

GAP Software Tutorial

Version 0.1 Build20051215

Han-Ming Wu and Chun-Houh Chen

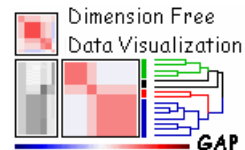
(吳漢銘，陳君厚)

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



0001001000100100100100

001001000100100



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

GAP Features

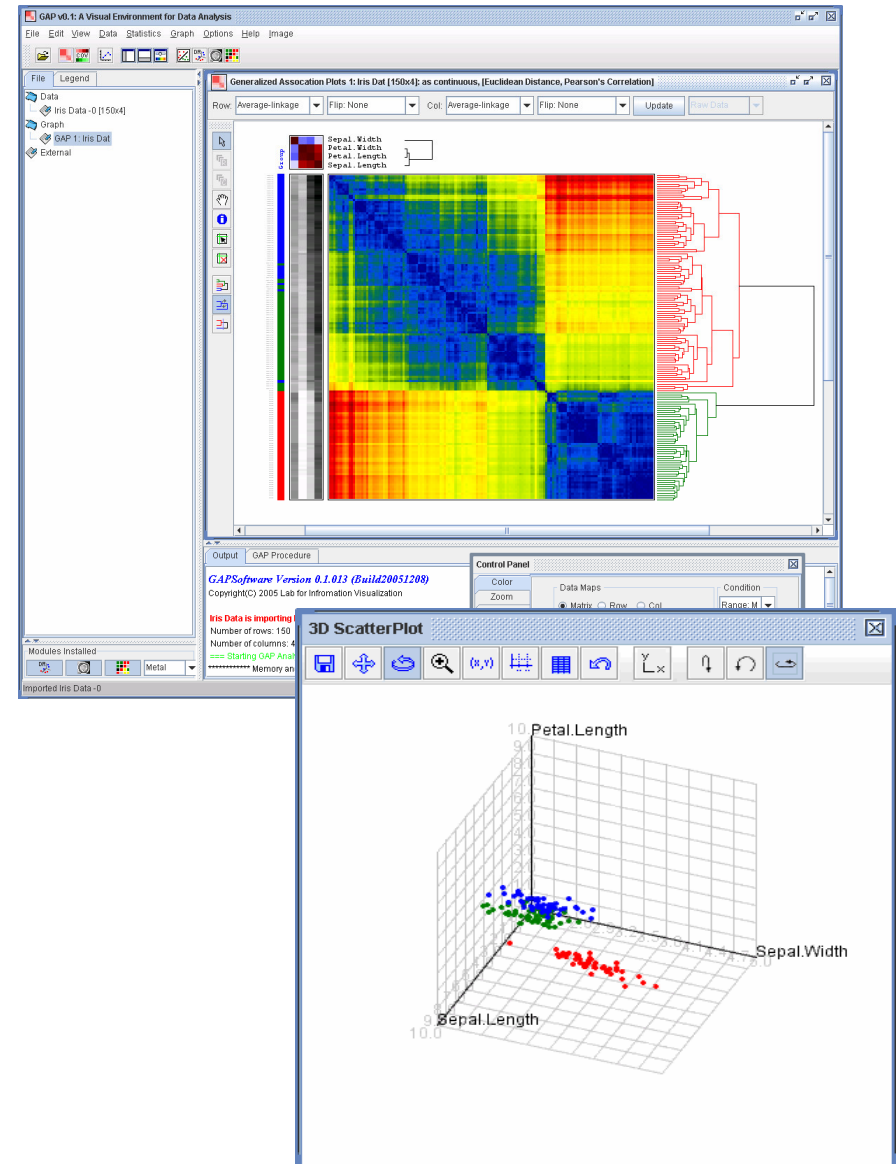


Generalized Association Plots

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

Statistical Plots

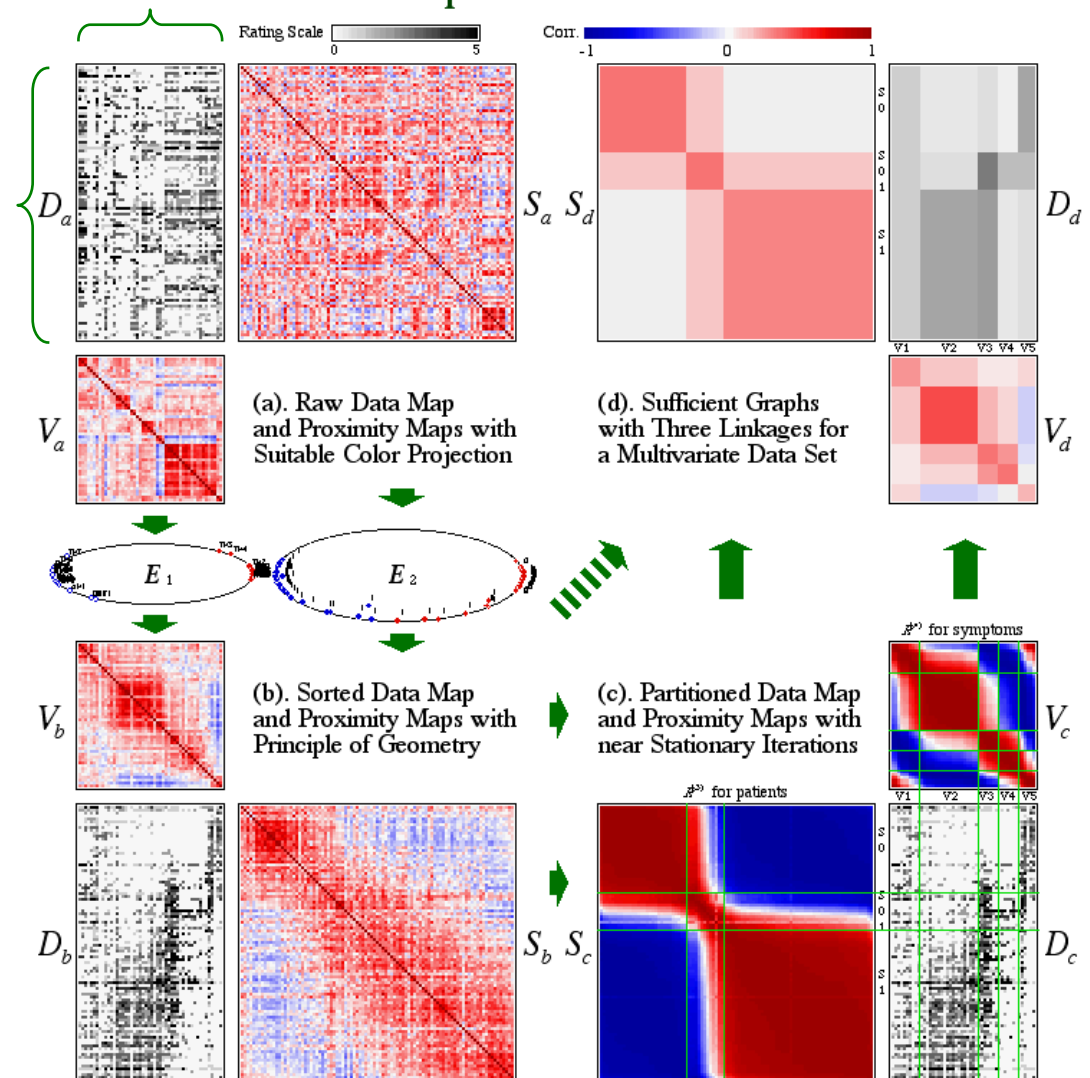
- 2D Scatterplot, 3D Scatterplot (Rotatable)



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 of Positive 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**).

A Complete GAP Procedure



Demo Data 2: Iris Data

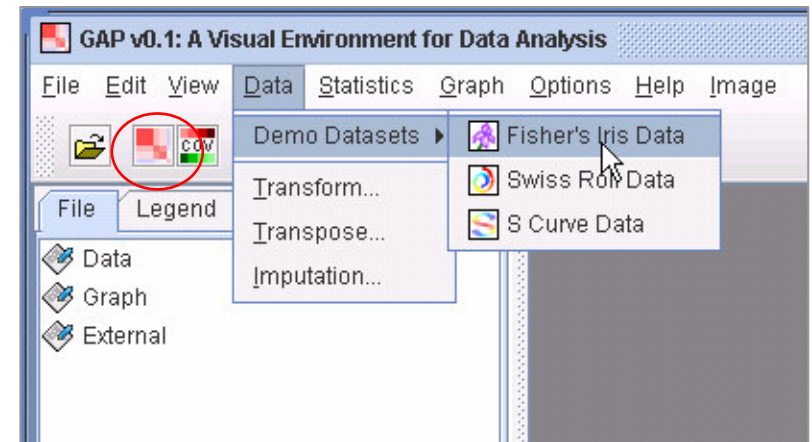
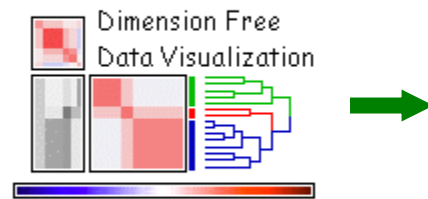
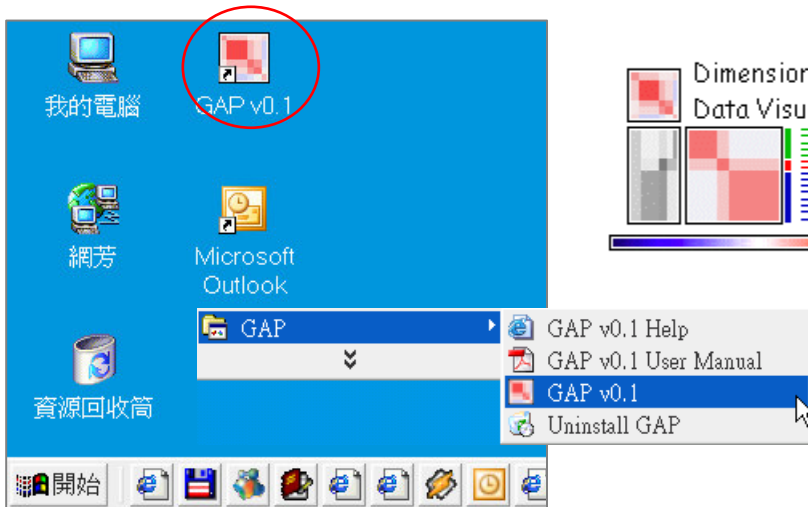
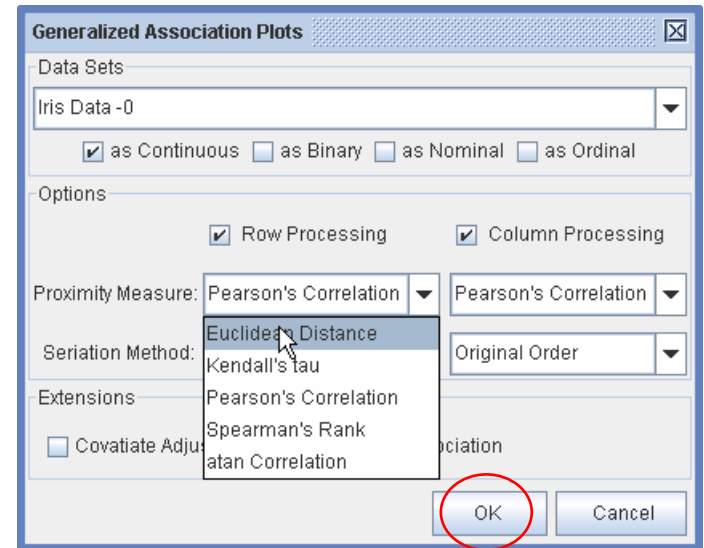
no.	Species	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
1	setosa	5.1	3.5	1.4	0.2
2	setosa	4.9	3.0	1.4	0.2
3	setosa	4.7	3.2	1.3	0.2
4	setosa	4.6	3.1	1.5	0.2
5	setosa	5.0	3.6	1.4	0.2
...					
76	versicolor	6.6	3.0	4.4	1.4
...					
150	virginica	5.9	3.0	5.1	1.8

