142	91.7	3.9769	4	307	21	396.9	18.72		
24	0.98843	0	8.14	0	0.538	5.813	100	4.0952	4	307	21	394.54	19.88		
25	0.75026	0	8.14	0	0.538	5.924	94.1	4.3996	4	307	21	394.33	16.3		
26	0.84054	0	8.14	0	0.538	5.599	85.7	4.4546	4	307	21	303.42	16.51		
27	0.67191	0	8.14	0	0.538	5.813	90.3	4.682	4	307	21	376.88	14.81		
28	0.95577	0	8.14	0	0.538	6.047	88.8	4.4534	4	307	21	306.38	17.28		
29	0.77299	0	8.14	0	0.538	6.495	94.4	4.4547	4	307	21	387.94	12.8		
30	1.00245	0	8.14	0	0.538	6.674	87.3	4.239	4	307	21	380.23	11.98		
31	1.13081	0	8.14	0	0.538	5.713	94.1	4.233	4	307	21	360.17	22.6		
32	1.35472	0	8.14	0	0.538	6.072	100	4.175	4	307	21	376.73	13.04		

Save the file as CSV



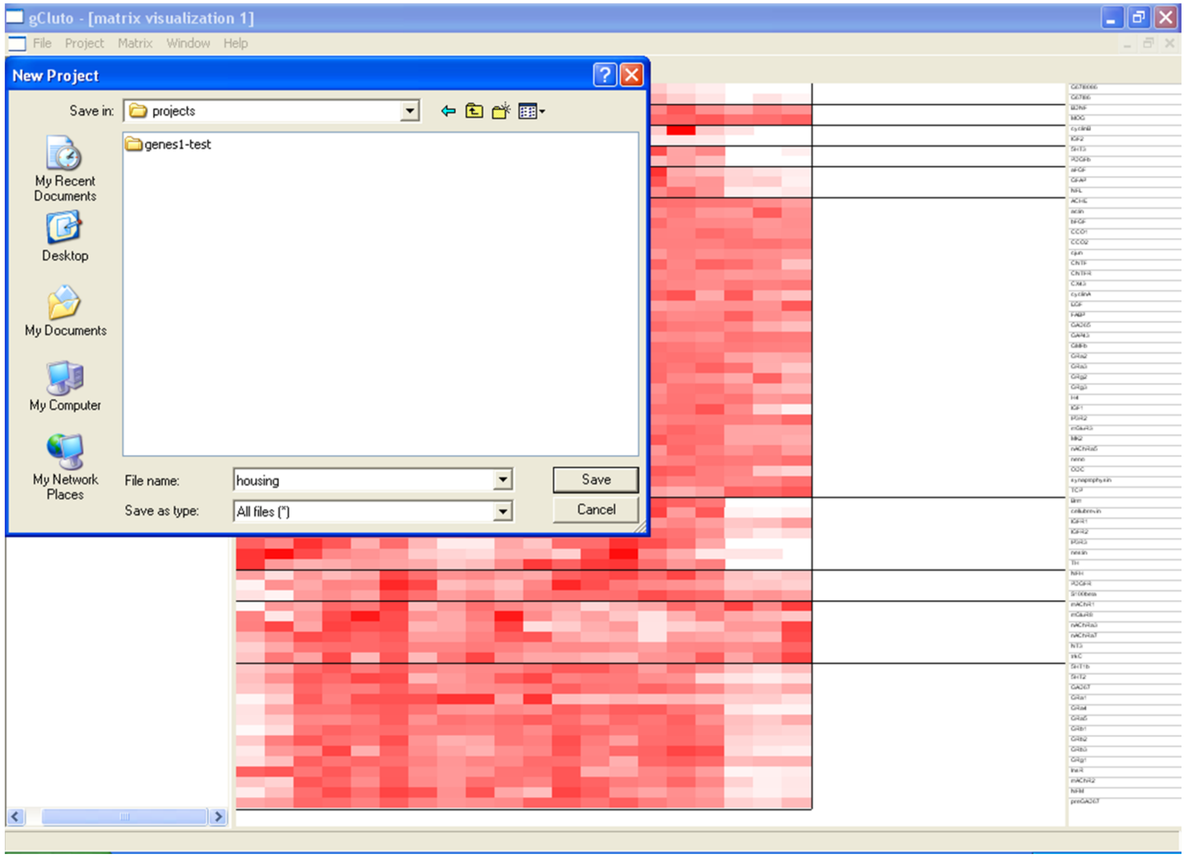
A1 = 0.00632

Save in: UCI

File name: housing cluster csv

Save as type: Text (Tab delimited)

