

PlotViz3

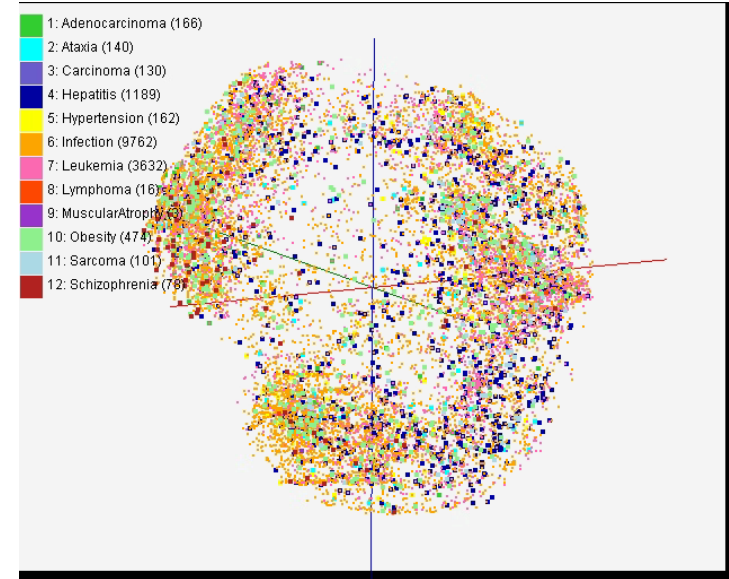
A cross-platform tool for visualizing large and high-dimensional data

Motivation and Introduction to use

Motivation of PlotViz3

► Large and high-dimensional data

- Appears in everywhere; science, business, web, ...
- Complex to understand and hard to interpret
- Dimension reduction is a way to summarize and visualize: PCA, MDS, GTM, SOM, etc

