The Fourth Asia Pacific Bioinformatics Conference
(APBC2006)
13-16 Feb, 2006, Taipei, Taiwan

GAP Software Tutorial

Version 0.1.014 Build20060207

Han-Ming Wu and Chun-Houh Chen
(吳漢銘，陳君厚)
http://gap.stat.sinica.edu.tw

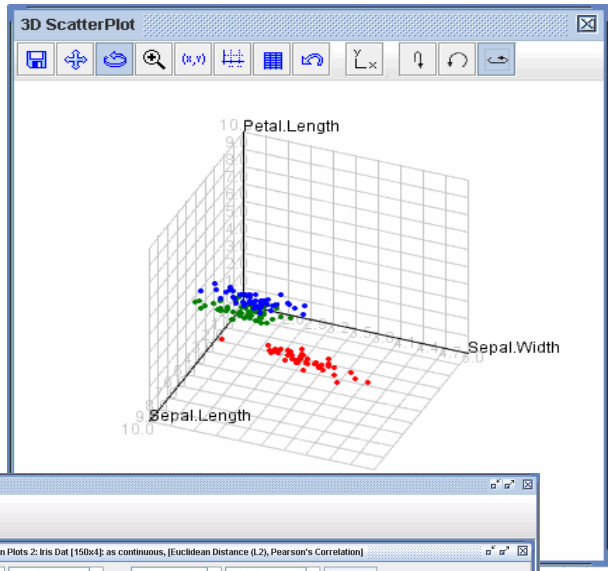
Dimension Free
Data Visualization
GAP

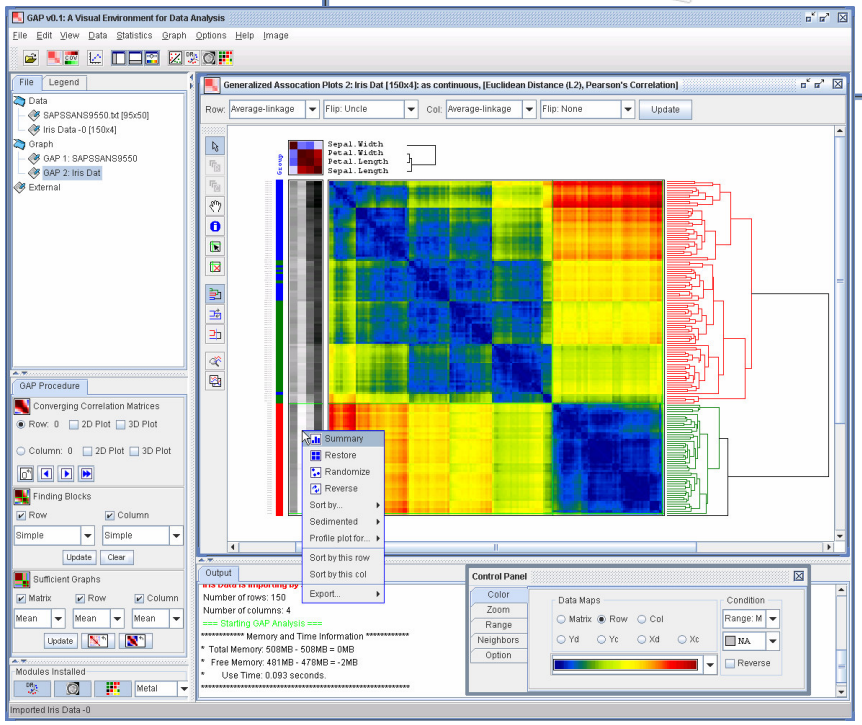
中央研究院 統計科學研究所
Institute of Statistical Science, Academia Sinica

GAP Features

Generalized Association Plots

- Input Data Type: continuous or binary.
- Various seriation algorithms and clustering analysis.
- Various display conditions.
- GAP with Covaraite Adjusted
- GAP with Nonlinear Association Analysis
- GAP with Missing Value Imputation



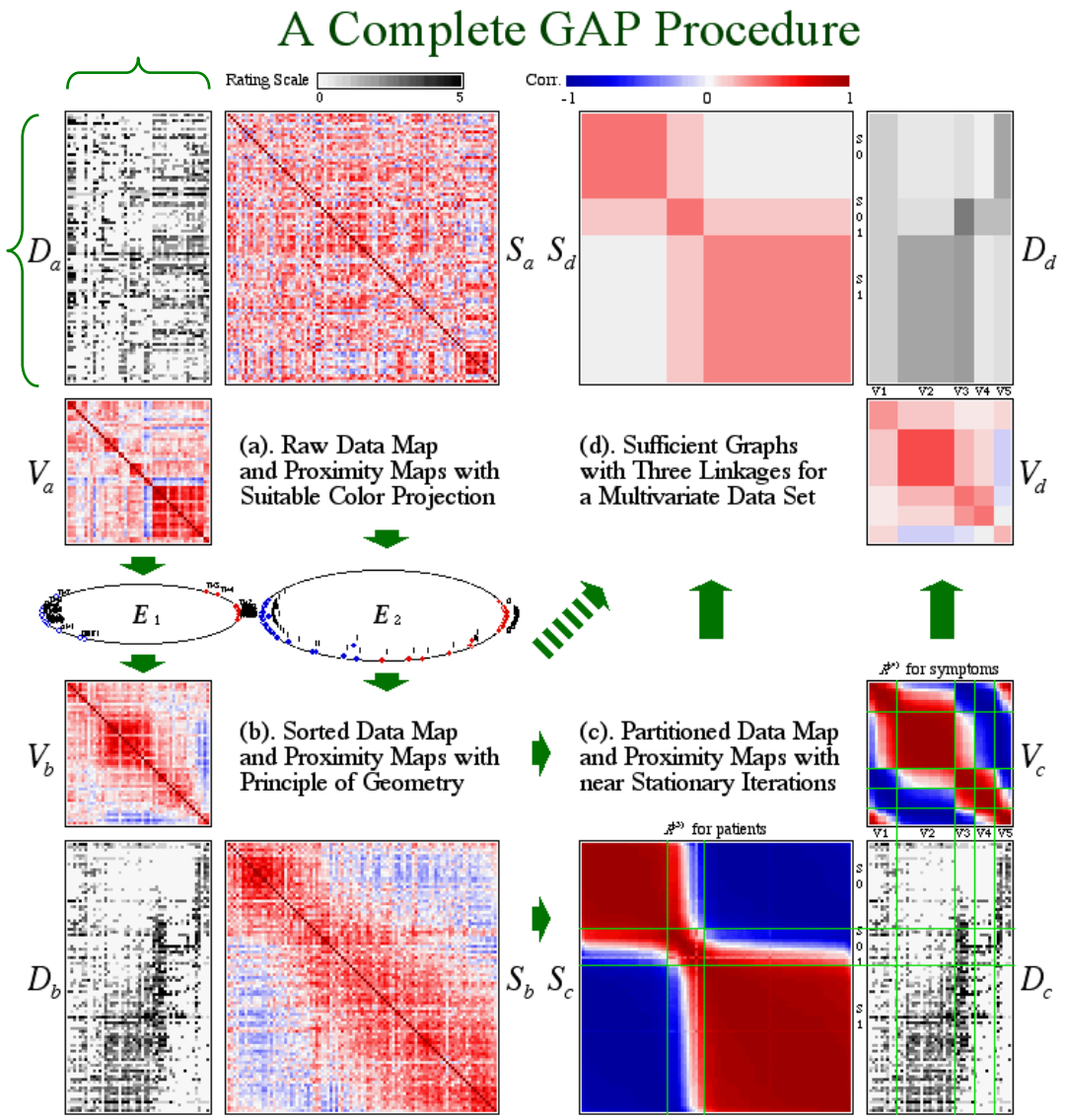


Statistical Plots

- 2D Scatterplot, 3D Scatterplot (Rotatable)

Demo Data 1: Schizo Data

- The data set consists of the Andreasen's positive and negative symptom scales (Andreasen (1983, 1984)) for **95** first-time hospitalized psychosis disorder patients. 95 patients: **69** schizophrenic and **26** bipolar disorders.
- Scale for Assessment ofPositive Symptoms (**SAPS**): **30 items**, four subgroups: Hallucinations (AH1-6), Delusions (DL1-12), Behavior (BE1-4) Thought disorder (TH1-8).
- Scale for Assessment of Negative Symptoms (**SANS**): **20 items**, five subgroups:
  Expression (NA1-7), Speech (NB1-4), Hygiene (NC1-3), Activity (ND1-4) and Inattentiveness (NE1-2).
- The available data set has 95 subjects (patients) with 50 variables (symptoms).
- All the symptoms are recorded on a six point scale (**0-5**).



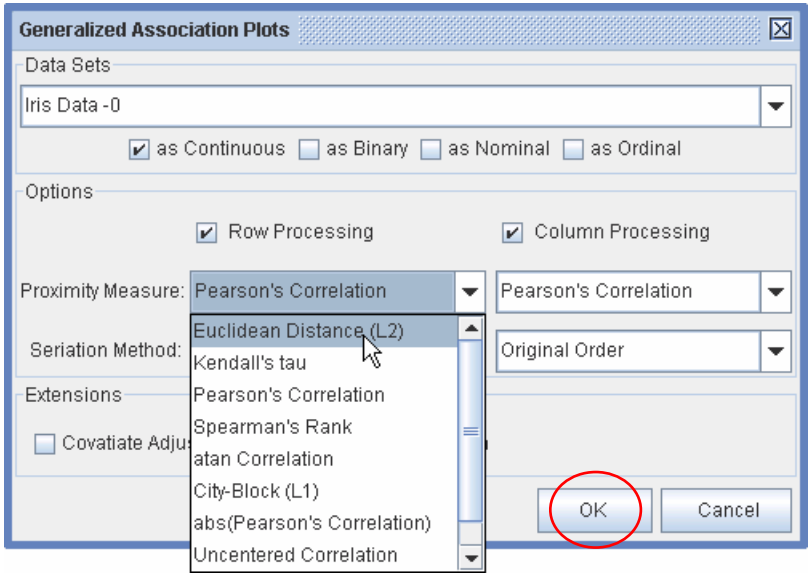
Demo Data 2: Iris Data

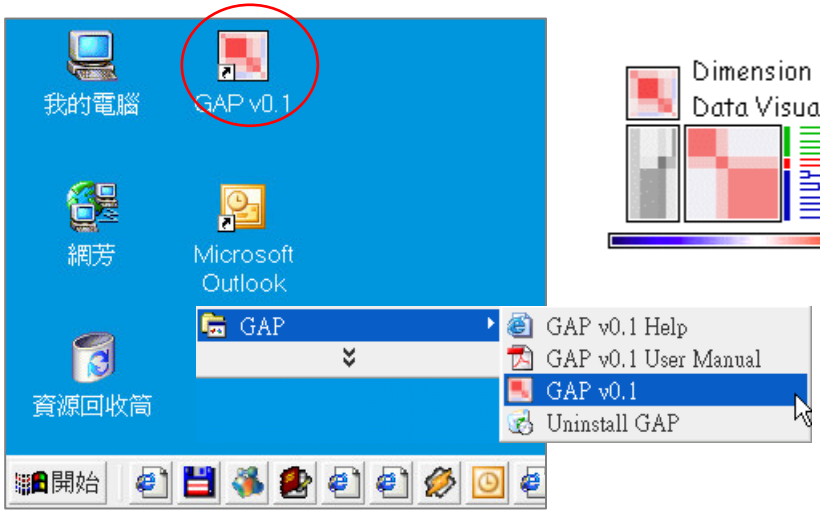
no.	Species	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
1	setosa	5.1	3.5	1.4	0.2
2	setosa	4.9	3.0	1.4	0.2
3	setosa	4.7	3.2	1.3	0.2
4	setosa	4.6	3.1	1.5	0.2
5	setosa	5.0	3.6	1.4	0.2
			...		
76	versicolor	6.6	3.0	4.4	1.4
			...		
150	virginica	5.9	3.0	5.1	1.8

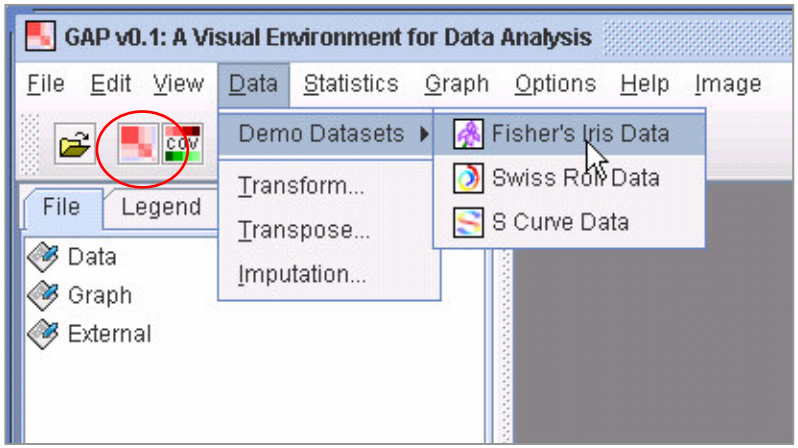
Iris Flowers

Iris Setosa Iris Versicolor Iris Virginica

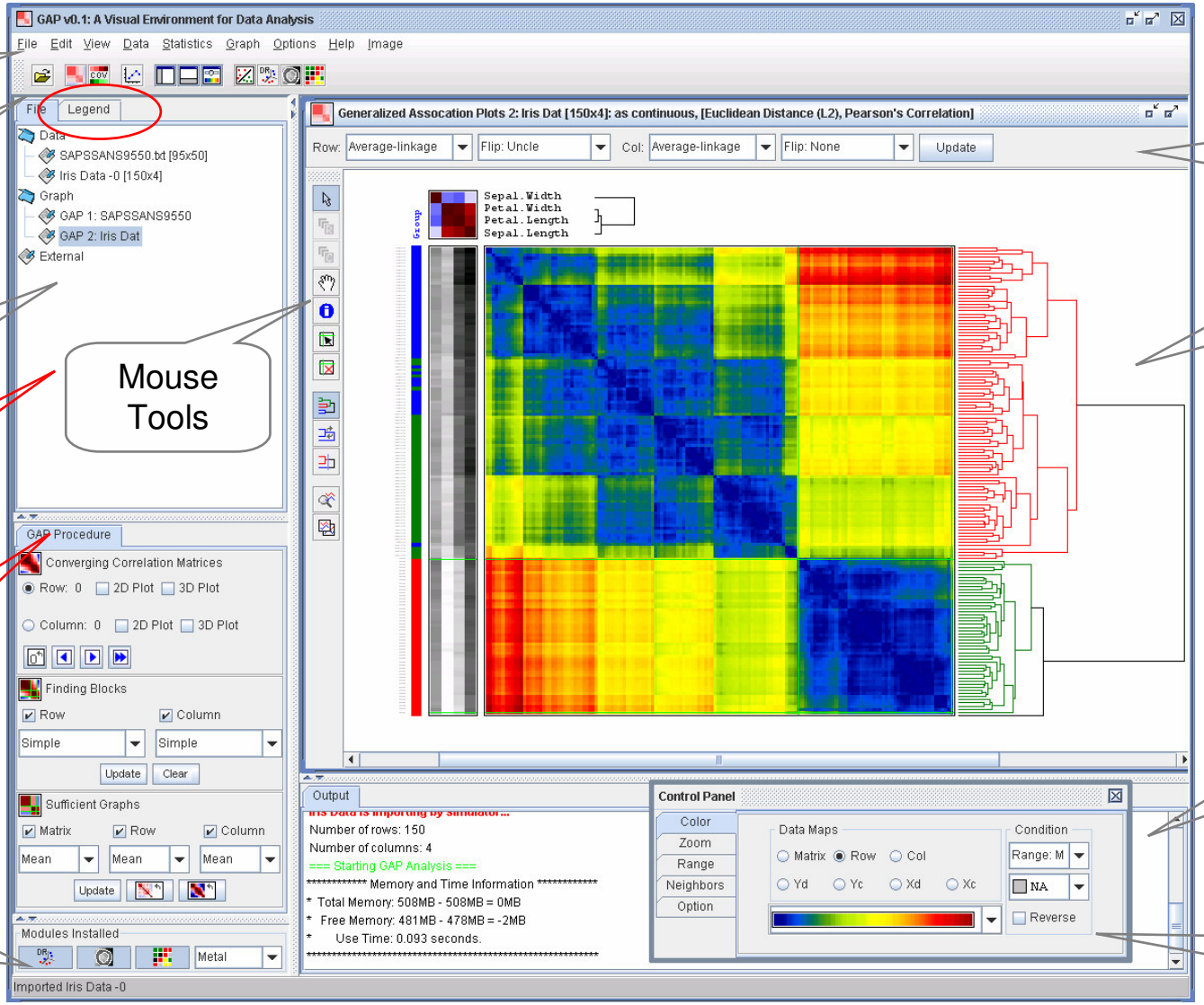
The sepal length, sepal width, petal length, and petal width are measured in centimeters on fifty iris specimens from each of three species, *Iris setosa*, *I. versicolor*, and *I. virginica*. Fisher (1936)







