

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

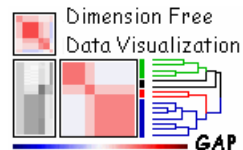
(吳漢銘，陳君厚)

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



0001001000100100100100

001001000100100



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

GAP Features

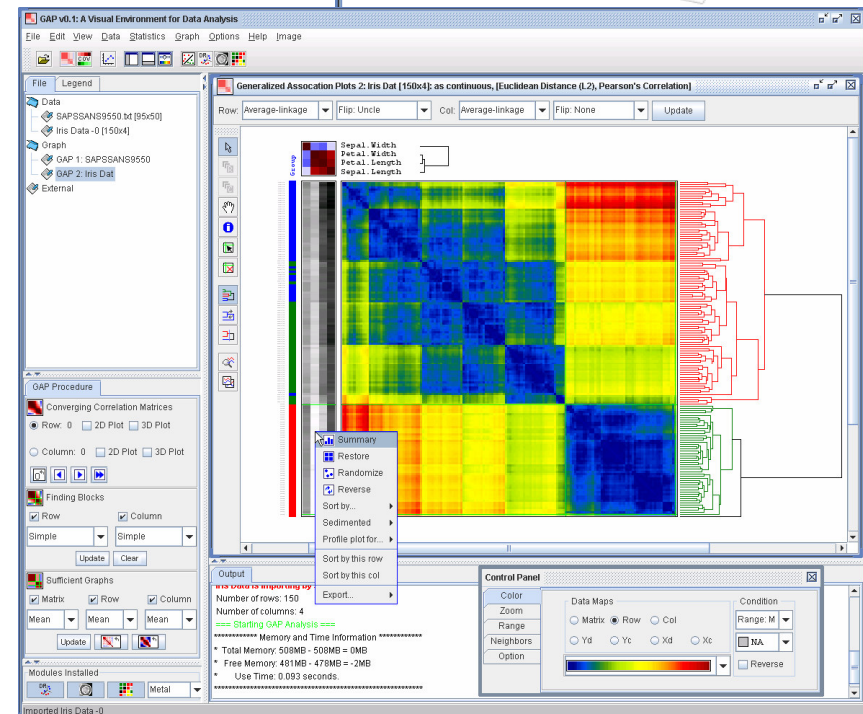
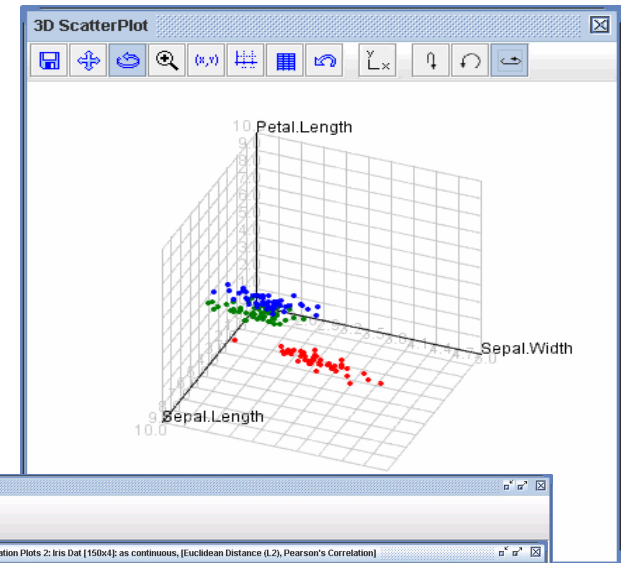


Generalized Association Plots

- 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

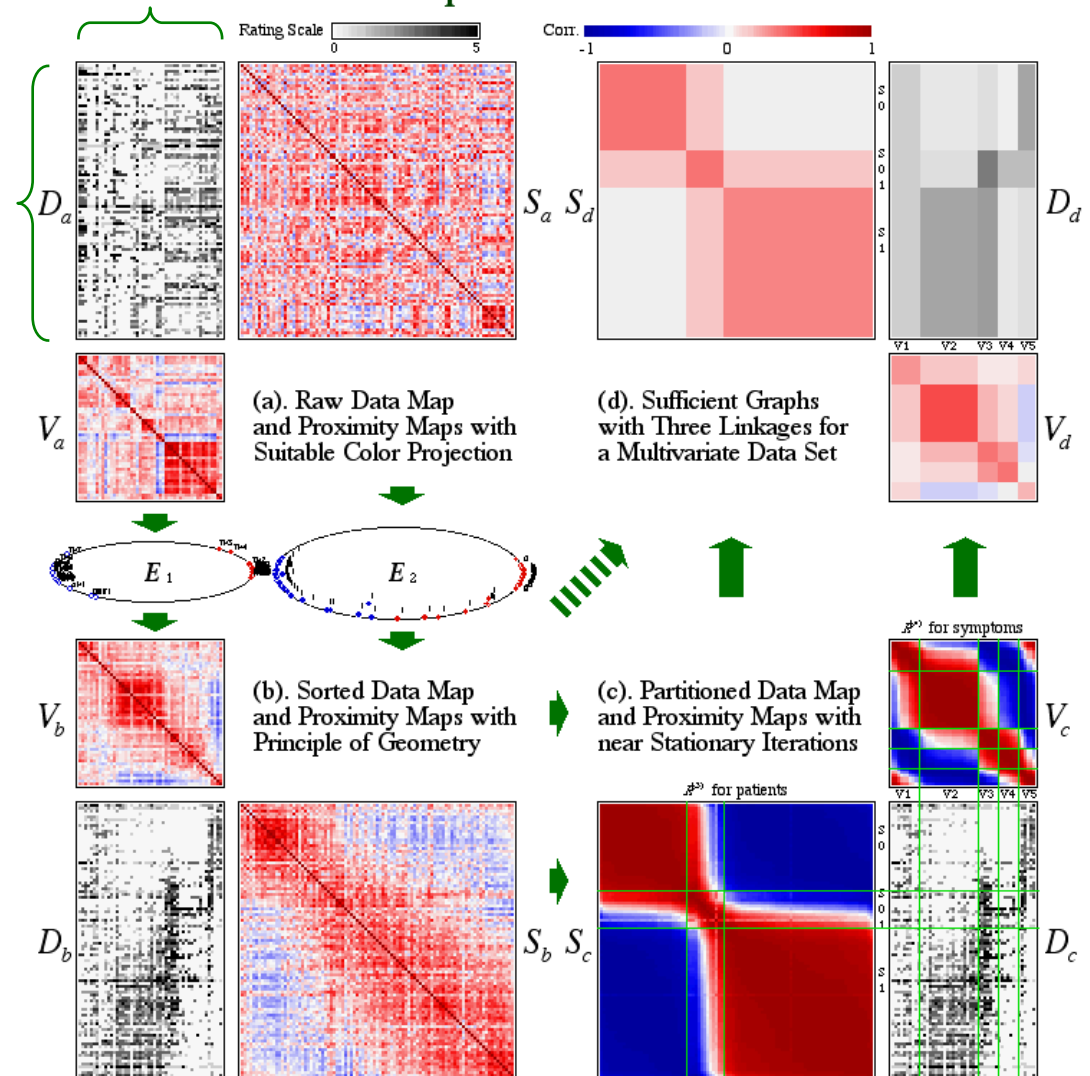
- 2D Scatterplot, 3D Scatterplot (Rotatable)



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

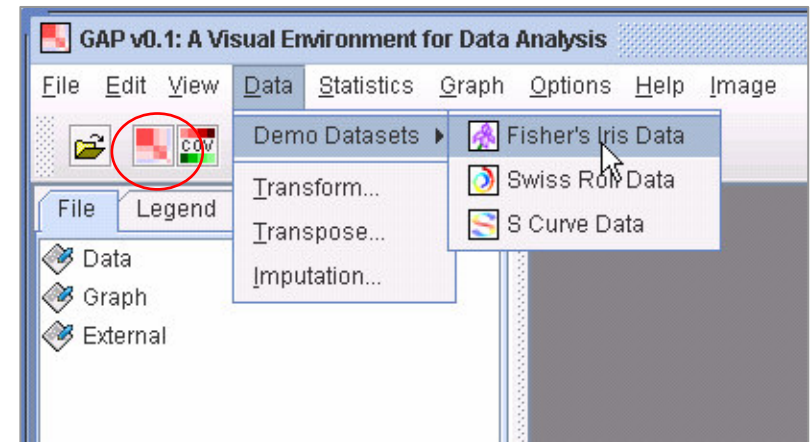
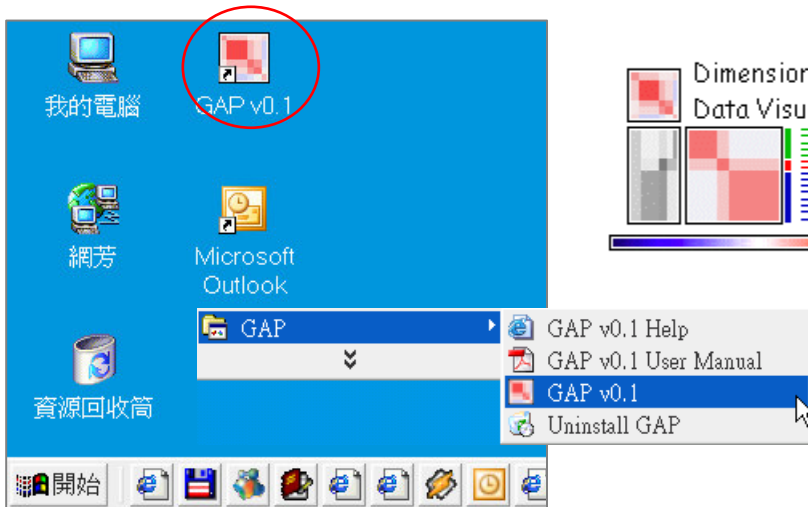
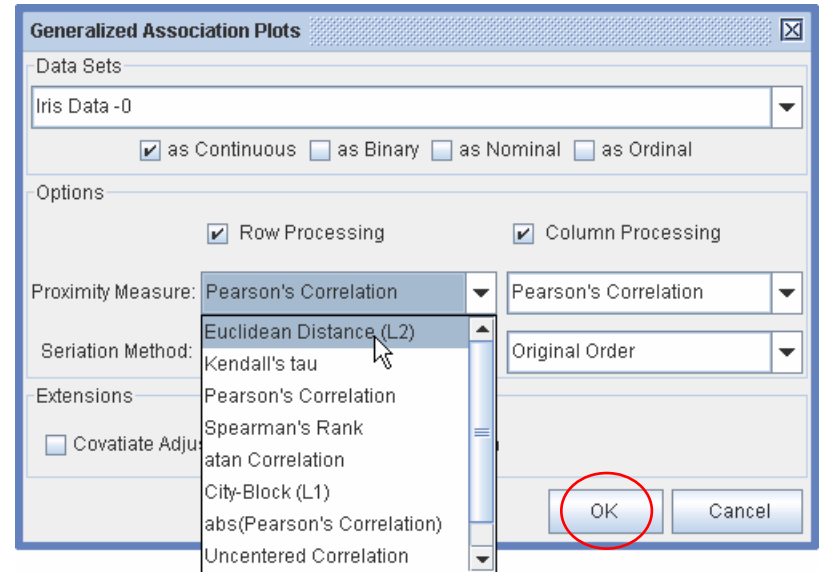
no.	Species	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
1	setosa	5.1	3.5	1.4	0.2
2	setosa	4.9	3.0	1.4	0.2
3	setosa	4.7	3.2	1.3	0.2
4	setosa	4.6	3.1	1.5	0.2
5	setosa	5.0	3.6	1.4	0.2
...					
76	versicolor	6.6	3.0	4.4	1.4
...					
150	virginica	5.9	3.0	5.1	1.8

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

The screenshot displays the GAP v0.1 software interface, which is a visual environment for data analysis. The main window is titled "Generalized Association Plots 2: Iris Dat [150x4]: as continuous, [Euclidean Distance (L2), Pearson's Correlation]". The interface includes a menu bar at the top with options like File, Edit, View, Data, Statistics, Graph, Options, Help, and Image. A toolbar is located below the menu bar. On the left side, there is a File/Data View showing a tree structure of data files, including "SAPSSANS9550.bt [95x50]", "Iris Data -0 [150x4]", and "GAP 2: Iris Dat". A color legend is visible in the top-left corner of the main plot area. The central plot area displays a heatmap with a dendrogram on the right side, representing the association between data points. Below the main plot, there is a "GAP Procedure" panel with options for "Converging Correlation Matrices" and "Finding Blocks". At the bottom, there is an "Output" window showing the results of the analysis, including the number of rows and columns, and a "Control Panel" with various settings for the data maps and condition. Callout boxes identify the following components: Menu Bar, Tool Bar, File/Data View, Mouse Tools, Color Legend, GAP Procedure, Modules Installed, Seriation and Tree Flip, GAP Maps Window, Output Window, and Control Panel.

