The S/Java Interface: Programming with Data on the Web

by

Mark Hansen

July 29, 1997



Joint work with

John Chambers (Bell Labs)

Jonathan Mattingly (Princeton University)

The Goal

To allow communities of our users to share statistical {ideas,data,techniques} over the Web.

We are already familiar with various repositories of code, data and papers...

- StatLib

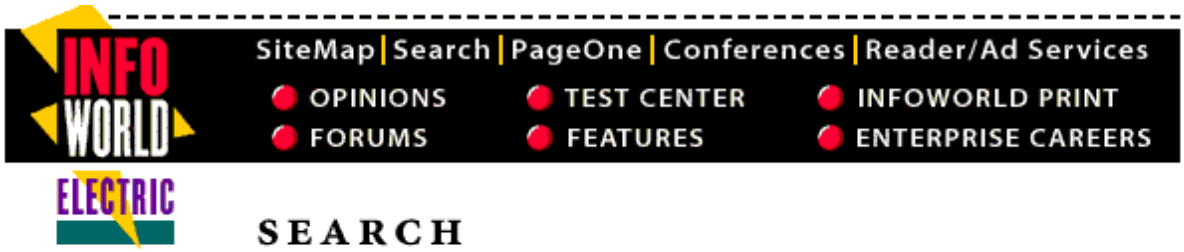
  http://www.stat.cmu.edu

- 

  http://www.netlib.org

- MCMC Preprint Service

  http://www.stats.bris.ac.uk/MCMC/



Tech Update Statistical data plots thicken with 3–D VRML

By Chris Jones

SAS Institute Inc., SPSS Inc., and Lucent Technologies Bell Laboratories are developing graphics technology that will enable users to create interactive models that have an improved capability for revealing patterns and trends in data.

Languages like Java hold tremendous potential for data visualization applications. The platform neutral nature of the language makes it somewhat appealing for distributing methods as well.

"...the ingenious ideas that make progress possible vanish into the inner workings of our machines, where only experts may be aware of their existence. Numerical algorithms, being exceptionally uninteresting and incomprehensible to the public, vanish exceptionally fast."

L. N. Trefethen

Increasingly, we are encountering specialized clients and servers for distributing methods

- GASP

  (Webster West, Todd Ogden, Tony Rossini, University of South Carolina)