GAP Main Window



Menu Bar

Tool Bar

File/Data View

Color Legend

GAP Procedure

Modules Installed

Mouse Tools

Seriation and Tree Flip

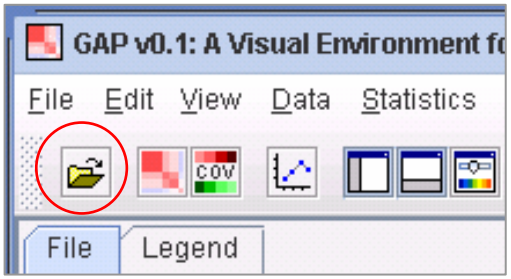
GAP Maps Window

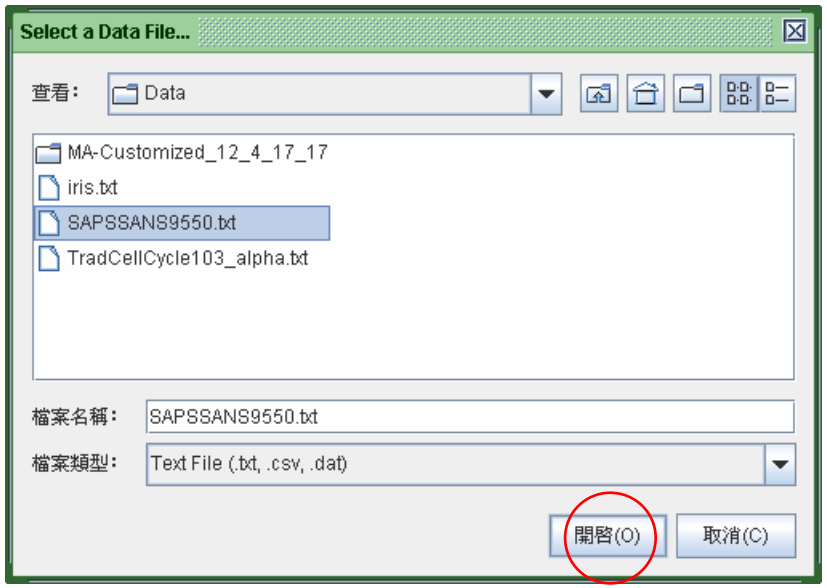
Output Window

Control Panel

Open/Import Data

Data Format: plain text with tab separator

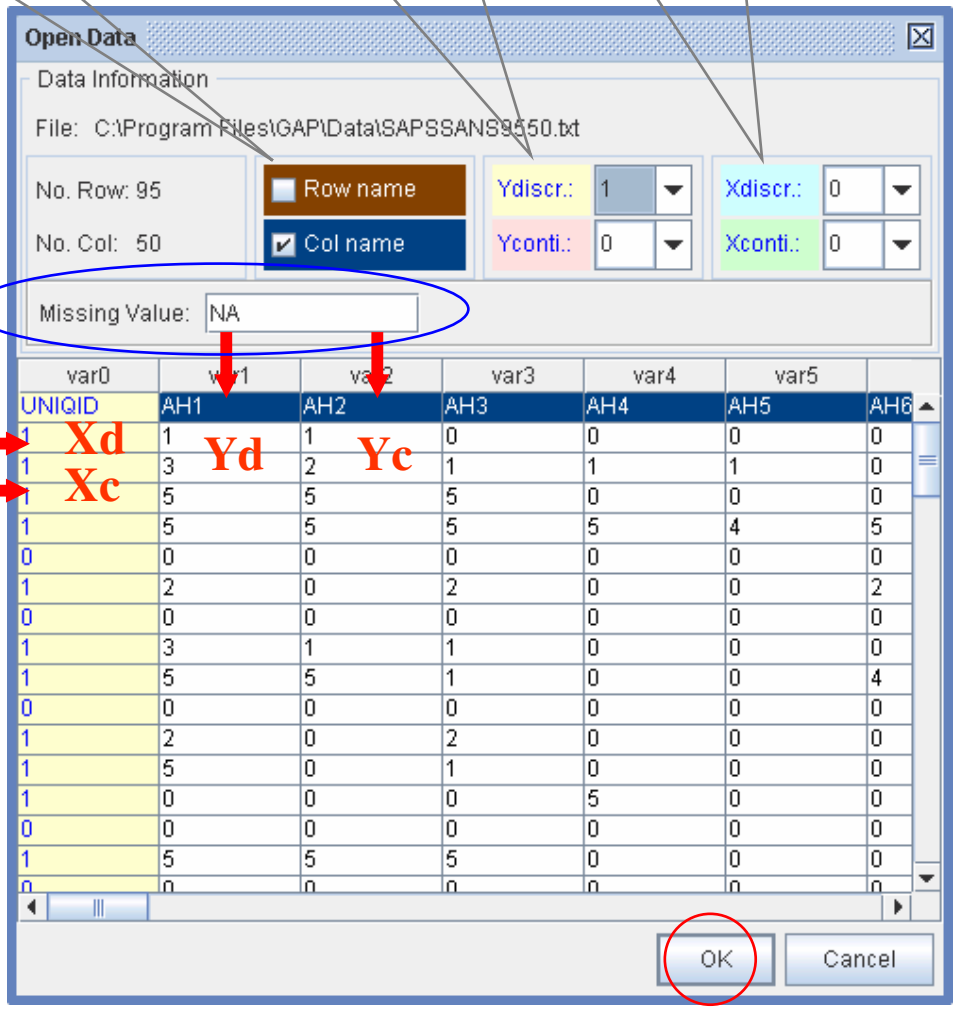




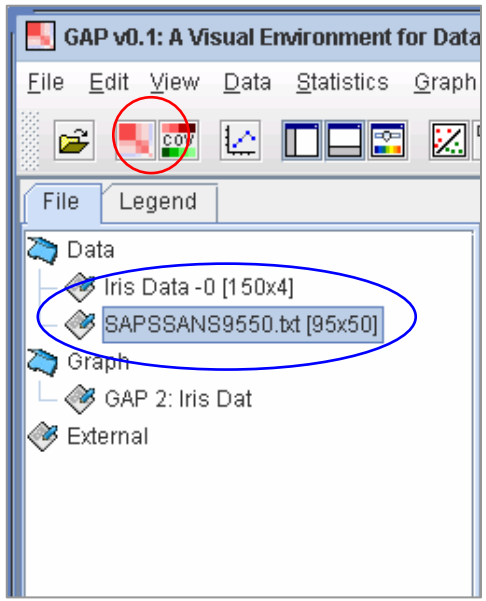
Row/Column Name

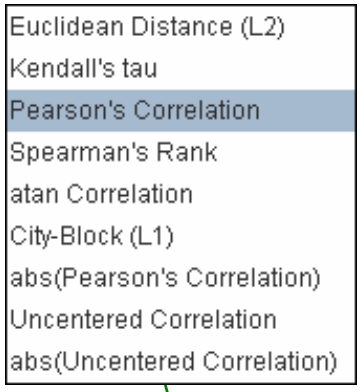
Number of Y Discrete/Continuous Covariates

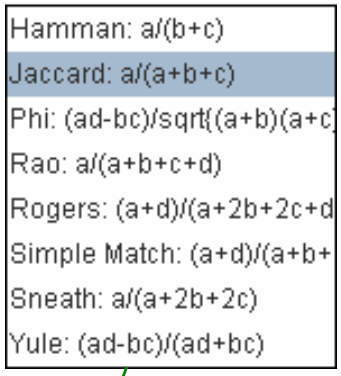
Number of X Discrete/Continuous Covariates



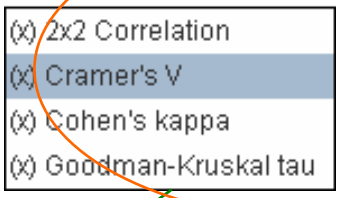
GAP Analysis Dialog

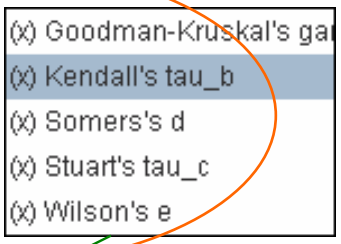


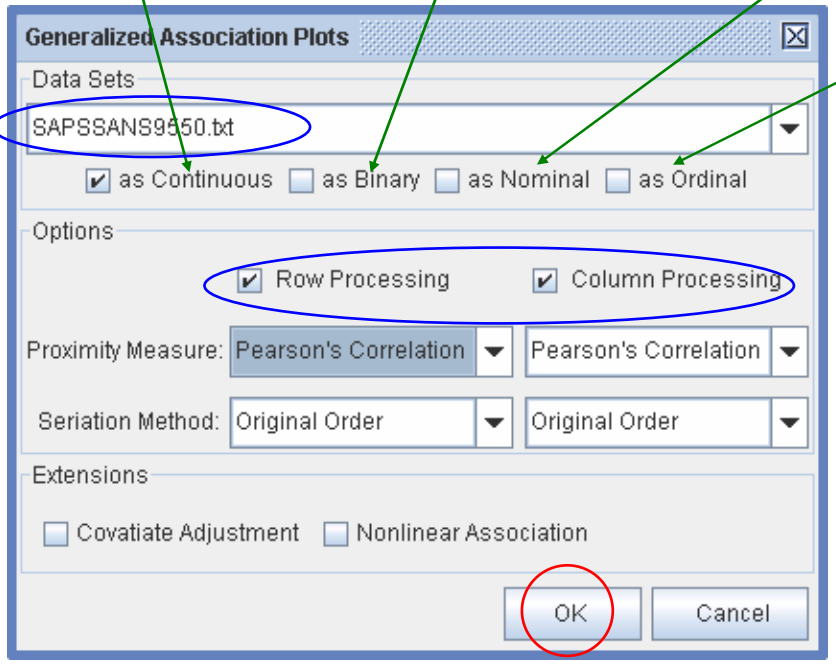




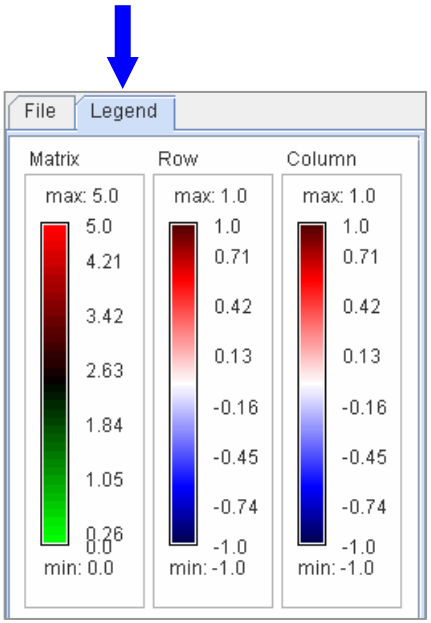
Not implemented Yet!

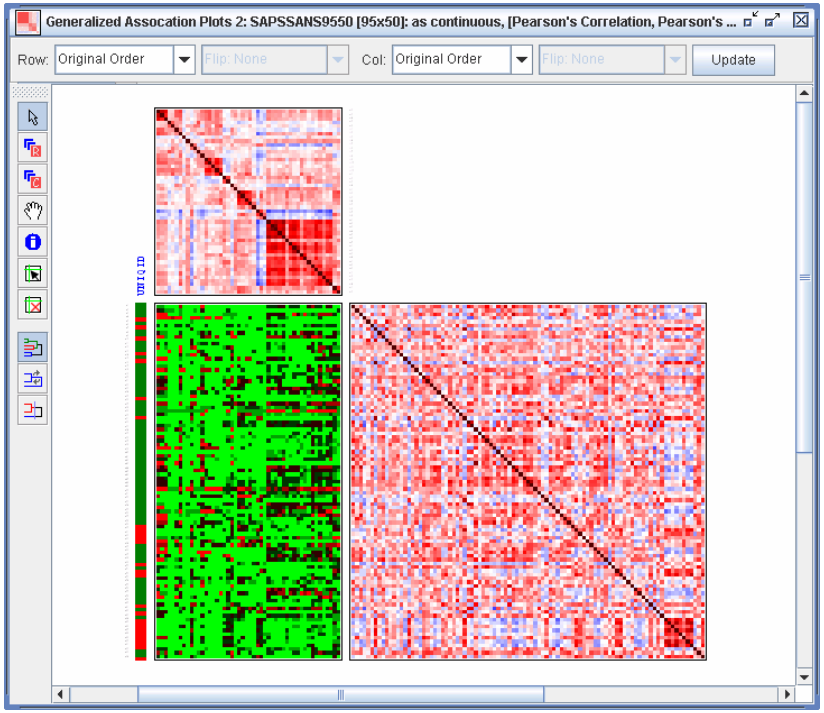




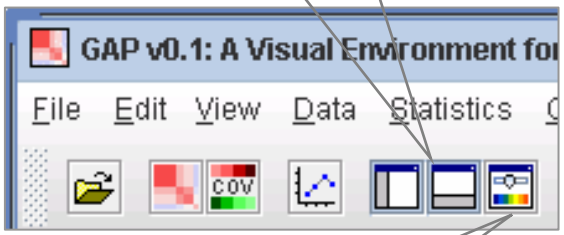


GAP Step 1:
Presentation of Raw Data Matrix

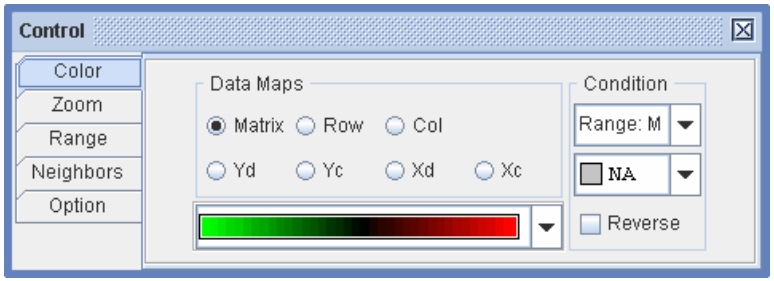




Show or Hide File/Data View, Output Panel



Show or Hide Control Panel



Control Panel: Color

Rainbow 130

Blue-white-Red 200

Green-Black-Red 38

Grey 256

Discrete 16

Biallel Marker

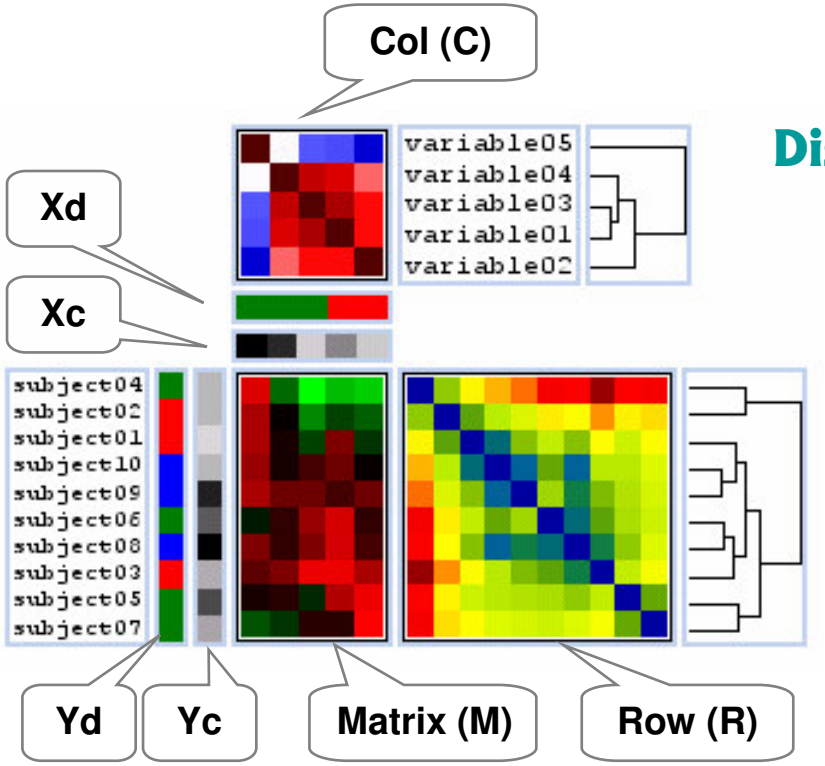
ACGT

Distance

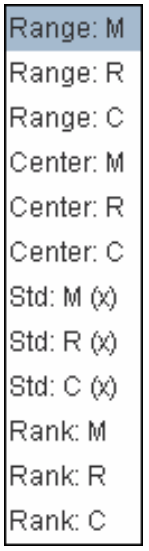
Correlation

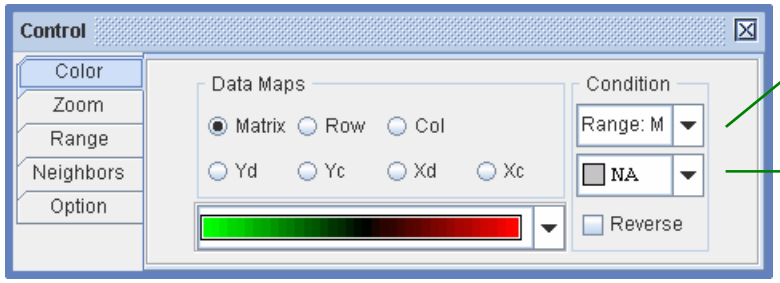
Gene Expression

Range

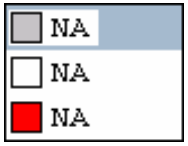


Display Conditions
(only for Matrix)