Menu Bar

Tool Bar

File/Data View

Mouse Tools

Color Legend

GAP Procedure

Modules Installed

Seriation and Tree Flip

GAP Maps Window

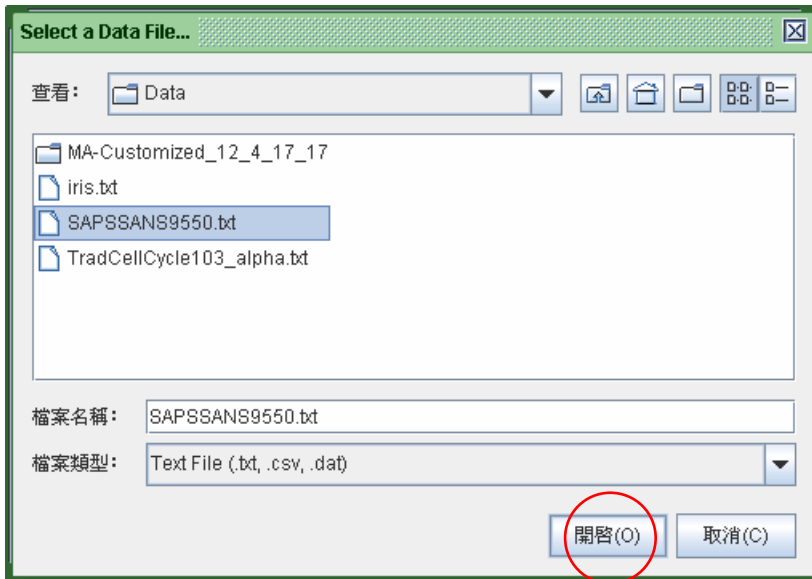
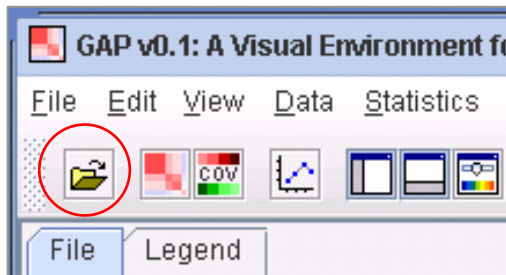
Output Window

Control Panel

Open/Import Data

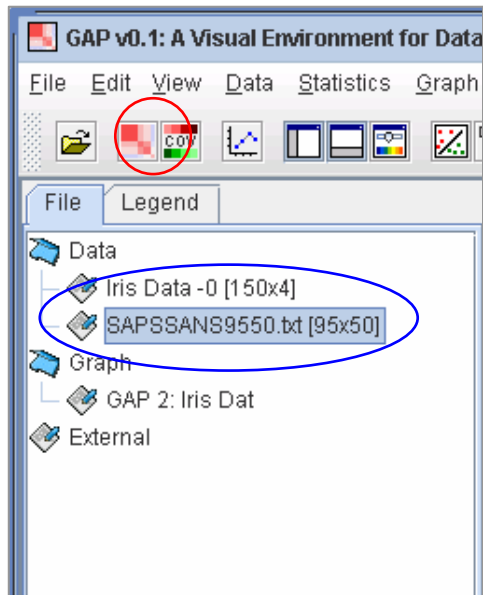
Data Format: plain text with tab separator

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



	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



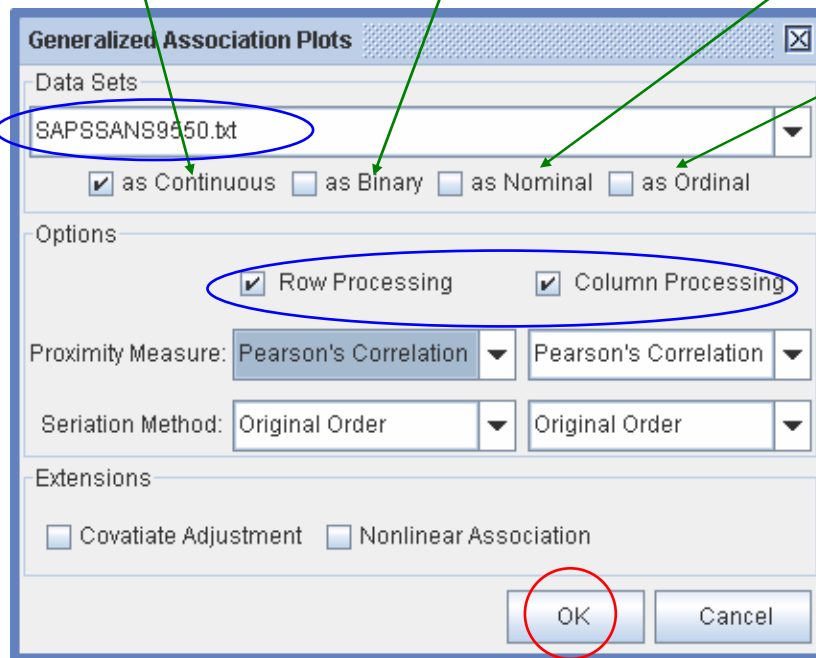
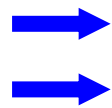
- Euclidean Distance (L2)
- Kendall's tau
- Pearson's Correlation
- Spearman's Rank
- atan Correlation
- City-Block (L1)
- abs(Pearson's Correlation)
- Uncentered Correlation
- abs(Uncentered Correlation)

- 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: $(ad-bc)/(ad+bc)$

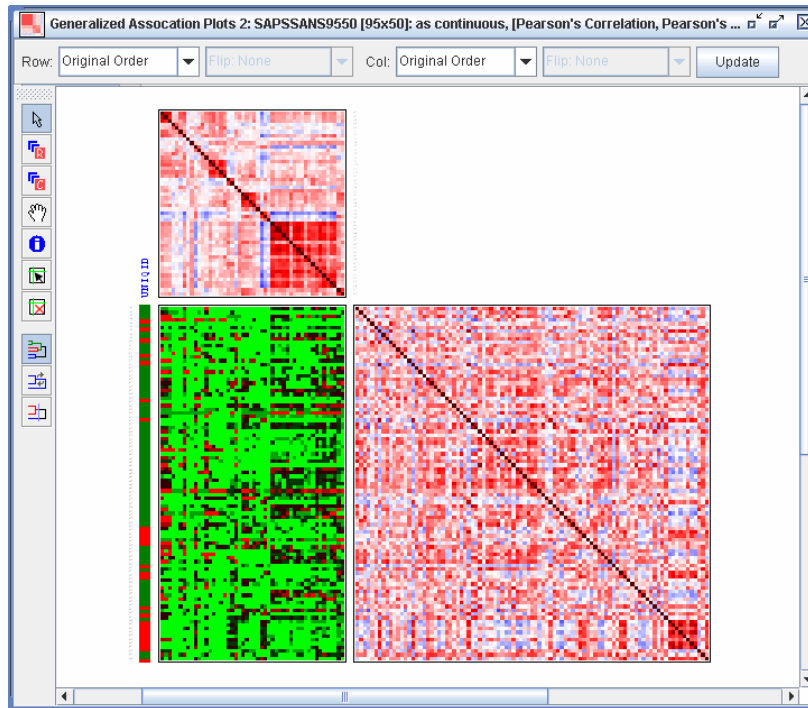
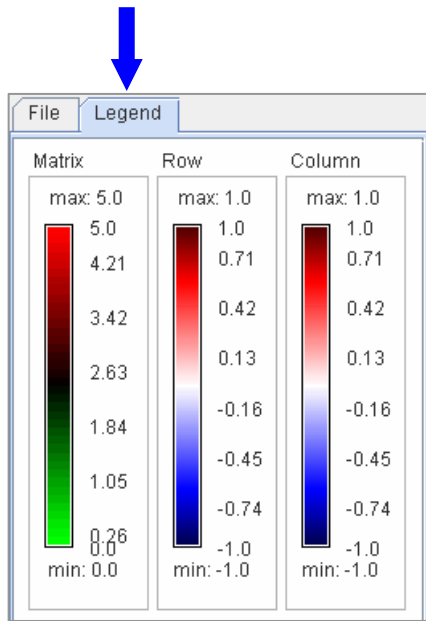
Not implemented Yet!

- (x) 2x2 Correlation
- (x) Cramer's V
- (x) Cohen's kappa
- (x) Goodman-Kruskal tau

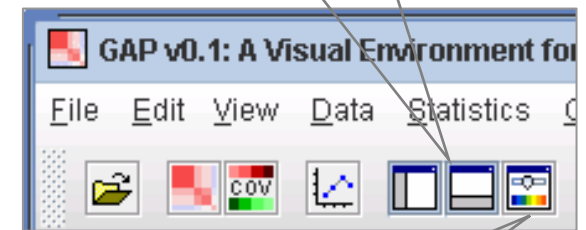
- (x) Goodman-Kruskal's gamma
- (x) Kendall's tau_b
- (x) Somers's d
- (x) Stuart's tau_c
- (x) Wilson's e



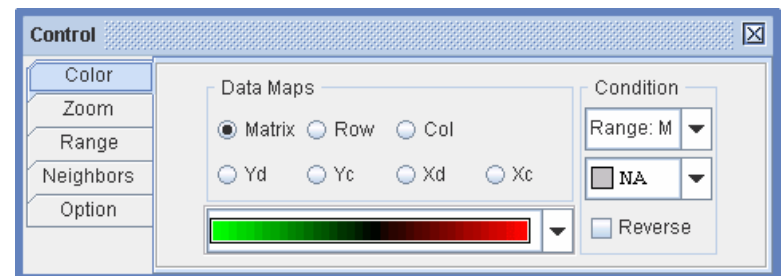
GAP Step 1: Presentation of Raw Data Matrix



Show or Hide File/Data View, Output Panel


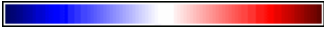







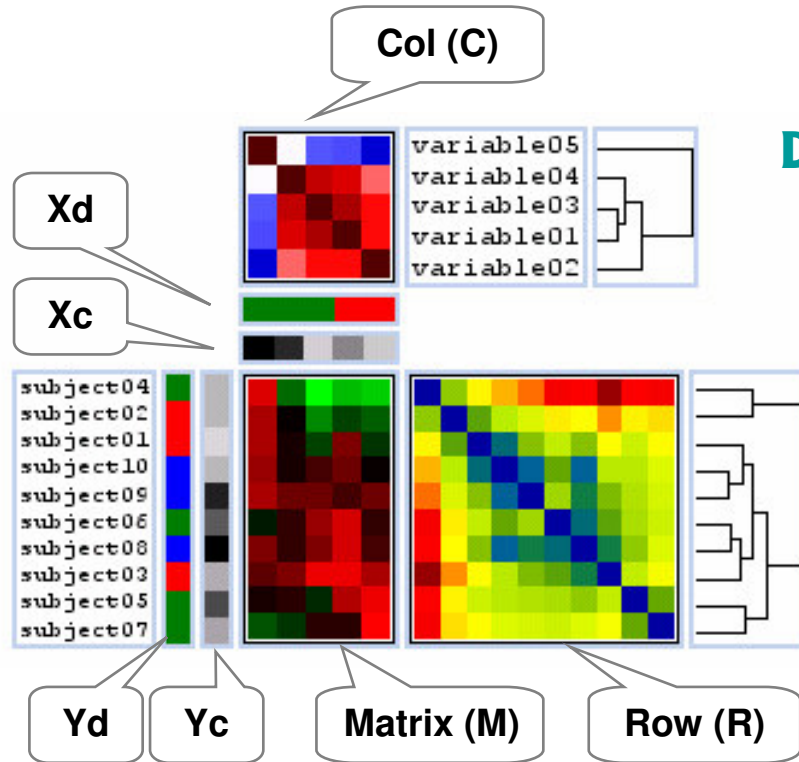
Show or Hide Control Panel



Control Panel: Color

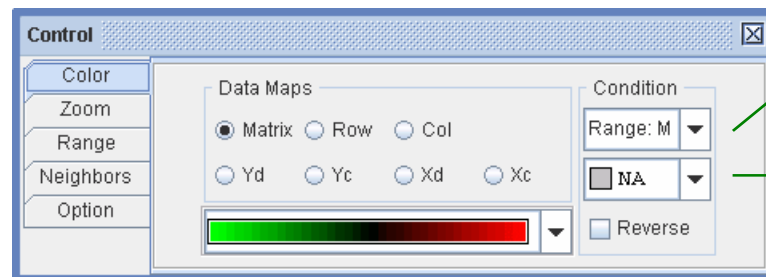


- Rainbow 130**
 Distance
- Blue-white-Red 200**
 Correlation
- Green-Black-Red 38**
 Gene Expression
- Grey 256**
 Range
- Discrete 16**