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

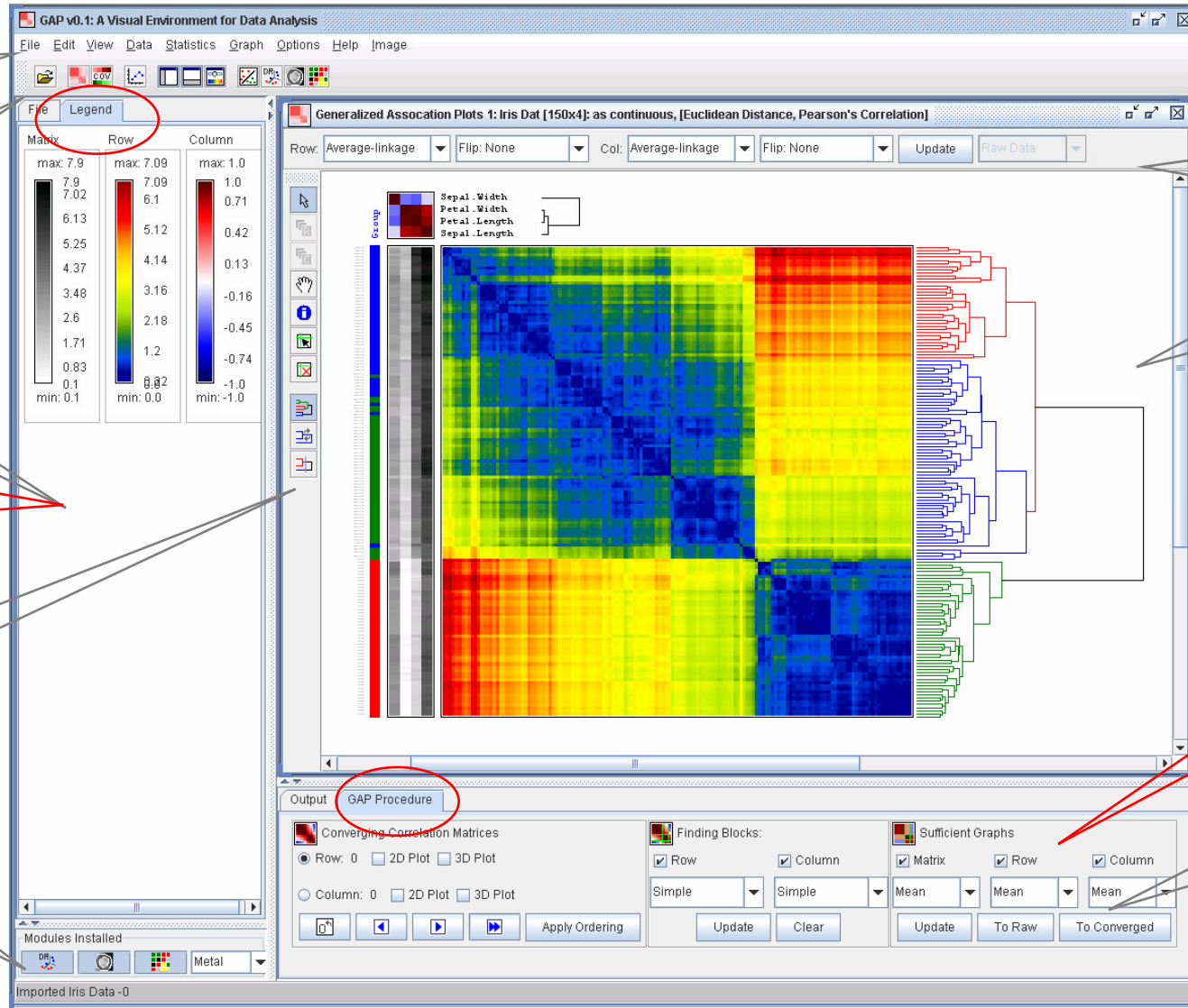
Tool Bar

File/Data View

Color Legend

Mouse Tools

Modules Installed



Serialization and Tree Flip

GAP Maps Window

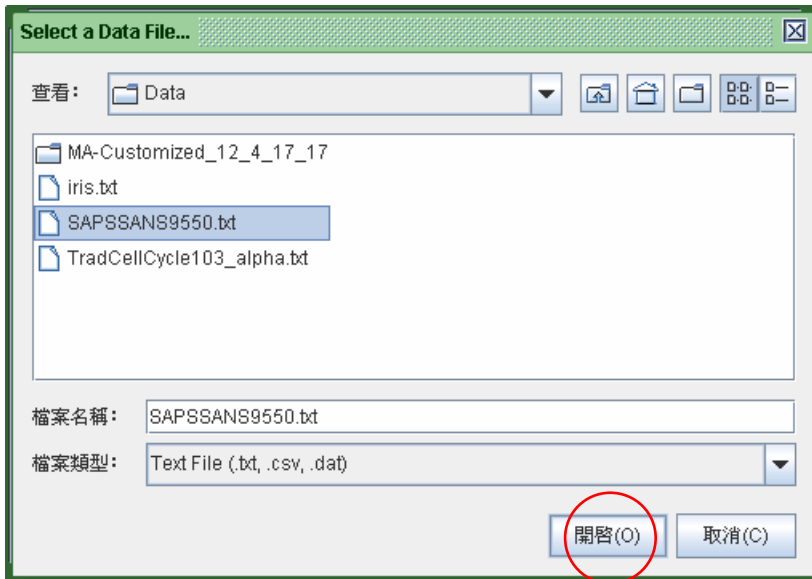
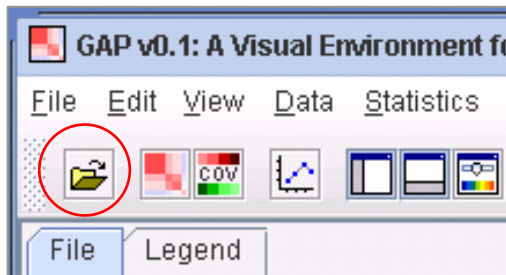
GAP Procedure

Output Window

Open/Import Data

Data Format: plain text with tab separator

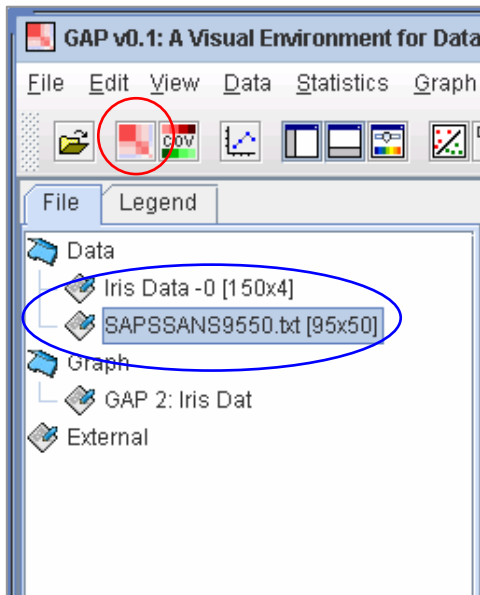
Row/Column Name
 Number of Y Discrete/Continuous Covariates
 Number of X Discrete/Continuous Covariates



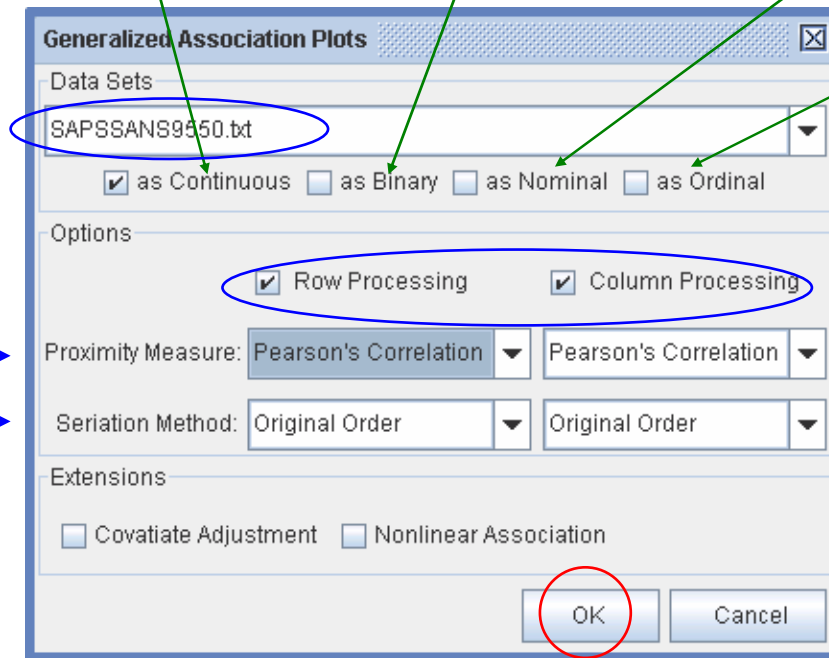
	var0	var1	var2	var3	var4	var5	var6
UNIQID		AH1	AH2	AH3	AH4	AH5	AH6
1	Xd	1	Yd	1	Yc	0	0
1	Xc	3	2	1	1	0	0
1		5	5	5	0	0	0
1		5	5	5	5	4	5
0		0	0	0	0	0	0
1		2	0	2	0	0	2
0		0	0	0	0	0	0
1		3	1	1	0	0	0
1		5	5	1	0	0	4
0		0	0	0	0	0	0
1		2	0	2	0	0	0
1		5	0	1	0	0	0
1		0	0	0	5	0	0
0		0	0	0	0	0	0
1		5	5	5	0	0	0
0		0	0	0	0	0	0

GAP Analysis Dialog

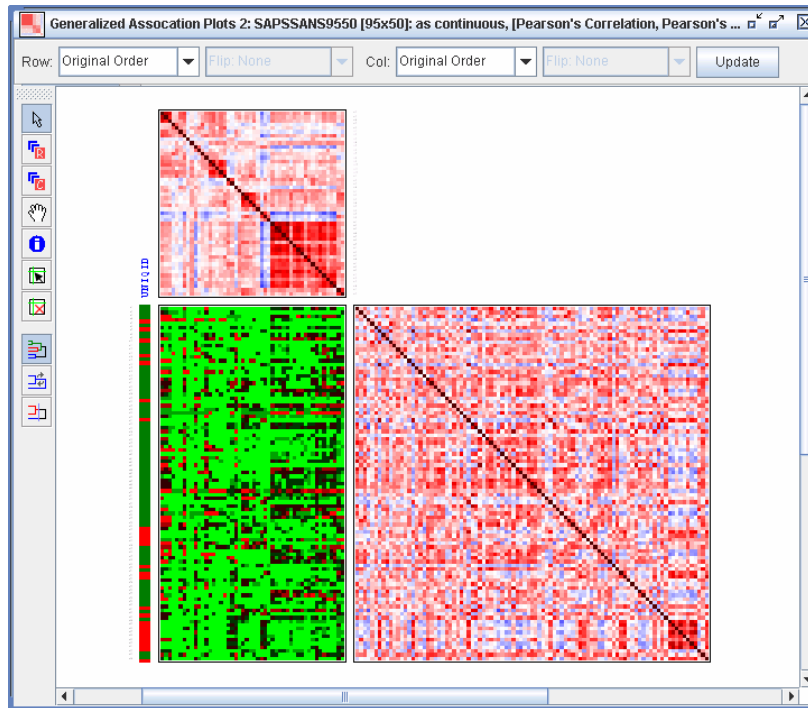
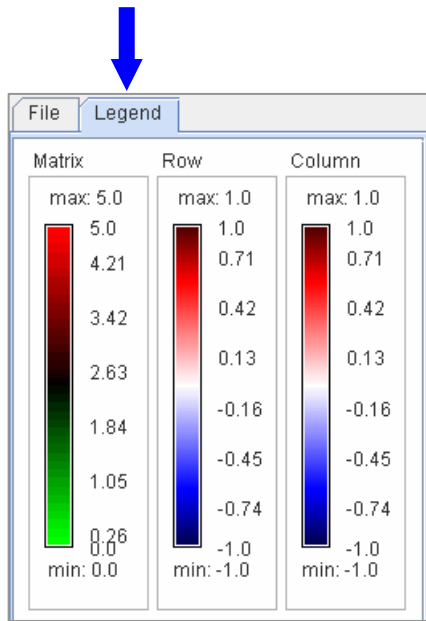
Not implemented Yet!