Create a new project in gCLUTO

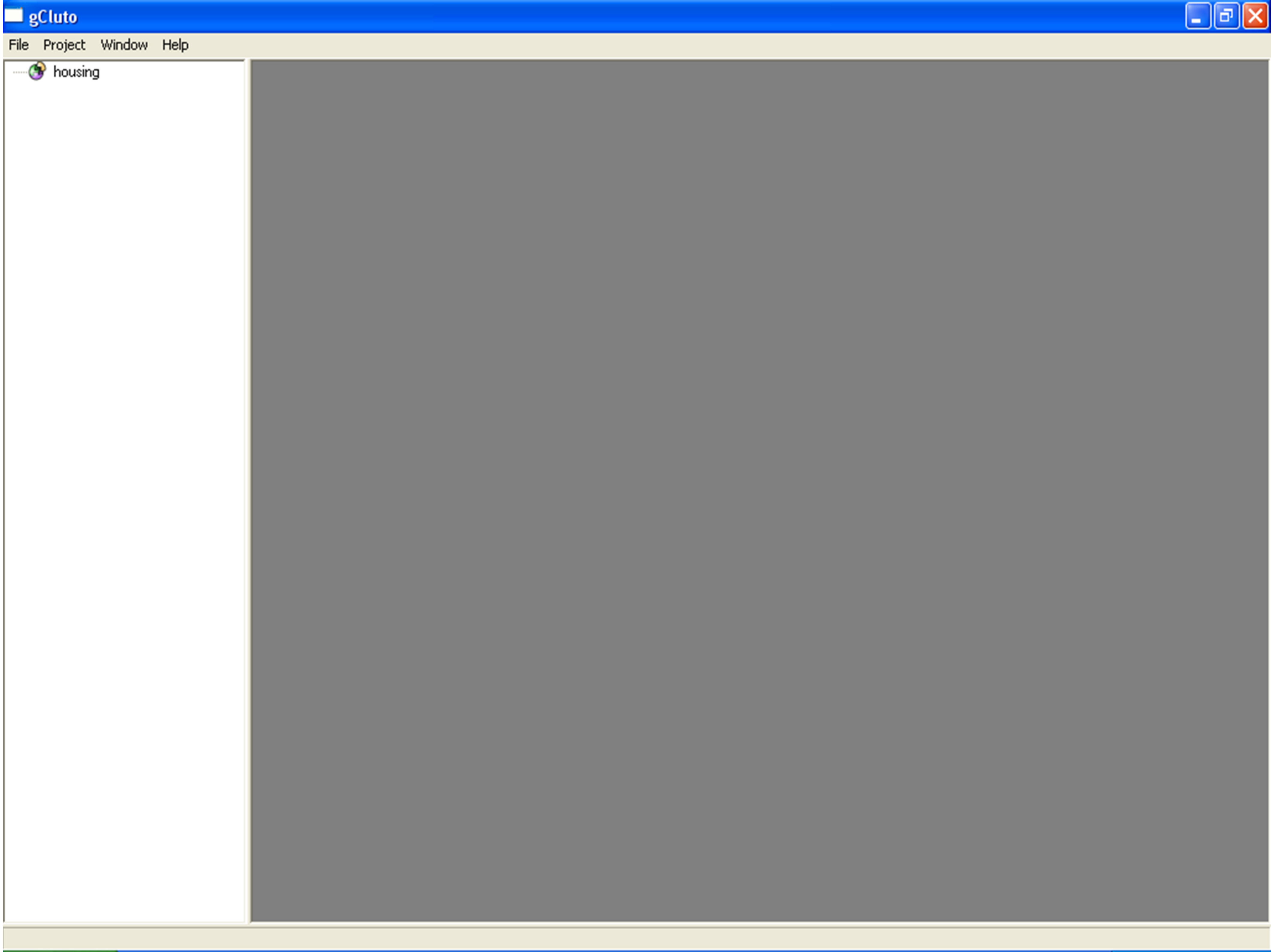