- Biallel Marker**

- ACGT**







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

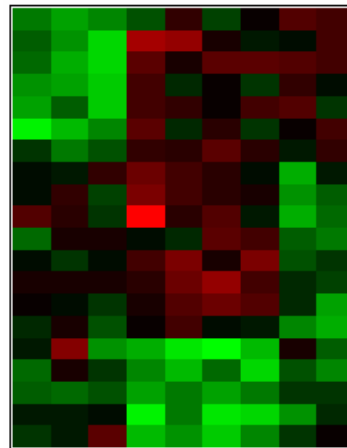


Missing Value

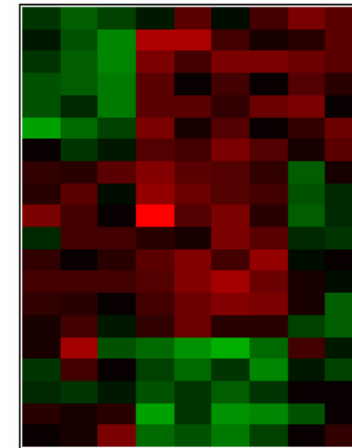
-  NA
-  NA
-  NA

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



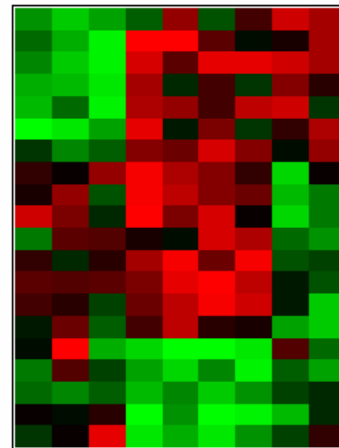
Center Matrix Condition

Bidirectional Color Spectrum

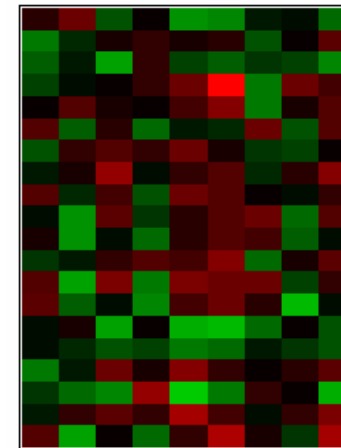
Blue-white-Red 200



Green-Black-Red 38

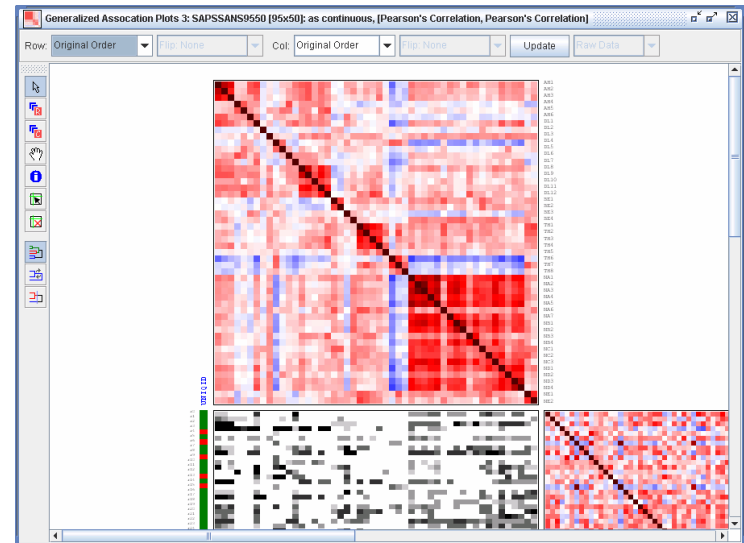
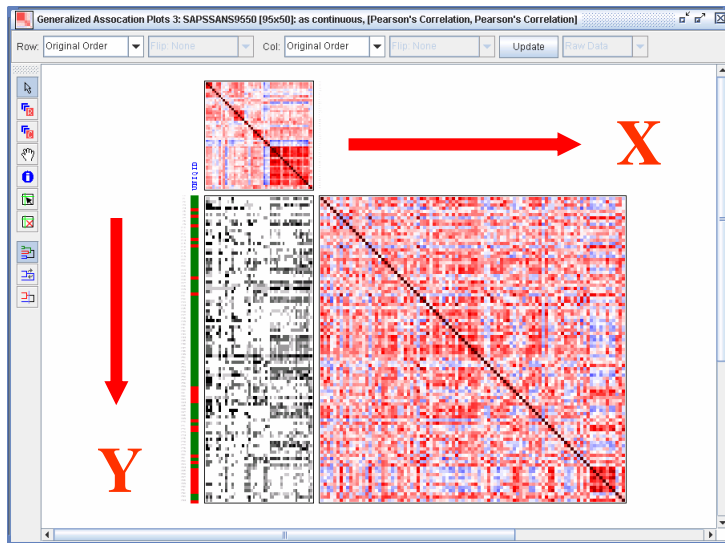


Rank Matrix Condition



Standardize Matrix Condition

Control Panel: Zoom



Actual Size

Fit Window

Fit Width

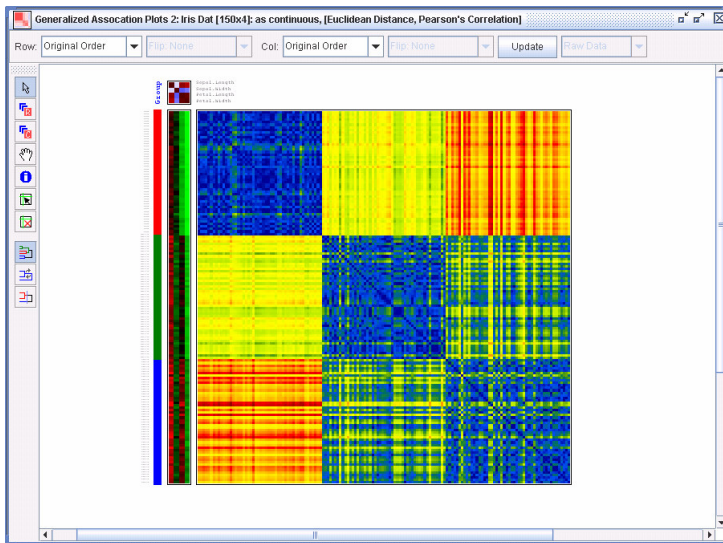
The Control panel is shown with three rows of zoom settings. Each row has a magnifying glass icon, a slider, and a percentage value. The first row has a 100% zoom level, the second row has a 100% zoom level, and the third row has a 76% zoom level. A blue circle highlights the sliders for the first two rows. Two blue arrows point to the zoom icons for the first and second rows.

Row	Zoom Level
1	100%
2	100%
3	76%

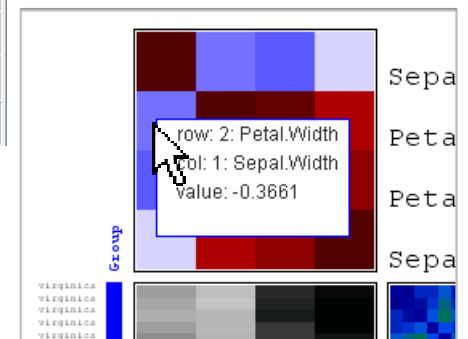
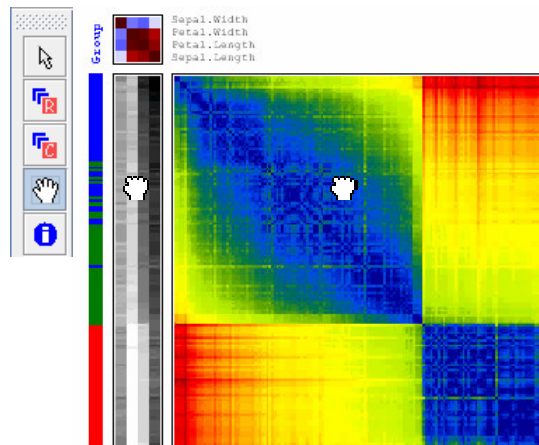
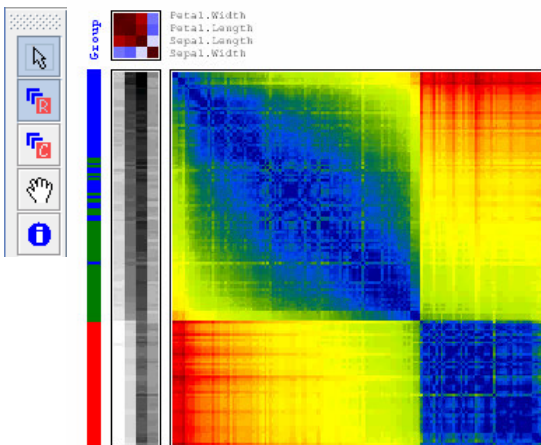
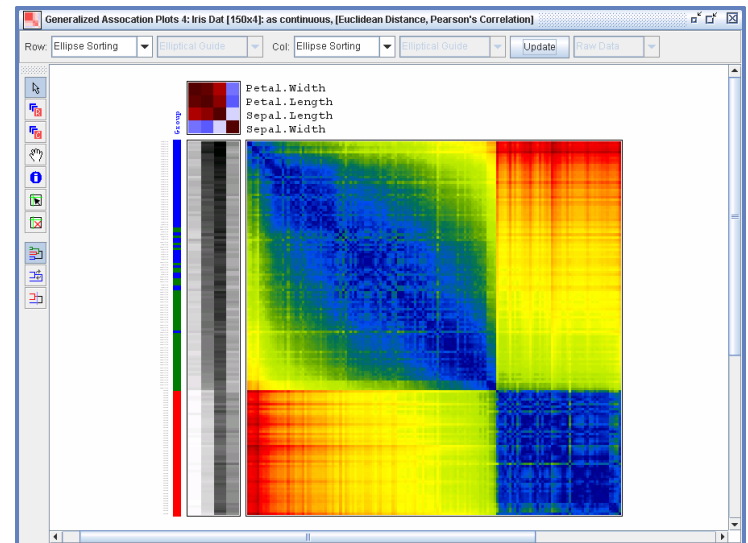
GAP Step 2: Ellipse Seriation

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

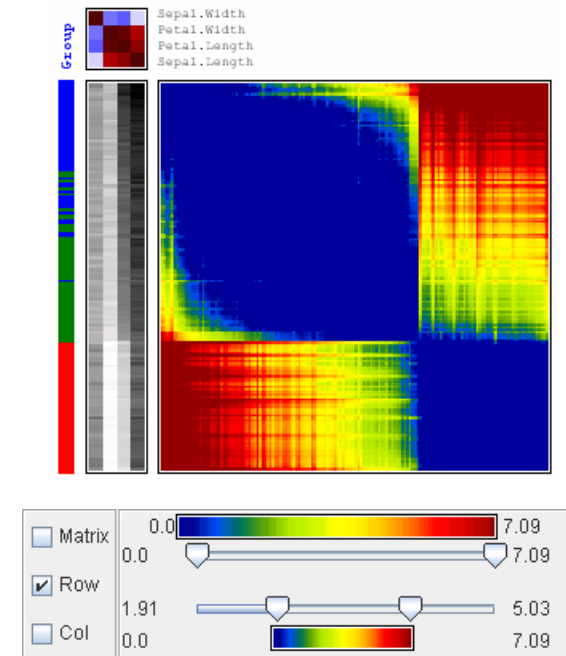
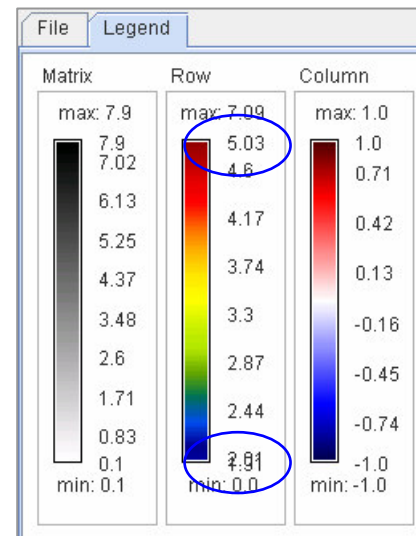
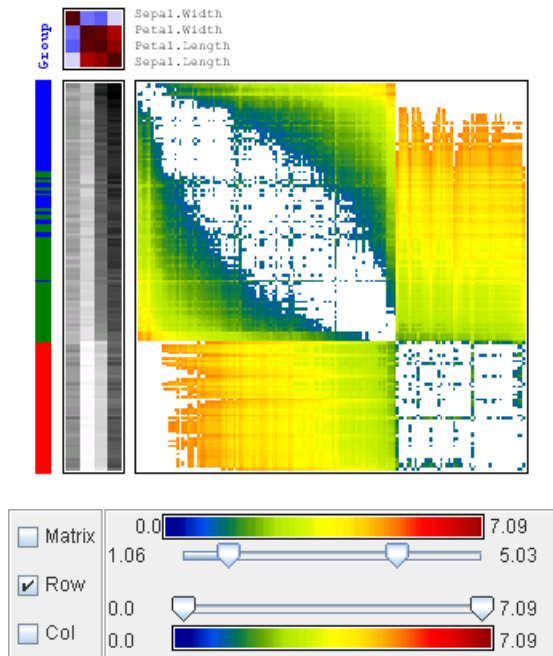
Row: **Ellipse Sorting** | Elliptical Guide | Col: **Ellipse Sorting** | Elliptical Guide | **Update** | Raw Data



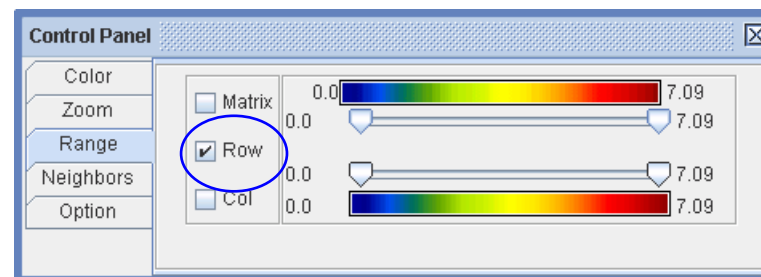
- Average-linkage
- Centroid-linkage
- Complete-linkage
- Ellipse Sorting**
- Original Order
- Single-linkage
- External
- Customized