  http://www.stat.sc.edu/rsrch/gasp/

- Graphics Production Library

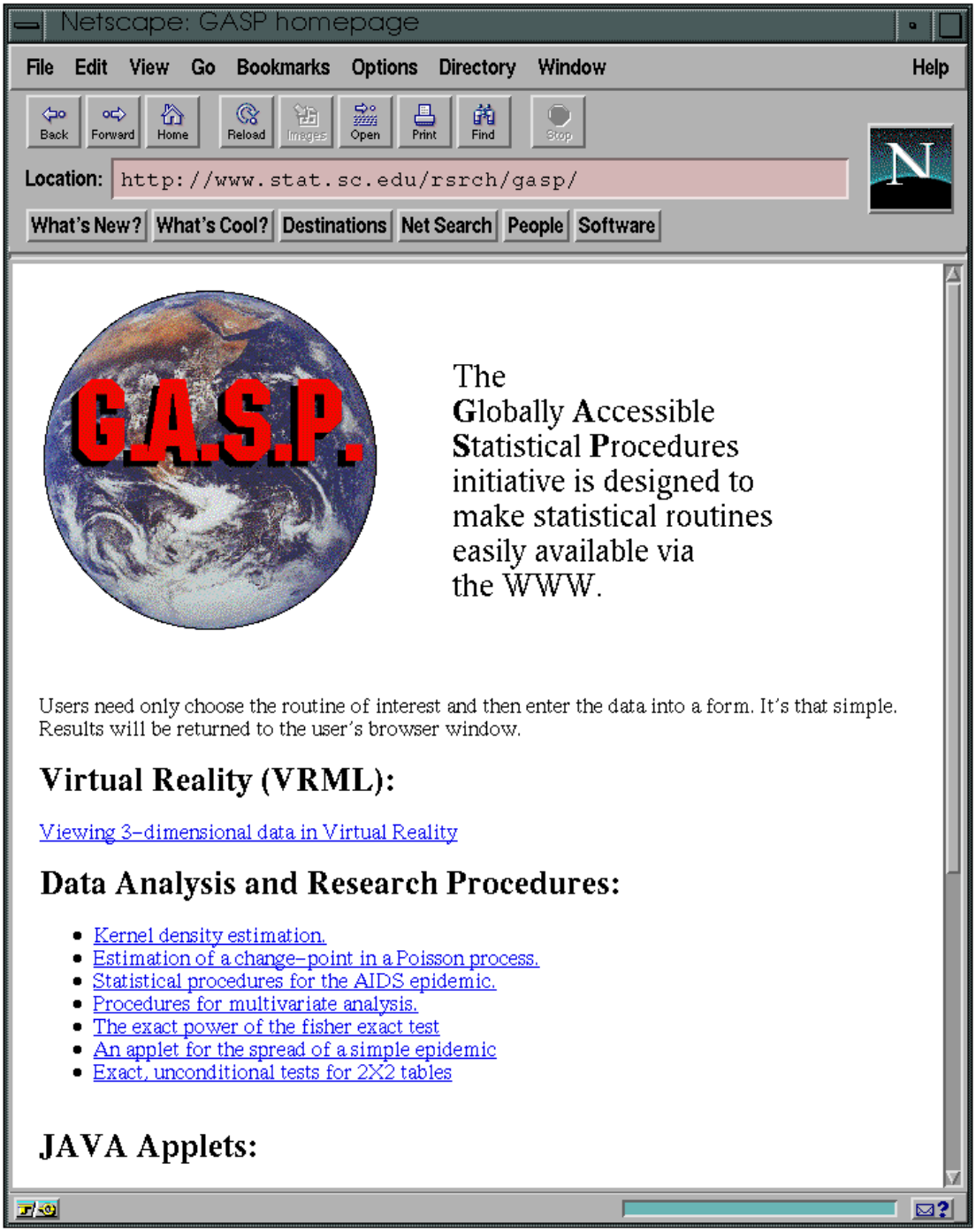
  (Dan Rope, SPSS)