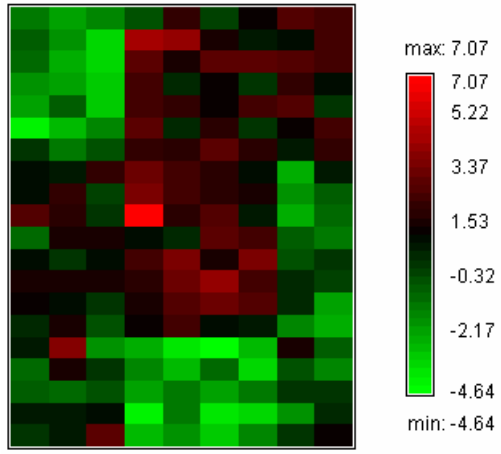


Missing Value

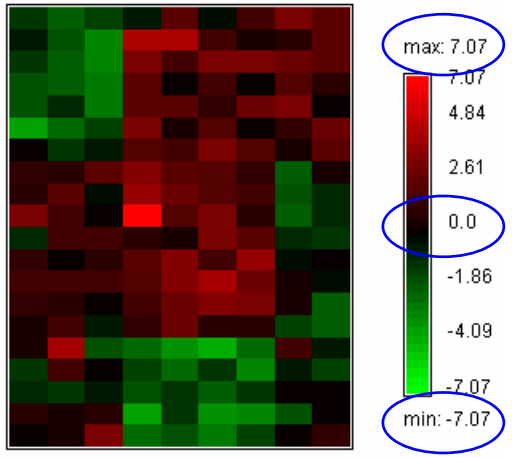


Display Conditions

	A	B	C	D	E	F	G	H	I
1	-1.37	-2.30	-1.80	-0.55	2.45	-0.13	1.49	3.03	2.48
2	-0.68	-2.11	-3.42	4.67	4.57	1.75	0.61	0.92	2.52
3	-1.19	-2.49	-3.66	3.14	1.70	3.29	3.33	2.92	2.48
4	-1.93	-2.28	-3.16	2.51	0.32	1.49	0.21	2.20	1.03
5	-2.21	-0.79	-3.29	2.55	2.44	1.45	2.68	3.03	0.19
6	-4.14	-2.91	-1.64	3.21	0.37	1.93	0.14	1.27	2.67
7	0.21	-1.36	-0.44	2.22	1.85	3.11	2.03	0.67	2.40
8	1.13	0.79	2.25	3.65	2.52	2.09	1.13	-2.59	0.67
9	0.95	2.33	-0.07	3.89	2.72	2.13	1.75	-2.17	-0.90
10	3.04	1.85	0.21	7.07	2.01	3.05	0.76	-2.58	-1.04
11	-1.02	1.65	1.53	0.95	0.60	3.12	2.52	-0.77	-1.40
12	1.21	0.24	1.04	2.50	3.69	1.81	3.98	-0.33	0.11
13	1.74	1.60	1.70	2.02	3.45	4.46	2.69	0.41	-0.09
14	1.34	1.06	0.06	1.81	2.90	3.64	3.04	0.49	-2.33
15	0.57	1.81	-0.47	1.40	2.70	0.99	0.82	-1.61	-2.56
16	0.61	4.22	-2.03	-2.61	-4.00	-4.64	-2.92	1.55	-0.71
17	-1.13	1.64	0.01	-1.77	-2.85	-1.24	-3.41	-0.59	-1.64
18	-0.86	-1.17	-0.41	-2.20	-1.30	-2.37	-1.41	0.08	0.25
19	0.75	0.66	1.04	-4.26	-1.41	-3.99	-3.53	-2.17	0.34
20	0.15	0.68	3.18	-2.86	-2.01	-3.18	-1.58	0.10	1.28