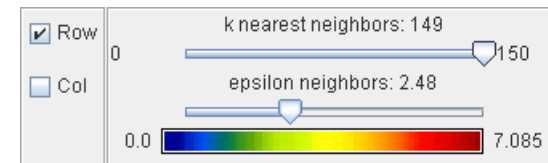
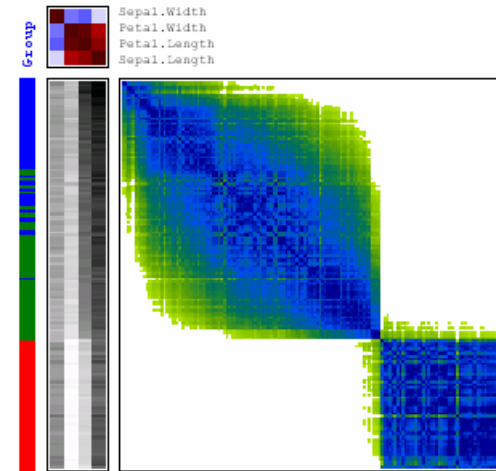
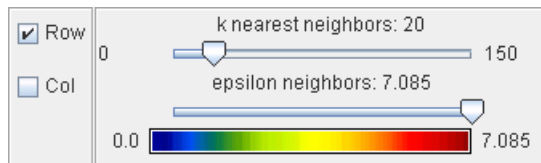
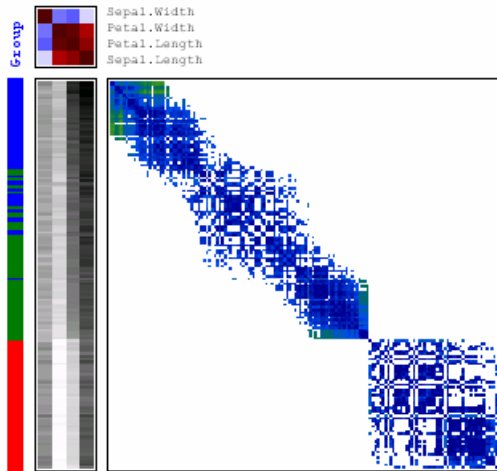
Control Panel: Range



“Resolution” of a Statistical Graph



Control Panel: Neighbors



Control Panel

Color

Zoom

Range

Neighbors

Option

Row

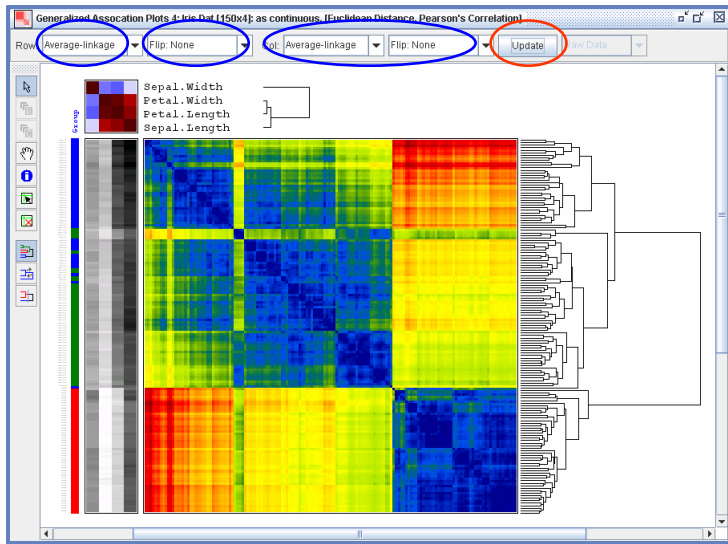
Col

k nearest neighbors: 149

epsilon neighbors: 7.085

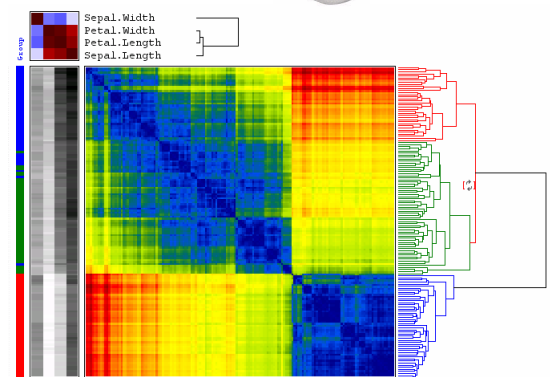
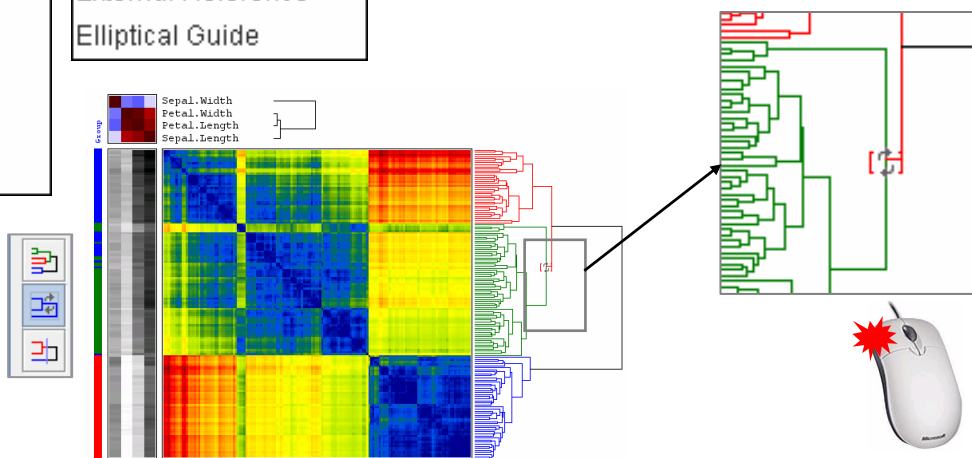
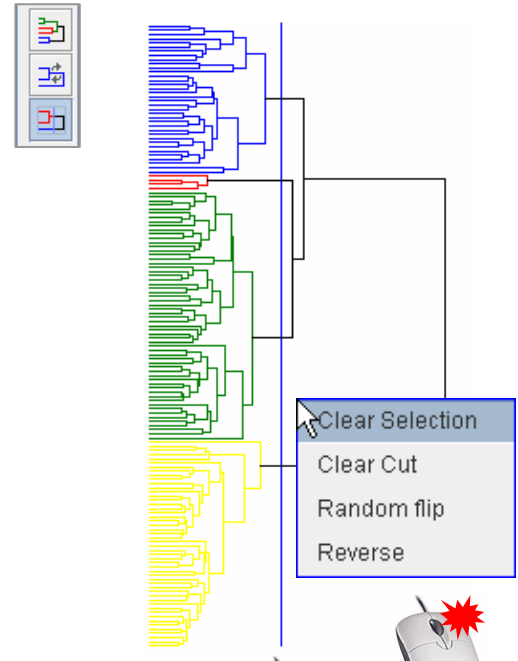
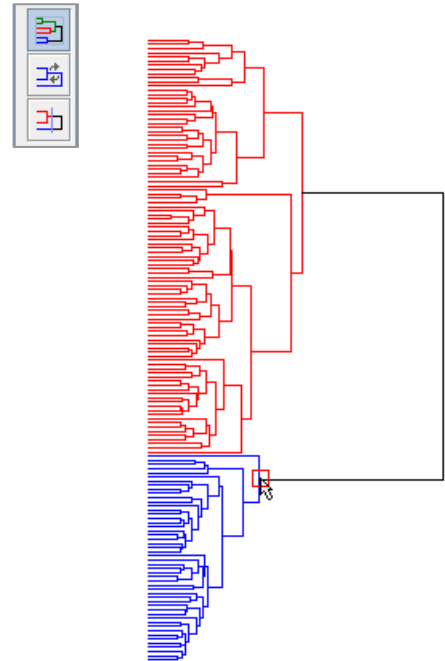
0.0 7.085

GAP Step 2: Tree Seriation/Flip

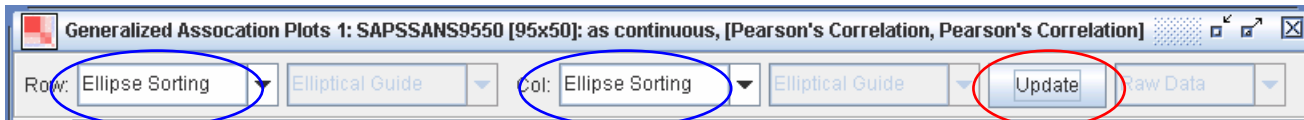


- Average-linkage
 - Centroid-linkage
 - Complete-linkage
 - Ellipse Sorting
 - Original Order
 - Single-linkage
 - External
 - Customized
- Flip: None
 - Flip: Uncle
 - Flip: GrandPa
 - External Reference
 - Elliptical Guide

Update



Pop-Up Menu: Sort by...

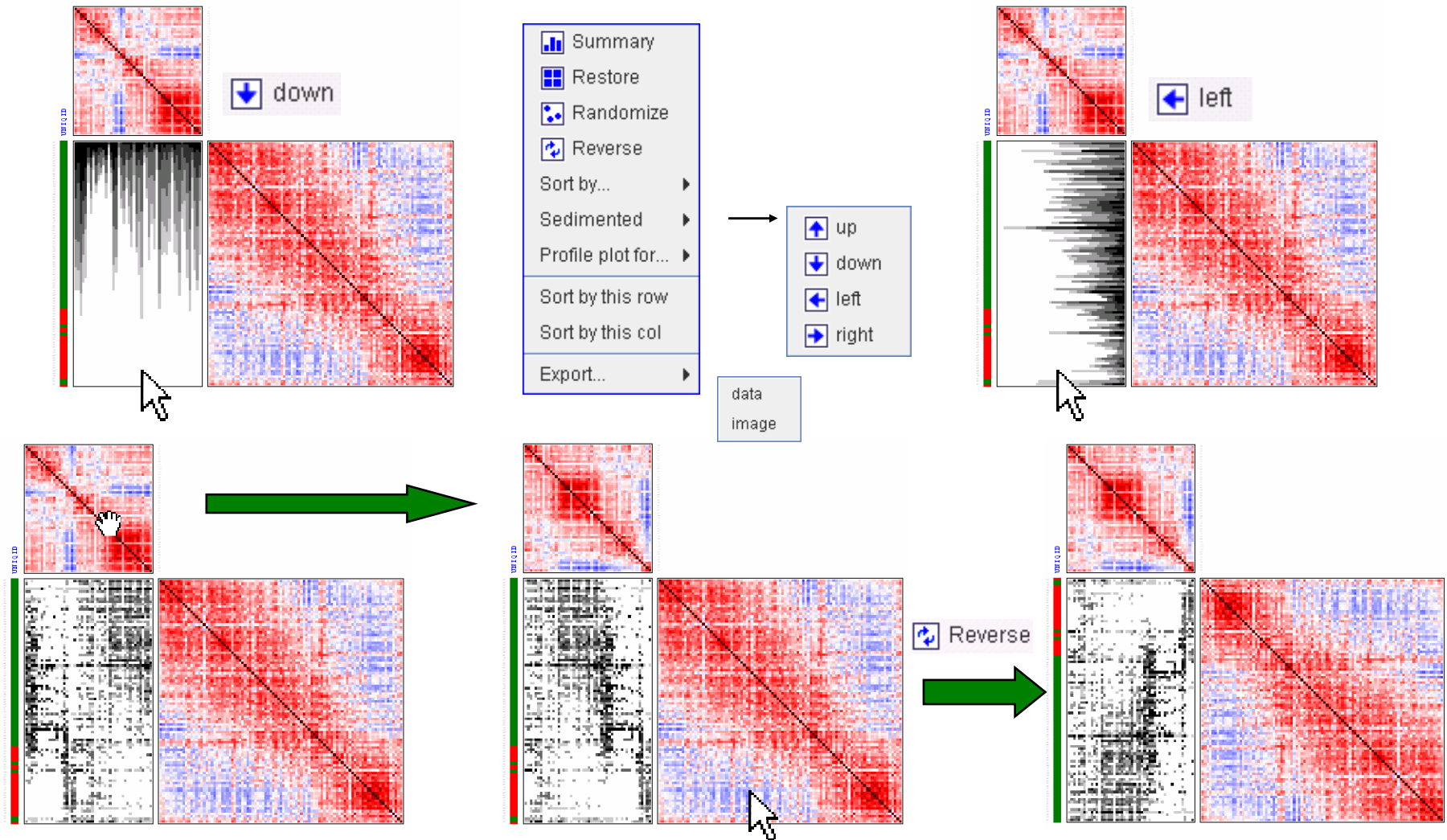


The figure illustrates the 'Sort by...' menu options through a sequence of six heatmaps. Each heatmap is accompanied by a mouse cursor icon and a corresponding menu button:

- Restore:** A mouse cursor points to the top-right corner of the heatmap. A button labeled 'Restore' is shown to the right.
- Randomize:** A mouse cursor points to the bottom-left corner of the heatmap. A button labeled 'Randomize' is shown to the right.
- Reverse:** A mouse cursor points to the bottom-left corner of the heatmap. A button labeled 'Reverse' is shown to the right.
- Sort by...:** A mouse cursor points to the bottom-left corner of the heatmap. A button labeled 'Sort by...' is shown to the right, with a sub-menu containing 'Means', 'Medians', 'Stds', and 'CVs'.