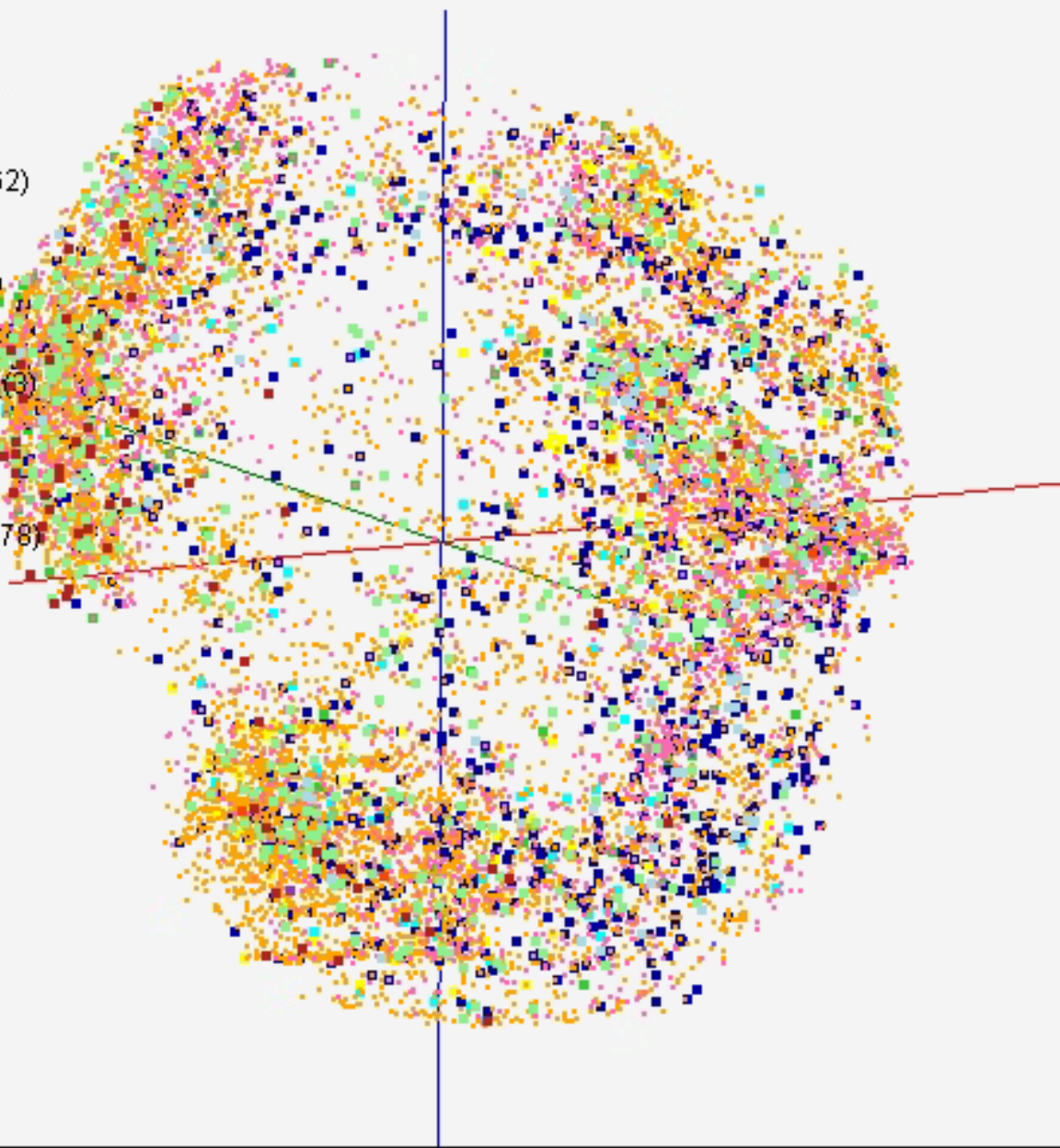


► Interactive 3D scatter plots

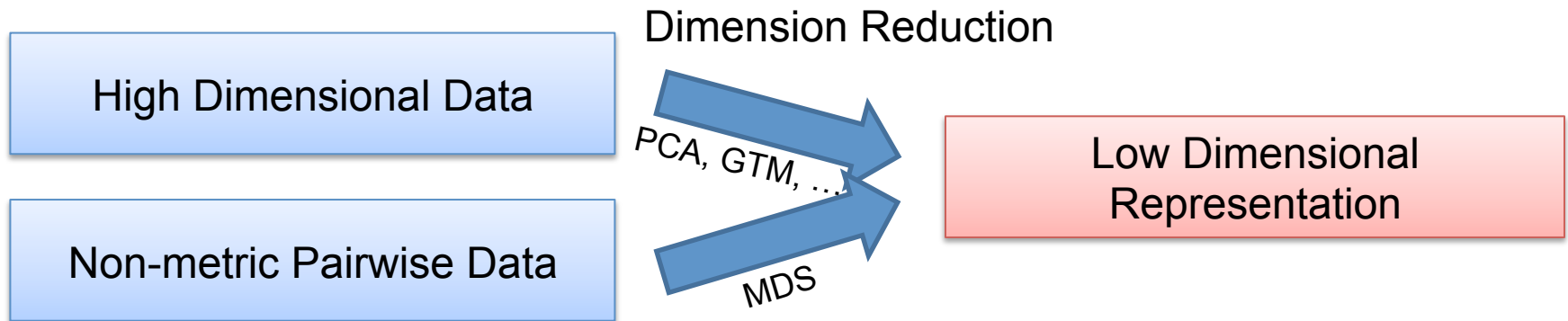
- Help to understand the overall structure
- Help to find cluster structures in data
- Exploit human visual perception
- Users can tag, use different colors, explore spaces, ...

Visualizing 166-dimensional PubChem data

- 1: Adenocarcinoma (166)
- 2: Ataxia (140)
- 3: Carcinoma (130)
- 4: Hepatitis (1189)
- 5: Hypertension (162)
- 6: Infection (9762)
- 7: Leukemia (3632)
- 8: Lymphoma (16)
- 9: MuscularAtrophy (3)
- 10: Obesity (474)
- 11: Sarcoma (101)
- 12: Schizophrenia (78)



Visualization by Dimension Reduction



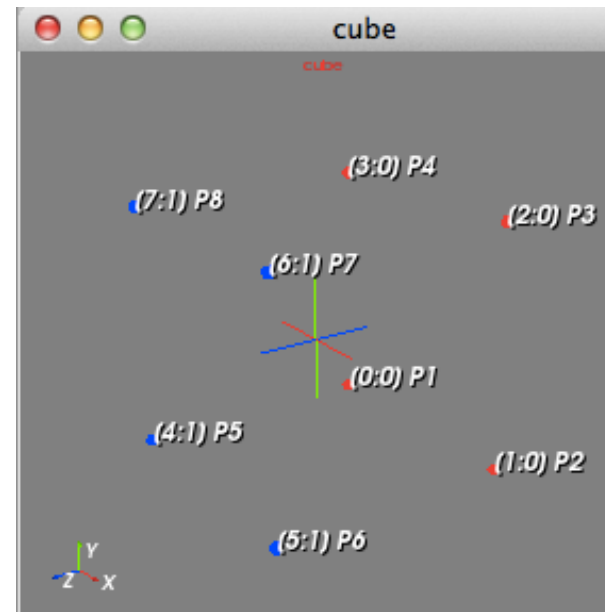
- ▶ Also known as low-dimensional embedding or non-linear mapping
- ▶ Provide a method to simplify data
- ▶ Preserve the raw data information as much as possible in a low dimensional space
- ▶ We focus on finding 2- or 3-dimensional representation in order to visualize high-dimensional data in a virtual 2D/3D space

PlotViz3 For 3D Scatter plots

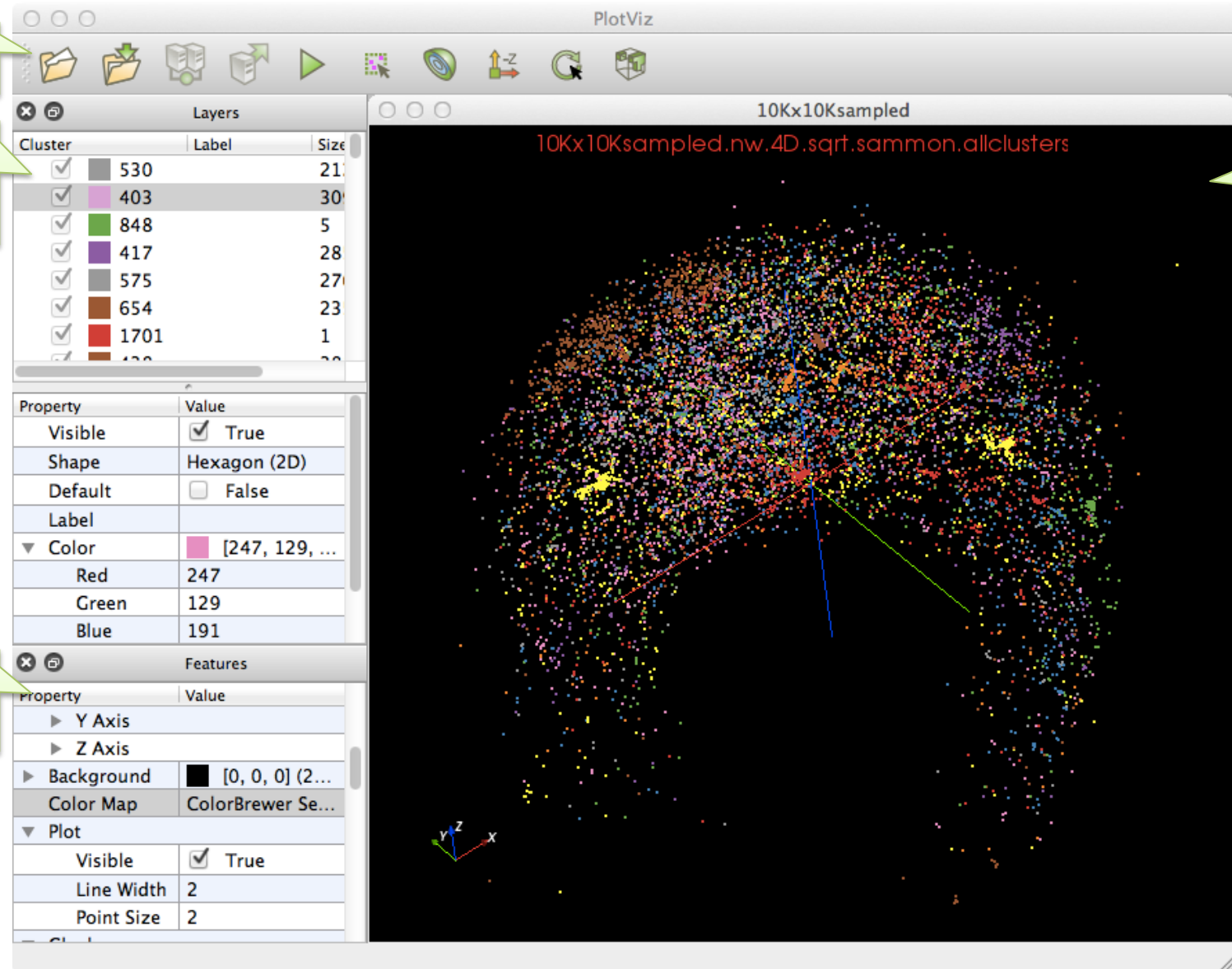
- ▶ PlotViz3 is designed to visualize 3D scatter plots
 - Data is displayed as a collection of points
 - Distance between points implies similarity
- ▶ Simple text input file
 - 5 column plain text data: ID, XYZ position, and class information
 - (Optional) 6th column for point labels
 - Example: Cube.txt (Available at <http://salsahpc.indiana.edu/pviz3/#screenshots>)