Range Matrix Condition

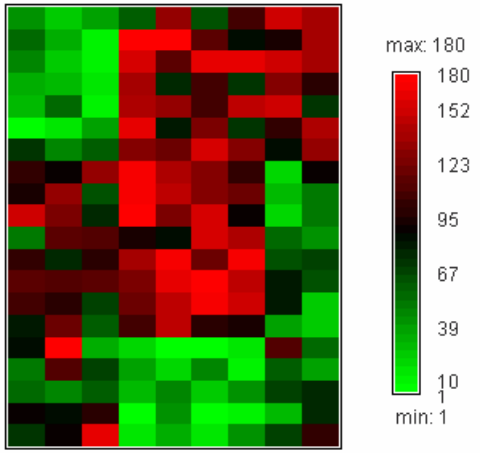


Center Matrix Condition

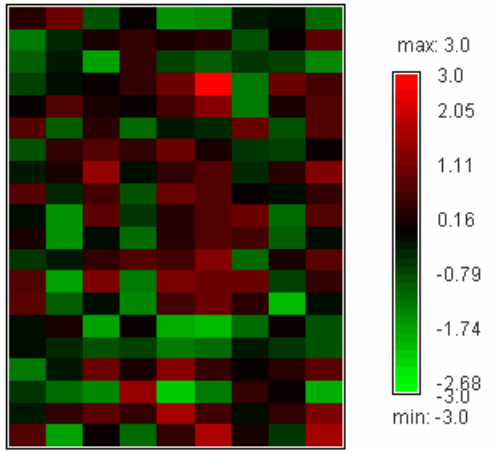
Bidirectional Color Spectrum

Blue-white-Red 200

Green-Black-Red 38

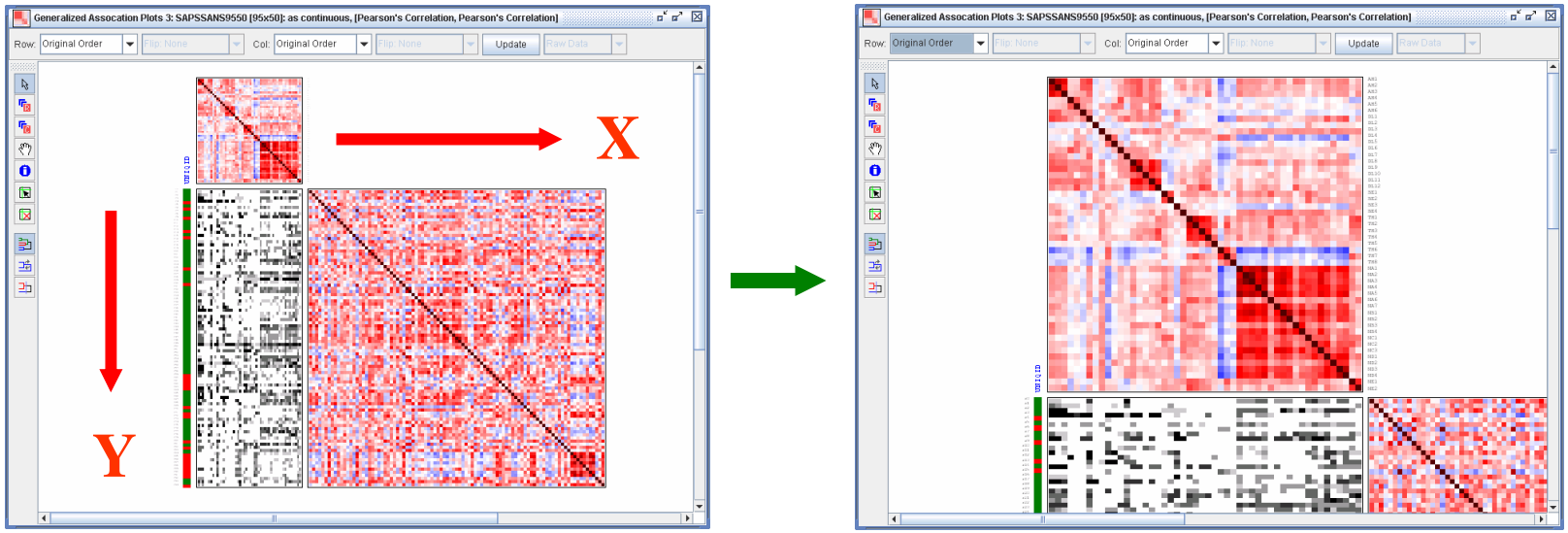


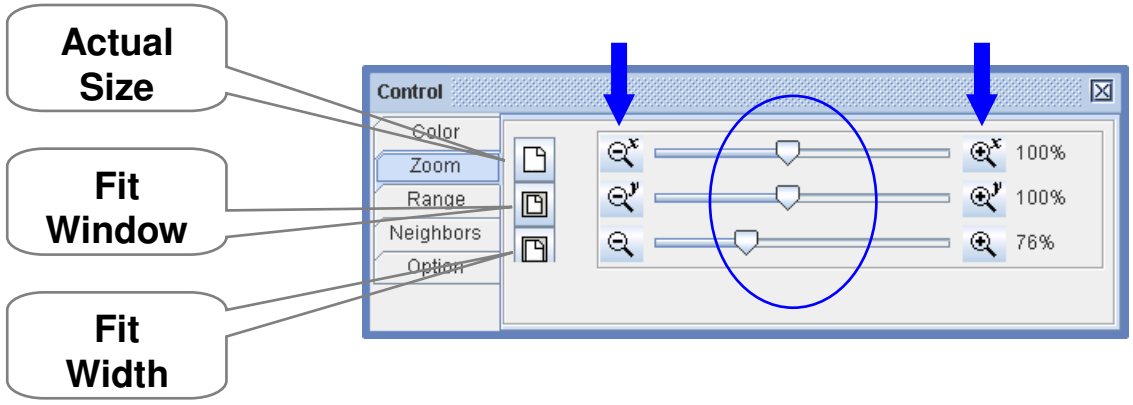
Rank Matrix Condition



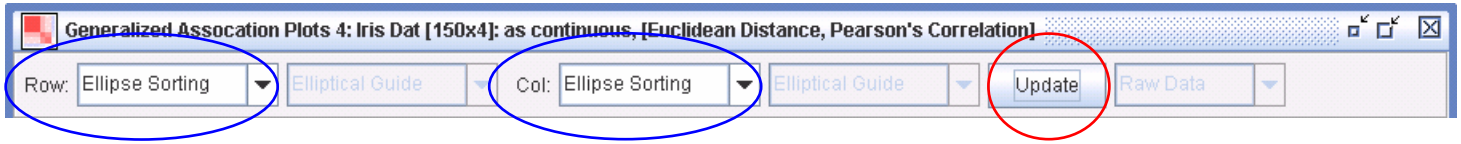
Standardize Matrix Condition

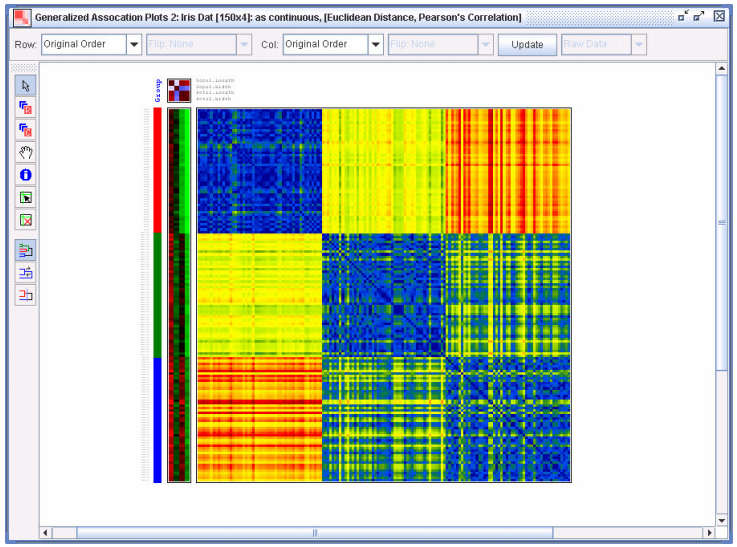
Control Panel: Zoom

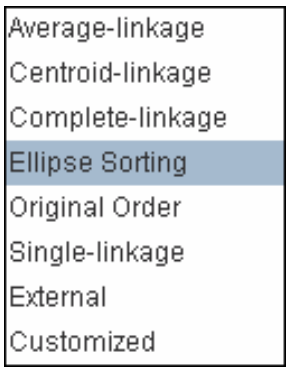


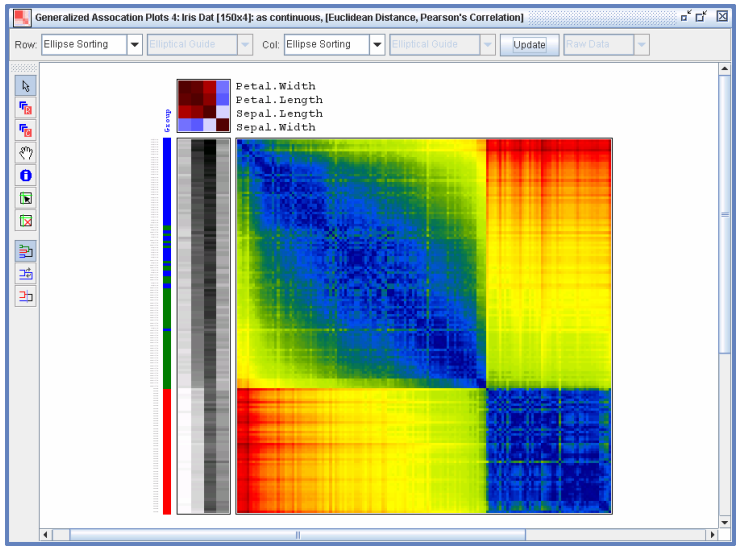


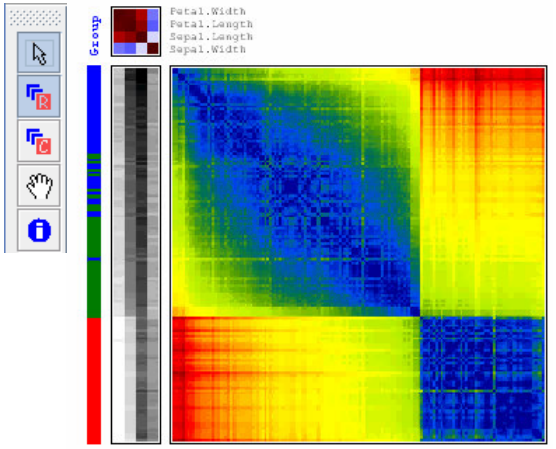
GAP Step 2: Ellipse Seriation

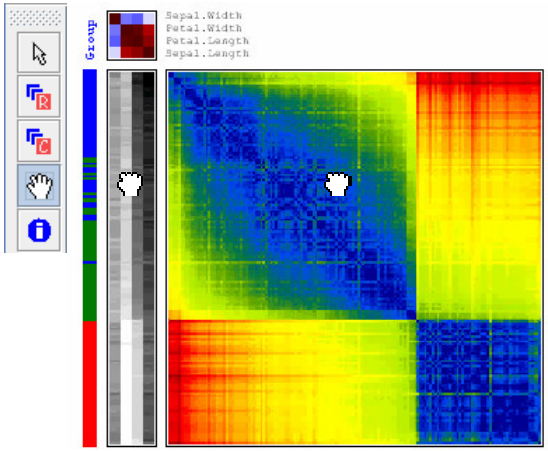


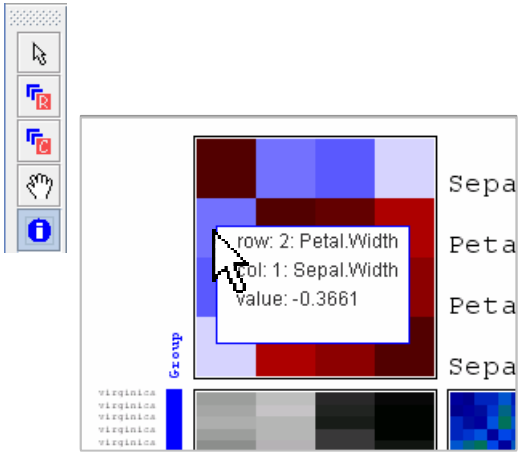




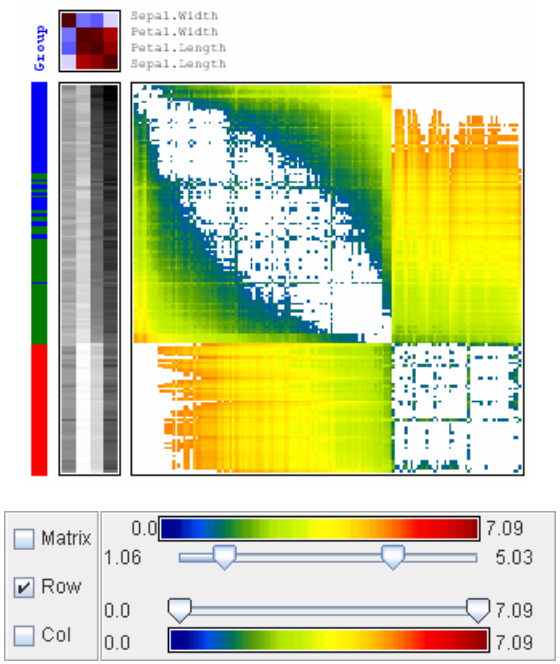


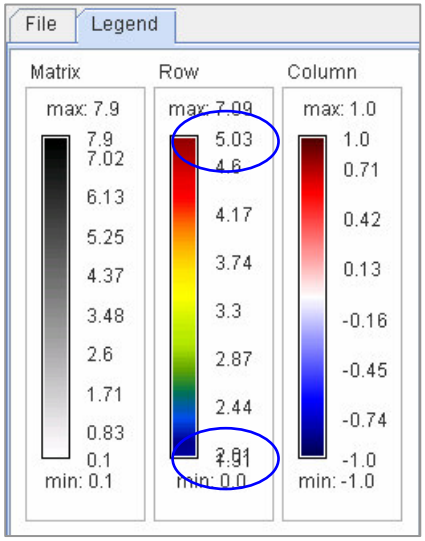


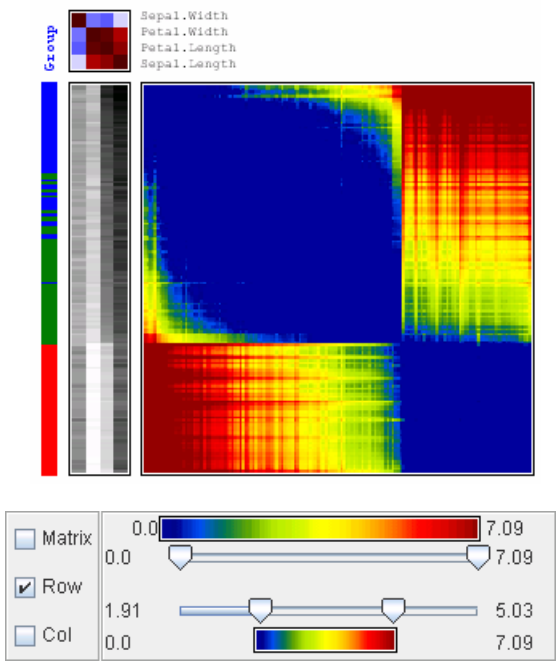




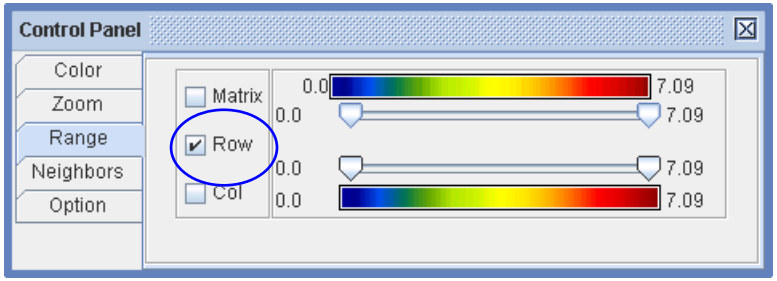
Control Panel: Range



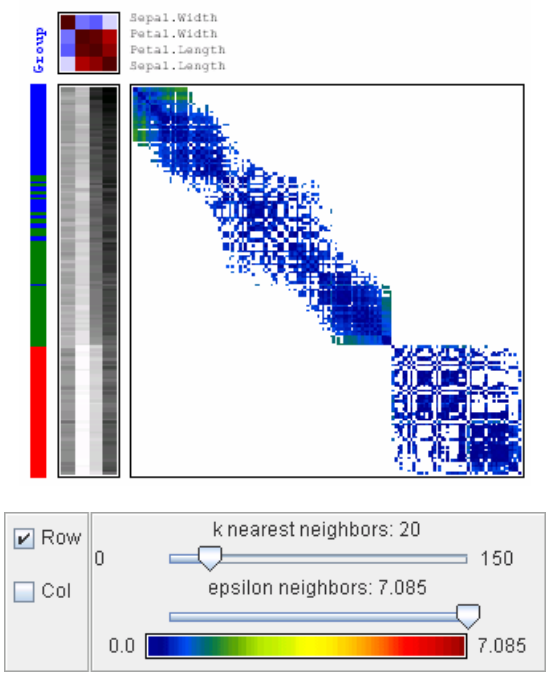


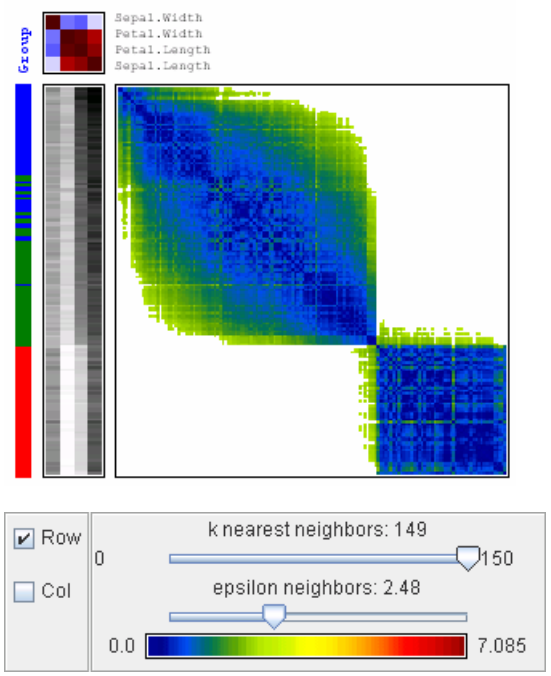


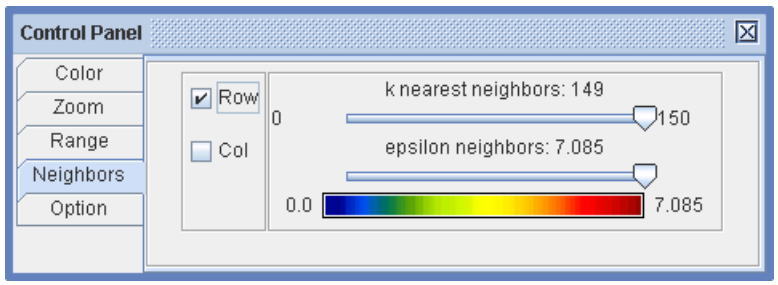
"Resolution" of a Statistical Graph