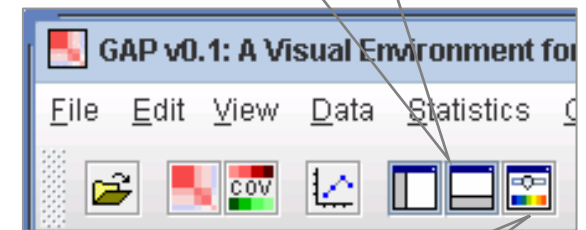
- Euclidean Distance
- Kendall's tau
- Pearson's Correlation
- Spearman's Rank
- atan Correlation
- Hamman: $a/(b+c)$
- Jaccard: $a/(a+b+c)$
- Phi: $(ad-bc)/\sqrt{(a+b)(a+c)}$
- Rao: $a/(a+b+c+d)$
- Rogers: $(a+d)/(a+2b+2c+d)$
- Simple Match: $(a+d)/(a+b+c+d)$
- Sneath: $a/(a+2b+2c)$
- Yule: $(ad-bc)/(ad+bc)$
- (x) 2x2 Correlation
- (x) Cramer's V
- (x) Cohen's kappa
- (x) Goodman-Kruskal tau
- (x) Goodman-Kruskal's gamma
- (x) Kendall's tau_b
- (x) Somers's d
- (x) Stuart's tau_c
- (x) Wilson's e



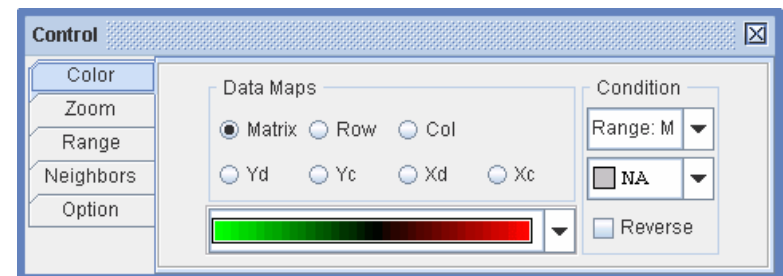
GAP Step 1: Presentation of Raw Data Matrix




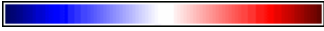





Show or Hide File/Data View, Output Panel

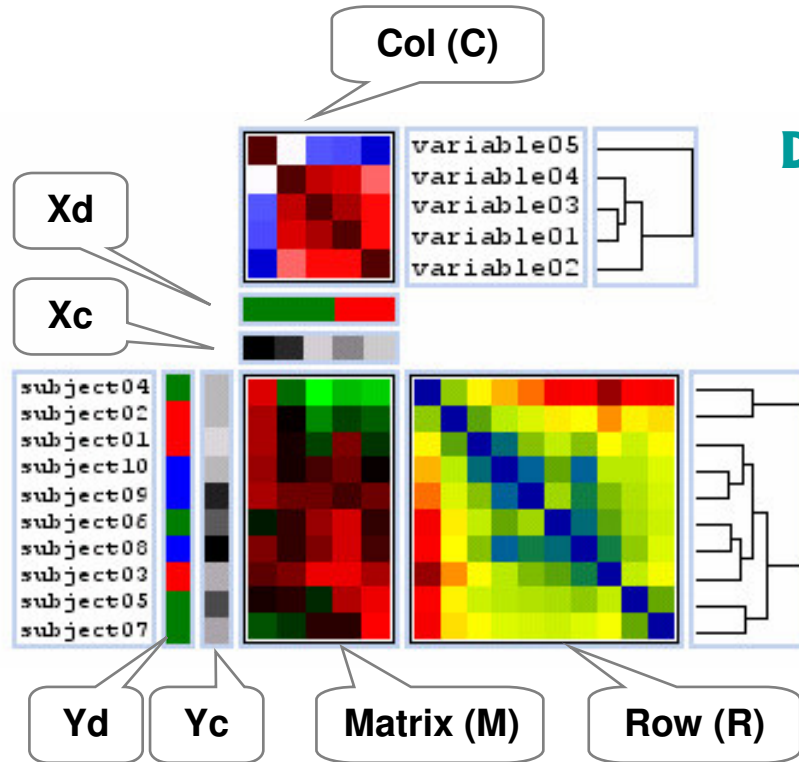


Show or Hide Control Panel



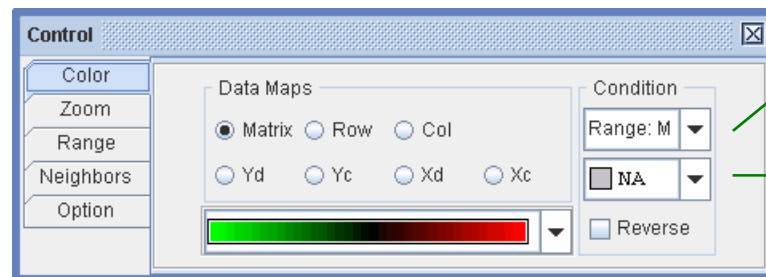
Control Panel: Color

- Rainbow 130**

 Distance
- Blue-white-Red 200**

 Correlation
- Green-Black-Red 38**

 Gene Expression
- Grey 256**

 Range
- Discrete 16**

 (No label)
- Biallel Marker**

 (No label)
- ACGT**

 (No label)






Display Conditions
(only for Matrix)

- Range: M
- Range: R
- Range: C
- Center: M
- Center: R
- Center: C
- Std: M (x)
- Std: R (x)
- Std: C (x)
- Rank: M
- Rank: R
- Rank: C

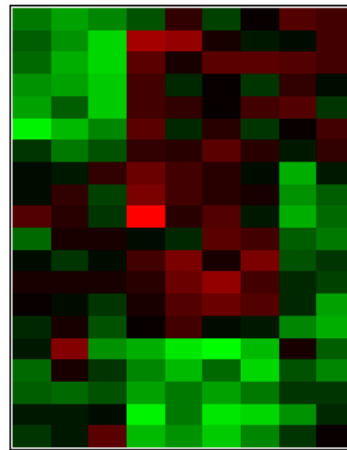


Missing Value

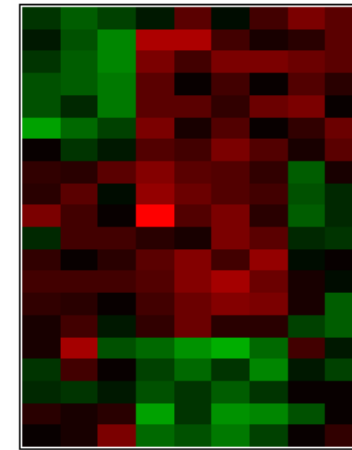
-  NA
-  NA
-  NA

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



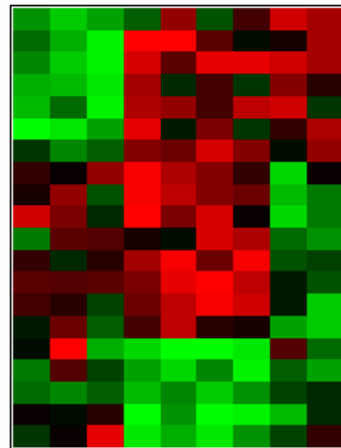
Center Matrix Condition

Bidirectional Color Spectrum

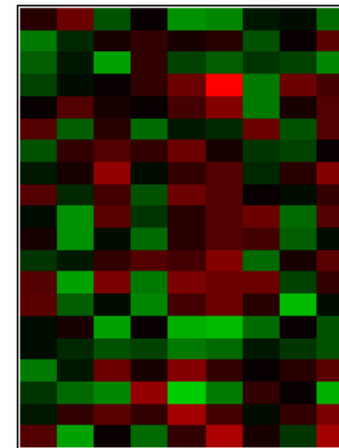
Blue-white-Red 200



Green-Black-Red 38

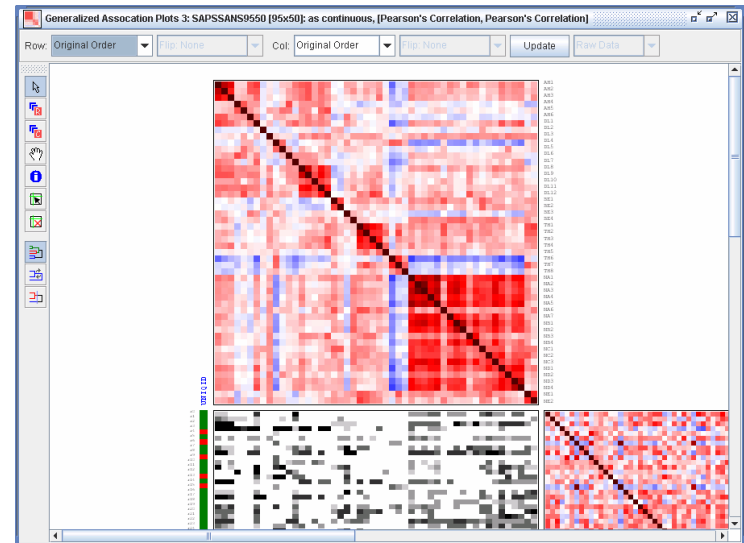
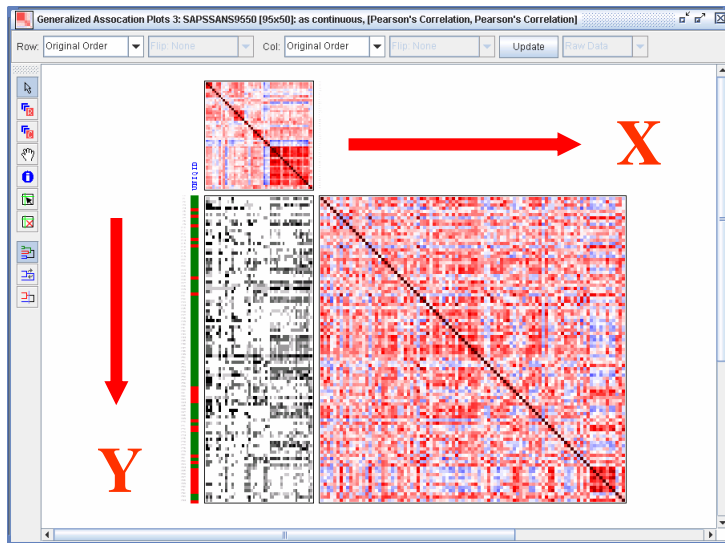


Rank Matrix Condition



Standardize Matrix Condition

Control Panel: Zoom



Actual Size

Fit Window

Fit Width

Control

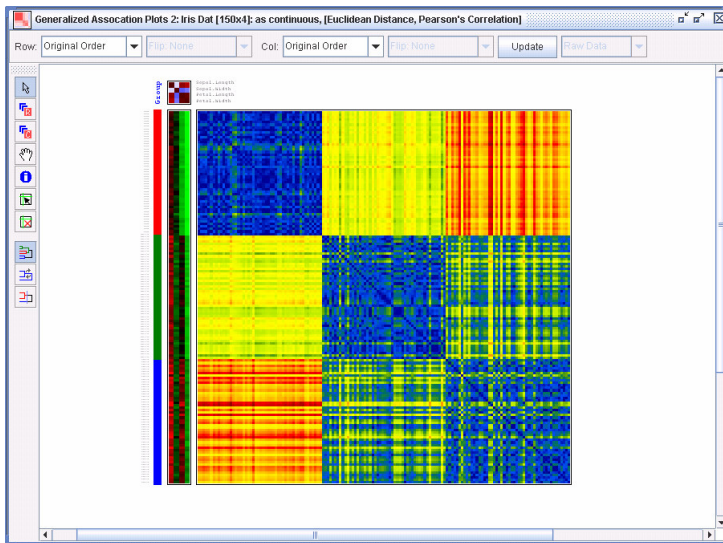
- Color
- Zoom
- Range
- Neighbors
- Option

Zoom In	Slider	Zoom Out	100%
Zoom In	Slider	Zoom Out	100%
Zoom In	Slider	Zoom Out	76%

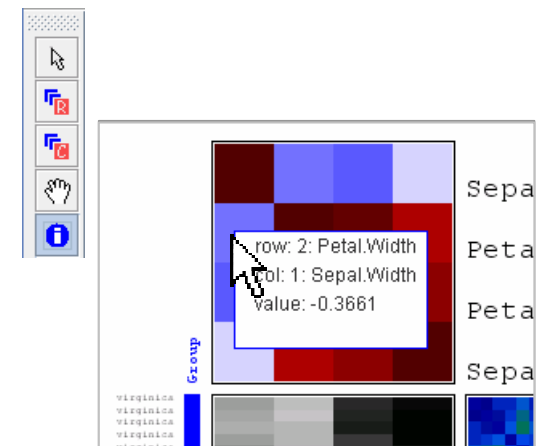
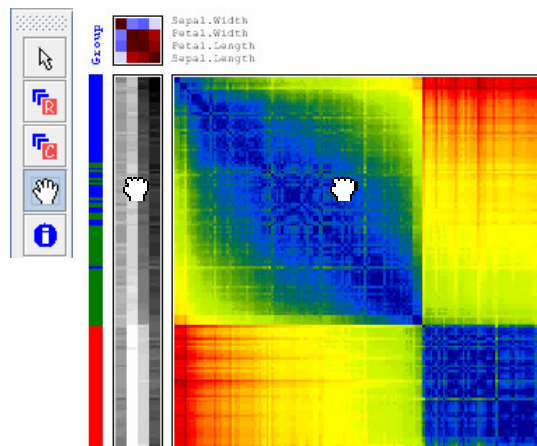
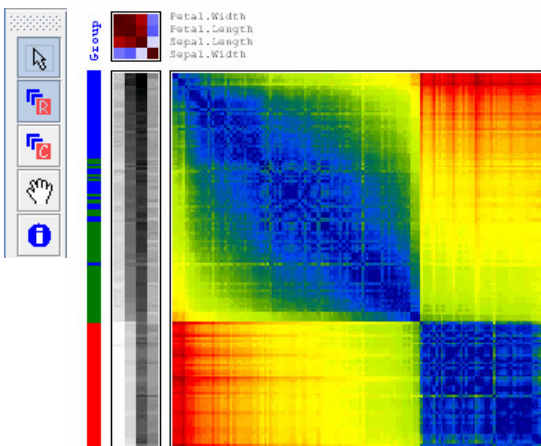
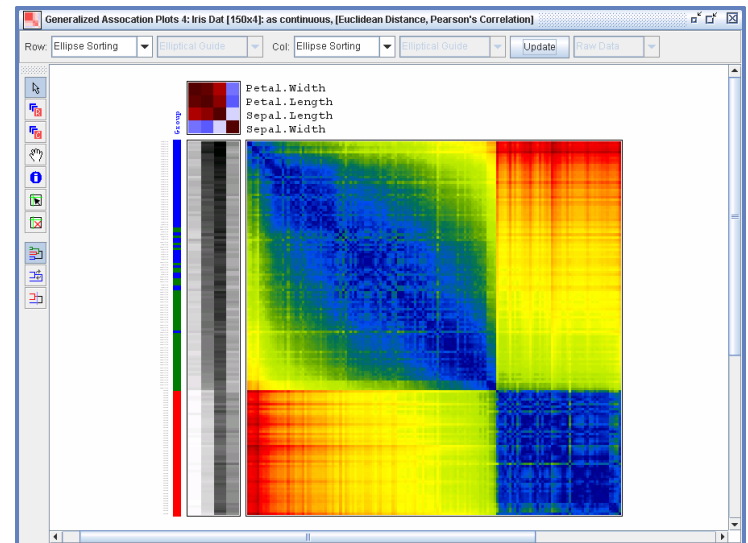
GAP Step 2: Ellipse Seriation