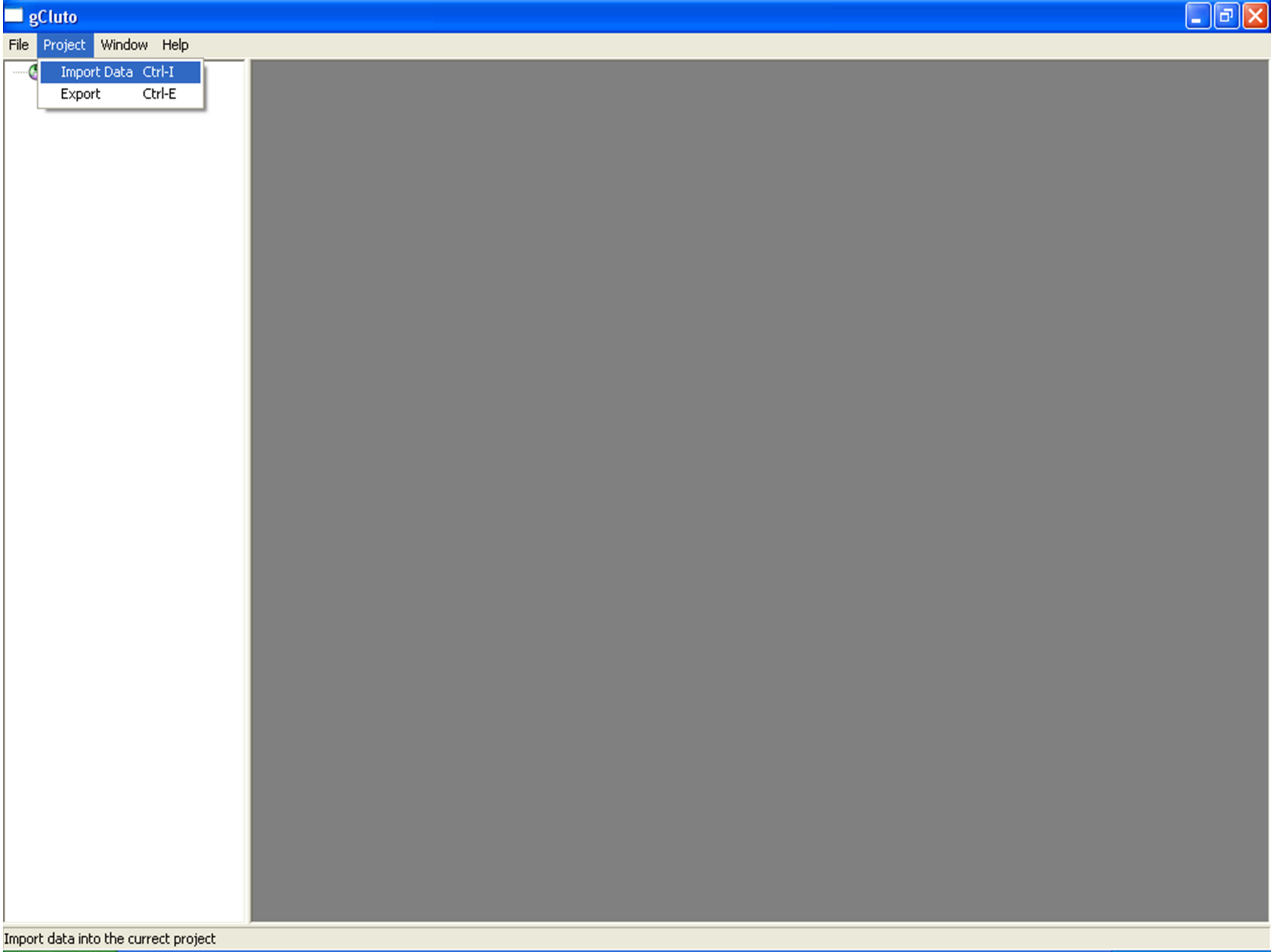
Save in: projects

genes1-test