On the left side of the sequence, a vertical menu is shown with the following options: 'Sort', 'Export data', and 'Export image'. A mouse cursor points to the 'Sort' option.

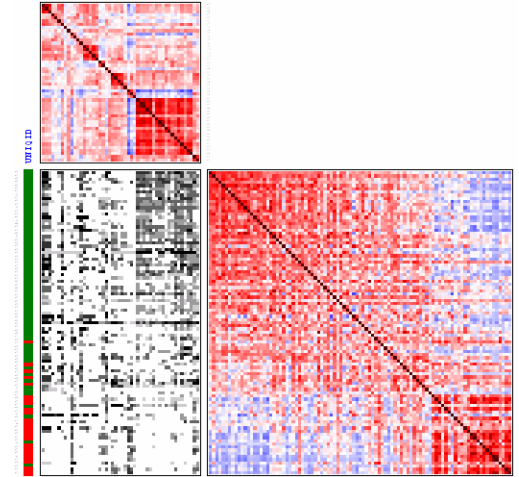
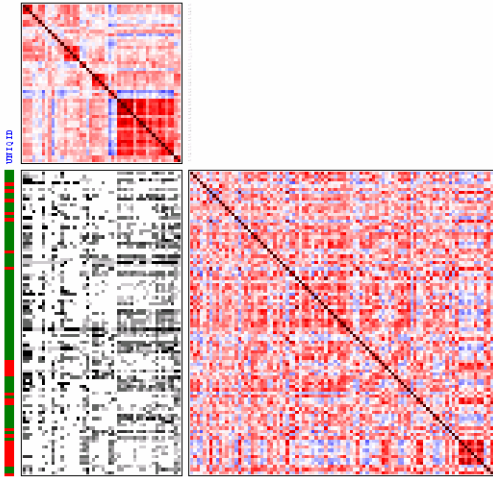
Pop-Up Menu: Sedimented plot



Mouse Tool: Find Similar Patterns

Generalized Association Plots 1: SAPSSANS9550 [95x50]: as continuous, [Pearson's Correlation, Pearson's Correlation]

Row: Original Order Flip: None Col: Original Order Flip: None **Update**



Input pattern

Finding Similar Patterns Dialog

Option

Subject: Subject

Pattern: 4 4 4 1 4 3 5 5 4 1 3 3 3 0

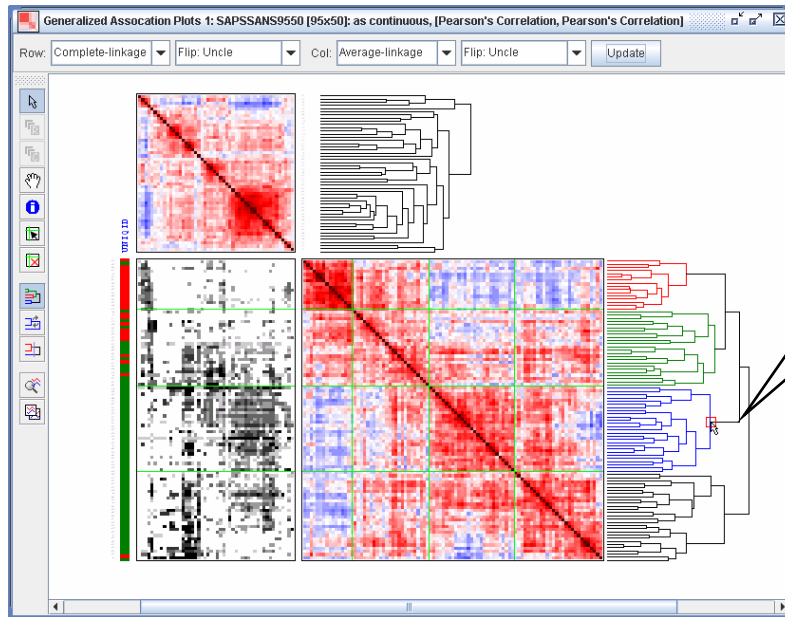
Data Type: Continuous

Similarity: Euclidean Distance

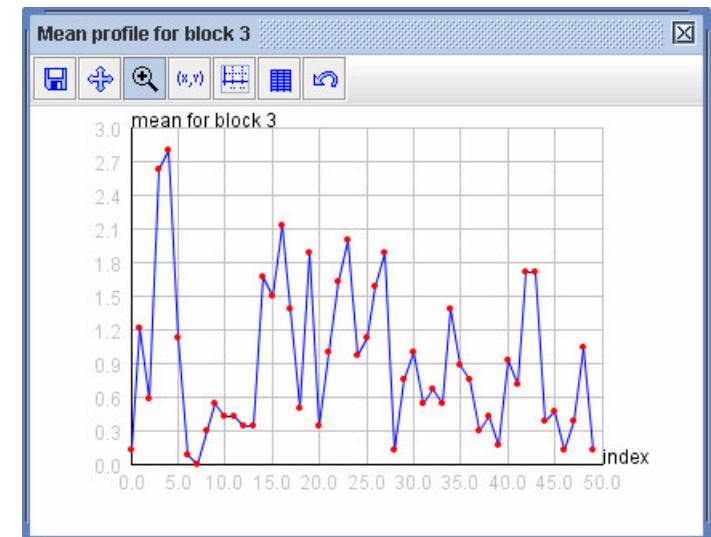
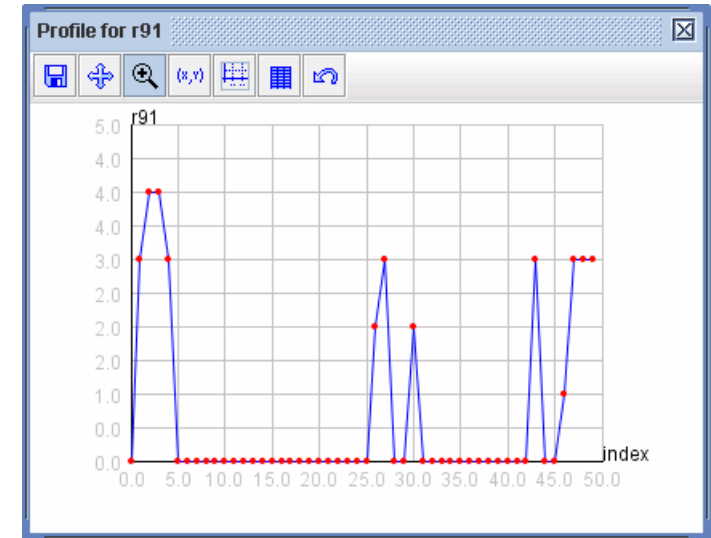
Preview

Compute Close

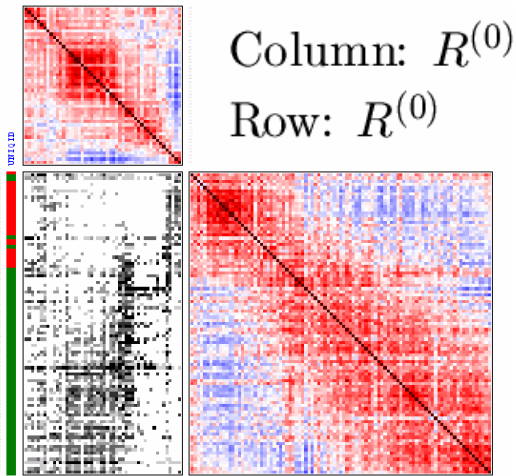
Pop-Up Menu: Profile plot



- Summary
- Restore
- Randomize
- Reverse
- Sort by... ▶
- Sedimented ▶
- Profile plot for... ▶
 - this row
 - this block (mean)
 - this block (all)
 - all blocks (mean)
 - all blocks (all)
- Sort by this row
- Sort by this col
- Export... ▶



Converging Correlation Matrices



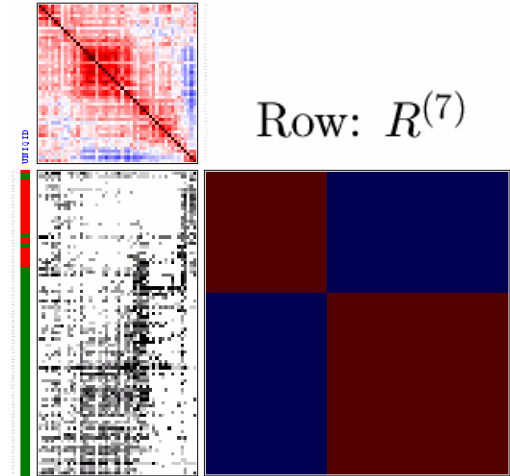
Converging Correlation Matrices

Row: 3 2D Plot 3D Plot

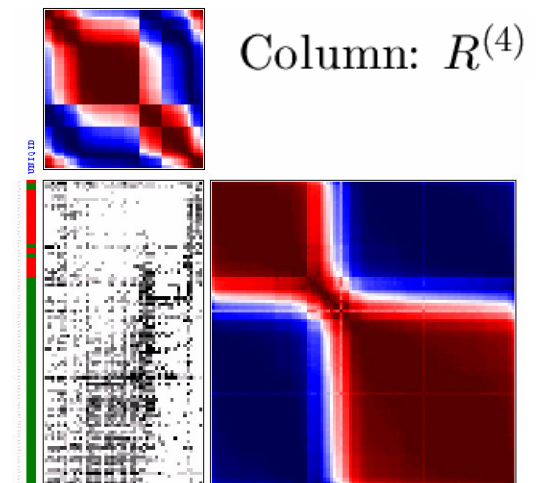
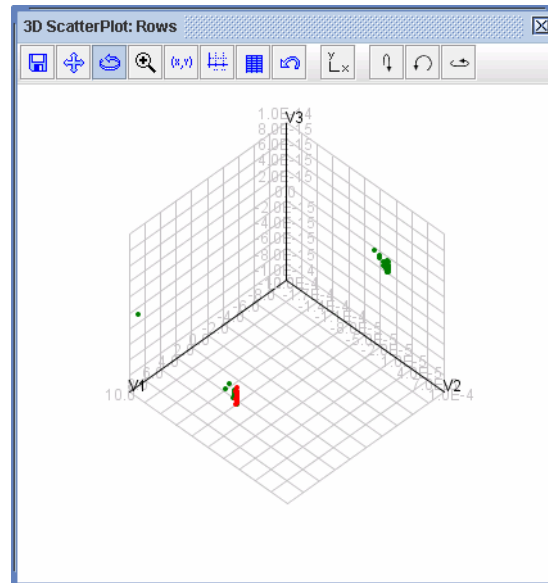
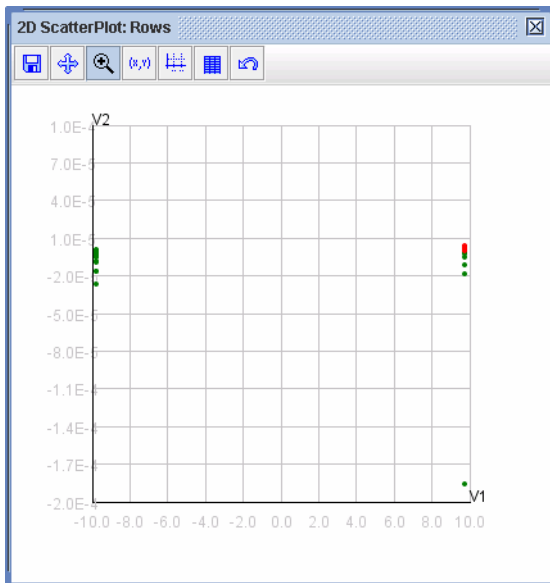
Column: 4 2D Plot 3D Plot

Navigation icons: Home, Left, Right, Stop

↑ ↑



Row: $R^{(7)}$



GAP Step 3: Partitions of Permuted Matrix Maps

Finding Blocks:

Row Column

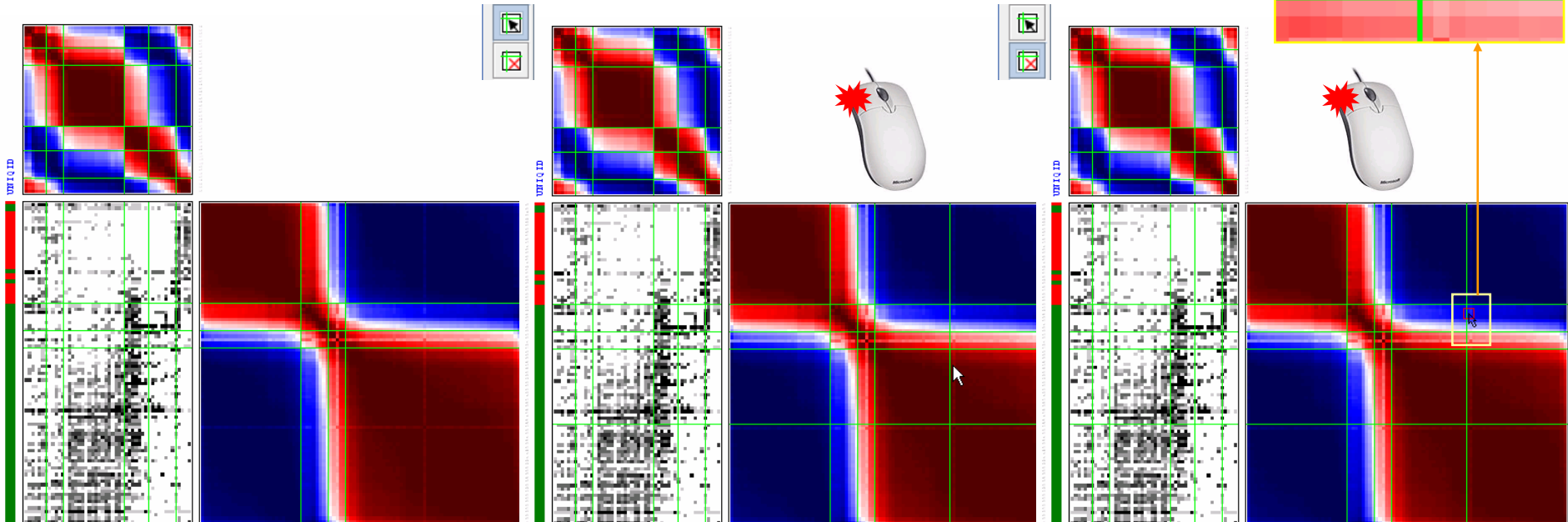
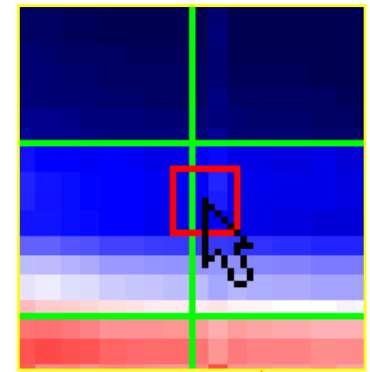
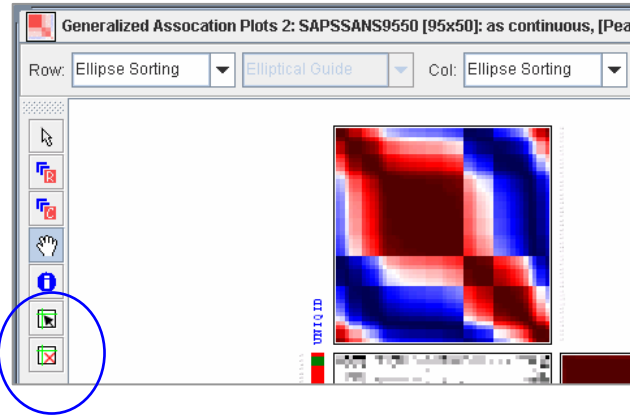
Simple Simple

Update **Clear**

- Simple
- Dendrogram
- max(BSS/WSS)

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

Row: Ellipse Sorting Elliptical Guide Col: Ellipse Sorting





GAP Step 4: Sufficient Graph

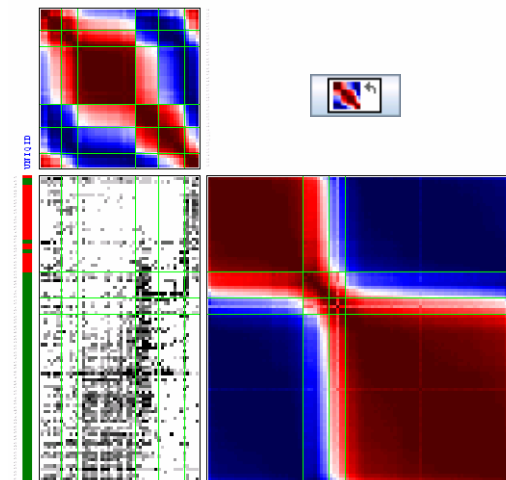
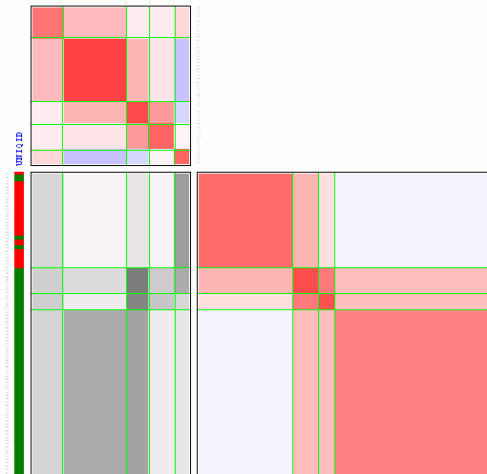
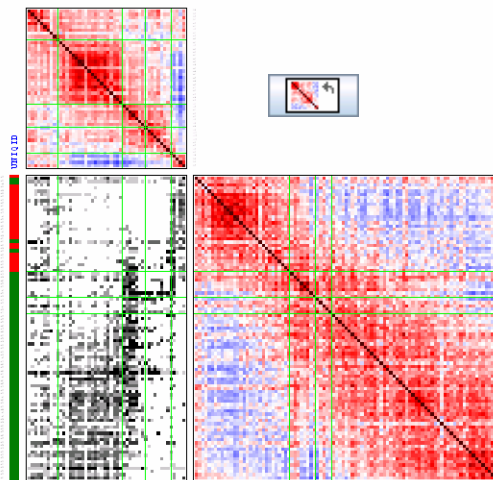
Sufficient Graphs

Matrix Row Column

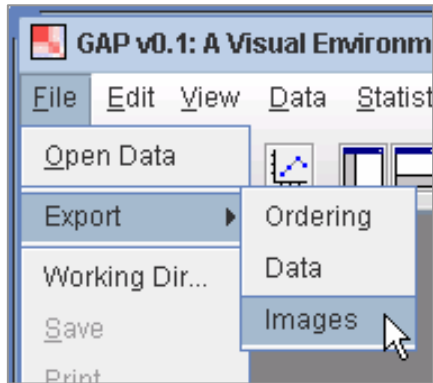
Mean Mean Mean

Update  

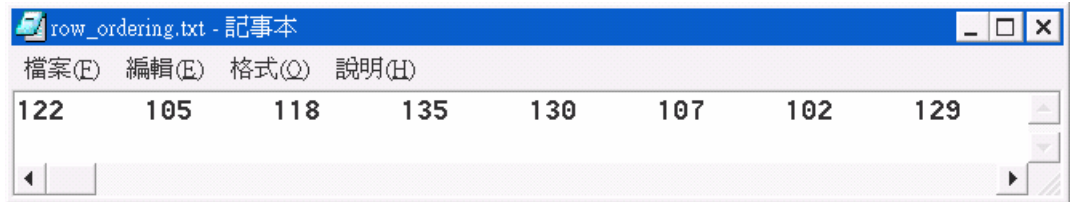
- Mean
- Median
- Std.
- CV



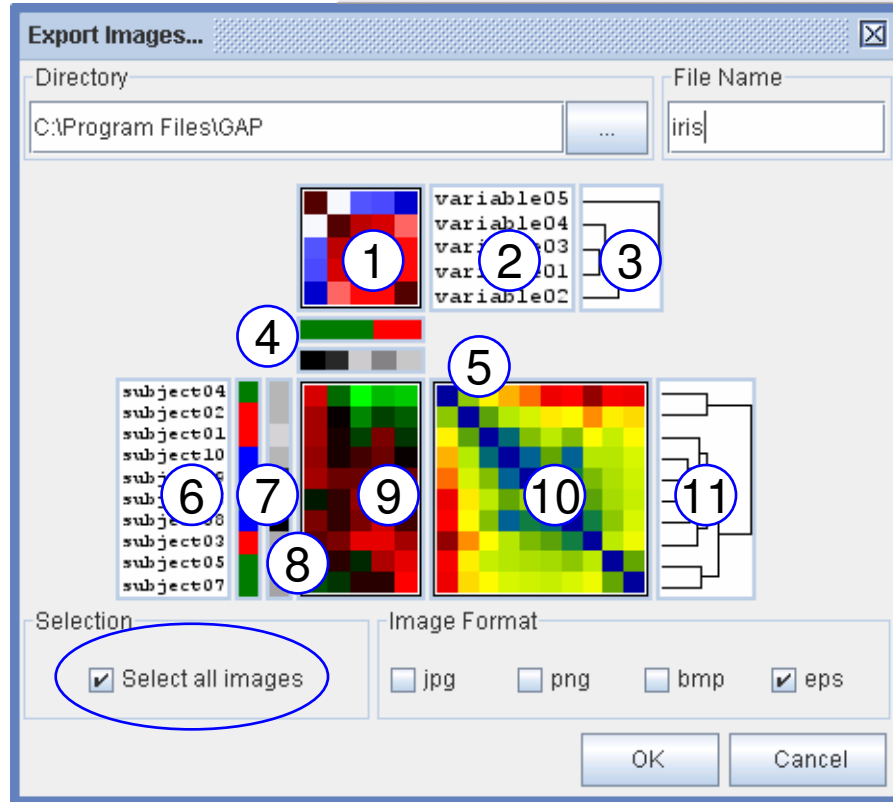
Export Ordering, Data, Images



row_ordering.txt
col_ordering.txt

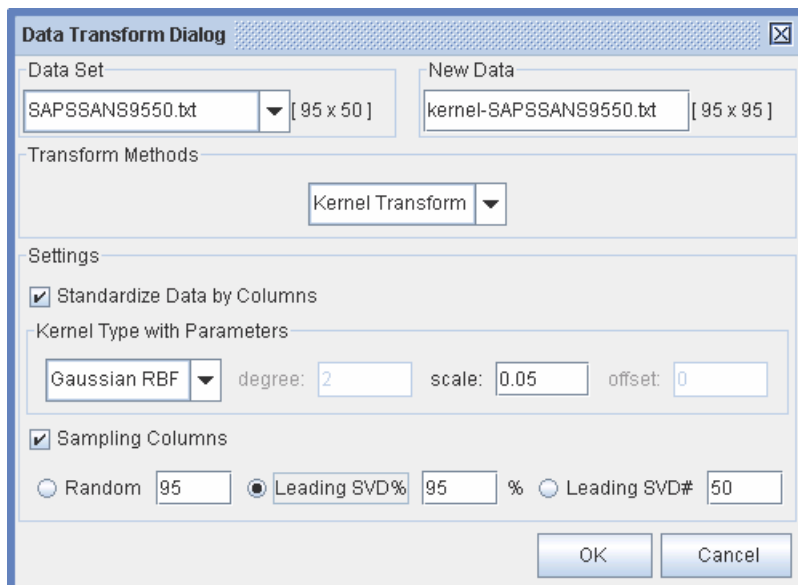
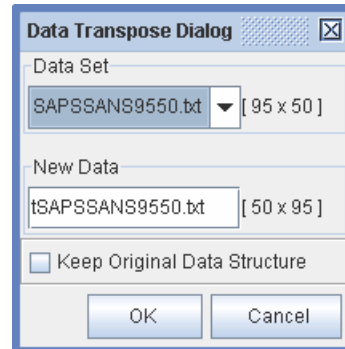
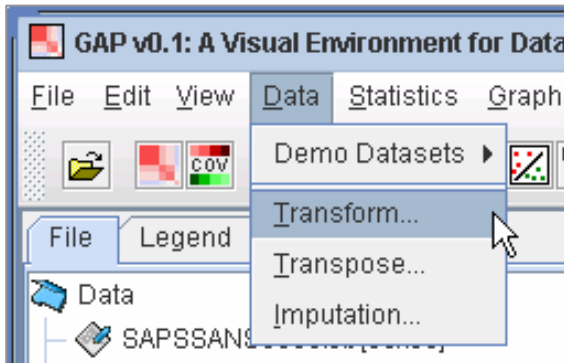


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	LINOID	TH3	TH4	TH2	TH5	TH1	NE2	NE1	DL2	BE1	BE2	NA6	NC1	DL3	NC2	NE2	NE1
2	#62	1	1	5	1	5	5	5	0	2	3	1	1	4	4	2	
3	#27	2	2	1	2	1	1	2	0	0	2	0	1	0	2	2	
4	#52	0	0	2	0	2	0	3	0	0	0	0	0	0	2	0	
5	#4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6	#89	0	0	1	2	2	2	1	0	0	0	0	0	0	0	0	
7	#84	0	0	1	0	2	1	1	0	0	2	1	1	1	1	1	
8	#88	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	
9	#60	2	0	3	1	3	3	3	0	2	0	0	0	0	3	0	
10	#90	0	0	3	0	2	0	2	0	0	0	0	0	0	0	0	
11	#87	0	0	0	0	0	1	1	0	2	3	2	2	0	1	0	
12	#91	0	0	0	0	0	2	3	3	3	3	0	2	0	3	0	
13	#86	0	0	0	0	0	0	3	0	2	3	0	0	0	0	0	
14	#15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
15	#6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
16	#72	0	0	0	0	0	1	3	0	1	0	1	0	1	0	2	
17	#90	0	0	0	0	2	1	1	0	2	2	1	0	0	0	0	
18	#86	0	0	0	0	0	1	3	0	2	0	0	0	0	2	1	
19	#71	0	0	0	0	0	3	4	0	0	3	0	0	0	0	0	
20	#9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
21	#94	0	0	0	2	0	0	3	0	0	0	0	0	0	3	0	
22	#51	5	5	5	0	5	0	2	0	0	2	2	0	3	2	2	
23	#91	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
24	#82	0	2	3	0	3	2	3	0	0	0	0	0	0	0	0	
25	#24	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	#90	0	0	0	0	2	0	3	0	1	0	0	0	2	0	0	
28	#13	0	0	0	0	0	3	3	0	0	0	0	0	1	0	0	
29	#9	3	3	0	4	2	3	0	0	2	2	0	0	2	0	4	
30	#83	0	3	4	4	0	0	1	0	0	2	0	0	0	0	0	
31	#89	0	0	3	0	0	2	0	0	3	3	2	1	0	2	0	
32	#29	4	4	3	1	4	1	1	0	0	1	1	0	0	1	1	
33	#19	0	0	0	0	2	2	2	0	1	0	0	3	0	2	0	
34	#84	3	4	4	2	5	5	4	0	2	4	2	0	4	0	0	
35	#42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
36	#75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
37	#33	0	0	0	0	0	2	0	0	0	0	0	0	2	0	0	
38	#3	0	3	0	0	0	0	0	0	0	1	0	0	0	0	0	
39	#73	2	3	0	0	3	2	2	0	0	2	2	0	0	4	0	
40	#86	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	
41	#10	0	0	0	0	0	2	0	0	0	0	0	0	0	3	0	
42	#20	3	4	3	0	4	5	4	0	2	4	0	3	0	3	0	
43	#1	0	0	0	0	0	2	0	0	0	2	0	0	0	3	0	
44	#56	0	2	1	2	2	1	2	0	0	0	0	0	0	3	1	
45	#11	0	0	0	0	0	1	1	1	3	0	3	0	2	0	0	
46	#7	0	3	0	0	0	1	0	0	0	0	0	0	0	3	0	
47	#57	0	3	0	0	0	1	1	0	0	2	0	0	0	0	3	
48	#70	0	0	0	0	0	2	2	0	0	0	2	0	3	0	0	
49	#40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
50	#44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