#ID	X	Y	Z	Cluster	Label
0	0.0	0.0	0.0	0	P1
1	1.0	0.0	0.0	0	P2
2	1.0	1.0	0.0	0	P3
3	0.0	1.0	0.0	0	P4
4	0.0	0.0	1.0	1	P5
5	1.0	0.0	1.0	1	P6
6	1.0	1.0	1.0	1	P7
7	0.0	1.0	1.0	1	P8

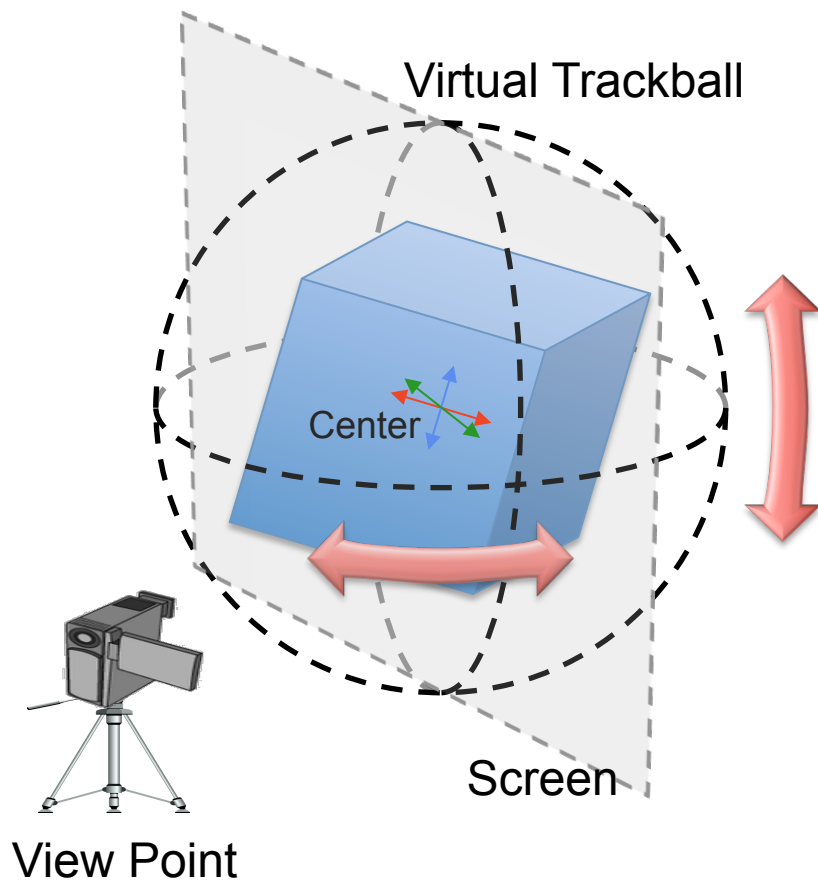
- Each point will be displayed in XYZ position and colored by its class information
- XML format is also available (See PlotViz3 website)




PlotViz3 Screenshot

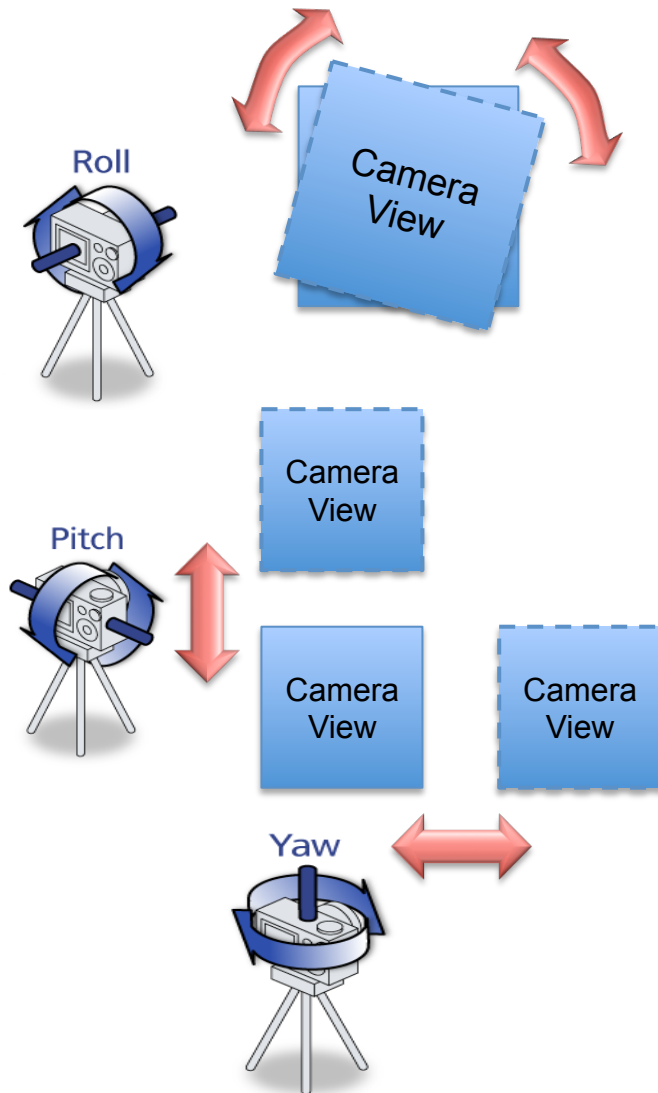


Virtual 3D Operations



- ▶ Virtual trackball rotation
 - Invisible trackball in the center of the scene
 - Mouse click and drag will rotate the virtual trackball
 - Zoom in and out with mouse wheel
- ▶ Change the center (origin) of the scene
 - Using “Pick Center” button  in the tool bar
 - Or, Shift + Mouse drag