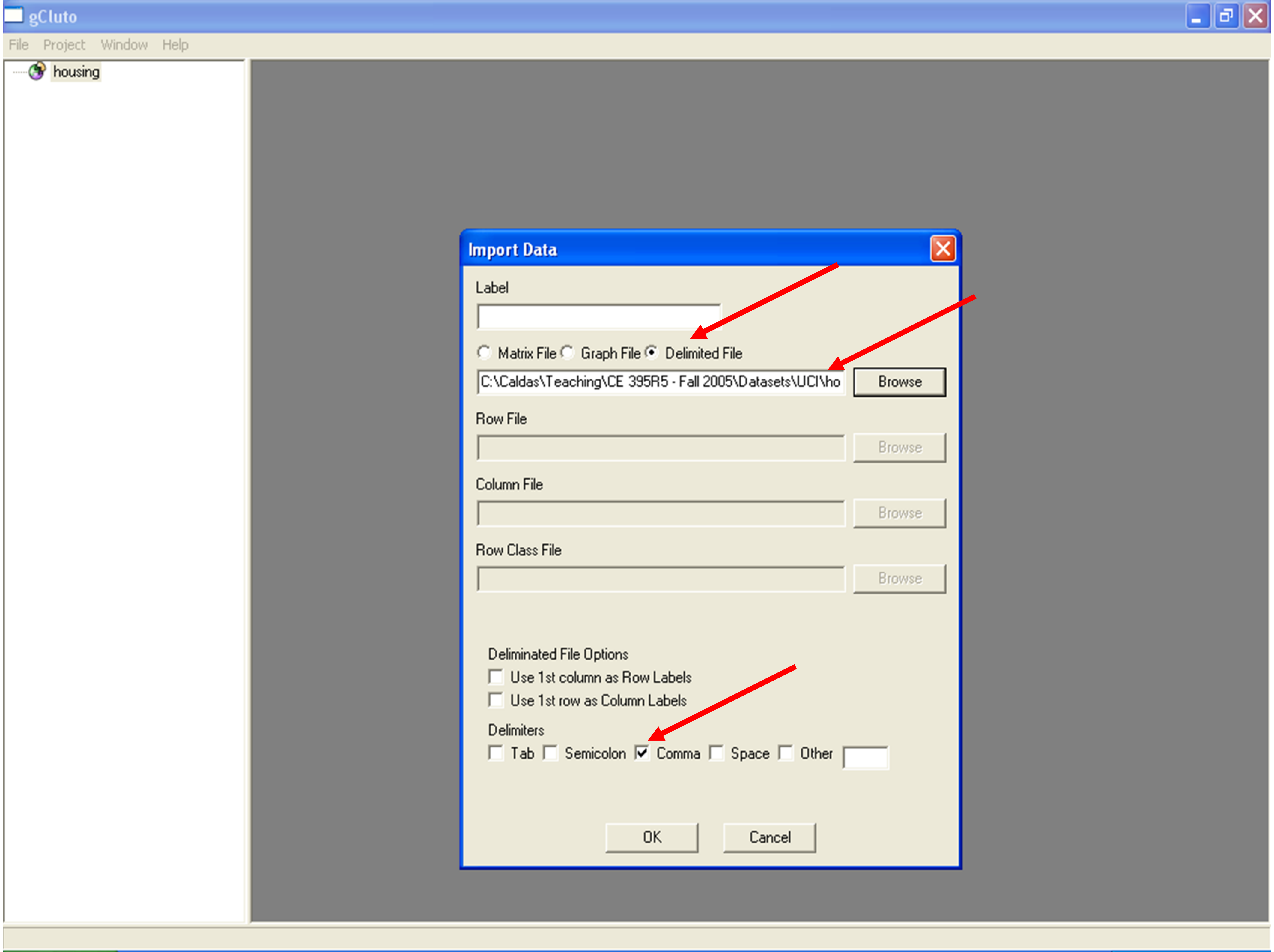
File name: housing

Save as type: All files (*)