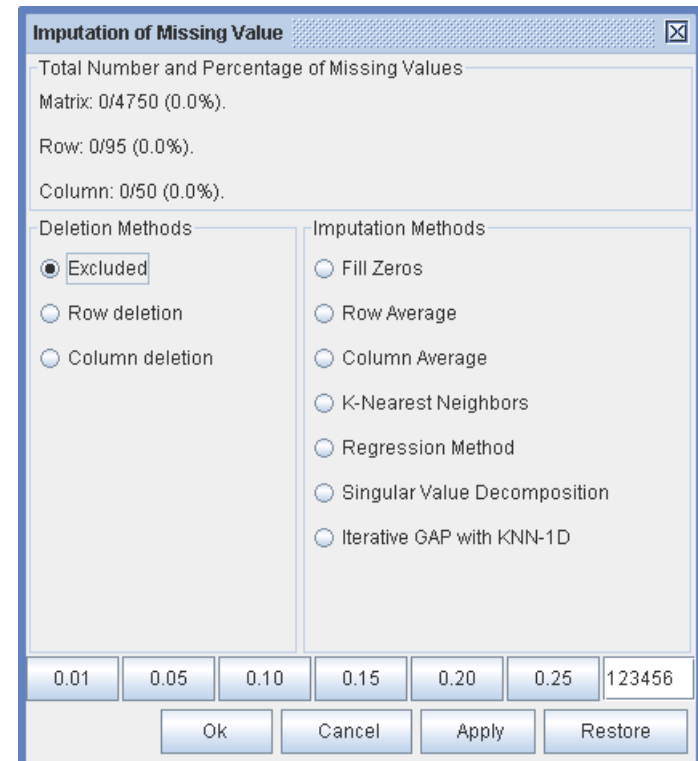


- 1 _gapCol.eps
- 2 _gapColName.eps
- 3 _TreeCol.eps
- 4 _gapXd.eps
- 5 _gapXc.eps
- 6 _gapRowName.eps
- 7 _gapYd.eps
- 8 _gapYc.eps
- 9 _gapMatrix.eps
- 10 _gapRow.eps
- 11 _TreeRow.eps

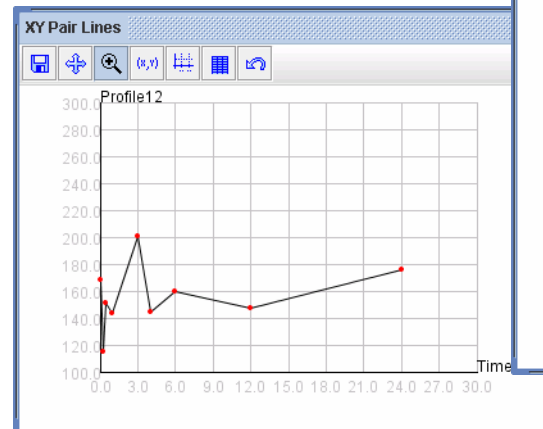
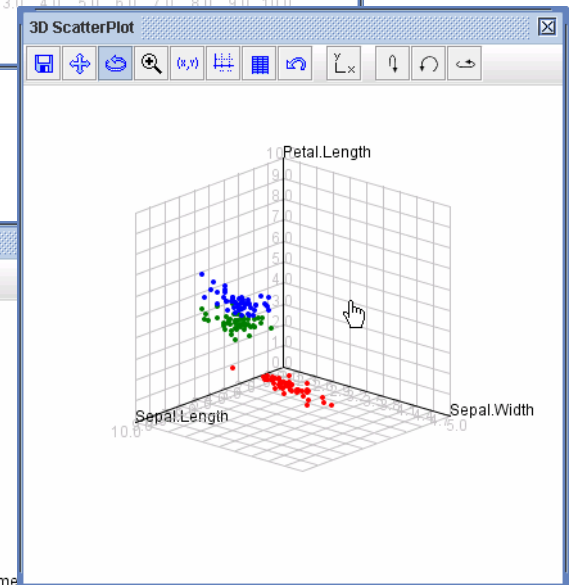
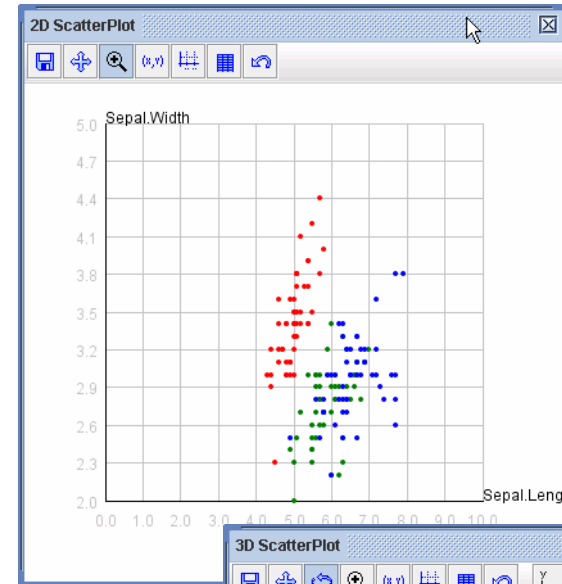
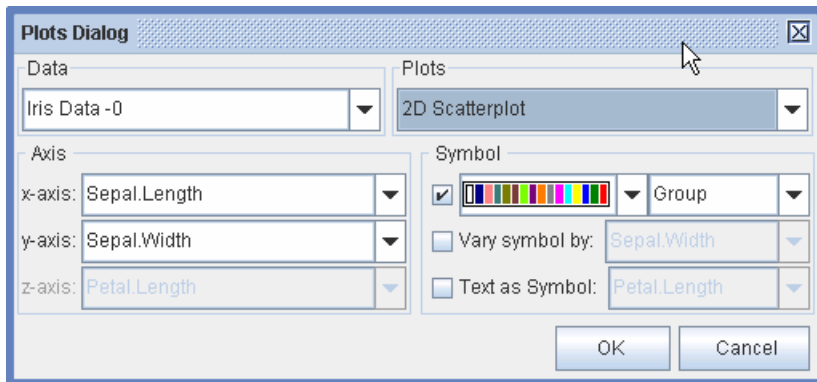
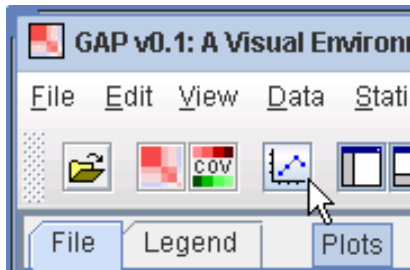
Data Processing



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

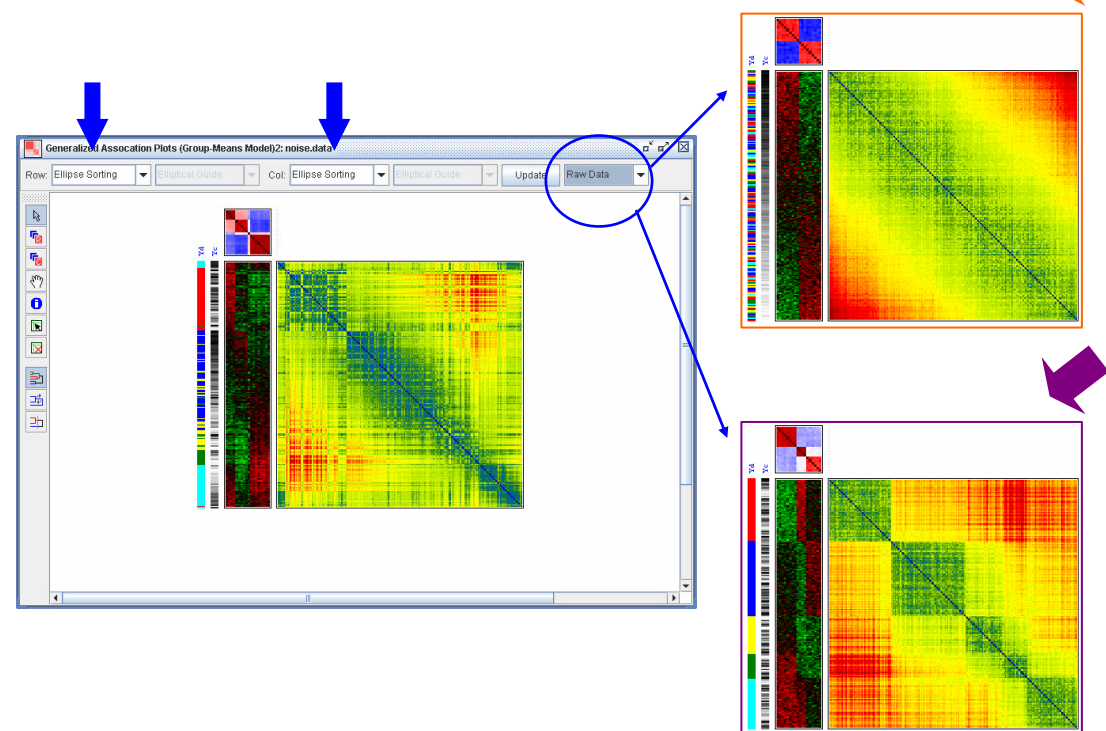
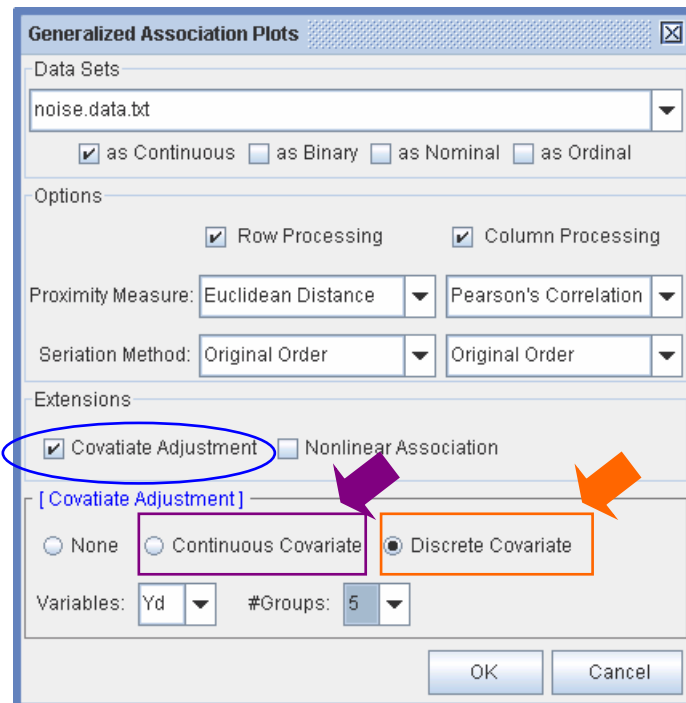
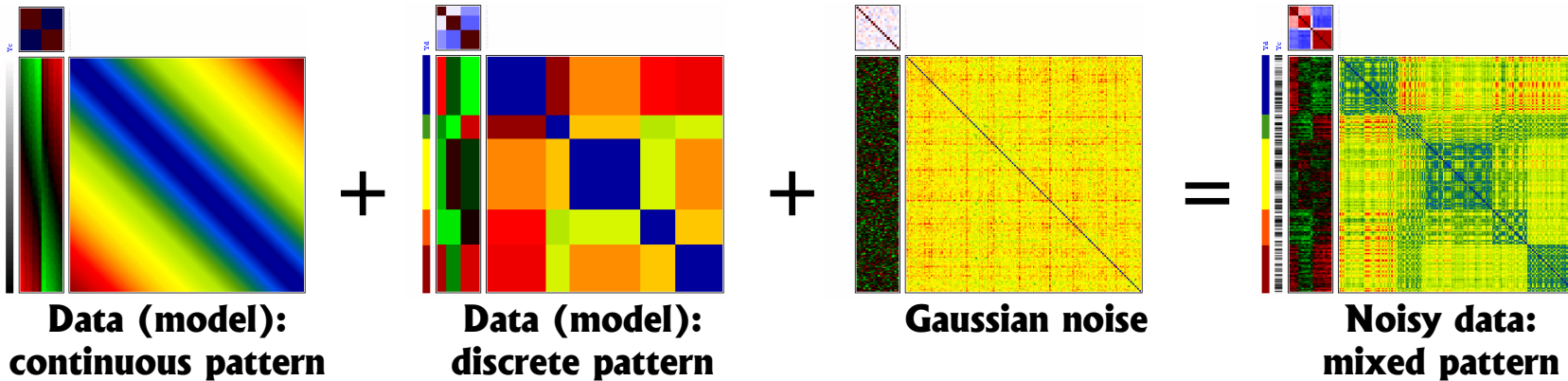