2D Operations



- ▶ Switching view point from 3D and 2D is sometimes useful
 - Press “2” for 2D and press “3” for 3D
- ▶ 2D view point
 - Mouse drag for pitching and yawing
 - No trackball rotation.
 - Ctrl+ Mouse drag for rolling

Example of Use I: Cube and Structured Dataset

Basic Examples I

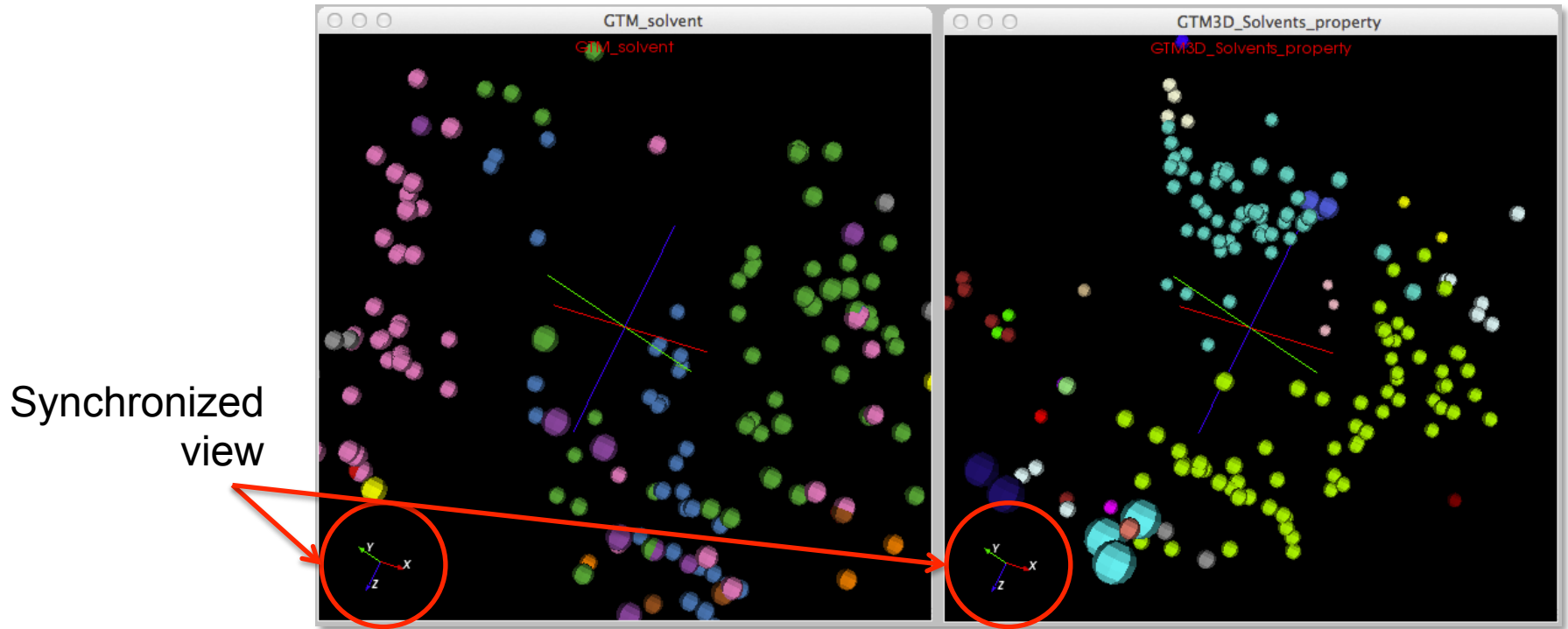
- ▶ cube.txt simple Cube
- ▶ PlotViz3_xml_simple_input.pviz or PlotViz3_simple_input.txt

Example of Use II: Proteomics and Synchronized Rotation

Basic Examples II

- ▶ 10Kx10Ksampled.nw.4D.sqrt.sammon.allclusters.pviz 10K
Proteomics COG Data

Synchronized Camera View



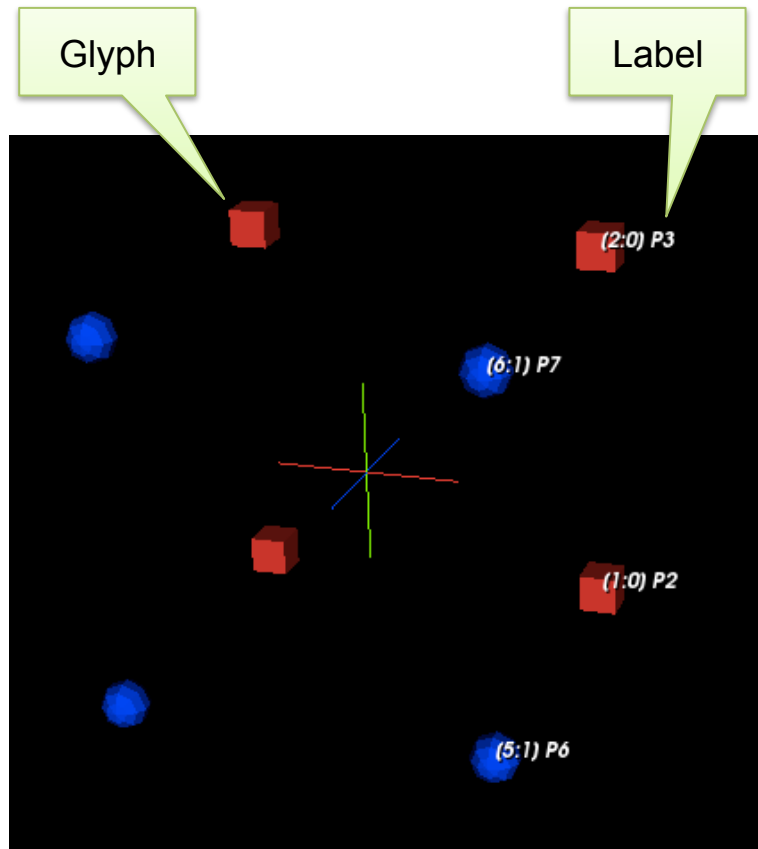
- ▶ Side-by-side analysis for multiple plots
 - Cameras in multiple plots (more than 2) can move in a synchronized way
 - Maintain same rotation angle, camera distance, ...
 - Turn on Features > Camera > Use SyncCamera