Generalized Association Plots 4: Iris Dat [150x4]: as continuous, [Euclidean Distance, Pearson's Correlation]

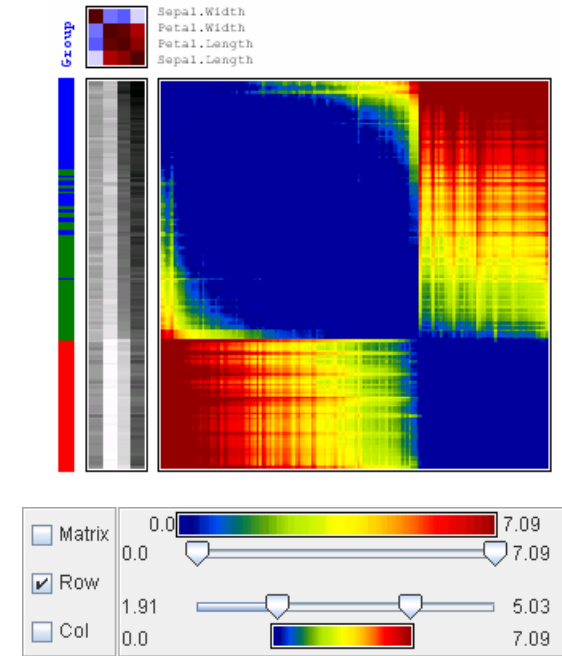
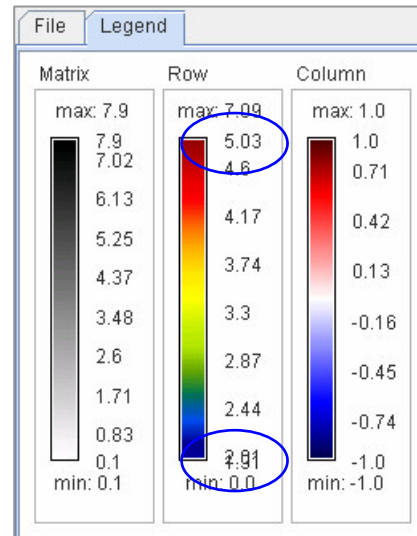
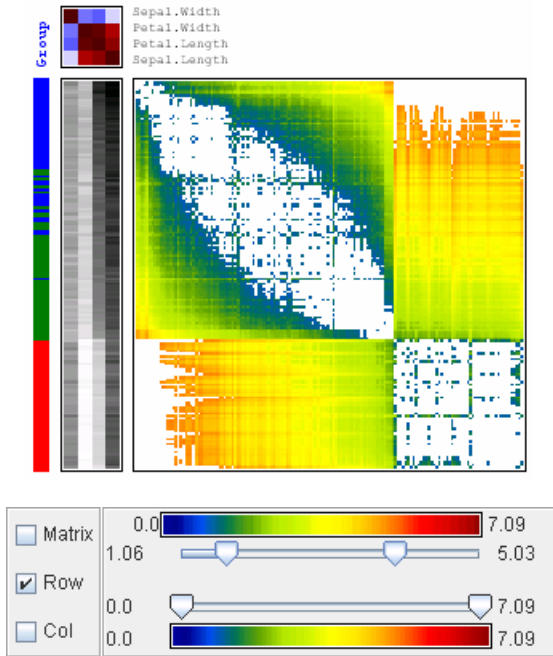
Row: **Ellipse Sorting** | Elliptical Guide | Col: **Ellipse Sorting** | Elliptical Guide | **Update** | Raw Data



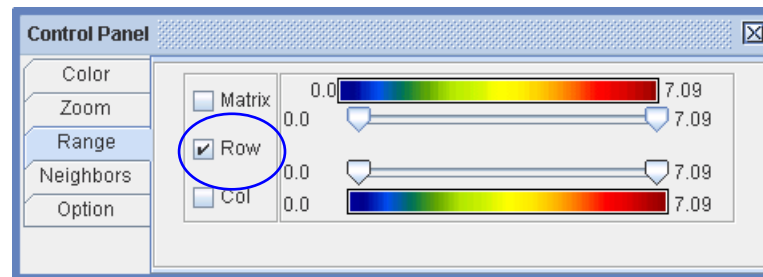
- Average-linkage
- Centroid-(x)
- Complete-linkage
- Ellipse Sorting**
- Original Order
- Single-linkage
- External



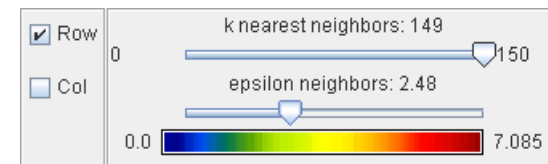
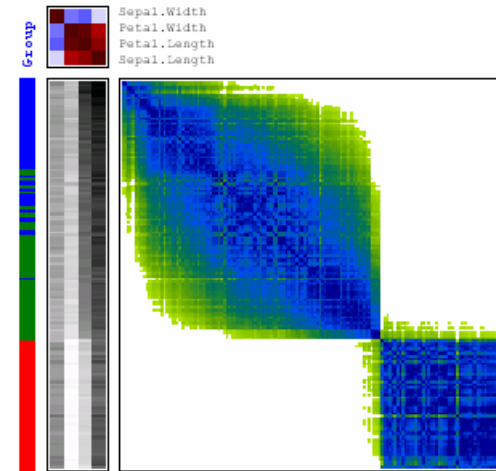
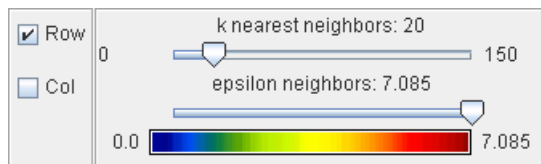
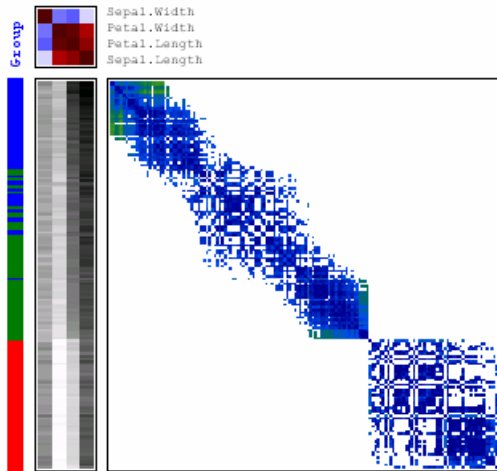
Control Panel: Range



“Resolution” of a Statistical Graph



Control Panel: Neighbors



Control Panel [X]

Color

Zoom

Range

Neighbors

Option

Row

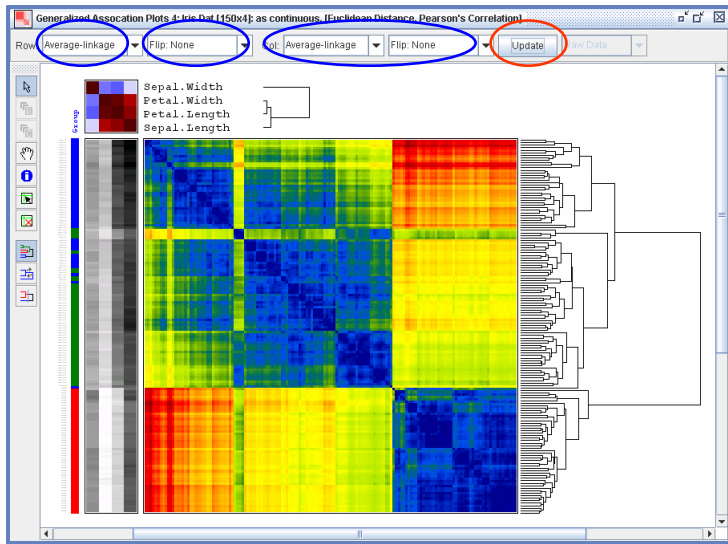
Col

k nearest neighbors: 149

epsilon neighbors: 7.085

0.0 7.085

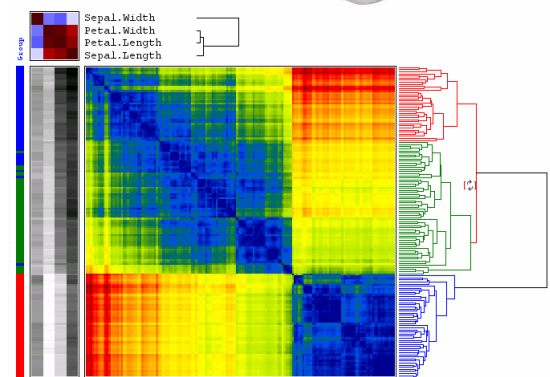
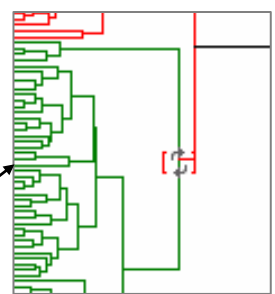
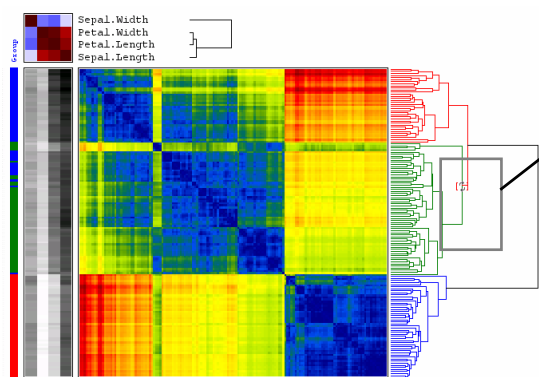
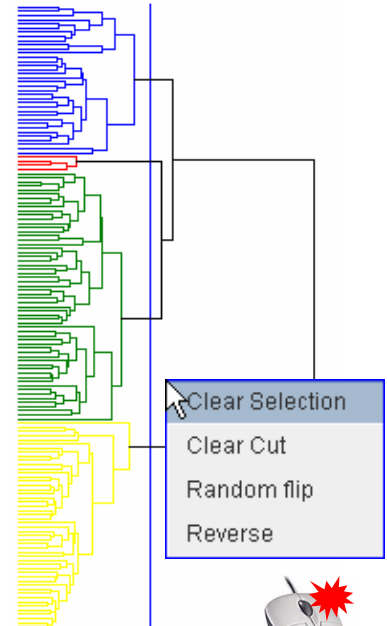
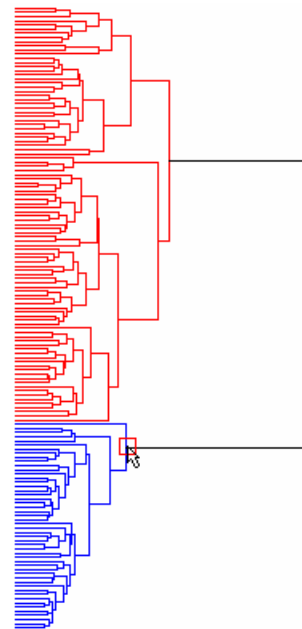
GAP Step 2: Tree Seriation/Flip