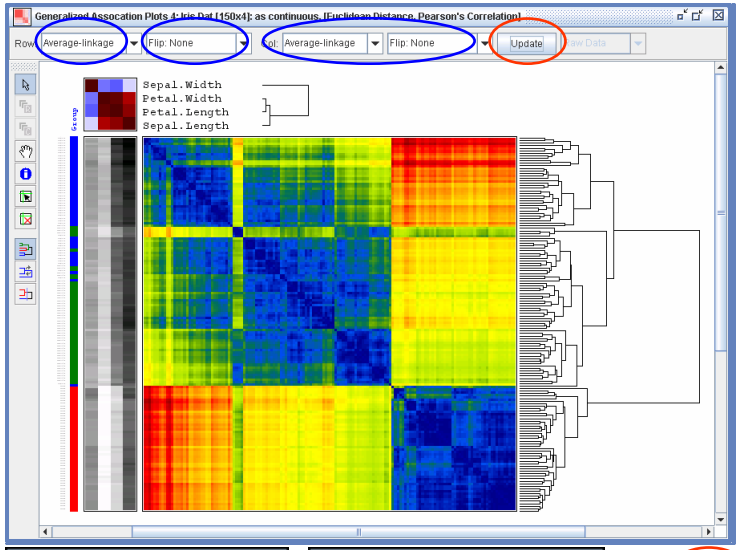
Control Panel: Neighbors

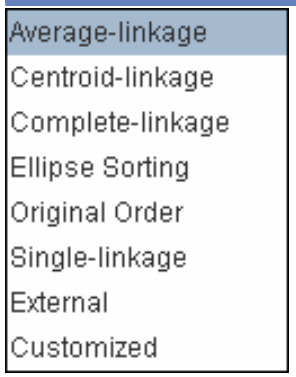


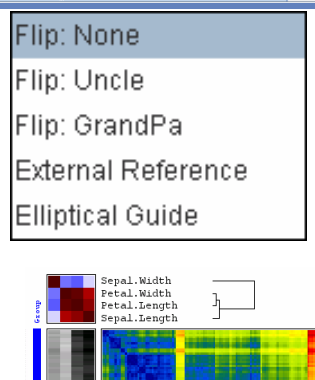


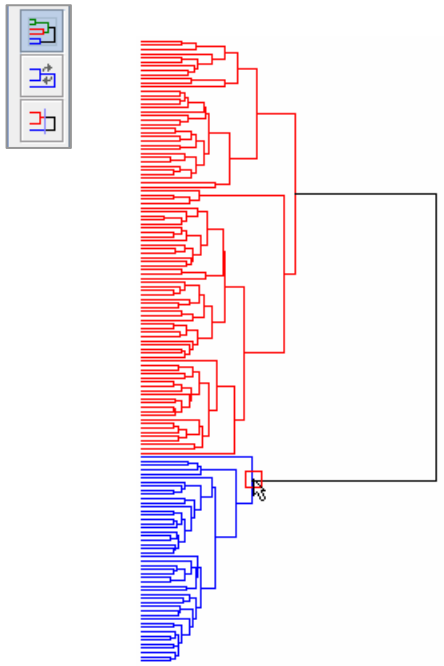


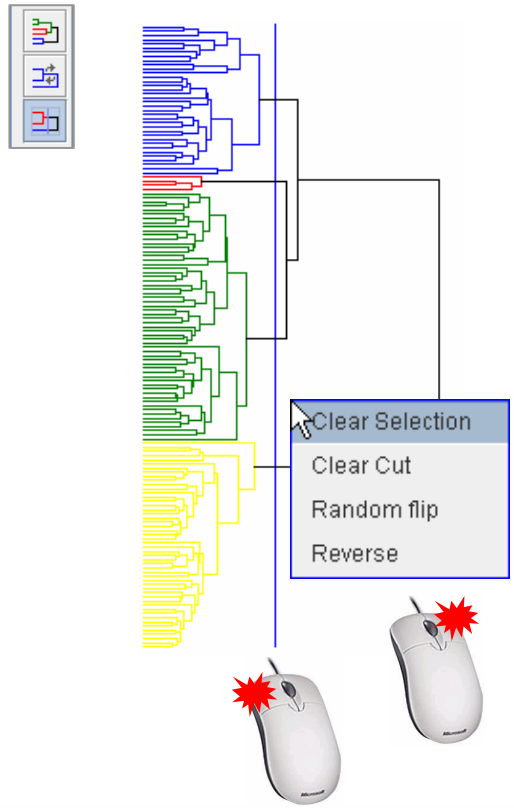
GAP Step 2: Tree Seriation/Flip

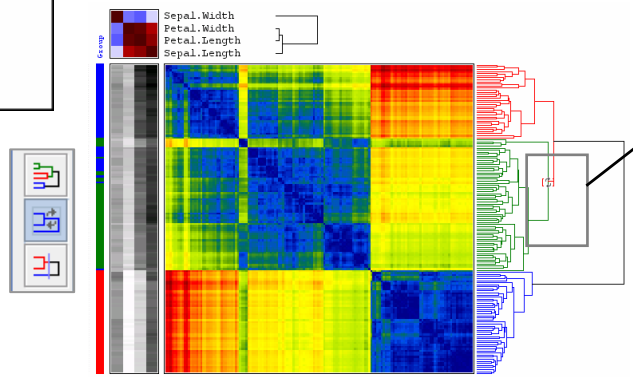


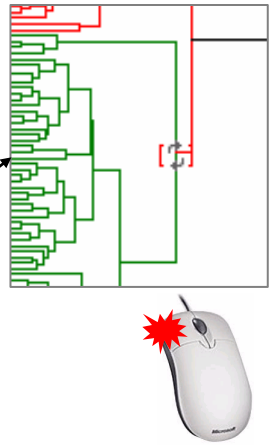


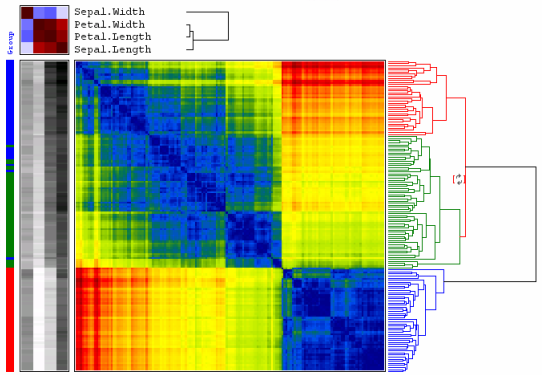




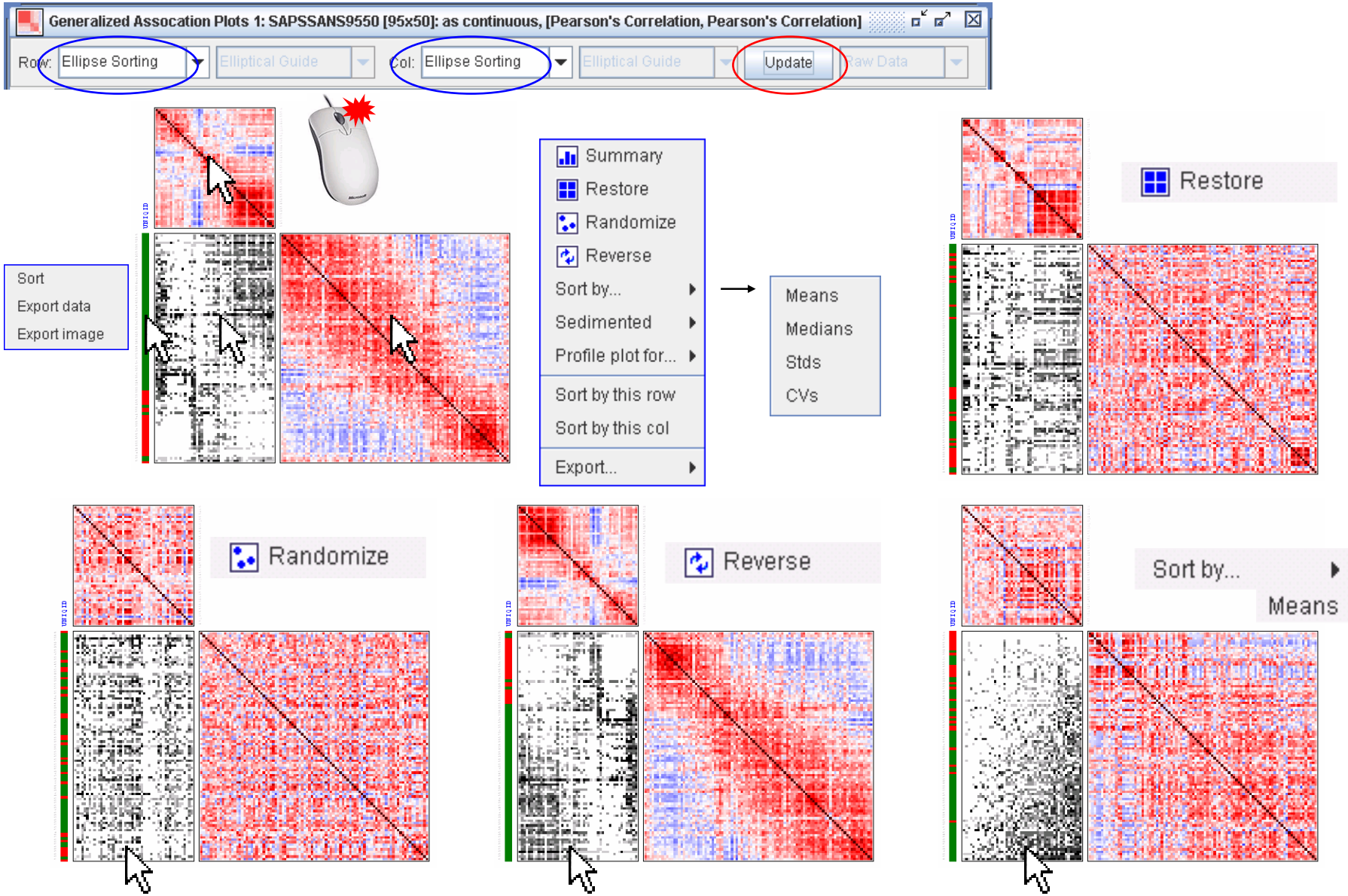




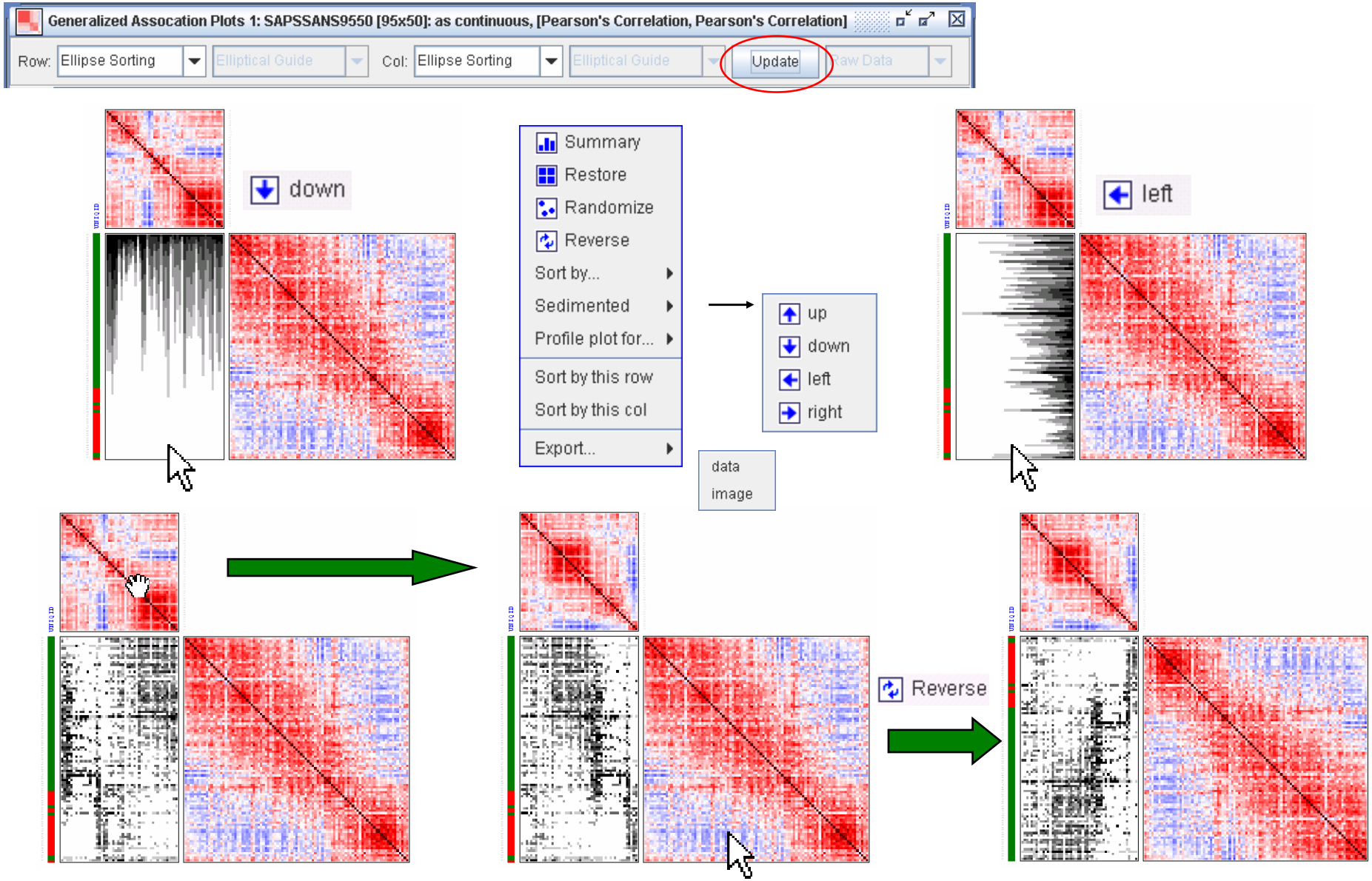




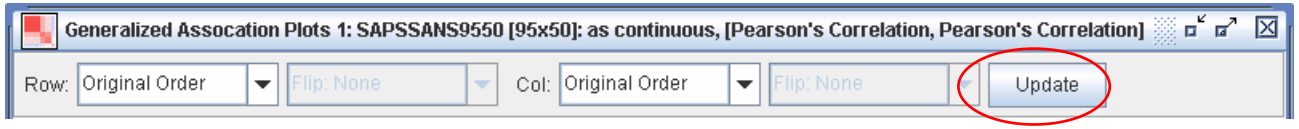
Pop-Up Menu: Sort by…

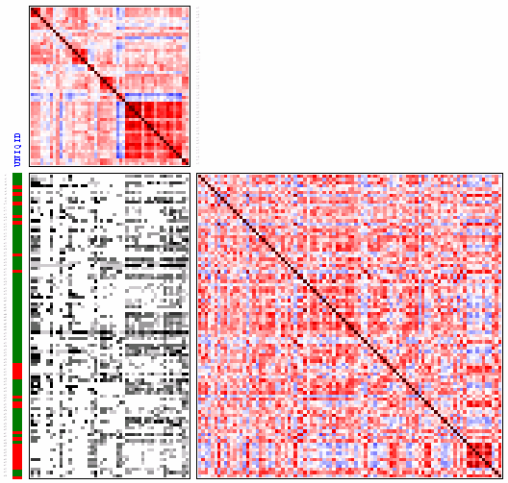


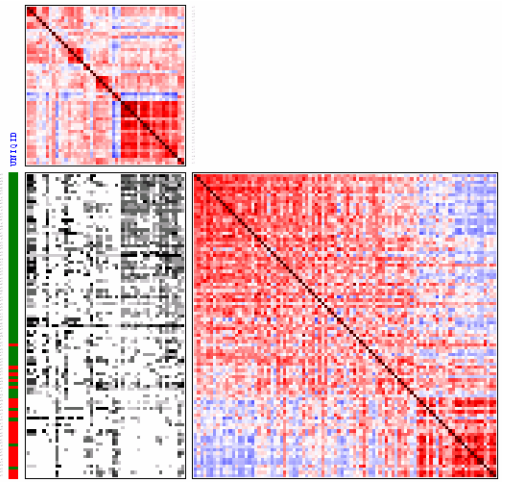
Pop-Up Menu: Sedimented plot



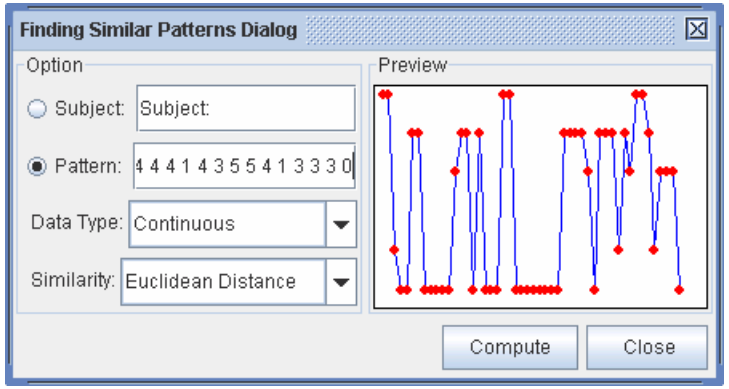
Mouse Tool: Find Similar Patterns



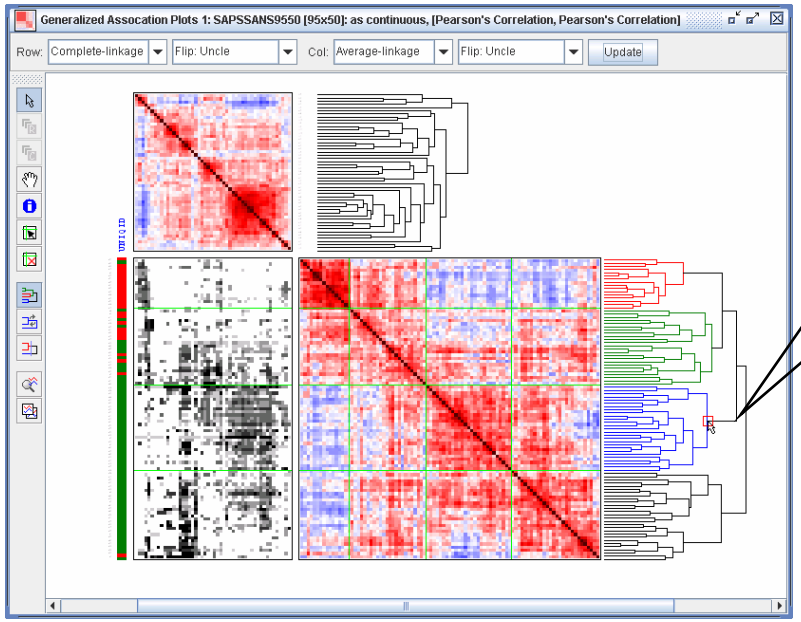




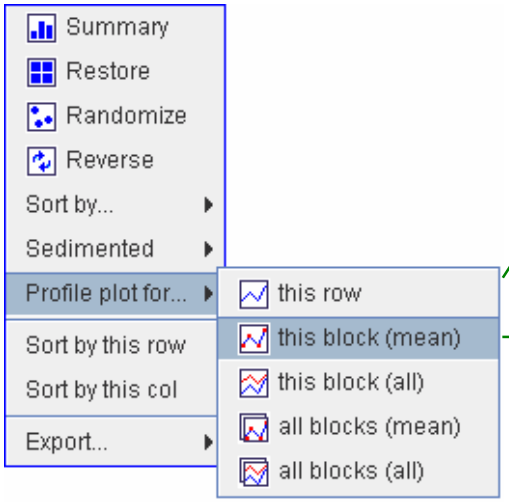
Input pattern

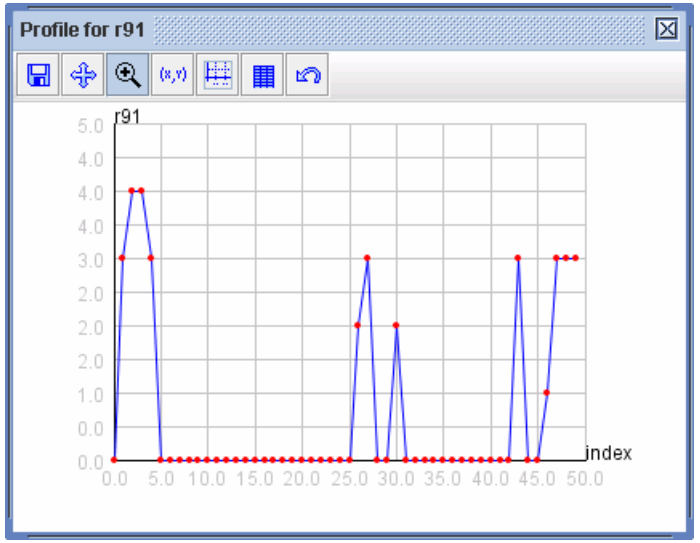


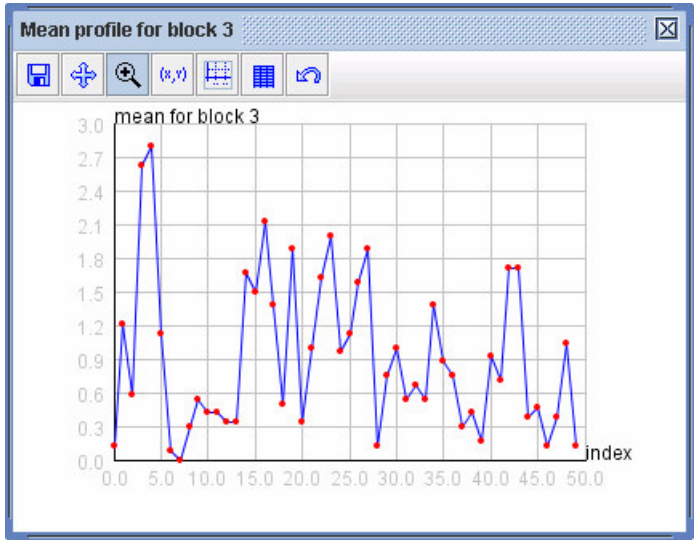
Pop-Up Menu: Profile plot



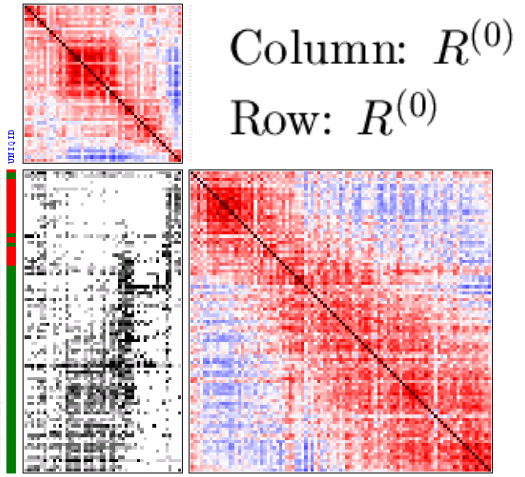
Select subtree