Synchronized Examples

- ▶ clusterFinal-M3-C3Dating-ReClustered.pviz
- ▶ DatingRating-OriginalLabels.pviz

Example of Use III: More Features and larger Proteomics Sample

Labels and Glyphs



► Labels

- Display labels associated with points
- Can be selective
- Turn on Features > Label > Visible
- Click Selection Mode in the tool bar and select points to show labels



► Glyphs

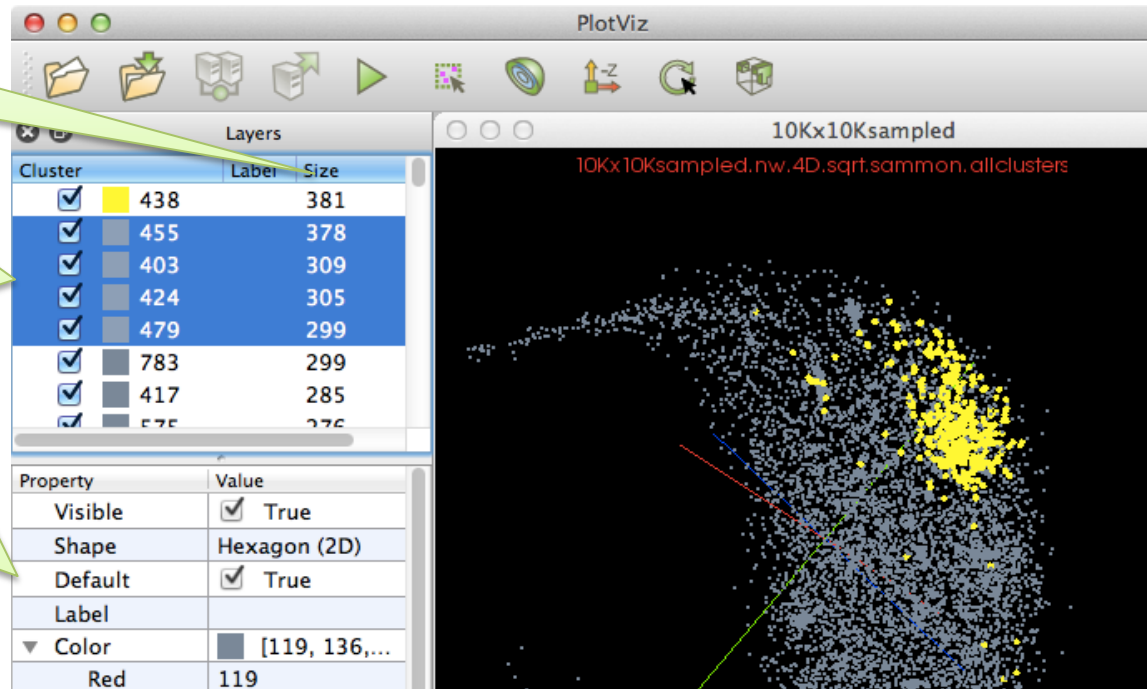
- Points can be visualized by using Glyphs (2D/3D objects)
- Can have different sizes and shapes based on cluster (class) information
- Turn on Features > Glyph > Visible
- Change shape and size in Layers panel in the side tool bar

Visualizing Large-scale Data

Click column header for sorting

Multi-selection by using Shift + Mouse click

“Default” cluster setting will avoid expensive glyph operation and color points



Sammon clusters results. 10,000 points. Available at Pviz3 website

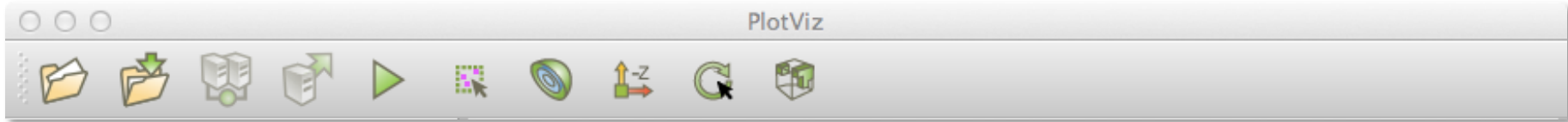
- Use sorting and multi-selection feature for browsing large number of clusters
- Set clusters as “Default” which will avoid some expensive operations (glyph, sorting, ...)
- “Default” clusters will be displayed as only points (not glyphs) and colored by grey.

Proteomics Example

- ▶ 100KProteomics.pviz 100K Proteomics COG data from MDS <http://salsabiology.blogspot.com/p/introduction.html>

Example of Use IV: Tools and Examples

Toolbar



File Open: Support both xml format and simple text format



Save as: Save current map and settings



Play: Rotate the plot 360 degree gradually.