gCluto
File
Project
Window
Help
housing



Import data into the currect project



Label

○ Matrix File ○ Graph File ● Delimited File

C:\Caldas\Teaching\CE 395R5 - Fall 2005\Datasets\UCI\ho

Row File

Column File

Row Class File

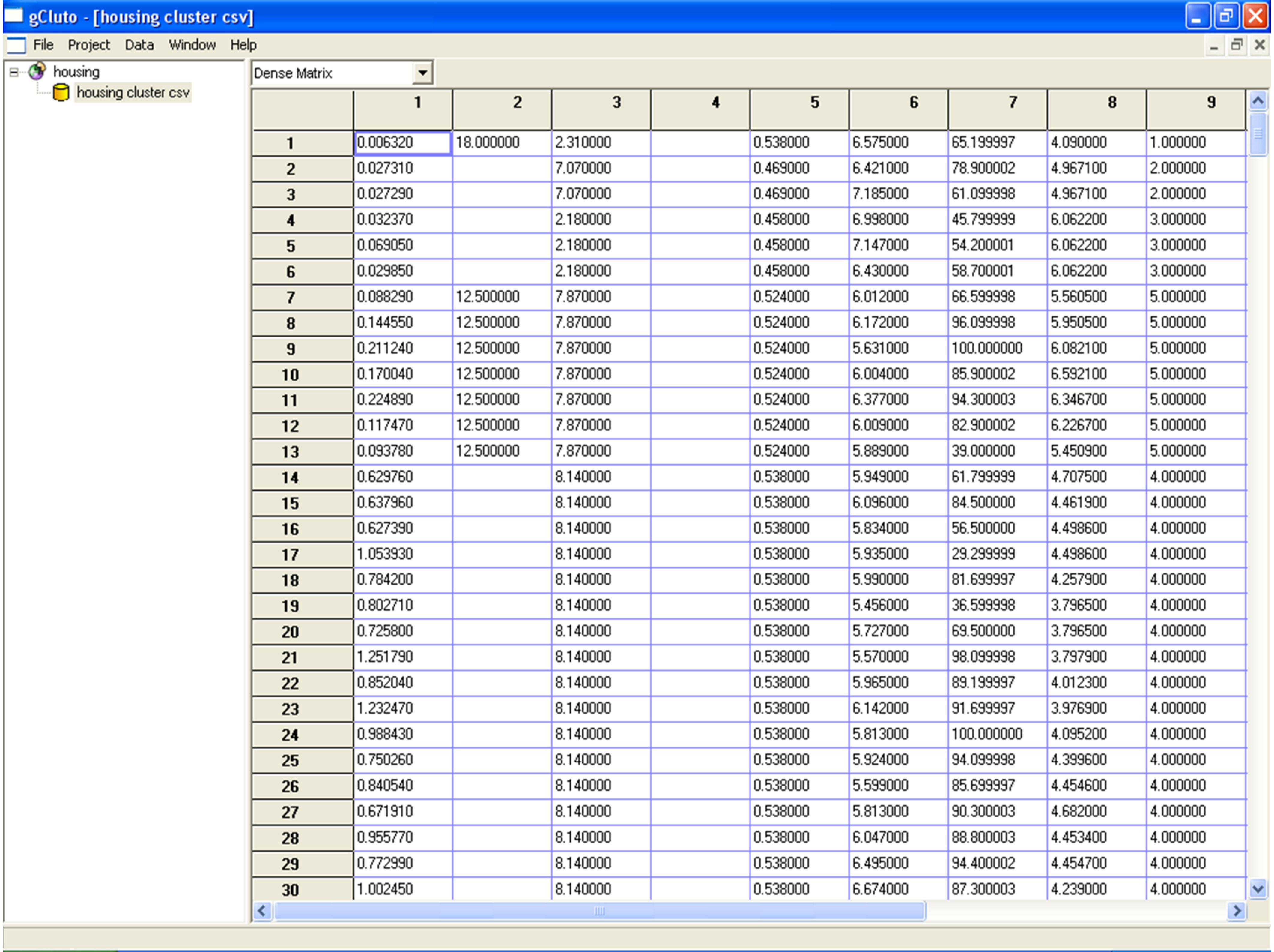
Deliminated File Options

☐ Use 1st column as Row Labels

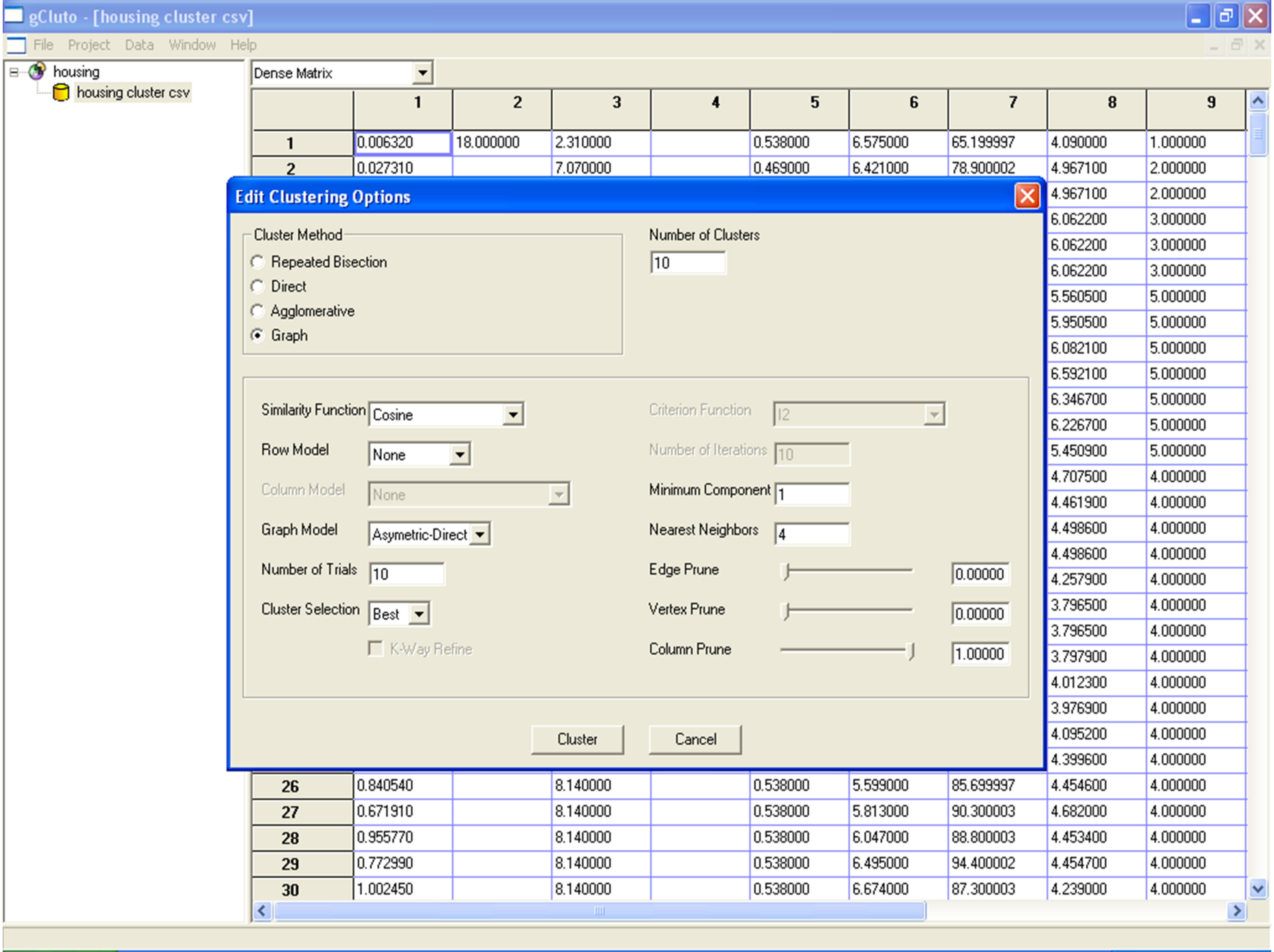
☐ Use 1st row as Column Labels

Delimiters

☐ Tab ☐ Semicolon ☑ Comma ☐ Space ☐ Other



	1	2	3	4	5	6	7	8	9
1	0.006320	18.000000	2.310000		0.538000	6.575000	65.199997	4.090000	1.000000
2	0.027310		7.070000		0.469000	6.421000	78.900002	4.967100	2.000000
3	0.027290		7.070000		0.469000	7.185000	61.099998	4.967100	2.000000
4	0.032370		2.180000		0.458000	6.998000	45.799999	6.062200	3.000000
5	0.069050		2.180000		0.458000	7.147000	54.200001	6.062200	3.000000
6	0.029850		2.180000		0.458000	6.430000	58.700001	6.062200	3.000000
7	0.088290	12.500000	7.870000		0.524000	6.012000	66.599998	5.560500	5.000000
8	0.144550	12.500000	7.870000		0.524000	6.172000	96.099998	5.950500	5.000000
9	0.211240	12.500000	7.870000		0.524000	5.631000	100.000000	6.082100	5.000000
10	0.170040	12.500000	7.870000		0.524000	6.004000	85.900002	6.592100	5.000000
11	0.224890	12.500000	7.870000		0.524000	6.377000	94.300003	6.346700	5.000000