Plots

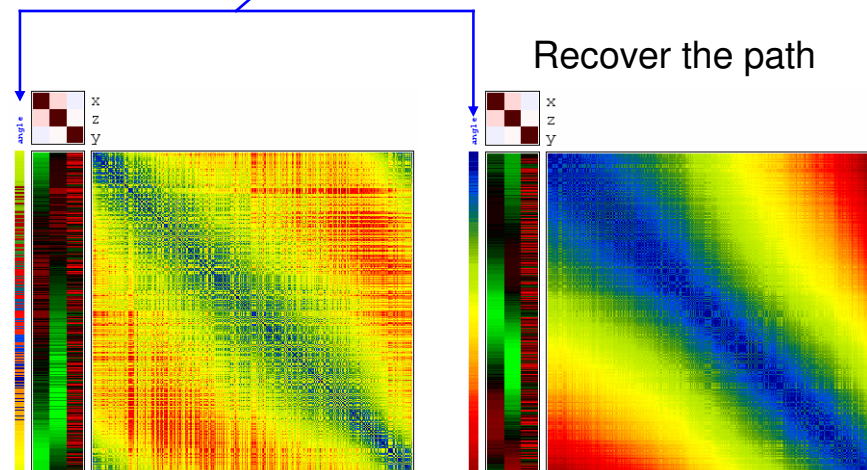
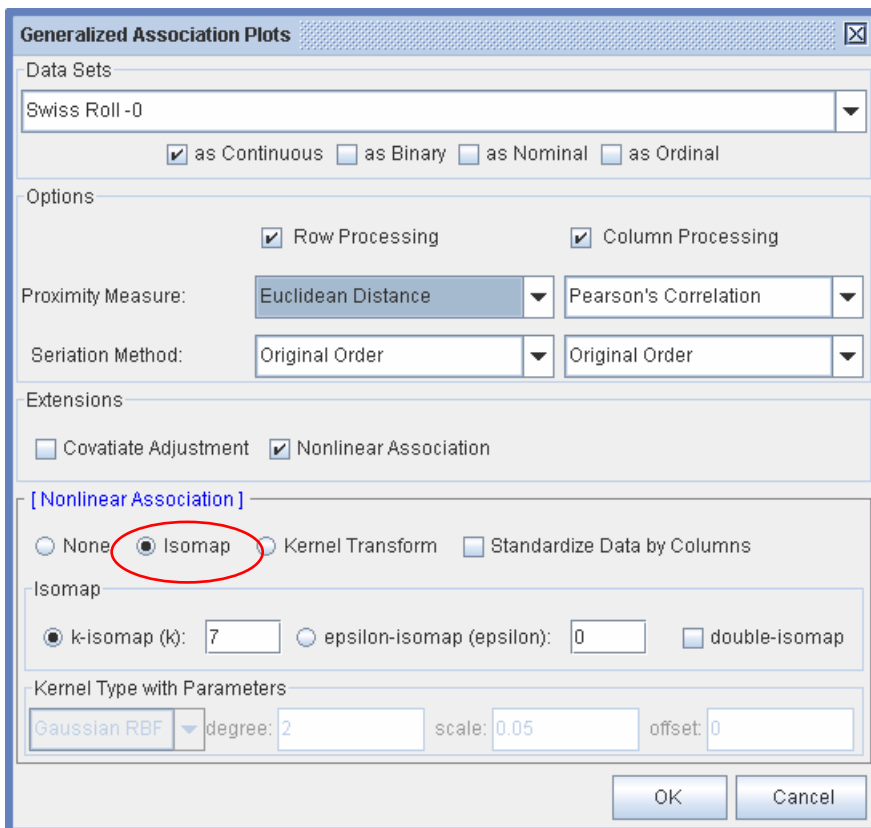
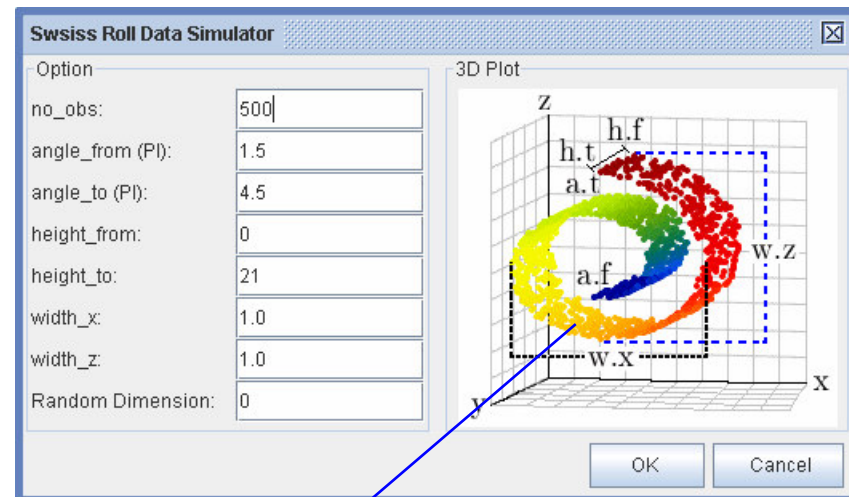
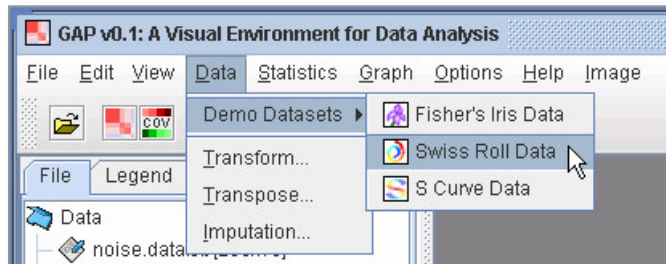


- 2D Scatterplot
- 3D Scatterplot
- XY PairLines Plot
- 2D Boxplot (X)
- 3D Boxplot (X)
- Histogram (X)

GAP with Covariate Adjustment



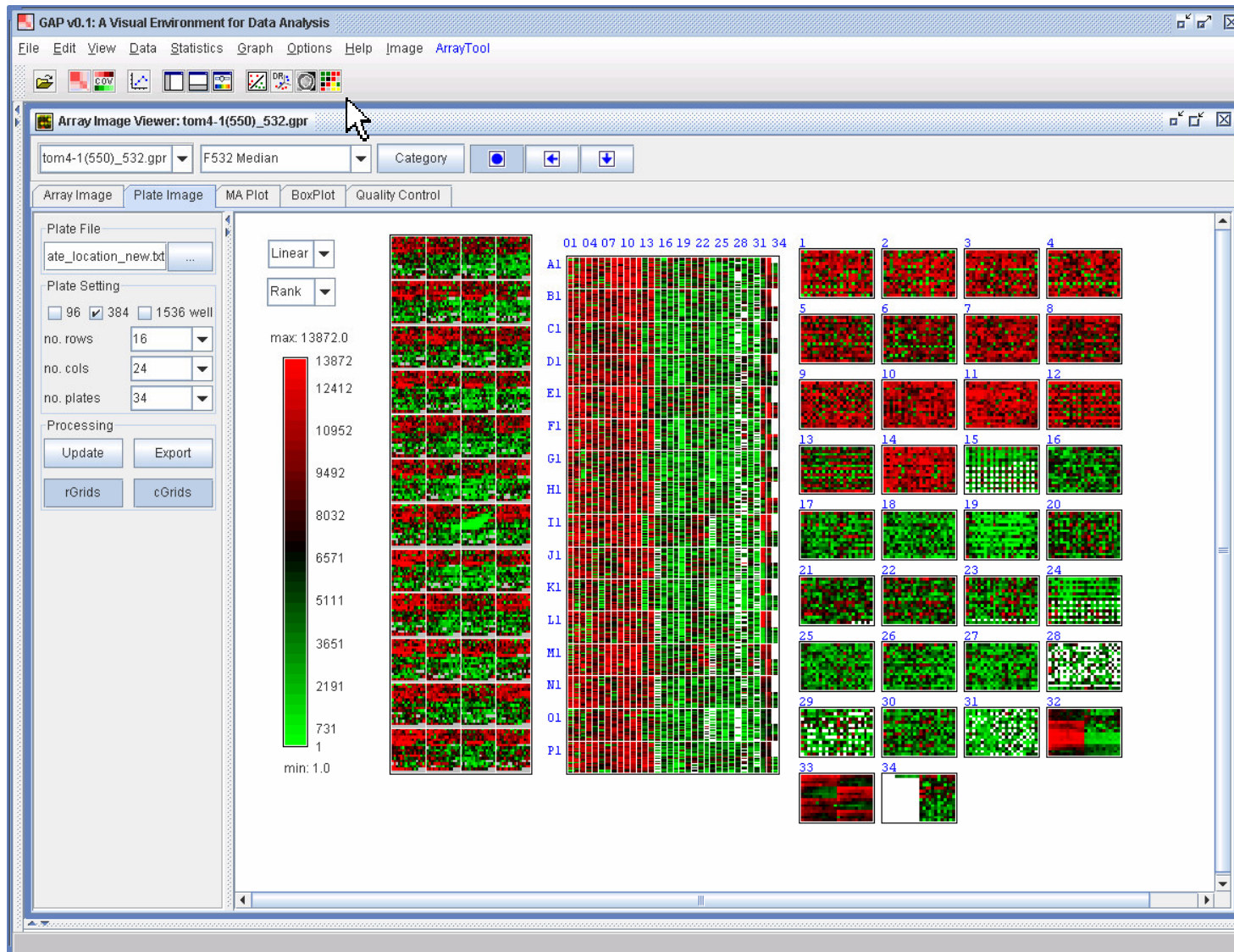
GAP with Nonlinear Association



GAP

isoGAP

GAP Module: Microarray Tool



Web Site

Lab for Information Visualization

The screenshot shows the homepage of the Lab for Information Visualization. It features a navigation menu with links for Home, Research, Members, Database, and Software. A sidebar on the left lists various research areas: Information Visualization (including GAP, SIR, and MDS), Psychiatry Research (Psychiatry), Bioinformatics (Microarray Data Analysis, SNPs), Talks/Seminar (Lecture Notes, Posters), and News/Conference. The main content area includes a profile for Chun-houh Chen, an Associate Research Fellow at the Institute of Statistical Science, Academia Sinica. Below the profile, there is a section for the 'Handbook of Computational Statistics (Volume III): Data Visualization' edited by Chun-houh Chen, Wolfgang Härdle, and Antony Unwin.

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

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

GAP Software

The screenshot displays the GAP Software website. It features a large 'GAP' logo with a Java logo below it. The text describes GAP as a Java-designed software for generalized association plots and exploratory data analysis, programmed for the Java runtime environment 1.5 (JRE version 1.5.0_04). The authors are identified as Han-Wine Wu and Chun-houh Chen at the Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, R. O. C. The official website is listed as <http://gap.stat.sinica.edu.tw/Software/GAP>. The current version is v0.1.014, Build 2006-02-07. The contact information is hmwu@stat.sinica.edu.tw. A list of features includes Clustering Analysis, Various Display Conditions, GAP with a Covariate Adjusted, Nonlinear Association Analysis, Missing Value Imputation, and Histogram, 2D Scatterplot, 3D Scatterplot (Rotatable).

GAP Forum

The screenshot shows the GAP Forum website. It includes a navigation menu with links for Index, User list, Register, and Login. A message states 'You are not logged in.' There is an announcement section describing GAP as a Java-designed software for general purposes matrix visualization (MV) and exploratory data analysis (EDA). Below the announcement is a forum table with columns for Forum, Topics, Posts, and Last post.

Forum	Topics	Posts	Last post
News & updates Announcements, news and updates about GAP. (Moderated by hmwu)	1	1	2006-02-05 17:02:23 by hmwu
General Questions about using GAP (Moderated by hmwu)	0	0	
Suggestions & Feedback Give us your thoughts and comments to improve GAP. (Moderated by hmwu)	0	0	
Bug Report Please report bugs here! (Moderated by hmwu)	0	0	

At the bottom of the forum page, it shows: Newest registered user: hqh1, Registered users online: 0, Guests online: 1, Total number of registered users: 2, Total number of topics: 1, Total number of posts: 1. The footer includes 'Powered by EumBB' and '© Copyright 2002-2005 Rickard Andersson'.

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