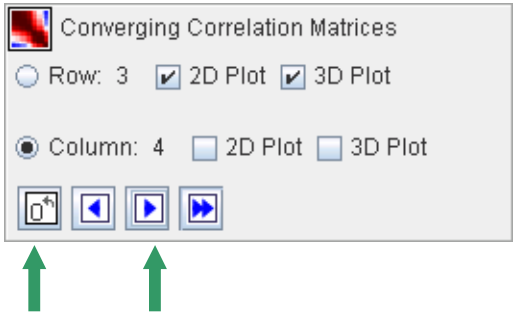


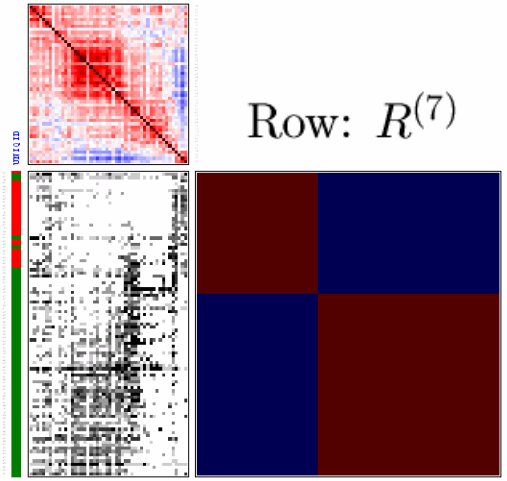
Converging Correlation Matrices



Column: $R^{(0)}$

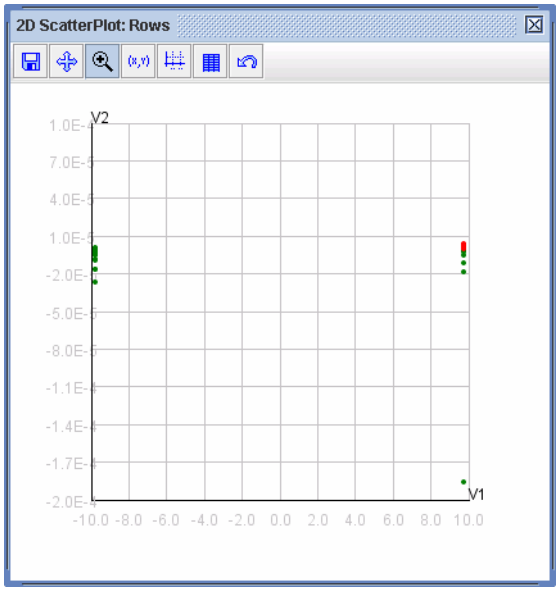
Row: $R^{(0)}$

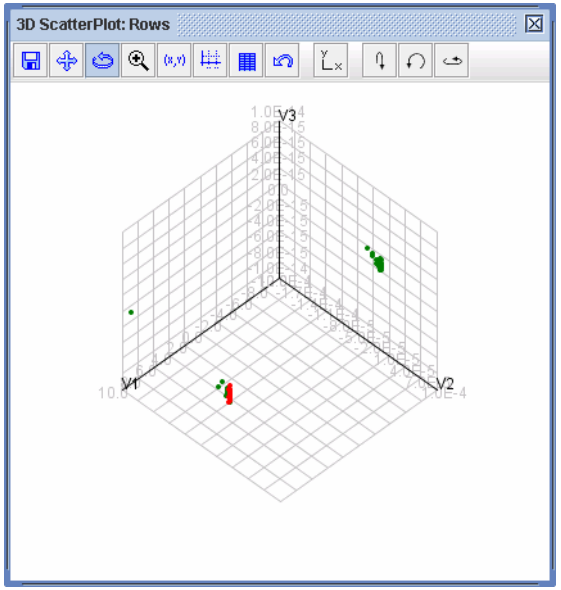


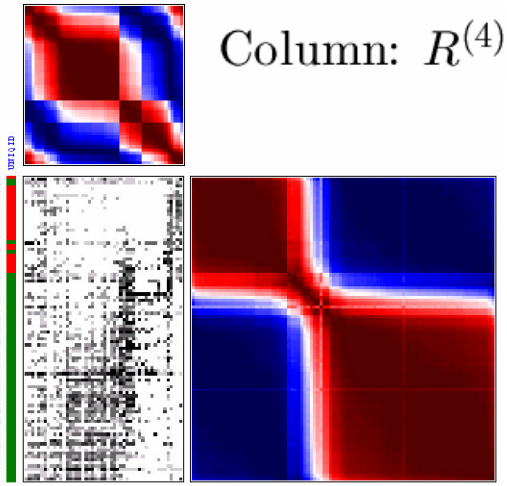


Row: $R^{(7)}$

Row: $R^{(7)}$

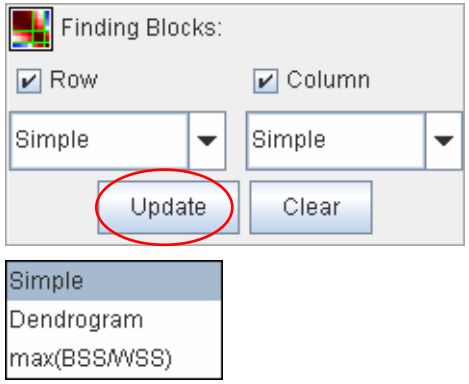


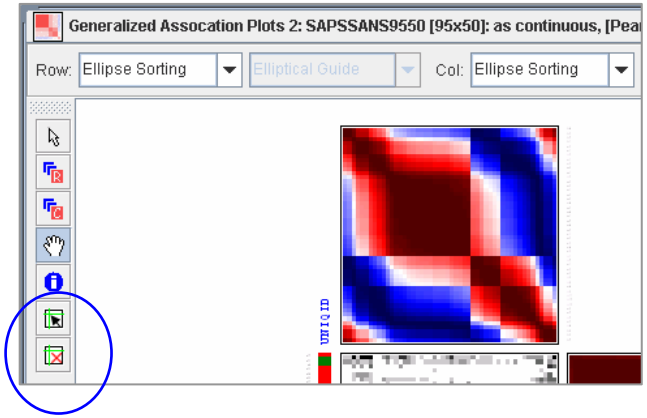


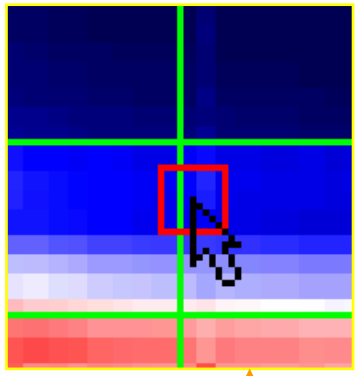


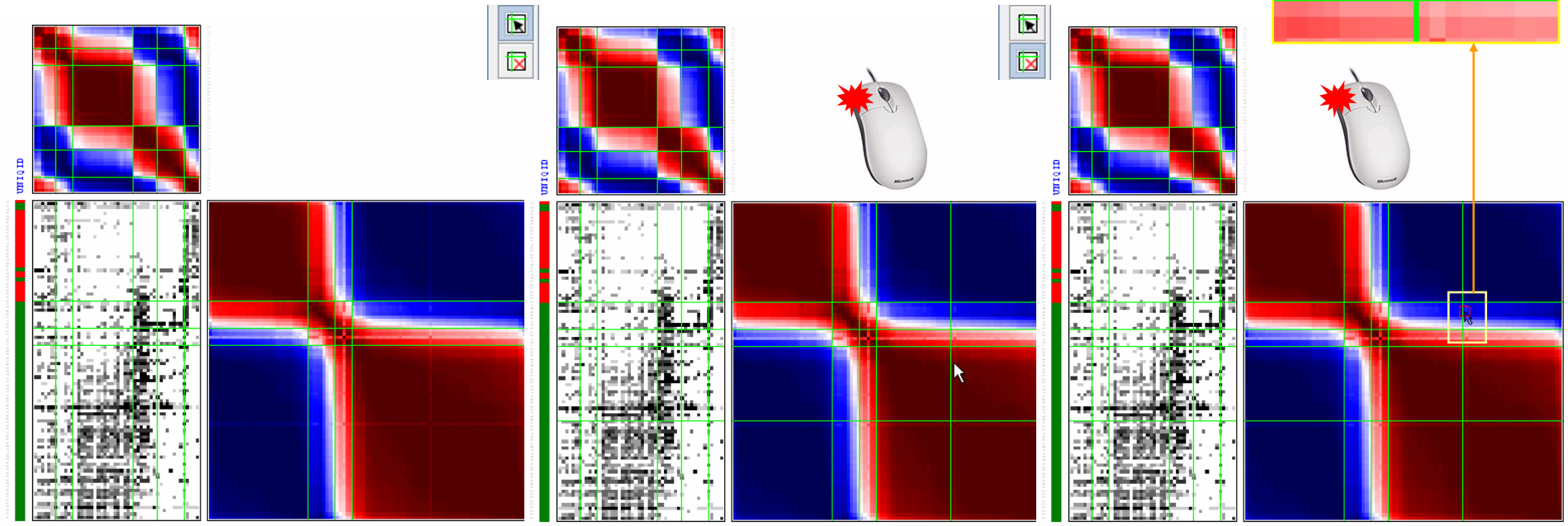
Column: $R^{(4)}$

GAP Step 3:
Partitions of Permuted Matrix Maps

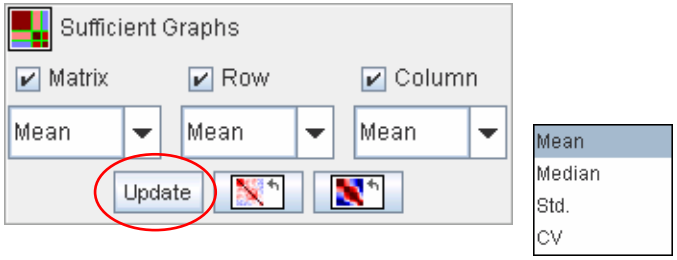


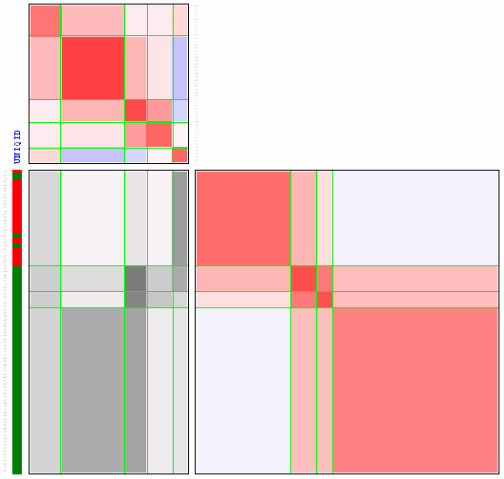


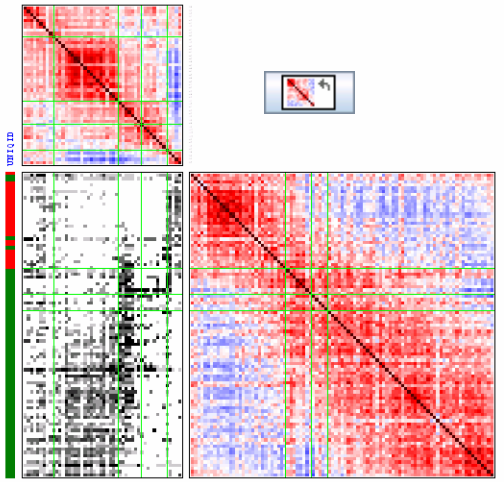


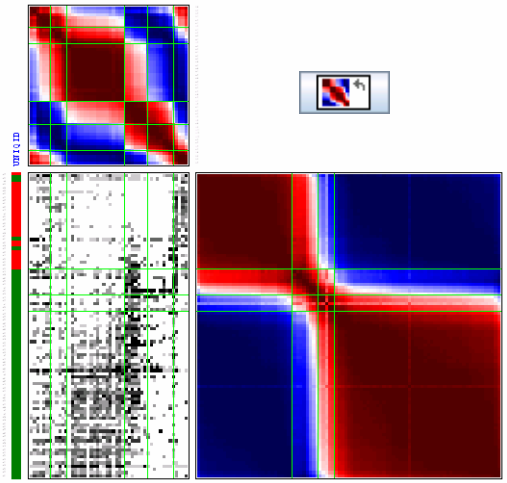


GAP Step 4: Sufficient Graph

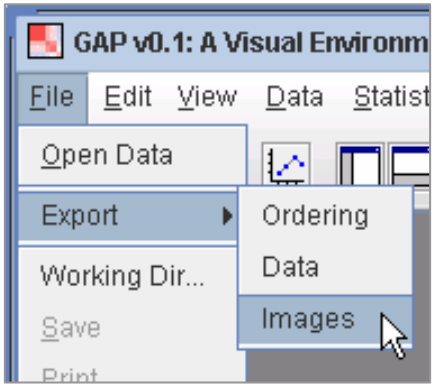






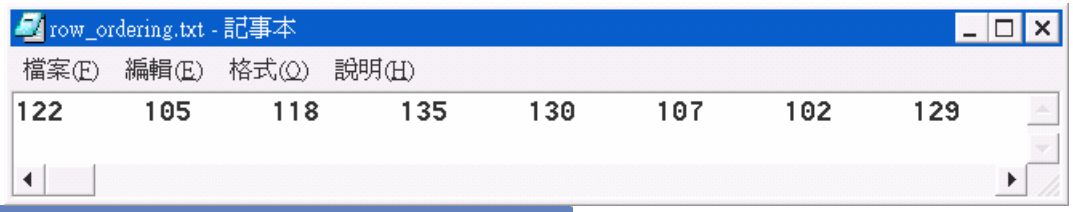


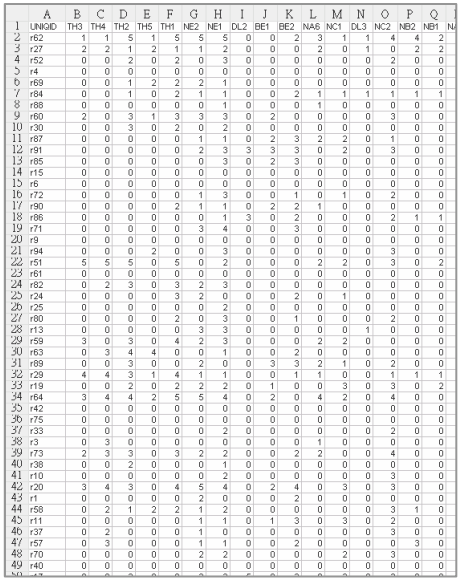
Export Ordering, Data, Images

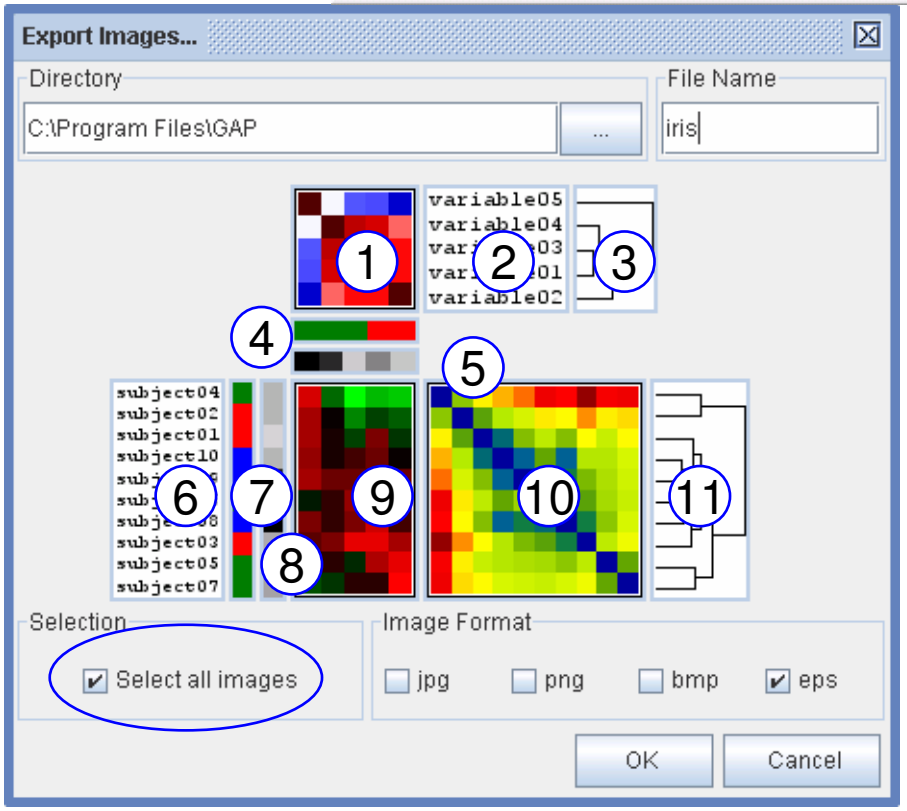


row_ordering.txt

col_ordering.txt

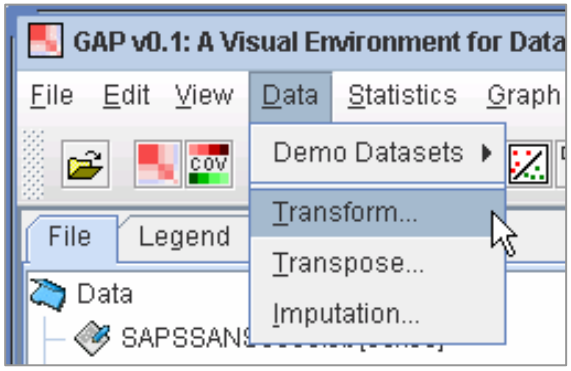