12	0.117470	12.500000	7.870000		0.524000	6.009000	82.900002	6.226700	5.000000
13	0.093780	12.500000	7.870000		0.524000	5.889000	39.000000	5.450900	5.000000
14	0.629760		8.140000		0.538000	5.949000	61.799999	4.707500	4.000000
15	0.637960		8.140000		0.538000	6.096000	84.500000	4.461900	4.000000
16	0.627390		8.140000		0.538000	5.834000	56.500000	4.498600	4.000000
17	1.053930		8.140000		0.538000	5.935000	29.299999	4.498600	4.000000
18	0.784200		8.140000		0.538000	5.990000	81.699997	4.257900	4.000000
19	0.802710		8.140000		0.538000	5.456000	36.599998	3.796500	4.000000
20	0.725800		8.140000		0.538000	5.727000	69.500000	3.796500	4.000000
21	1.251790		8.140000		0.538000	5.570000	98.099998	3.797900	4.000000
22	0.852040		8.140000		0.538000	5.965000	89.199997	4.012300	4.000000
23	1.232470		8.140000		0.538000	6.142000	91.699997	3.976900	4.000000
24	0.988430		8.140000		0.538000	5.813000	100.000000	4.095200	4.000000
25	0.750260		8.140000		0.538000	5.924000	94.099998	4.399600	4.000000
26	0.840540		8.140000		0.538000	5.599000	85.699997	4.454600	4.000000
27	0.671910		8.140000		0.538000	5.813000	90.300003	4.682000	4.000000
28	0.955770		8.140000		0.538000	6.047000	88.800003	4.453400	4.000000
29	0.772990		8.140000		0.538000	6.495000	94.400002	4.454700	4.000000
30	1.002450		8.140000		0.538000	6.674000	87.300003	4.239000	4.000000