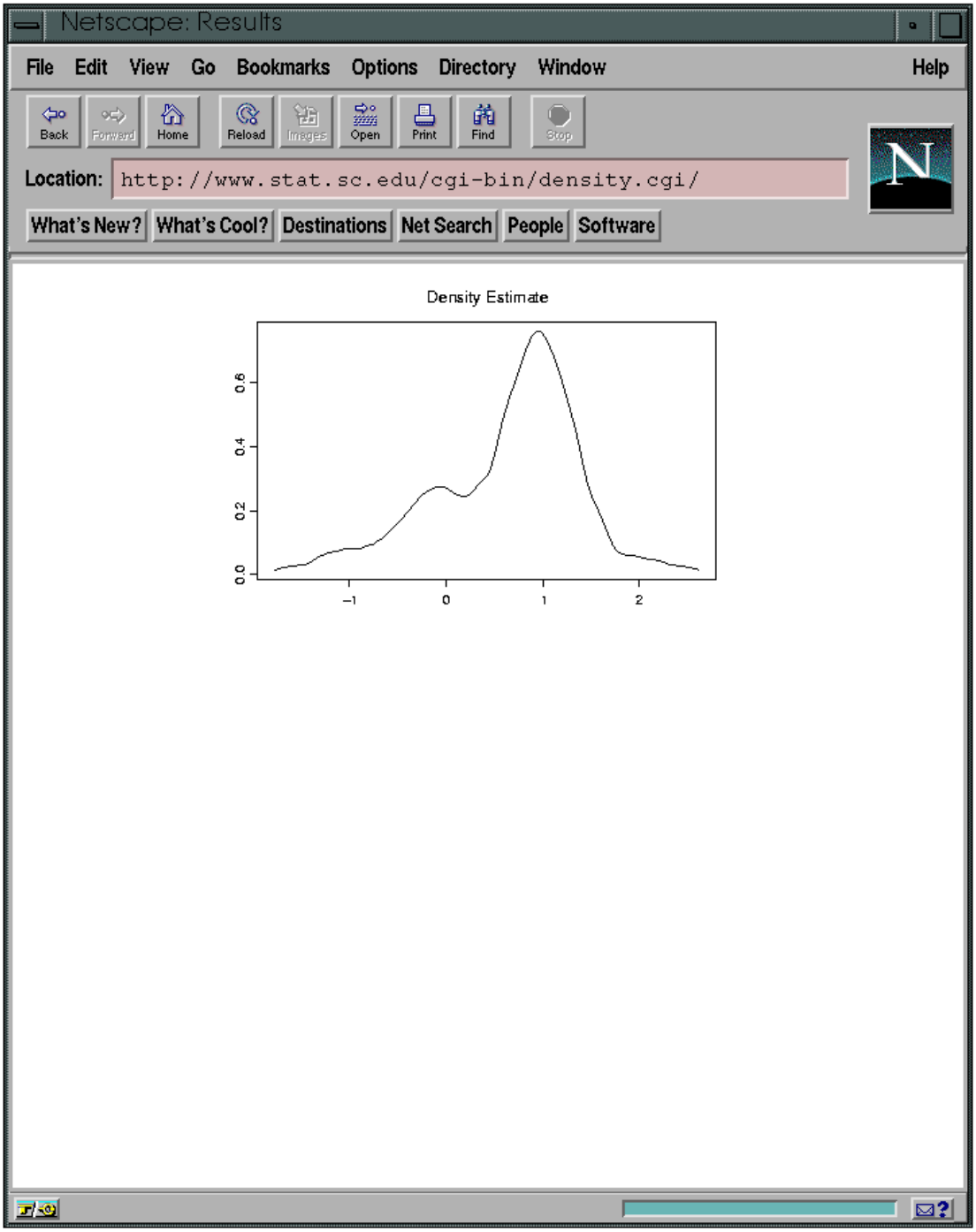
  http://www.mnsinc.com/dan_rope/gpl/