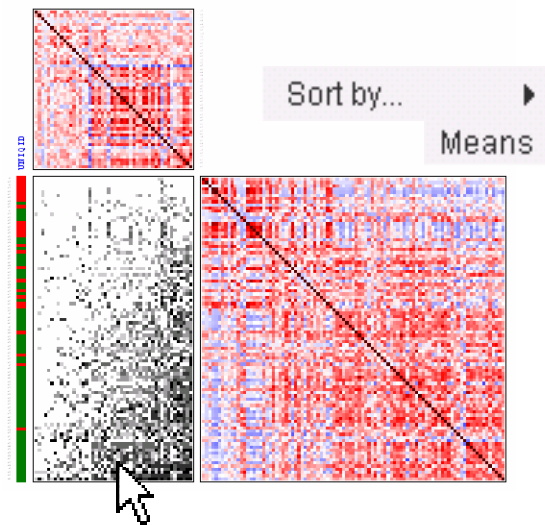
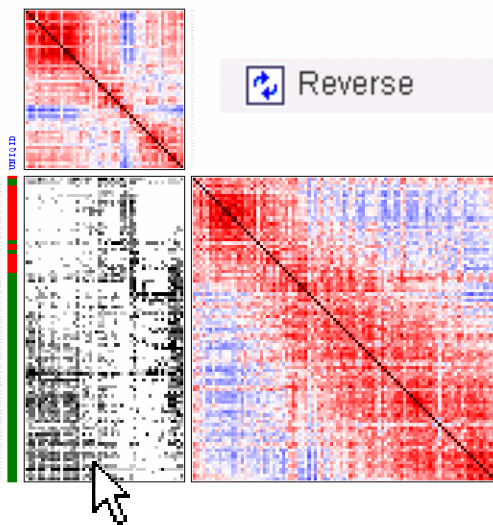
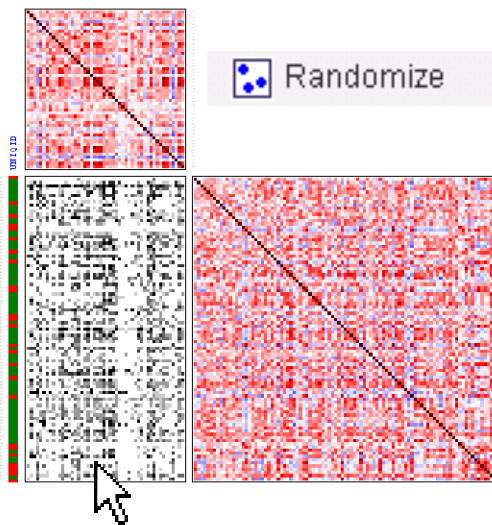
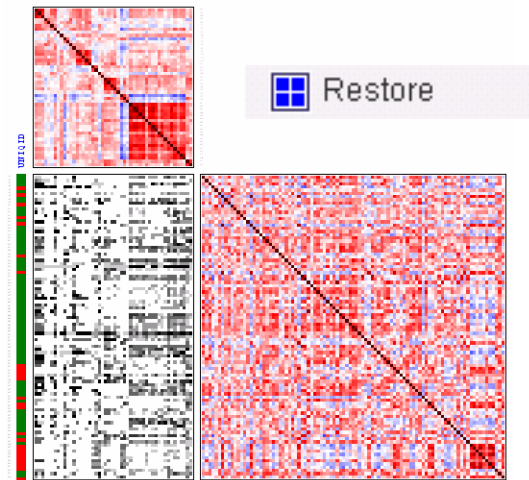
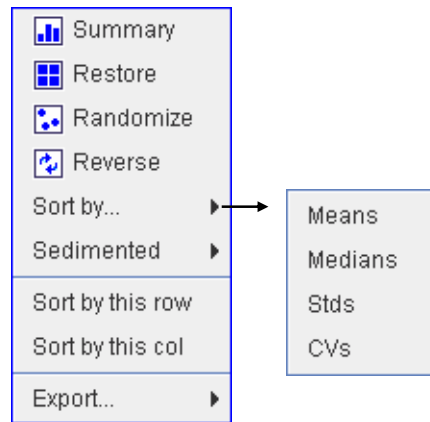
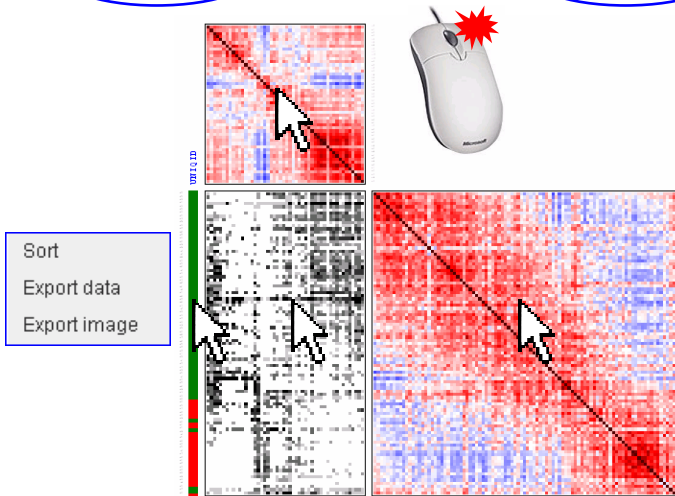
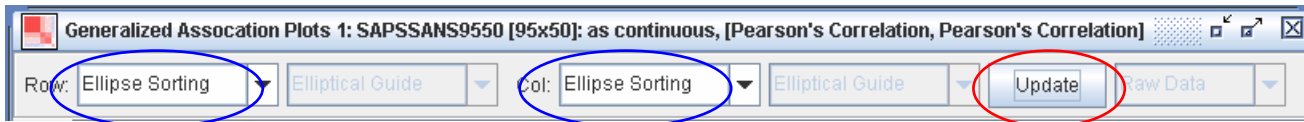
- Average-linkage
- Centroid-(x)
- Complete-linkage
- Ellipse Sorting
- Original Order
- Single-linkage
- External

- Flip: None
- Flip: Uncle
- Flip: GrandPa
- External Reference
- Elliptical Guide

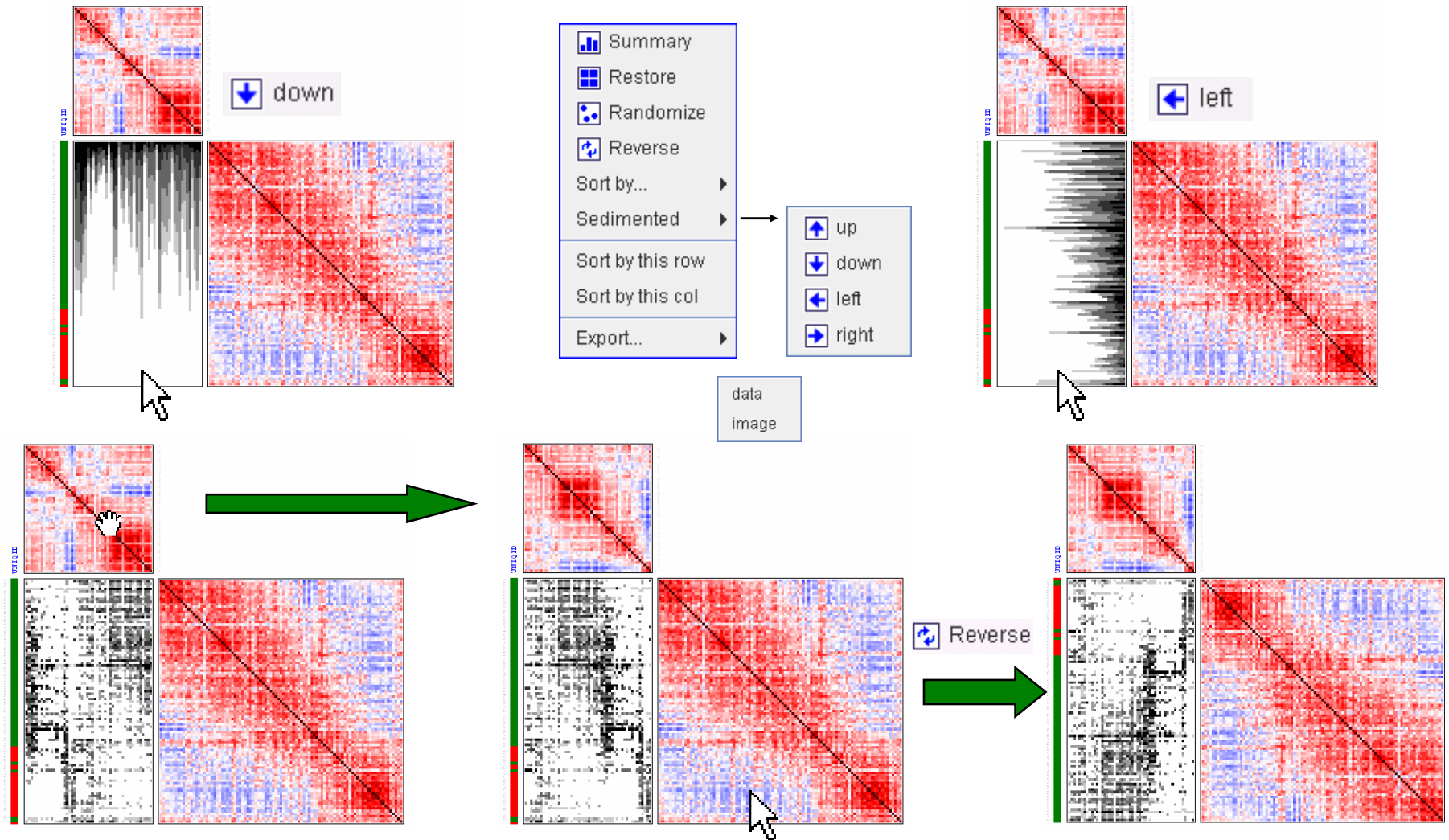
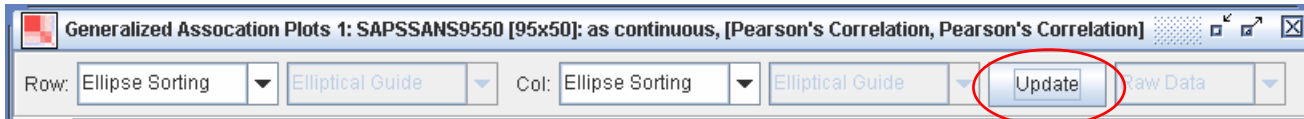
Update



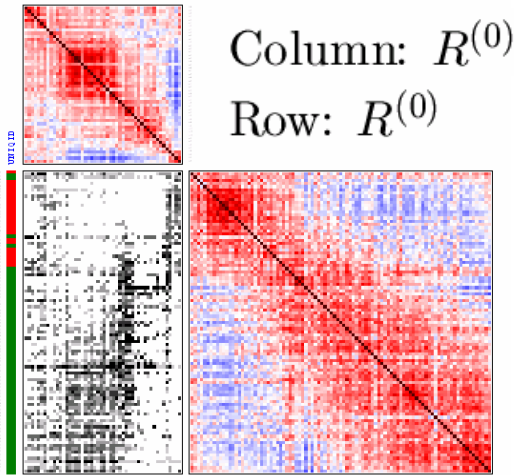
Pop-Up Menu



Pop-Up Menu

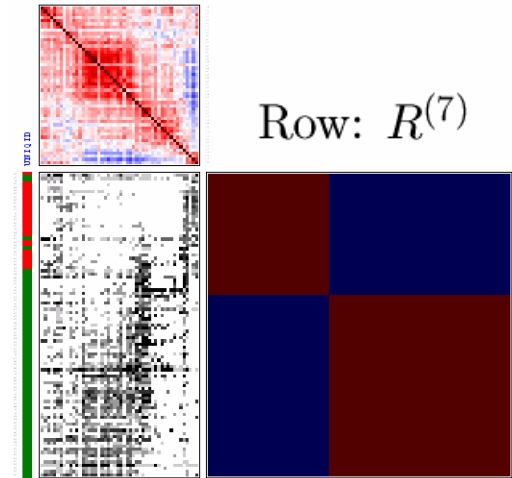


Converging Correlation Matrices

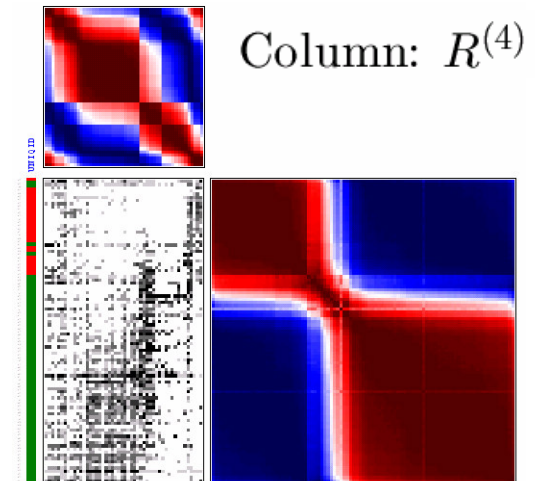
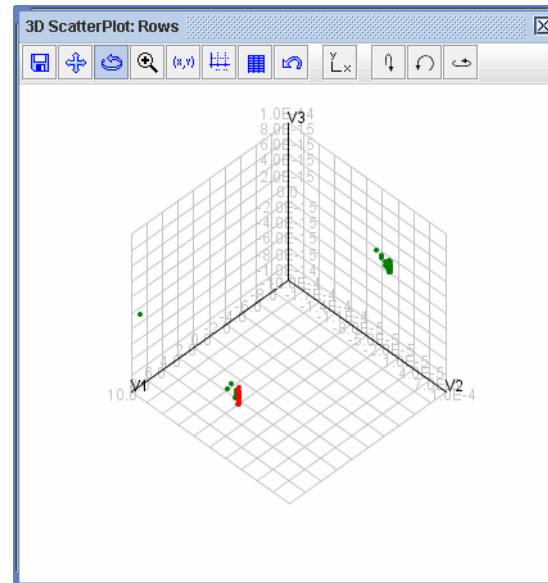
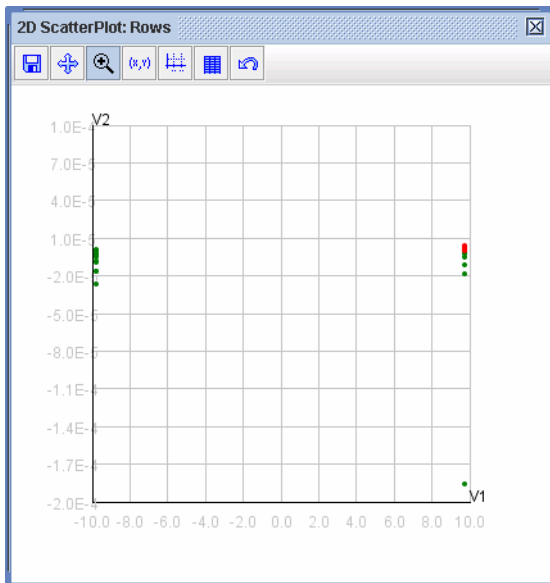