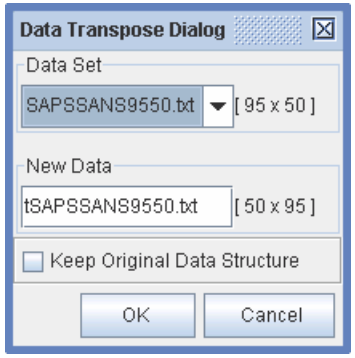


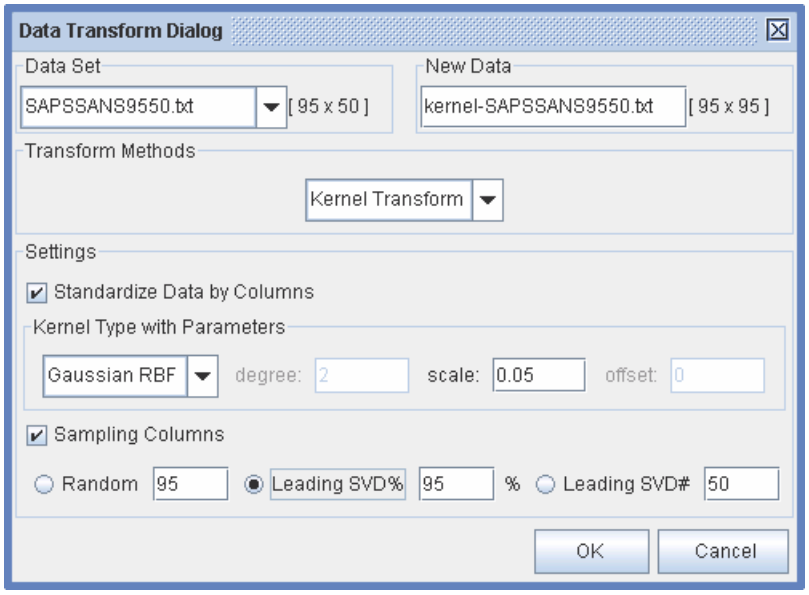


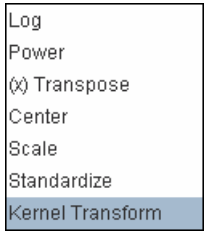
1. _gapCol.eps
2. _gapColName.eps
3. _TreeCol.eps
4. _gapXd.eps
5. _gapXc.eps
6. _gapRowName.eps
7. _gapYd.eps
8. _gapYc.eps
9. _gapMatrix.eps
10. _gapRow.eps
11. _TreeRow.eps

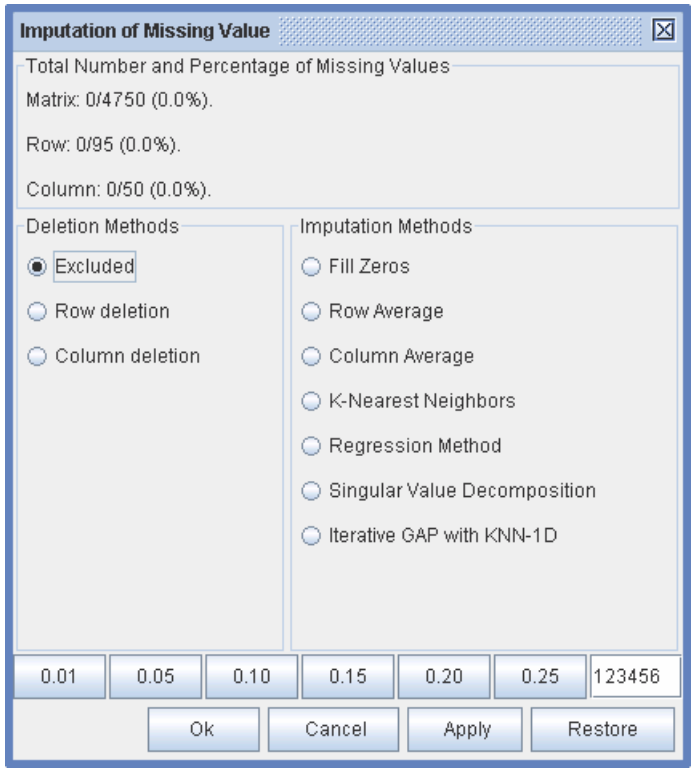
Data Processing



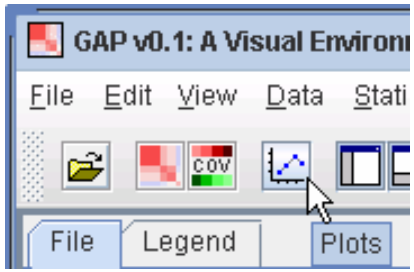


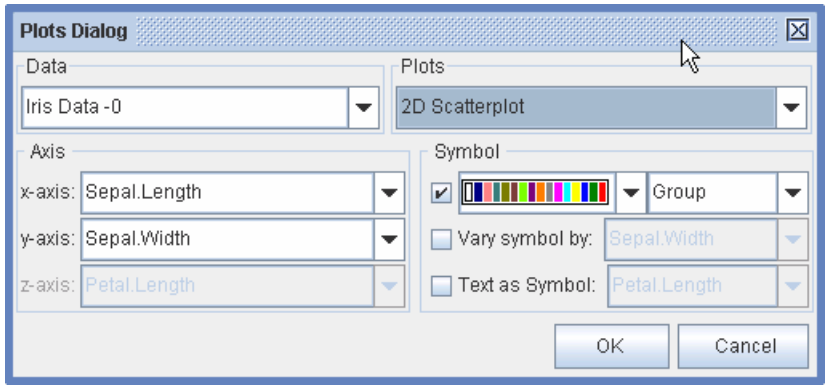


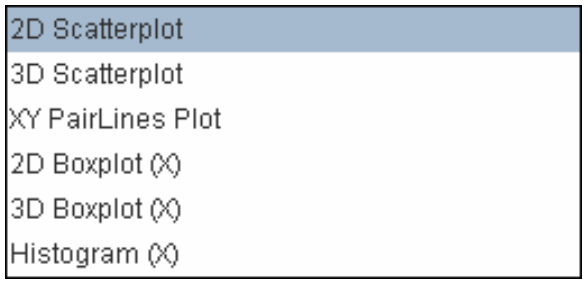


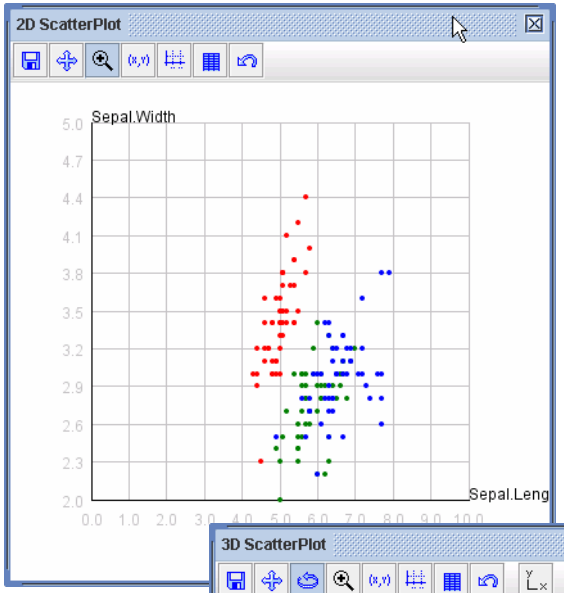


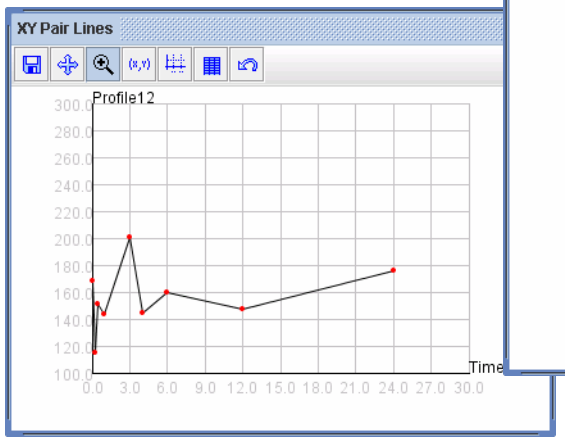
Plots

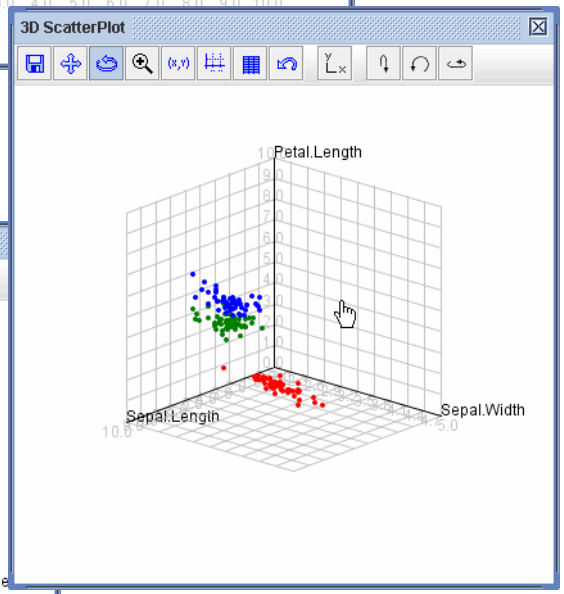

GAP v0.1: A Visual Environ
File Edit View Data Stati
cov
File Legend Plots


Plots Dialog
Data
Iris Data -0
Plots
2D Scatterplot
Axis
x-axis: Sepal.Length
y-axis: Sepal.Width
z-axis: Petal.Length
Symbol
Group
Vary symbol by: Sepal.Width
Text as Symbol: Petal.Length
OK
Cancel

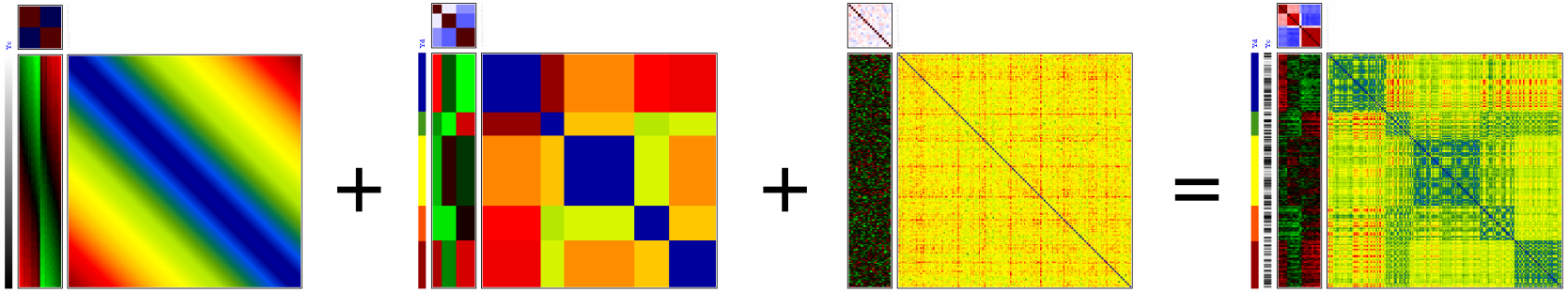

2D Scatterplot
3D Scatterplot
XY PairLines Plot
2D Boxplot (X)
3D Boxplot (X)
Histogram (X)