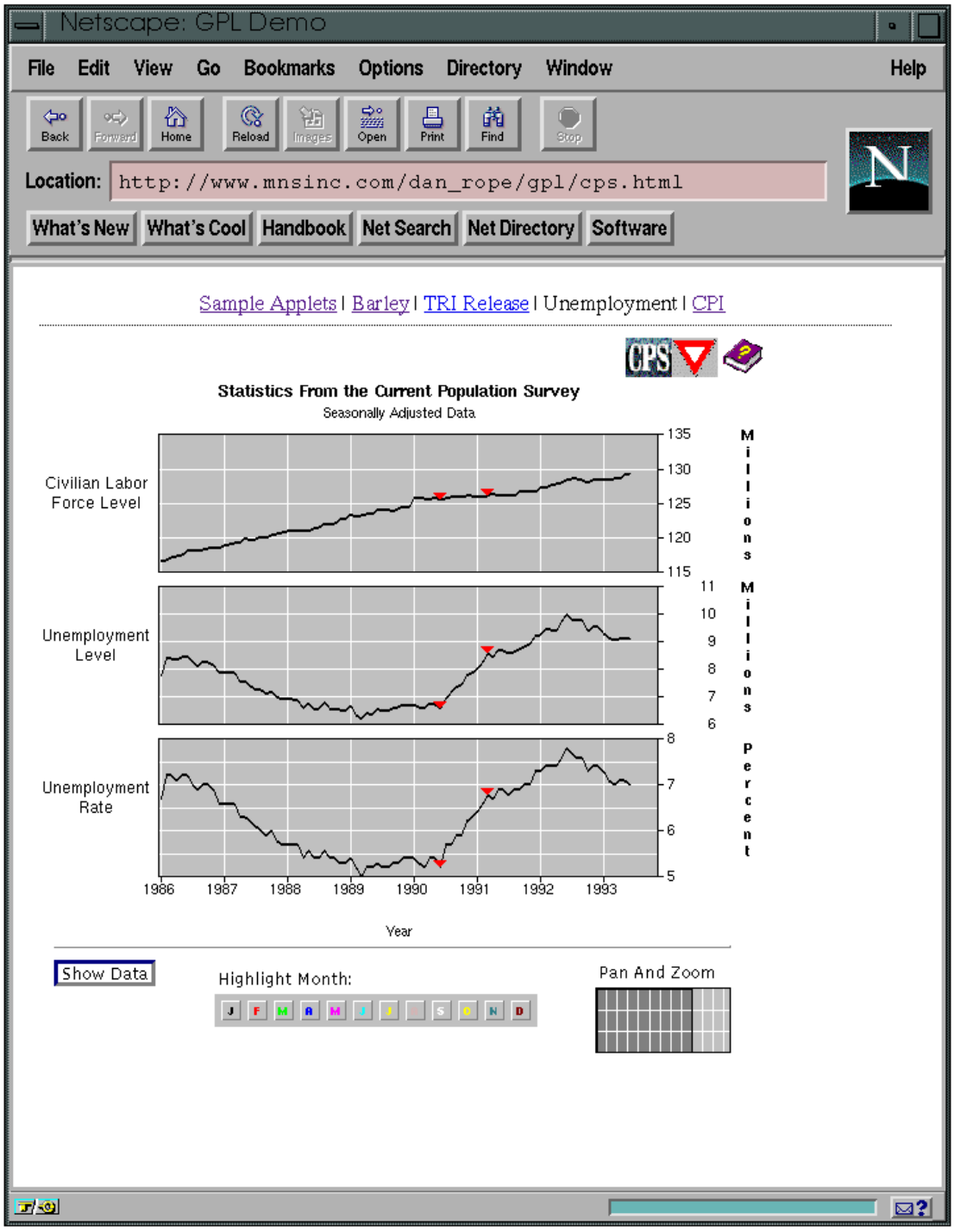
- Support Vector Demo

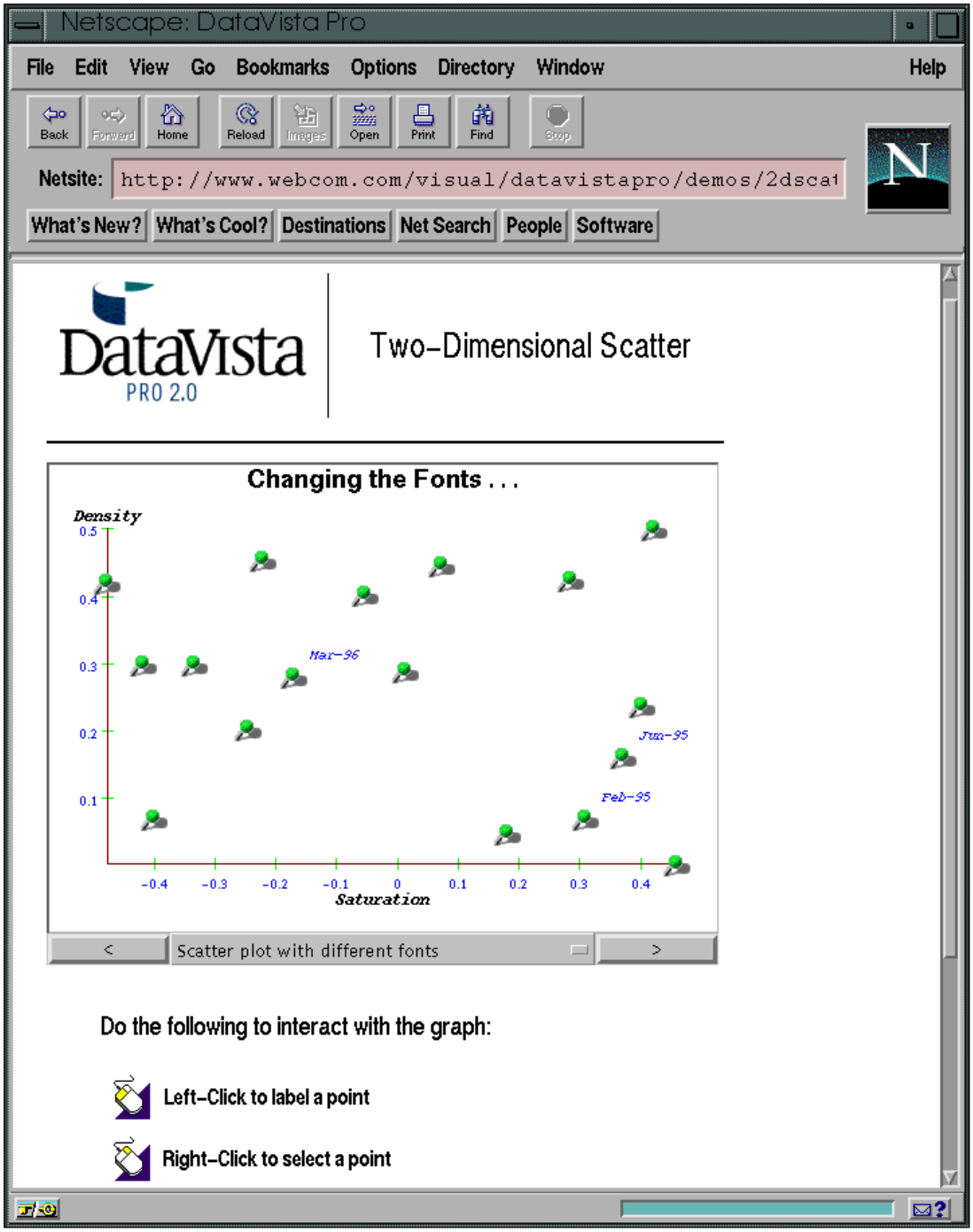
  (Vladimir Vapnik and Chris Burges AT&T Research and Bell Labs)