Converging Correlation Matrices

Row: 3 2D Plot 3D Plot
 Column: 4 2D Plot 3D Plot



Row: $R^{(7)}$



GAP Step 3: Partitions of Permuted Matrix Maps



Finding Blocks:

Row Column

Simple Simple

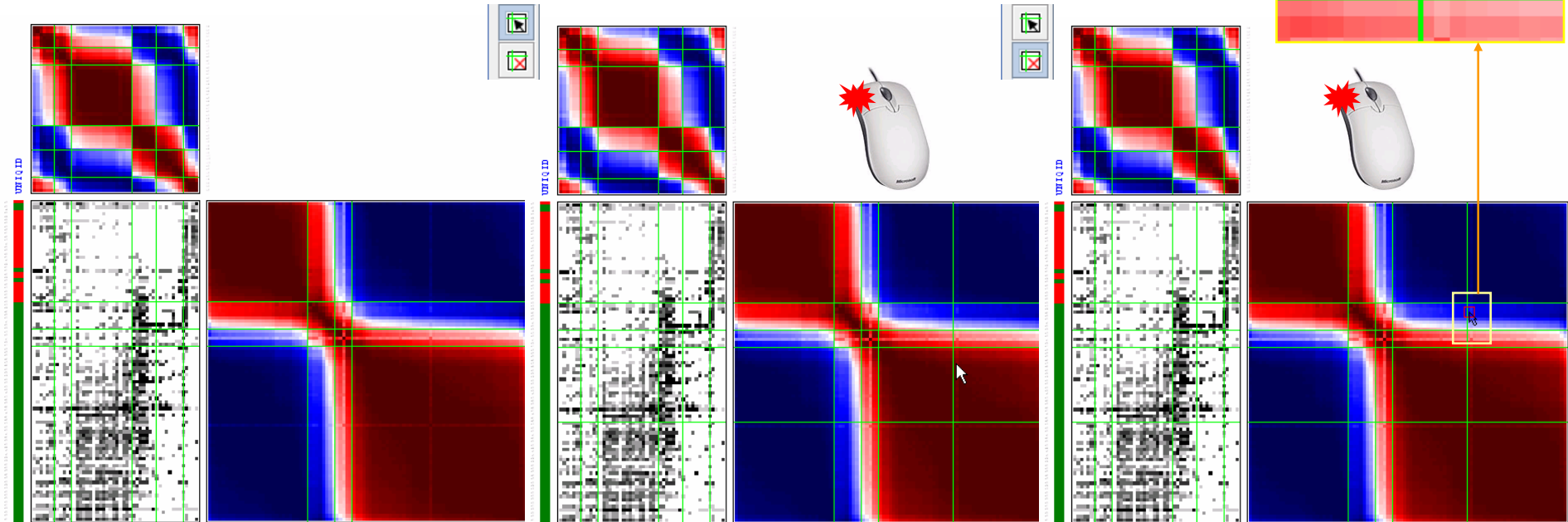
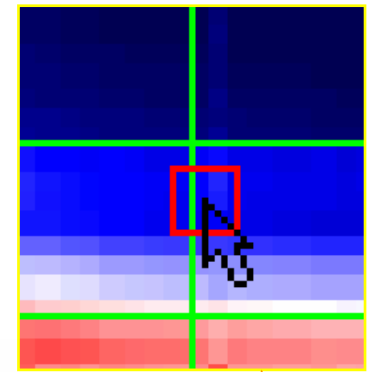
Update **Clear**

- Simple
- Dendrogram
- max(BSS/WSS)

Generalized Association Plots 2: SAPSSANS9550 [95x50]: as continuous, [Pea

Row: Ellipse Sorting Elliptical Guide Col: Ellipse Sorting

UIQ ID



GAP Step 4: Sufficient Graph

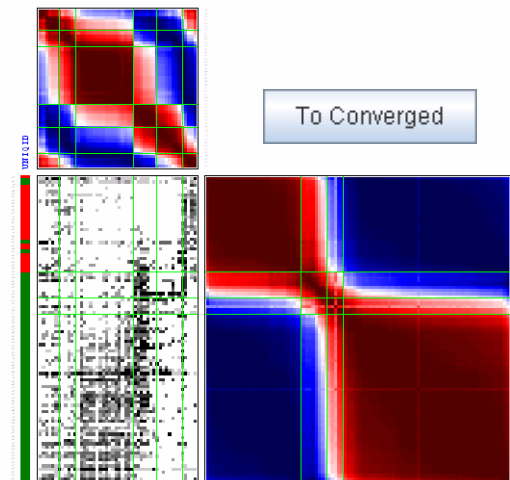
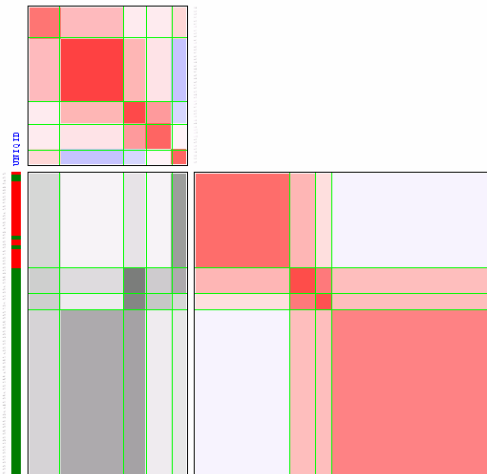
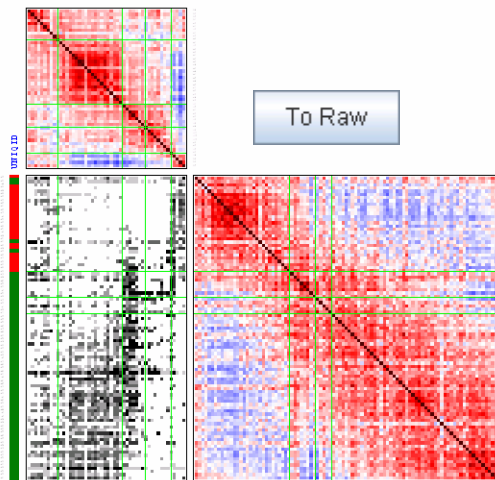
Sufficient Graphs

Matrix Row Column

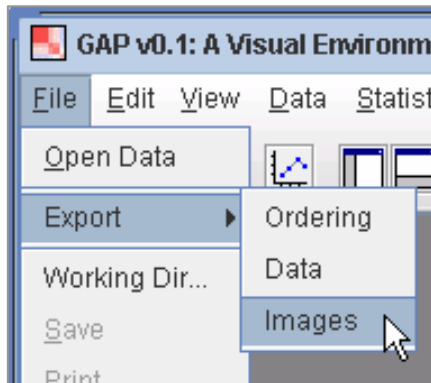
Mean Mean Mean

Update To Raw To Converged

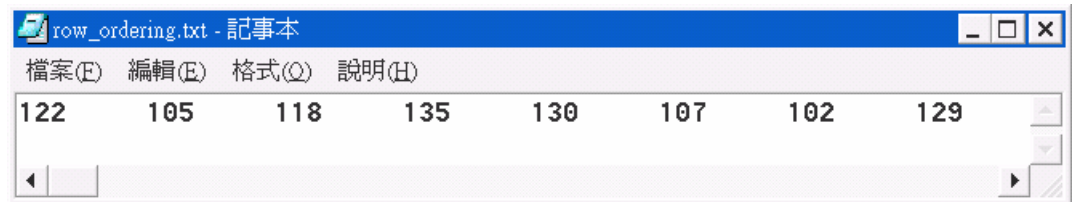
- Mean
- Median
- Std.
- CV



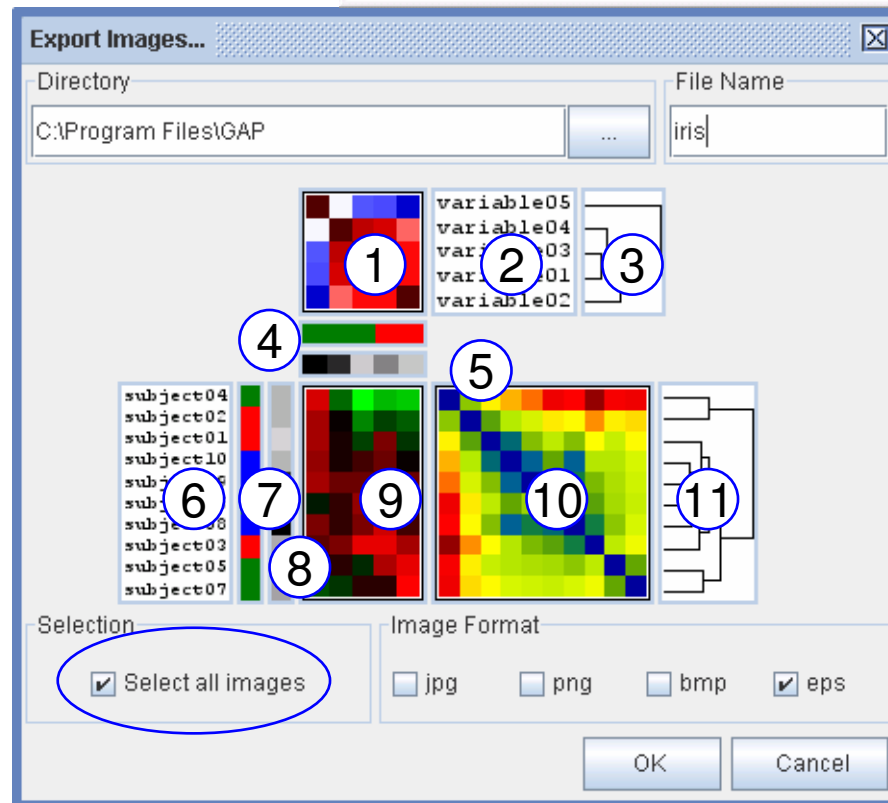
Export Ordering, Data, Images



row_ordering.txt
col_ordering.txt

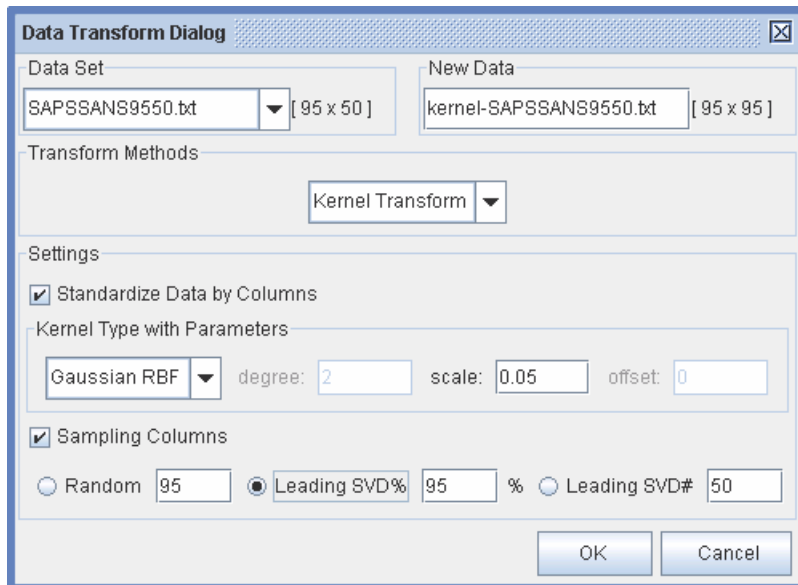
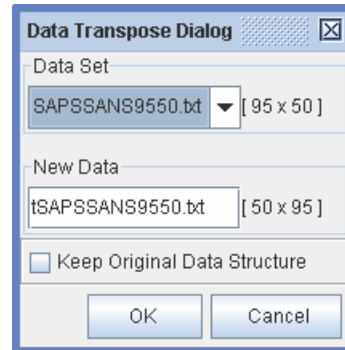
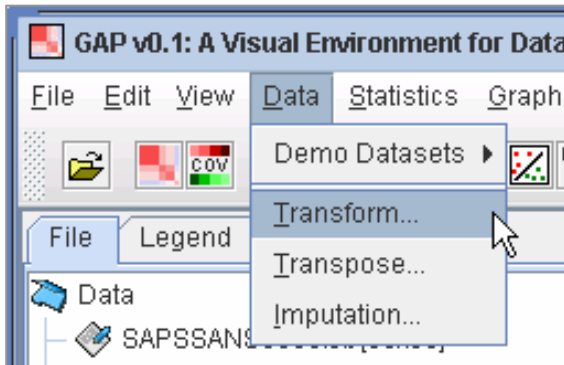