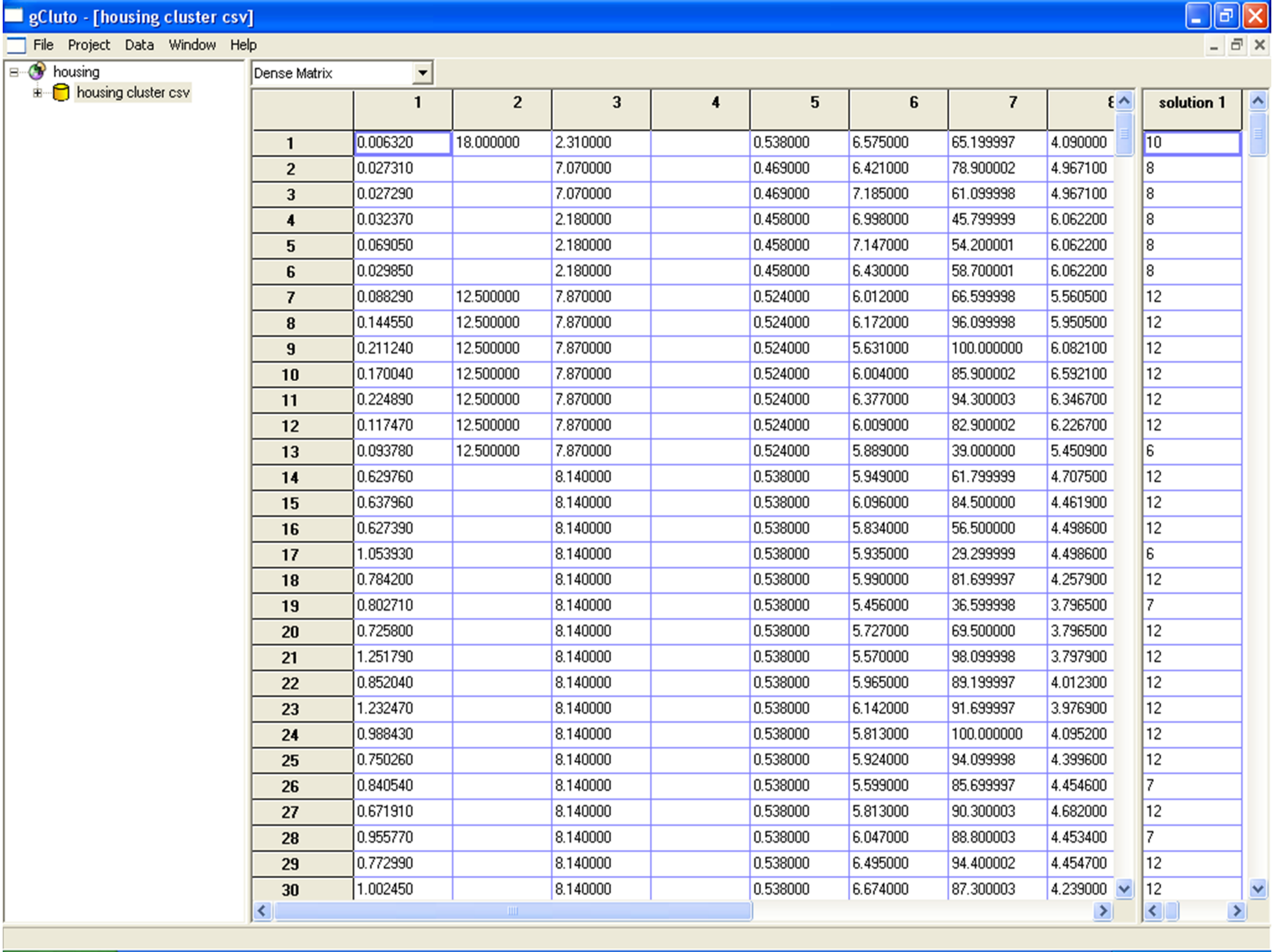
	1	2	3	4	5	6	7	8	9
1	0.006320	18.000000	2.310000		0.538000	6.575000	65.199997	4.090000	1.000000
2	0.027310		7.070000		0.469000	6.421000	78.900002	4.967100	2.000000
								4.967100	2.000000
								6.062200	3.000000
								6.062200	3.000000
								6.062200	3.000000
								5.560500	5.000000
								5.950500	5.000000
								6.082100	5.000000
								6.592100	5.000000
								6.346700	5.000000
								6.226700	5.000000
								5.450900	5.000000
								4.707500	4.000000
								4.461900	4.000000
								4.498600	4.000000
								4.498600	4.000000
								4.257900	4.000000
								3.796500	4.000000
								3.796500	4.000000
								3.797900	4.000000
								4.012300	4.000000
								3.976900	4.000000
								4.095200	4.000000
								4.399600	4.000000
26	0.840540		8.140000		0.538000	5.599000	85.699997	4.454600	4.000000
27	0.671910		8.140000		0.538000	5.813000	90.300003	4.682000	4.000000
28	0.955770		8.140000		0.538000	6.047000	88.800003	4.453400	4.000000
29	0.772990		8.140000		0.538000	6.495000	94.400002	4.454700	4.000000
30	1.002450		8.140000		0.538000	6.674000	87.300003	4.239000	4.000000