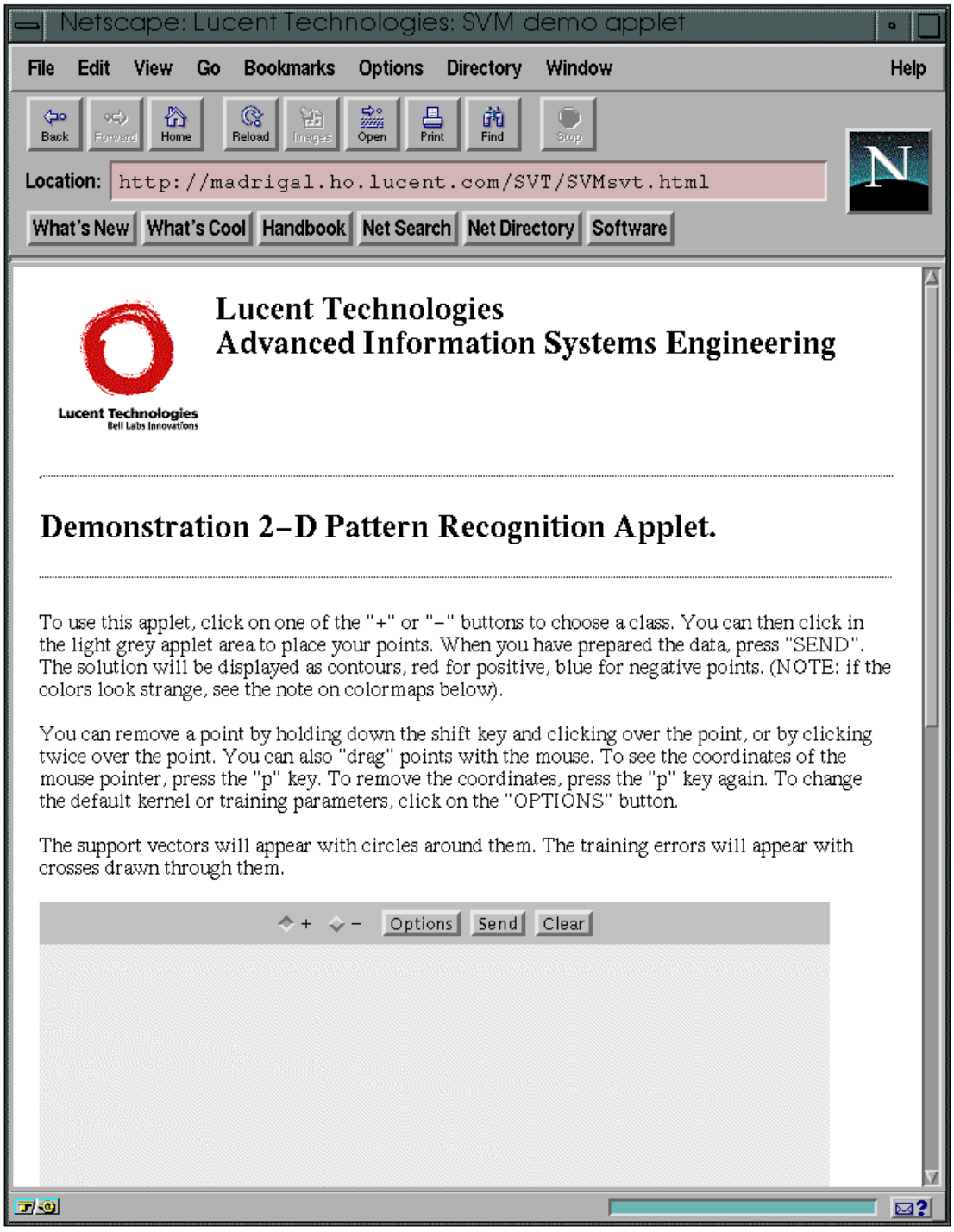
  http://madrigal.ho.lucent.com/SVT/SVMsvt.html