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	LNQID	TH3	TH4	TH2	TH5	TH1	NE2	NE1	DL2	BE1	BE2	NA6	NC1	DL3	NC2	NE2	NE1
2	162	1	1	5	1	5	5	5	0	2	3	1	1	4	4	2	
3	227	2	2	1	2	1	1	2	0	0	2	0	1	0	2	2	
4	152	0	0	2	0	2	0	3	0	0	0	0	0	0	2	0	
5	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6	189	0	0	1	2	2	2	1	0	0	0	0	0	0	0	0	
7	184	0	0	1	0	2	1	1	0	0	2	1	1	1	1	1	
8	188	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	
9	160	2	0	3	1	3	3	3	0	2	0	0	0	0	3	0	
10	130	0	0	3	0	2	0	2	0	0	0	0	0	0	0	0	
11	187	0	0	0	0	0	1	1	0	2	3	2	2	0	1	0	
12	191	0	0	0	0	0	2	3	3	3	3	0	2	0	3	0	
13	186	0	0	0	0	0	0	3	0	2	3	0	0	0	0	0	
14	115	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
15	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
16	172	0	0	0	0	0	1	3	0	1	0	1	0	1	0	2	
17	190	0	0	0	0	2	1	1	0	2	2	1	0	0	0	0	
18	186	0	0	0	0	0	1	3	0	1	0	2	0	0	2	1	
19	171	0	0	0	0	0	3	4	0	0	3	0	0	0	0	0	
20	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
21	194	0	0	0	2	0	0	3	0	0	0	0	0	0	3	0	
22	151	5	5	5	0	5	0	2	0	0	2	2	0	3	2	2	
23	191	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
24	182	0	2	3	0	3	2	3	0	0	0	0	0	0	0	0	
25	124	0	0	0	0	3	2	0	0	2	0	1	0	0	0	0	
26	25	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	
27	180	0	0	0	0	2	0	3	0	1	0	0	0	2	0	0	
28	113	0	0	0	0	0	3	3	0	0	0	0	0	1	0	0	
29	189	3	3	0	3	4	2	3	0	0	2	2	0	2	0	4	
30	183	0	3	4	4	0	0	1	0	0	2	0	0	0	0	0	
31	189	0	0	3	0	0	2	0	0	3	3	2	1	0	2	0	
32	129	4	4	3	1	4	1	1	0	0	1	1	0	0	1	1	
33	119	0	0	0	0	2	2	2	0	0	2	0	0	3	0	2	
34	184	3	4	4	2	5	5	4	0	2	4	2	0	4	0	0	
35	142	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
36	175	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
37	133	0	0	0	0	0	2	0	0	0	0	0	0	2	0	0	
38	13	0	3	0	0	0	0	0	0	0	1	0	0	0	0	0	
39	173	2	3	0	3	2	2	0	0	2	2	0	0	4	0	0	
40	136	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	
41	110	0	0	0	0	0	2	0	0	0	0	0	0	3	0	0	
42	120	3	4	3	0	4	5	4	0	2	4	0	3	0	3	0	
43	11	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	
44	156	0	2	1	2	2	1	2	0	0	0	0	0	0	3	1	
45	111	0	0	0	0	0	1	1	0	1	3	0	3	0	2	0	
46	137	0	3	0	0	0	1	0	0	0	0	0	0	0	3	0	
47	157	0	3	0	0	0	1	1	0	0	2	0	0	0	0	3	
48	170	0	0	0	0	0	2	2	0	0	0	0	2	0	3	0	
49	140	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

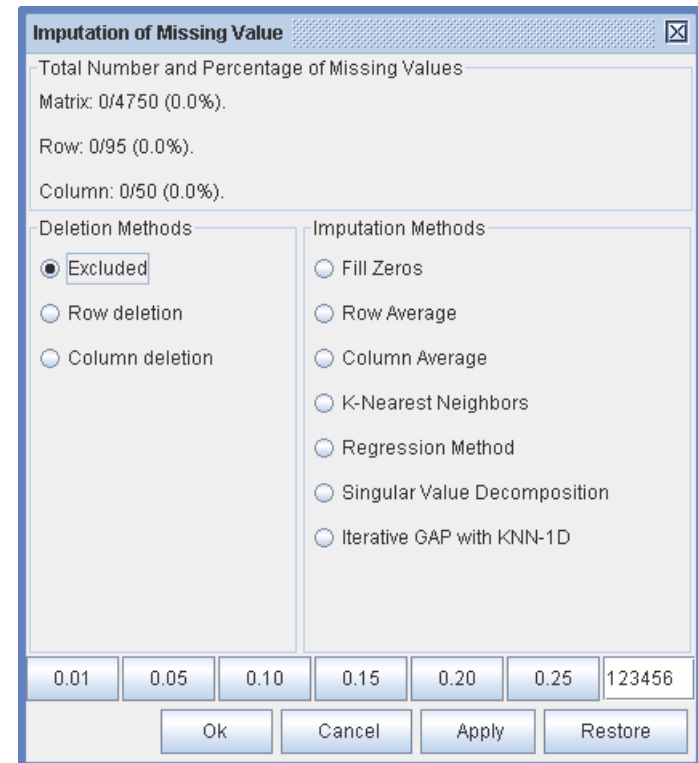


- 1 _gapCol.eps
- 2 _gapColName.eps
- 3 _TreeCol.eps
- 4 _gapXd.eps
- 5 _gapXc.eps
- 6 _gapRowName.eps
- 7 _gapYd.eps
- 8 _gapYc.eps
- 9 _gapMatrix.eps
- 10 _gapRow.eps
- 11 _TreeRow.eps

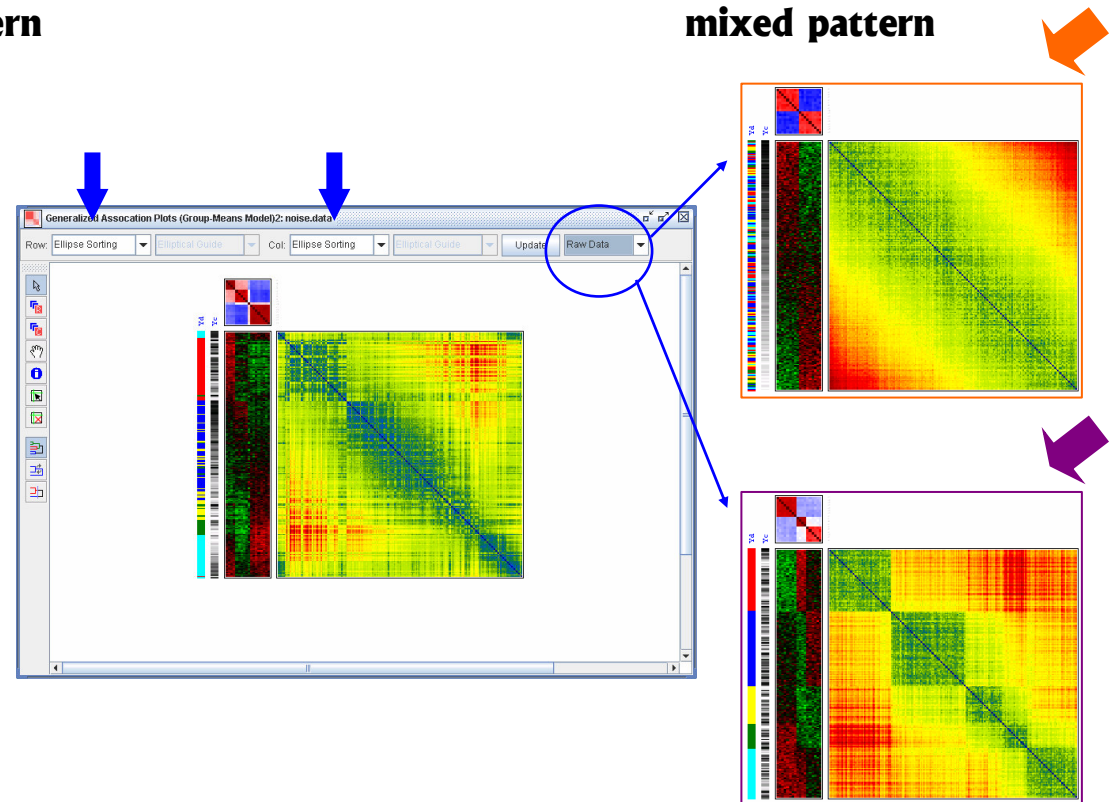
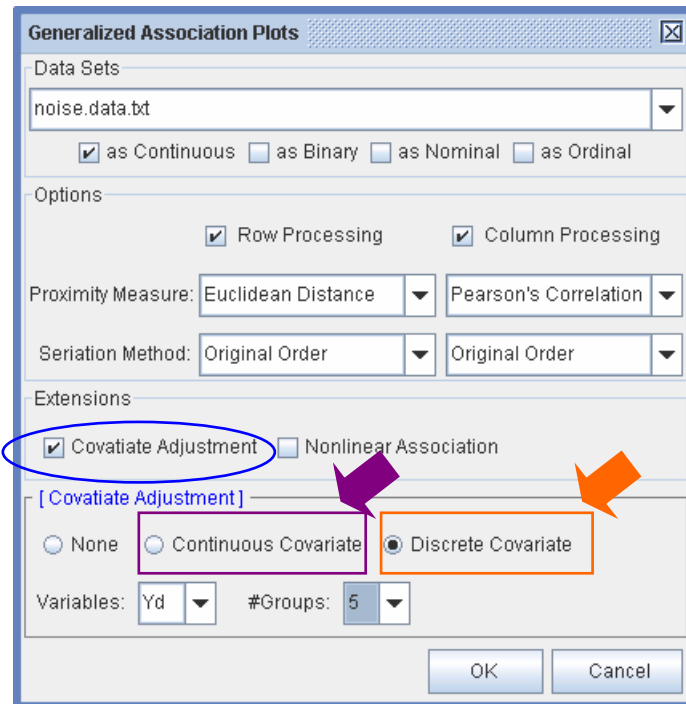
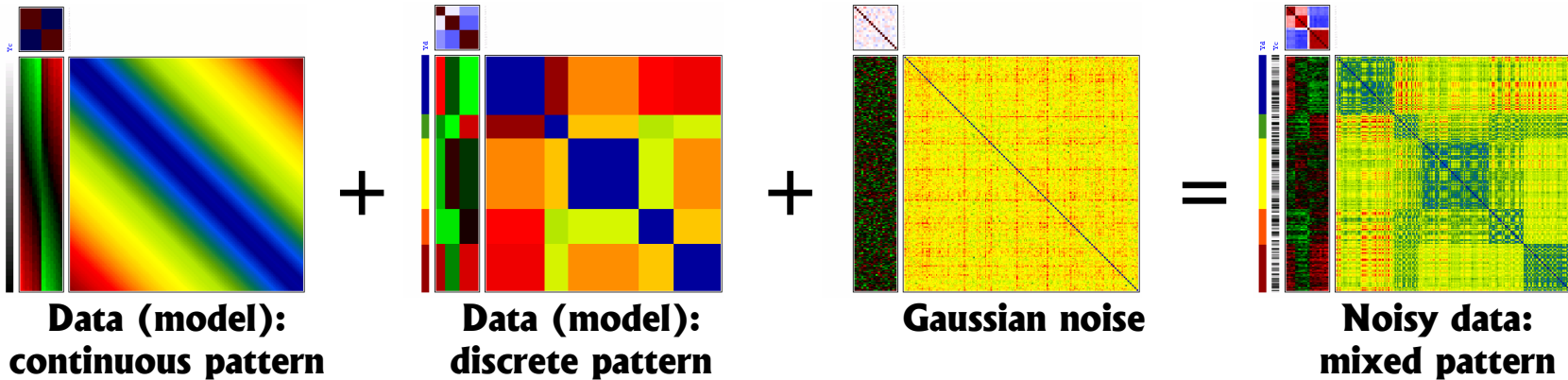
Data Processing