Cluster Method
○ Repeated Bisection
○ Direct
○ Agglomerative
● Graph

Number of Clusters: 10

Similarity Function: Cosine
Row Model: None
Column Model: None
Graph Model: Asymetric-Direct
Number of Trials: 10
Cluster Selection: Best
☐ K-Way Refine

Criterion Function: I2
Number of Iterations: 10
Minimum Component: 1
Nearest Neighbors: 4
Edge Prune: 0.00000
Vertex Prune: 0.00000
Column Prune: 1.00000



	1	2	3	4	5	6	7	8	solution 1
1	0.006320	18.000000	2.310000		0.538000	6.575000	65.199997	4.090000	10
2	0.027310		7.070000		0.469000	6.421000	78.900002	4.967100	8
3	0.027290		7.070000		0.469000	7.185000	61.099998	4.967100	8
4	0.032370		2.180000		0.458000	6.998000	45.799999	6.062200	8
5	0.069050		2.180000		0.458000	7.147000	54.200001	6.062200	8
6	0.029850		2.180000		0.458000	6.430000	58.700001	6.062200	8
7	0.088290	12.500000	7.870000		0.524000	6.012000	66.599998	5.560500	12
8	0.144550	12.500000	7.870000		0.524000	6.172000	96.099998	5.950500	12
9	0.211240	12.500000	7.870000		0.524000	5.631000	100.000000	6.082100	12
10	0.170040	12.500000	7.870000		0.524000	6.004000	85.900002	6.592100	12
11	0.224890	12.500000	7.870000		0.524000	6.377000	94.300003	6.346700	12
12	0.117470	12.500000	7.870000		0.524000	6.009000	82.900002	6.226700	12
13	0.093780	12.500000	7.870000		0.524000	5.889000	39.000000	5.450900	6
14	0.629760		8.140000		0.538000	5.949000	61.799999	4.707500	12
15	0.637960		8.140000		0.538000	6.096000	84.500000	4.461900	12
16	0.627390		8.140000		0.538000	5.834000	56.500000	4.498600	12
17	1.053930		8.140000		0.538000	5.935000	29.299999	4.498600	6
18	0.784200		8.140000		0.538000	5.990000	81.699997	4.257900	12
19	0.802710		8.140000		0.538000	5.456000	36.599998	3.796500	7
20	0.725800		8.140000		0.538000	5.727000	69.500000	3.796500	12
21	1.251790		8.140000		0.538000	5.570000	98.099998	3.797900	12
22	0.852040		8.140000		0.538000	5.965000	89.199997	4.012300	12
23	1.232470		8.140000		0.538000	6.142000	91.699997	3.976900	12
24	0.988430		8.140000		0.538000	5.813000	100.000000	4.095200	12
25	0.750260		8.140000		0.538000	5.924000	94.099998	4.399600	12
26	0.840540		8.140000		0.538000	5.599000	85.699997	4.454600	7
27	0.671910		8.140000		0.538000	5.813000	90.300003	4.682000	12
28	0.955770		8.140000		0.538000	6.047000	88.800003	4.453400	7
29	0.772990		8.140000		0.538000	6.495000	94.400002	4.454700	12
30	1.002450		8.140000		0.538000	6.674000	87.300003	4.239000	12