Select points: Select points to show labels



Show only half sphere

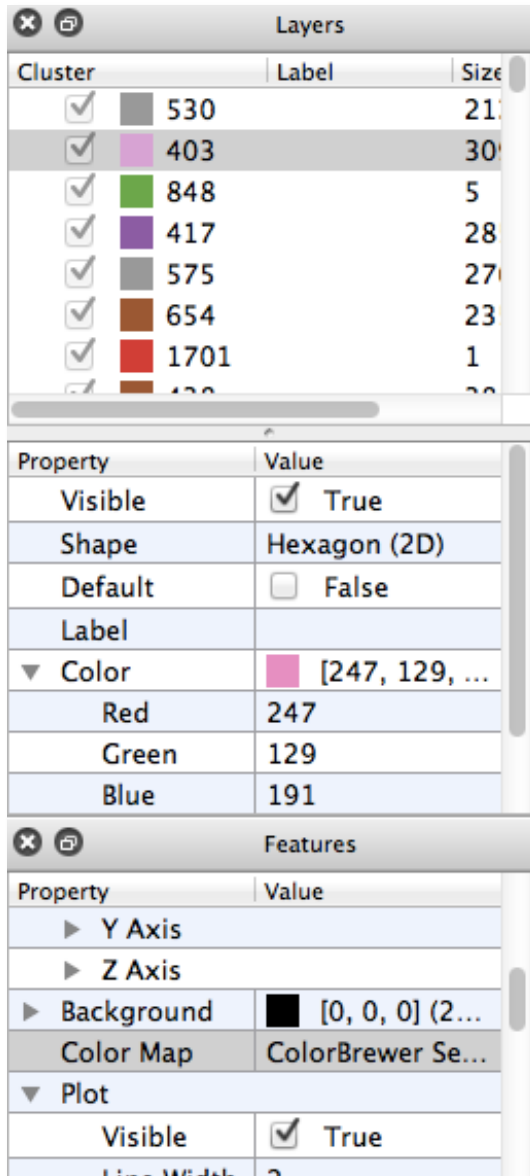


Reset: Reset view point



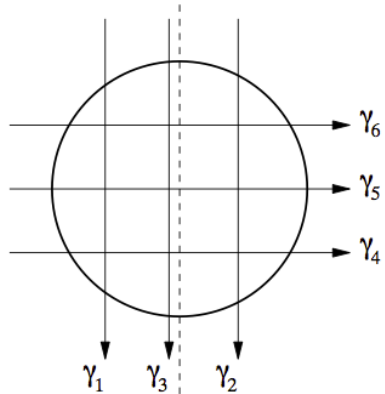
Select custom origin: Change origin of point of view.

Sidebar Tools

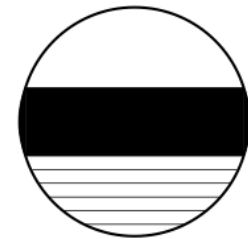
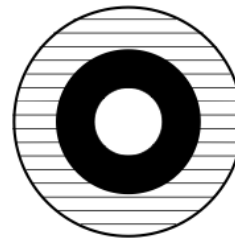
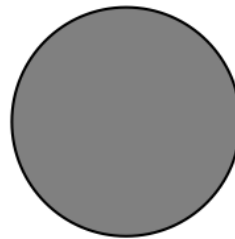


- ▶ Cluster information
 - Visibility
 - Color
 - Shape (point or 2D/3D glyph)
 - Label
 - Glyph size
- ▶ Plot setting
 - Background color
 - Point size
 - Show/hide legend
 - Show/hide labels
 - Show/hide glyph
 - ...

GTM Example: Oil flow data



Cross-section of oil pipe and 6 gamma-beam locations



oil



water



gas



homogeneous mix

3 possible flow configurations; homogeneous, annular, and stratified flow

(Source: Svénson, J., GTM: the generative topographic mapping. PhD thesis, Aston University, 1998)

► Oil flow data

- Measured in 12 variables (dimensions): 2 different wave lengths from 6 gamma beams
- Performed 1,000 observations with different flow settings
- Each flow can be one of 3 flow configurations (i.e., 3 classes): homogeneous, annular, and stratified flows
- Want to find cluster structures in 3-dimensional space

GTM Example: Oil flow data (2)

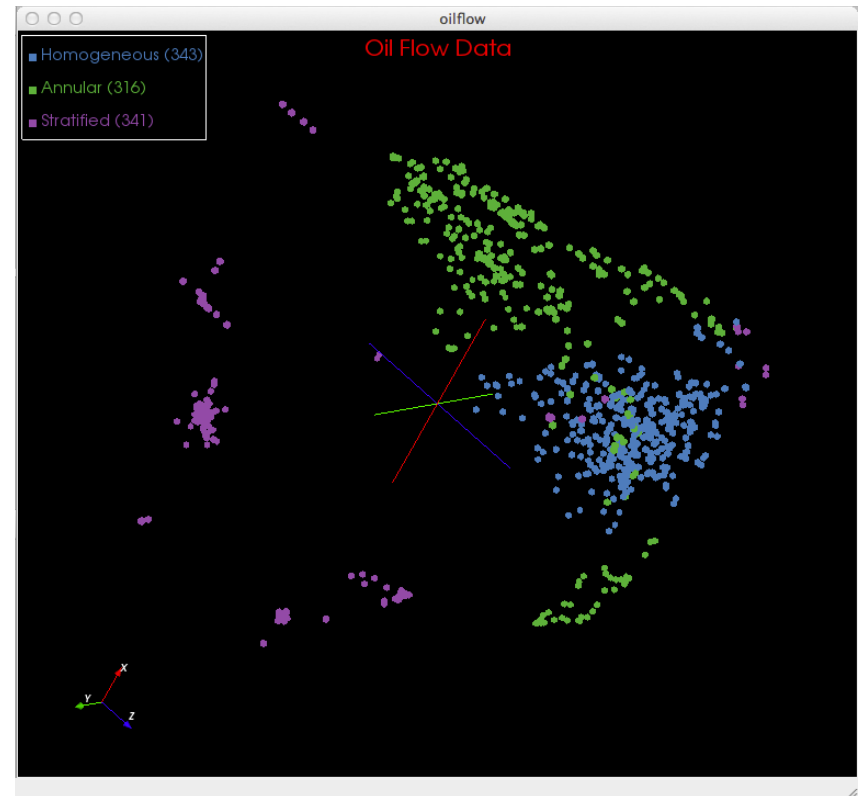
- ▶ Use GTM to find a low dimensional mapping
 - Find an optimal 3-dimensional representation of the original 12-dimensional data.
 - Visualize the data in 3D to see cluster structures
- ▶ Parallel GTM implementation
 - Available at <http://salsahpc.indiana.edu/dagtm/>
 - An extension to use better optimization method, Deterministic Annealing (DA), is also available (but, this is out of scope in this talk)
 - Oil flow data is included in ./data directory

GTM Example: Oil flow data (3)

► Example:

```
# run Parallel GTM to find 3-dimensional mapping
$ mpiexec -n 4 dagtm -i Oil.h5 -o Oil.out.h5
# convert output to PlotViz input file
$ ../bin/hdf2pviz Oil.out.h5
Saved ... Oil.out.txt
# Open Oil.out.txt
  in PlotViz3
```

- This example file is available at
<http://salsahpc.indiana.edu/pviz3/#screenshots>



MDS Example: Trading data

- ▶ Trading data in 1986 (Cold war era)
 - Data from the New Geographical Digest (1986), analyzed in Cox and Cox (2001)
 - 20 countries traded with other countries in 1986
 - Pairwise matrix representing dissimilarities (0.0 = close, 1.0 = distant)