2D ScatterPlot
Sepal.Width
Sepal.Leng

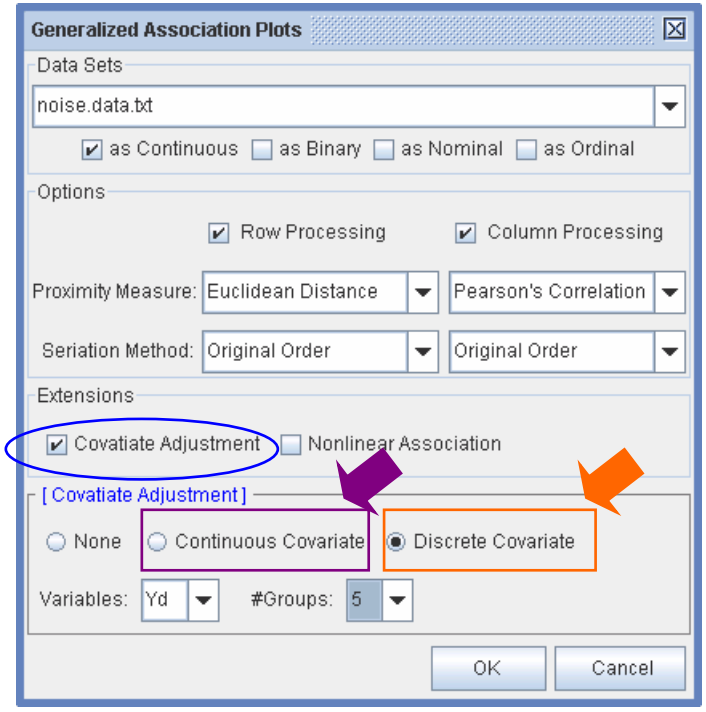

XY Pair Lines
Profile12
Time

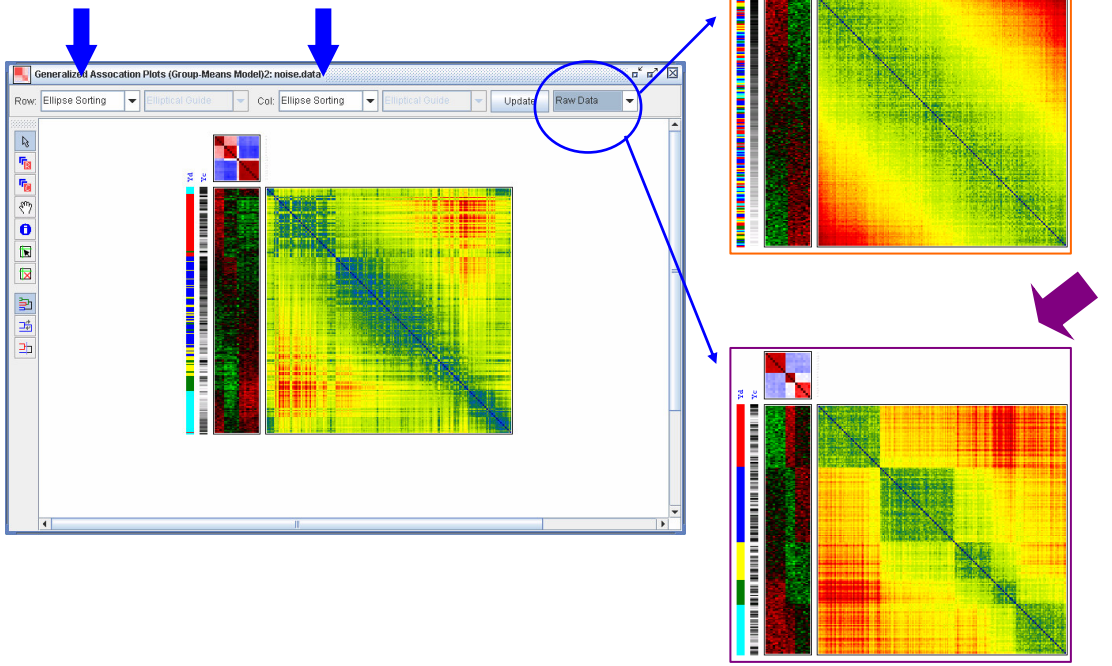

3D ScatterPlot
Petal.Length
Sepal.Length
Sepal.Width

GAP with Covariate Adjustment

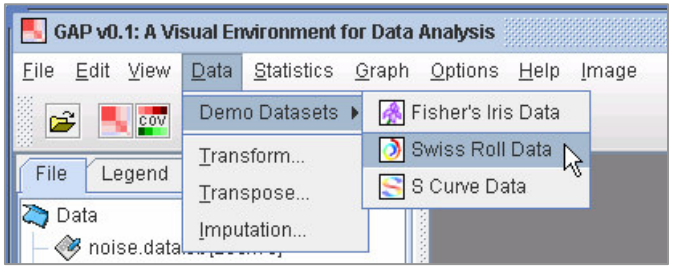


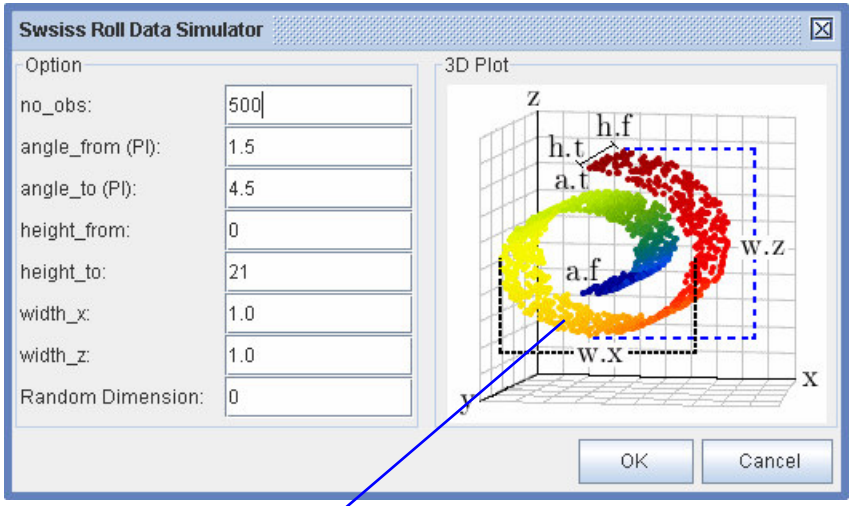
Data (model): continuous pattern + **Data (model): discrete pattern** + **Gaussian noise** = **Noisy data: mixed pattern**

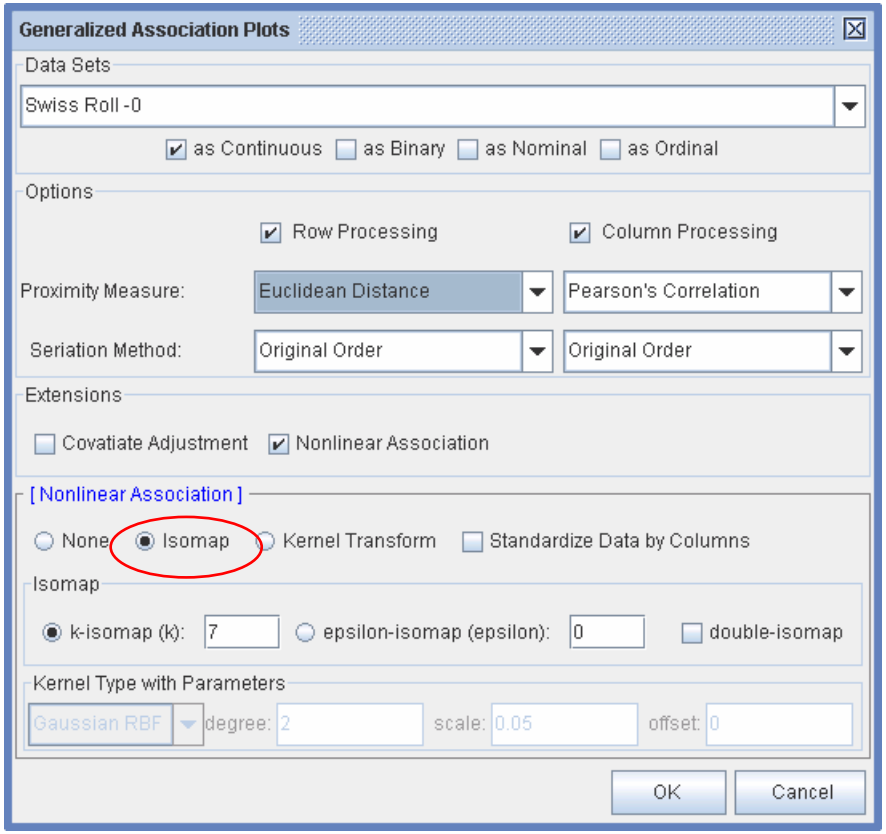




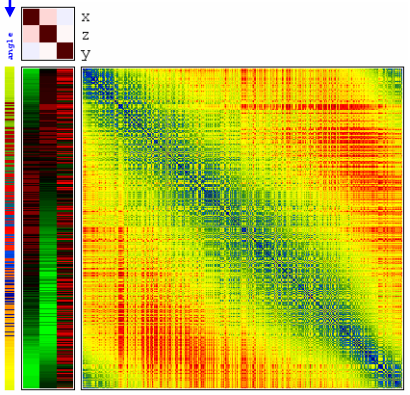
GAP with Nonlinear Association



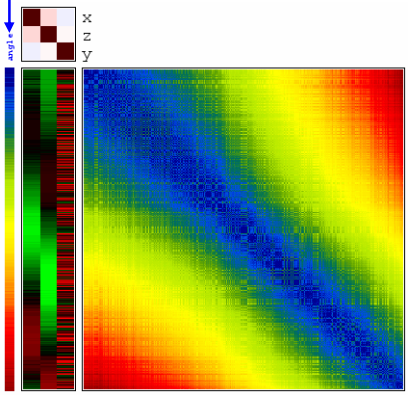




Recover the path

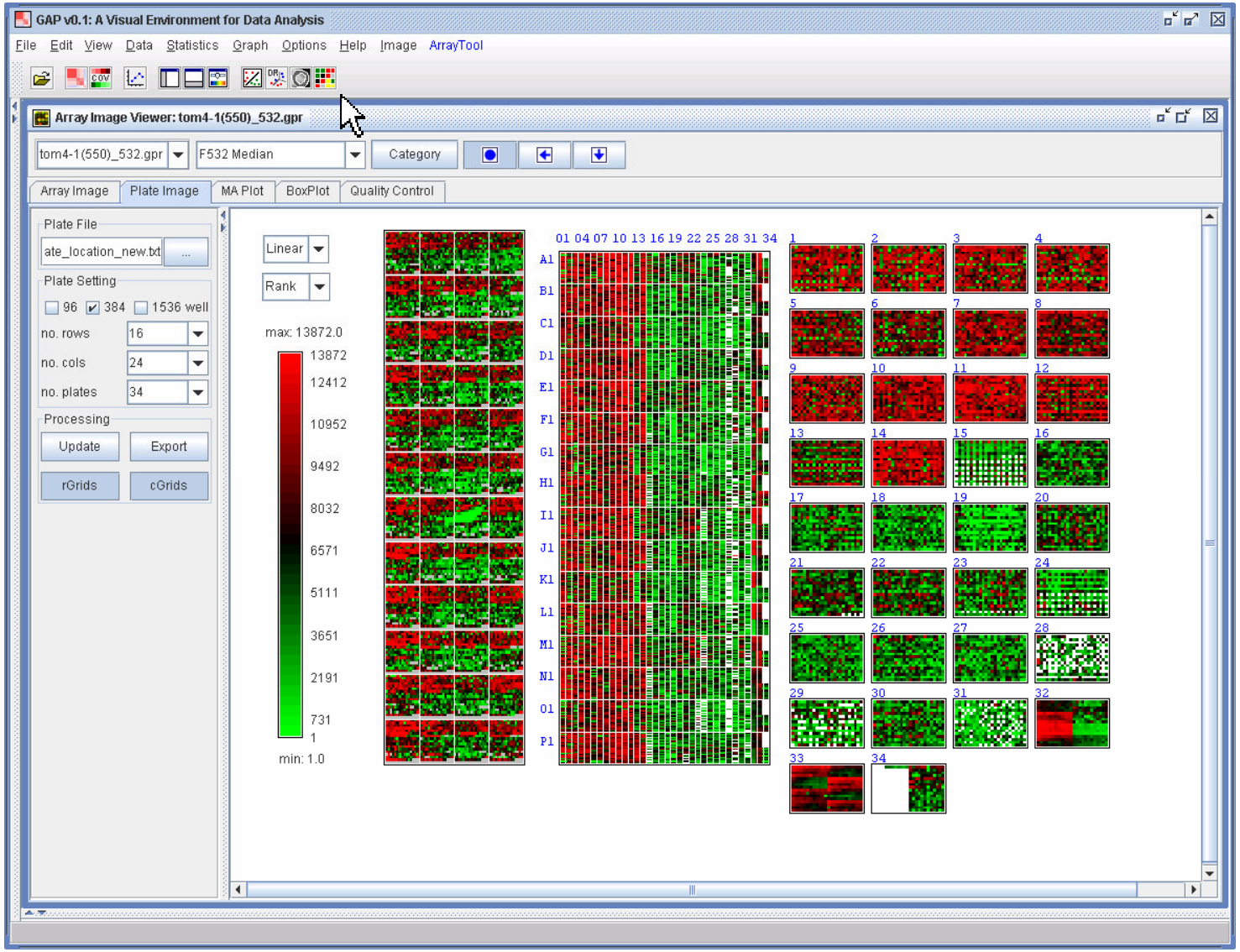


GAP



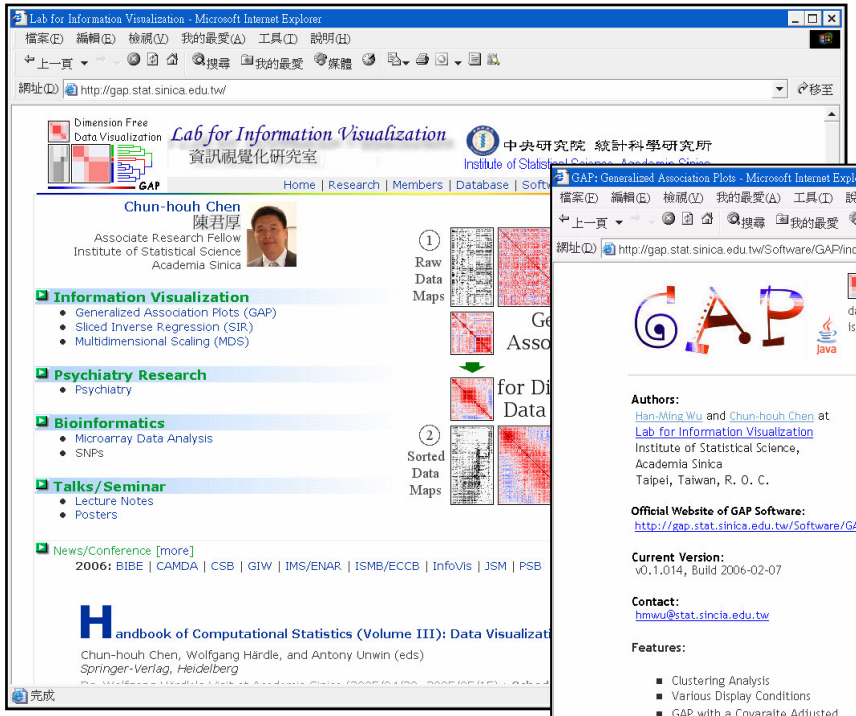
isoGAP

GAP Module: Microarray Tool



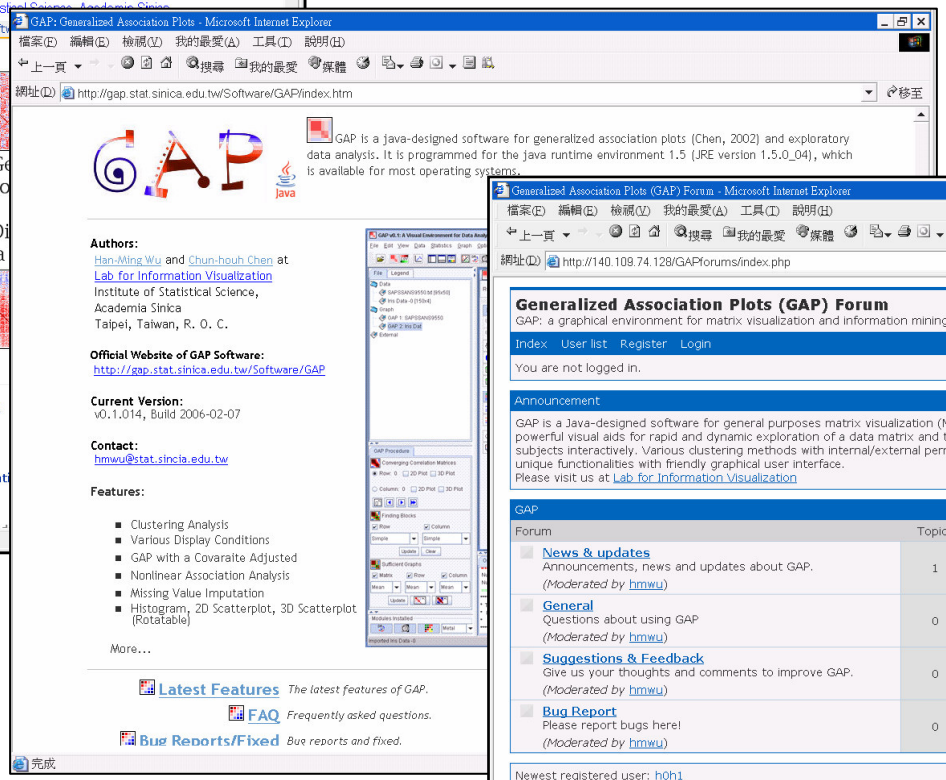
Web Site

Lab for Information Visualization



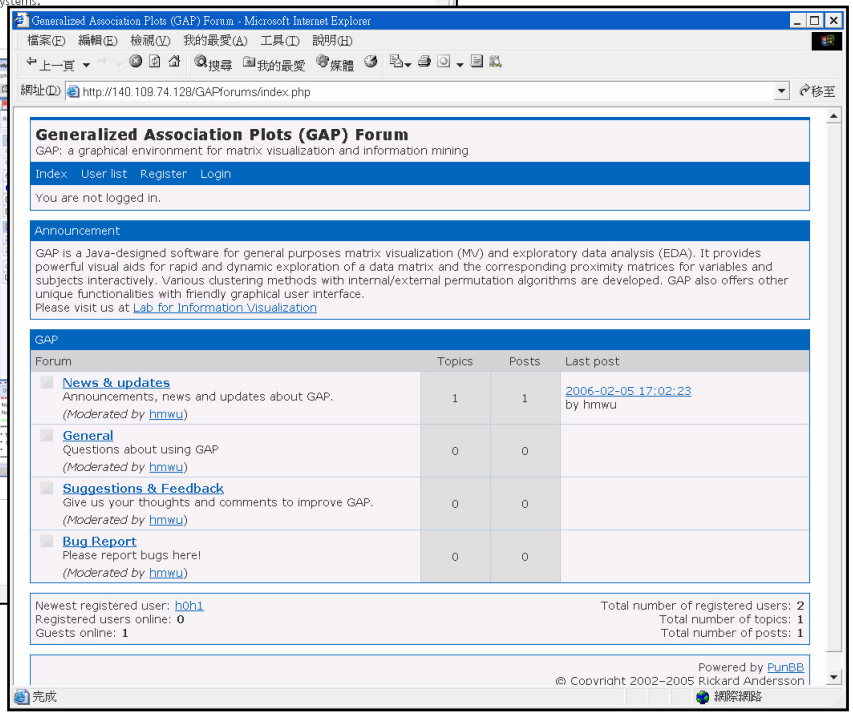
http://gap.stat.sinica.edu.tw

GAP Software



http://gap.stat.sinica.edu.tw/Software/GAP

GAP Forum



http://gap.stat.sinica.edu.tw/GAPforums