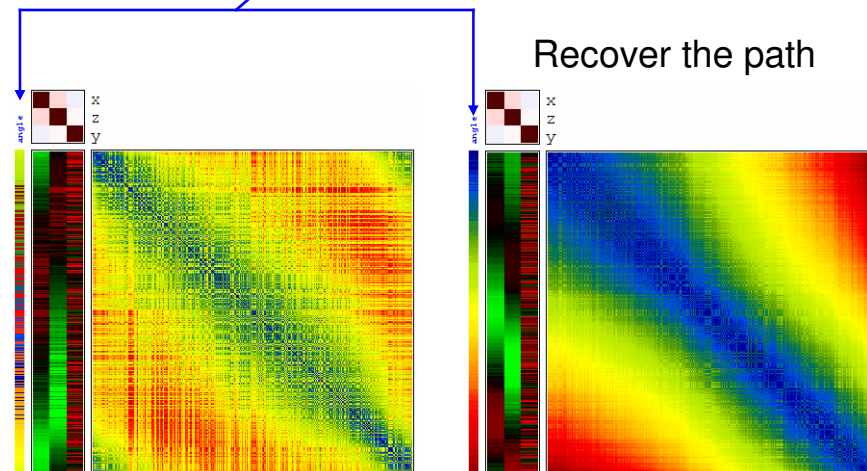
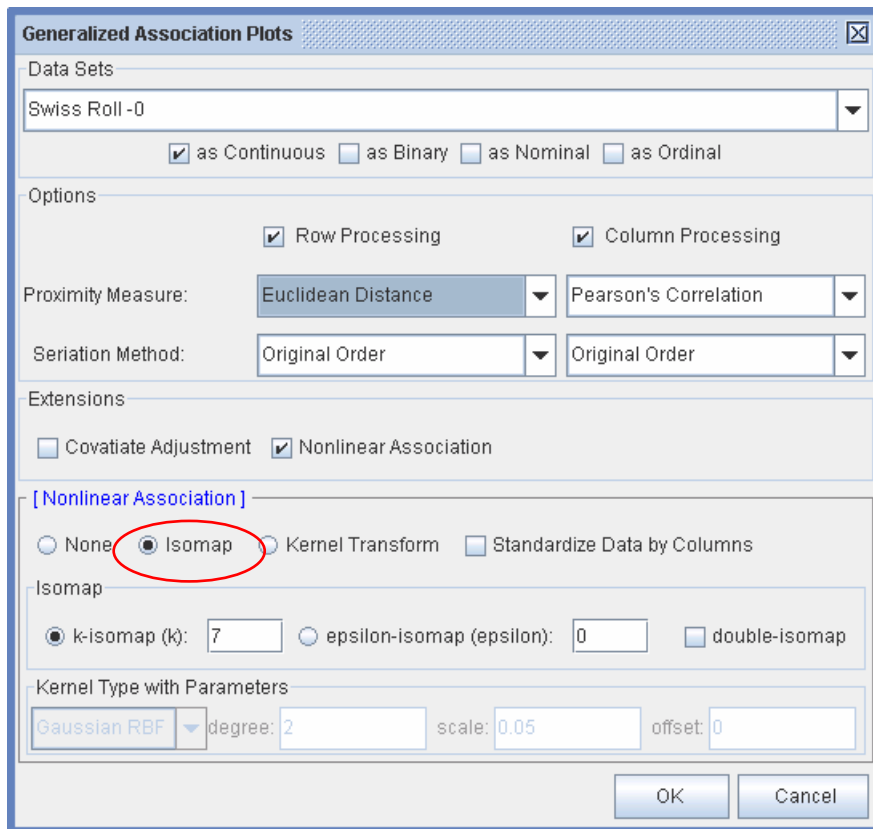
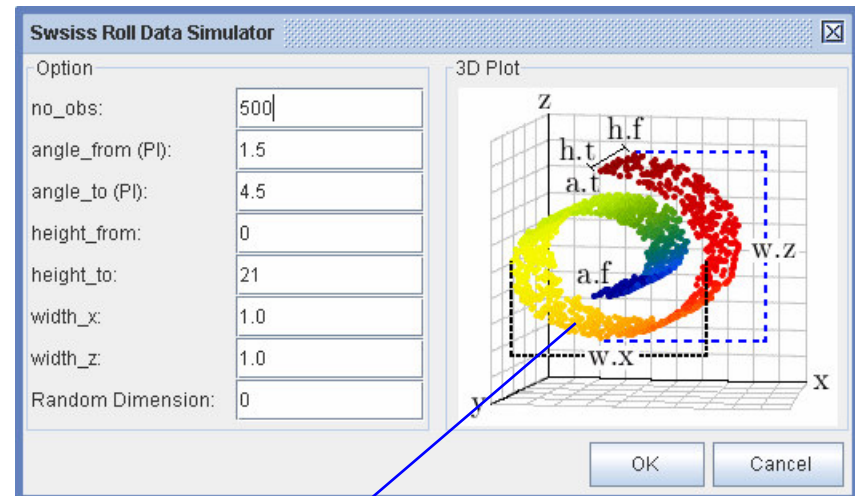
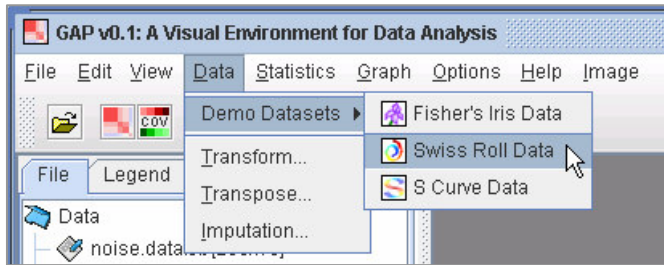
- Log
- Power
- (x) Transpose
- Center
- Scale
- Standardize
- Kernel Transform



GAP with Covariate Adjustment



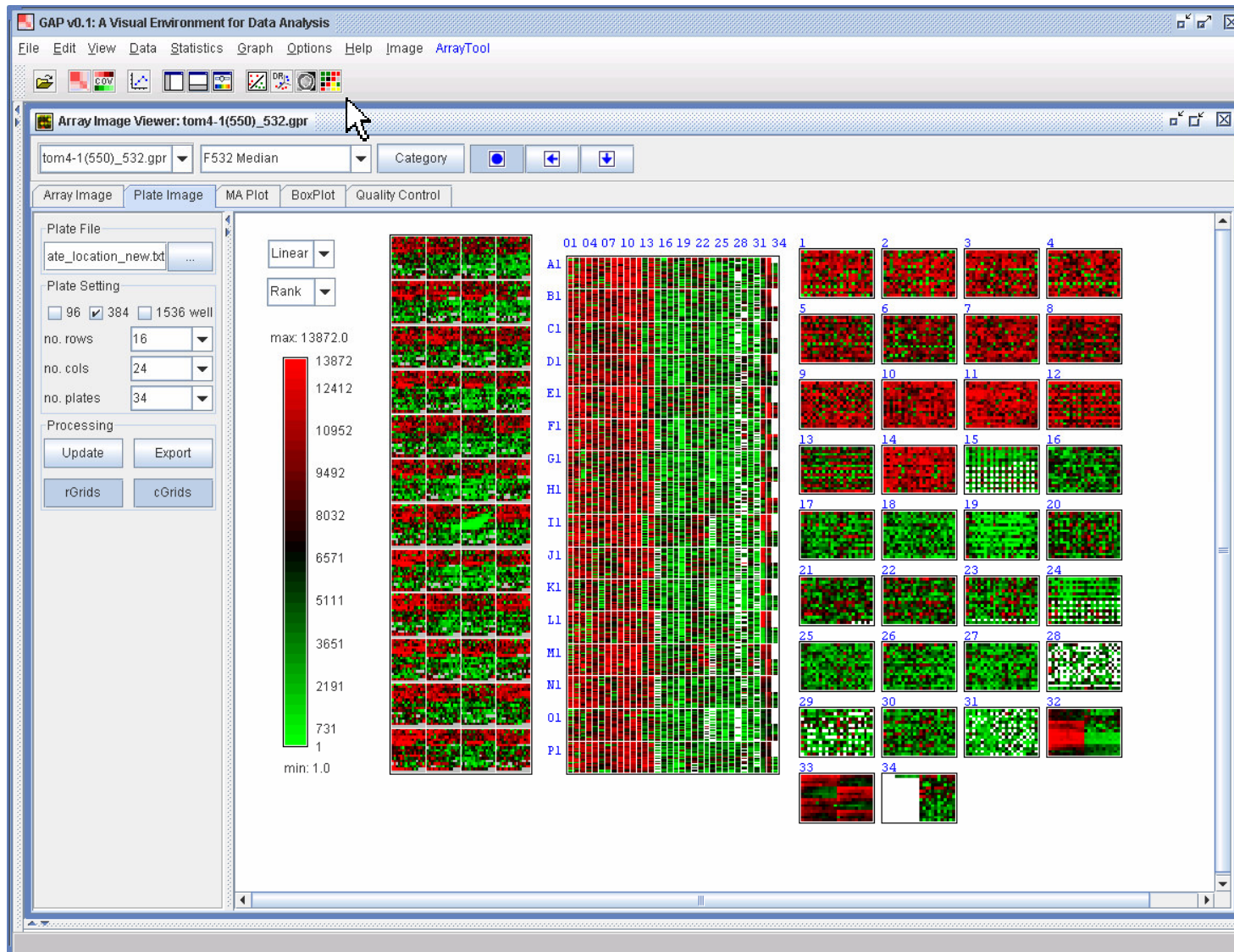
GAP with Nonlinear Association



GAP

isoGAP

GAP Module: Microarray Tool



Web Site

Lab for Information Visualization - Microsoft Internet Explorer

檔案(E) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

地址(D) http://gap.stat.sinica.edu.tw

Dimension Free Data Visualization **Lab for Information Visualization** 資訊視覺化研究室

中央研究院 統計學研究所 Institute of Statistical Science Academia Sinica

Home | Research | Members | Database | Software | Link

Chun-houh Chen 陳君厚
Associate Research Fellow
Institute of Statistical Science
Academia Sinica

① Raw Data Maps

② Sorted Data Maps

Generalized Association Plots (GAP) for Dimensional Data Visualization

- Information Visualization**
 - Generalized Association Plots (GAP)
 - Sliced Inverse Regression (SIR)
 - Multidimensional Scaling (MDS)
- Psychiatry Research**
 - Psychiatry
- Bioinformatics**
 - Microarray Data Analysis
 - SNPs
- Talks/Seminar**
 - Lecture Notes
 - Posters
- News/Conference [more]**
 - 2005: BIBE | CAMDA | CSB | GIW | IMS/ENAR | ISMB/ECCB | InfoVis | JSM | PSB

Handbook of Computational Statistics (Volume III): Data Visualization
Chun-houh Chen, Wolfgang Härdle, and Antony Unwin (eds)
Springer-Verlag, Heidelberg
Dr. Wolfgang Härdle's Visit at Academia Sinica (2005/04/30~2005/05/15) : Schedule in Taiwan

iasc Dec. 15-17, 2005 (IASC-ARS 2005)
The 5th Asian Conference on Statistical Computing
The University of Hong Kong, Hong Kong

完成

- Chen, C. H. (2002), "Generalized Association Plots for Information Visualization: The applications of the convergence of iteratively formed correlation matrices." *Statistica Sinica* 12, 1-23.

GAP: Generalized Association Plots - Microsoft Internet Explorer

檔案(E) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

地址(D) http://gap.stat.sinica.edu.tw/Software/GAP/index.htm

GAP GAP is a java-designed software for generalized association plots (Chen, 2002) and exploratory data analysis. It is programmed for the java runtime environment 1.5 (JRE version 1.5.0_04), which is available for most operating systems.

Author: Han-Ming Wu in Chen's Lab for Information Visualization
e-mail: hmwu@stat.sinica.edu.tw
Institute of Statistical Science, Academia Sinica
Official Website of GAP Software: <http://gap.stat.sinica.edu.tw/Software/GAP>
[Current Version of GAP Software: v0.1.012, Build 2005-11-21]
[Last Updated of This Page: 2005/12/09]

Generalized Association Plots

- Clustering Analysis
- Various Display Conditions
- GAP with a Covariate Adjusted
- Nonlinear Association Analysis
- Missing Value Imputation

Statistical Plots

- Histogram, 2D Scatterplot, 3D Scatterplot (Rotatable)

More...

Latest Features

FAQ

Bug Reports/Fixed

Forum

GAP v0.1.0 Visual Environment for Data Analysis

Generalized Association Plots: A Java GUI (HTML) on continuous, Discretized Data, Pearson's Correlation

Color: Data Maps: Matrix Row Col

Range: Y0 Y1 X0 X1

Neighbors: NA

Option: Reverse

Output: GAP Procedure

GAP Software Version 0.1.012 (Build 20051120)

Copyright © 2005 Lab for Information Visualization

This data is simulated by simulator...

Number of rows: 100

Number of columns: 4

Memory and Time Information: *****

完成

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

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