	Arge	Aust	Braz	Cana	Chin	Czec	Egyp	E.Ge	Fran	Hung	Indi	Ital	Japa	N.Ze	Pola	Swed	USA	USSR	U.K.
Aust	0.6																		
Braz	0.5	0.5																	
Cana	0.7	0.3	0.6																
Chin	0.8	0.8	0.8	0.8															
Czec	1.0	1.0	1.0	1.0	0.8														
Egyp	0.6	0.6	0.7	0.7	0.8	0.8													
E.Ge	0.8	0.9	0.8	1.0	0.8	0.5	0.7												
Fran	0.5	0.6	0.7	0.7	1.0	1.0	0.4	0.8											
Hung	0.8	0.8	0.8	1.0	0.8	0.6	0.7	0.5	0.8										
Indi	0.5	0.4	0.5	0.5	0.7	0.8	0.4	0.7	0.5	0.7									
Ital	0.8	0.8	0.6	1.0	1.0	1.0	0.7	0.8	0.7	0.8	0.8								
Japa	0.8	0.7	0.8	0.8	0.8	1.0	0.9	1.0	0.9	1.0	0.9	0.9							
N.Ze	0.7	0.5	0.6	0.4	0.6	1.0	0.7	1.0	0.7	1.0	0.5	1.0	0.8						
Pola	0.8	0.8	0.8	1.0	0.8	0.6	0.7	0.5	0.8	0.0	0.7	0.8	1.0	1.0					
Swed	0.6	0.5	0.6	0.6	1.0	1.0	0.5	0.8	0.4	0.8	0.4	0.8	0.9	0.6	0.8				
USA	0.7	0.7	0.7	0.8	0.7	1.0	0.8	0.9	0.7	0.9	0.7	0.6	0.5	0.7	0.9	0.8			
USSR	1.0	0.9	1.0	0.9	1.0	0.6	1.0	0.6	0.9	0.7	1.0	0.9	0.8	1.0	0.7	1.0	0.8		
U.K.	0.8	0.7	0.8	0.9	0.8	1.0	0.7	0.9	0.7	0.9	0.8	0.7	0.5	0.8	0.9	0.8	0.3	0.8	
W.Ge	0.8	0.8	0.8	0.8	0.9	0.7	0.7	0.8	0.7	0.9	0.8	0.7	0.6	0.7	0.9	0.8	0.5	0.6	0.6

MDS Example: Trading data (2)

► R smacof package

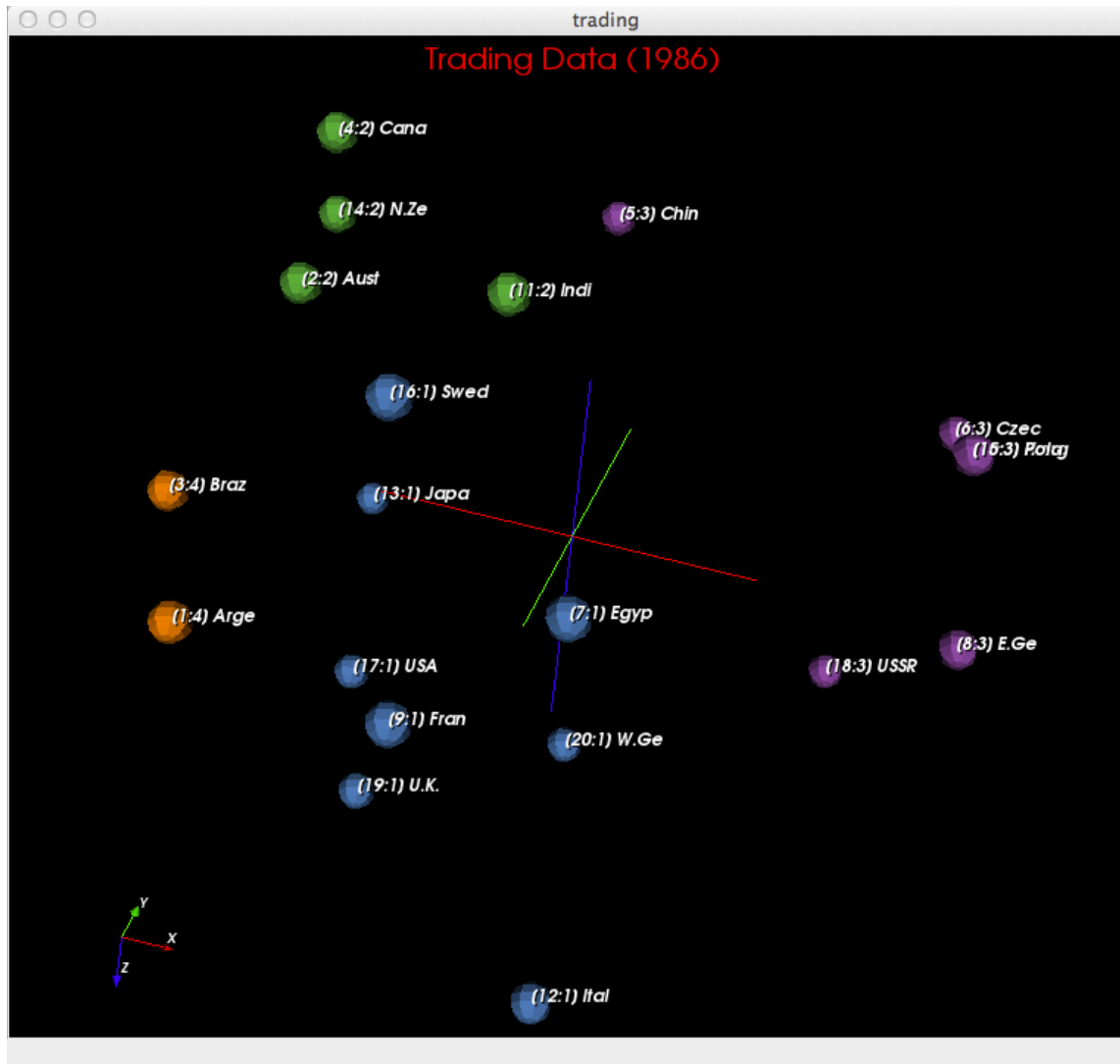
- Trading data: 20 countries' pairwise trading closeness
- For more details, see

<http://cran.r-project.org/web/packages/smacof/vignettes/smacof.pdf>

► Example

```
# In R, run MDS
> library(smacof)
> data("trading") # load data
> M <- smacofSym(trading,
  ndim=3) # run MDS
# Save output for PlotViz3
> write.table(
  cbind(1:20, M$conf, 1, row.names(M$conf)),
  'trading.txt', row.names=F, col.names=F, quote=F)
# Open trading.txt in PlotViz3
```

MDS Example: Trading data (3)



- ▶ Clusters
 - Commonwealth country (green): New Zealand, Australia, Canada, India
 - South America (orange): Brazil, Argentina
 - Communist (purple): USSR, Czechoslovakia, Poland, E. Germany, China
 - Others (blue): USA, Japan, some western Europe

Examples

- ▶ Oilflowdata-p41-42.pdf discussing Oil flow data pages 41-42 from http://eprints.aston.ac.uk/1245/1/NCRG_98_024.pdf
- ▶ OilFlowData.out.pviz or OilFlowData.out.txt Oil Flow data from <http://salsahpc.indiana.edu/pviz3/#screenshots>

Example of Use V: Final Examples

More Examples II

- ▶ TradingData-p24-25.pdf discussing Trading data pages 24-25 from <http://www.jstatsoft.org/v31/i03/paper>
- ▶ TradingData.pviz or TradingData.txt Trading data from <http://salsahpc.indiana.edu/pviz3/#screenshots>
- ▶ GTM3D_Solvents_property.pviz from <http://salsahpc.indiana.edu/pviz3/#screenshots>
- ▶
- ▶ 446041_sequences_100K_fixed_collage.pviz 446K Fungi points from MDS <http://salsabiology.blogspot.com/p/introduction.html>
- ▶
- ▶ 100KMetagenomics.pviz 100K Metagenomics data from MDS <http://salsabiology.blogspot.com/p/introduction.html>

Download and Other Examples

- Available at <http://salsahpc.indiana.edu/pviz3>

The screenshot shows the PlotViz3 website. The header features the 'PlotViz3' logo and the tagline 'A cross-platform tool for visualizing large and high-dimensional data'. The SALSAPHPC logo is in the top right. A green navigation bar contains links for Home, Screenshots, Downloads, Papers, and Contacts. A left sidebar lists various resources, with two red callout boxes highlighting 'Download Program (Mac and Windows)' and 'Examples and samples'. The main content area includes an 'Introduction' section and a 3D scatter plot of data points.

PlotViz3
A cross-platform tool for visualizing large and high-dimensional data

SALSAPHPC

Home Screenshots Downloads Papers Contacts

Download Program (Mac and Windows)

Examples and samples

Introduction

Large-scale high dimensional data visualization is highly valuable for scientific discovery in many fields of data mining and information retrieval. PlotViz is a 3D data point browser that visualizes large volume of 2- or 3-dimensional data as points in a virtual space on a computer screen and enable users to explore the virtual space interactively. PlotViz was initially designed to consume outputs of

Legend:

- 1: Hepatitis C (HCV)
- 2: Hepatitis B (HBV)
- 3: Hepatitis A (HAV)
- 4: Hepatitis D (HDV)
- 5: Hepatitis E (HEV)
- 6: Hepatitis F (HFV)
- 7: Hepatitis G (HGV)
- 8: Hepatitis I (HIV)
- 9: Hepatitis J (HJV)
- 10: Hepatitis K (HKV)
- 11: Hepatitis L (HLV)
- 12: Hepatitis M (HMV)

<http://salsahpc.indiana.edu/pviz3>

PlotViz3

A cross-platform tool for visualizing large and high-dimensional data

[Home](#)

Screenshots

Downloads

Papers

Contacts

Menu

Introduction

Key Features and Functions

Screenshots

Download and Installation

Instructions for developers

Samples

Inputfile Format

Program Functions

Papers

Supports and Contacts

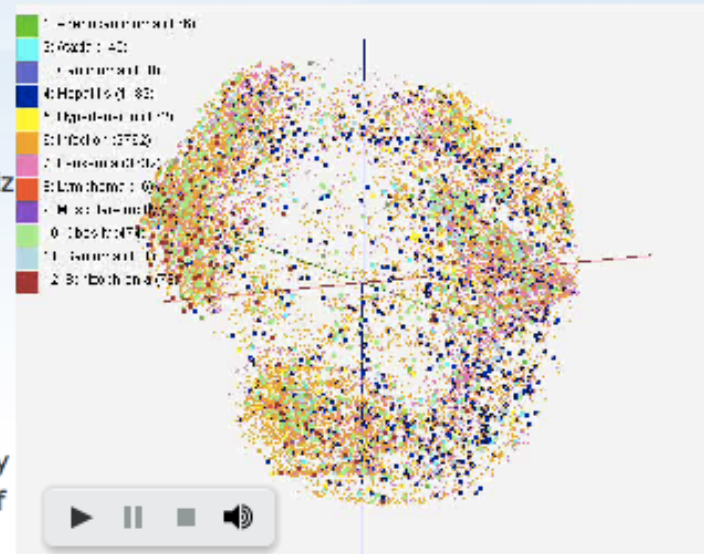
Links

PlotViz

SALSA Project

Introduction

Large-scale high dimensional data visualization is highly valuable for scientific discovery in many fields of data mining and information retrieval. PlotViz is a 3D data point browser that visualizes large volume of 2- or 3-dimensional data as points in a virtual space on a computer screen and enable users to explore the virtual space interactively. PlotViz was initially designed to consume outputs of dimension reduction algorithms



for visualizing high-dimensional data in a lower-dimensional space, such as Multi-dimensional Scaling (MDS) and Generative Topographic Mapping (GTM). Used together with such dimension reduction algorithms, PlotViz can help users to discover intrinsic structures of high-dimensional data and browse large volumes of data points interactively and efficiently in a virtual 3D space.

PlotViz3 Software Dependency

▶ Visualization Toolkit (VTK)

- Open-source visualization library
- Used in many advanced visualization applications: ParaView, VisIt, VisTrails, 3DSlicer, etc.



▶ Qt (GUI Framework)

- Cross-platform application framework
- Supports OSX, Windows, Linux, etc.
- Used for application GUIs



▶ Boost

- C++ library
- Used for multi-threading, memory management, event handling, ...