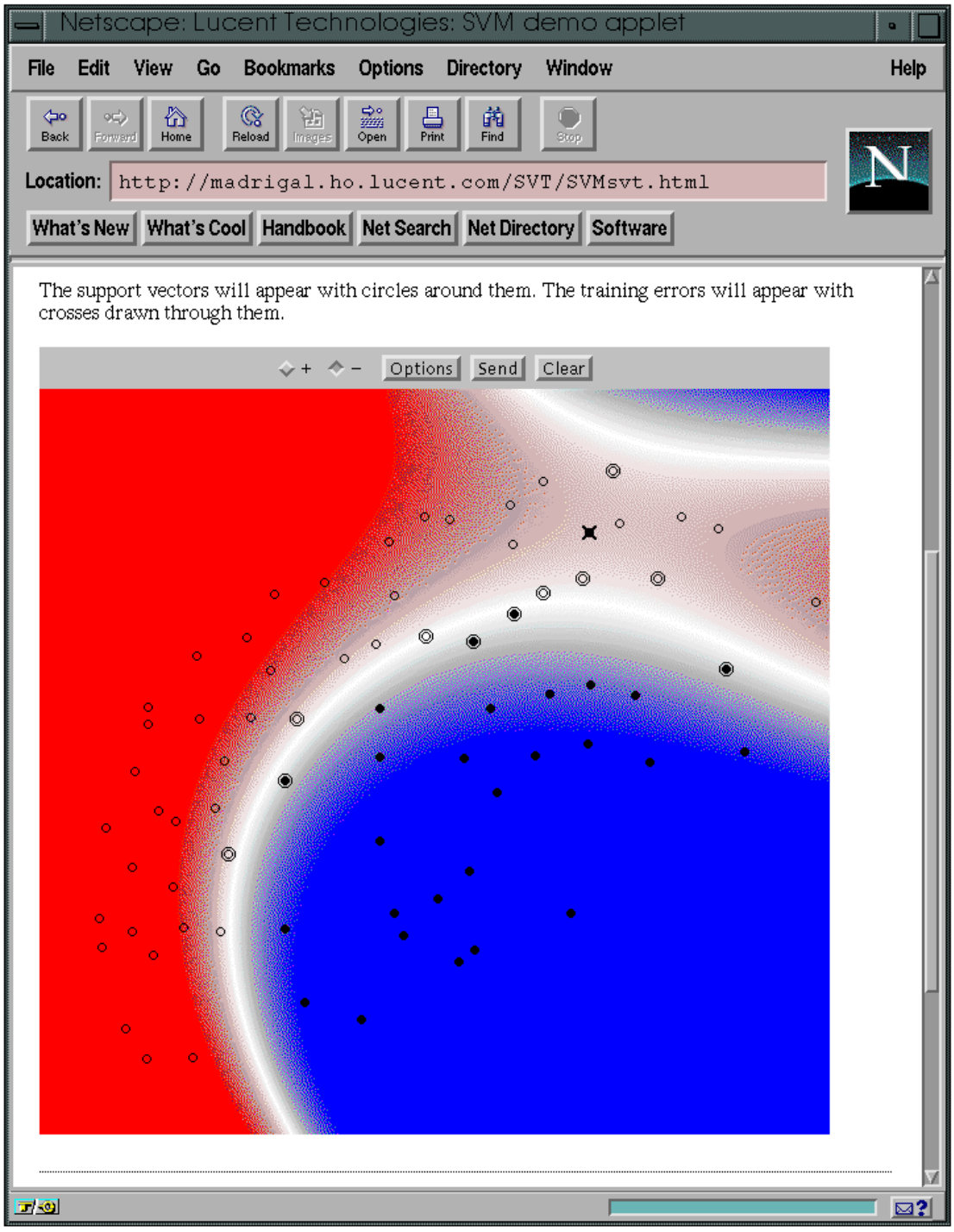

Netscape: GASP homepage
File Edit View Go Bookmarks Options Directory Window Help
Back Forward Home Reload Images Open Print Find Stop
Location: http://www.stat.sc.edu/rsrch/gasp/
What's New? What's Cool? Destinations Net Search People Software
G.A.S.P.
The
Globally Accessible
Statistical Procedures
initiative is designed to
make statistical routines
easily available via
the WWW.
Users need only choose the routine of interest and then enter the data into a form. It's that simple.
Results will be returned to the user's browser window.
Virtual Reality (VRML):
Viewing 3-dimensional data in Virtual Reality
Data Analysis and Research Procedures:
Kernel density estimation.
Estimation of a change-point in a Poisson process.
Statistical procedures for the AIDS epidemic.
Procedures for multivariate analysis.
The exact power of the fisher exact test
An applet for the spread of a simple epidemic
Exact, unconditional tests for 2X2 tables
JAVA Applets:


Netscape: Results
File Edit View Go Bookmarks Options Directory Window Help
Back Forward Home Reload Images Open Print Find Stop
Location: http://www.stat.sc.edu/cgi-bin/density.cgi/
What's New? What's Cool? Destinations Net Search People Software
Density Estimate
0.0 0.2 0.4 0.6
−1 0 1 2


Netscape: GPL Demo
File Edit View Go Bookmarks Options Directory Window Help
Back Forward Home Reload Images Open Print Find Stop
Location: http://www.mnsinc.com/dan_rope/gpl/cps.html
What's New What's Cool Handbook Net Search Net Directory Software
Sample Applets | Barley | TRI Release | Unemployment | CPI
Statistics From the Current Population Survey
Seasonally Adjusted Data
Civilian Labor Force Level
Unemployment Level
Unemployment Rate
135
130
125
120
115
11
10
9
8
7
6
8
7
6
5
Millions
Millions
Percent
1986 1987 1988 1989 1990 1991 1992 1993
Year
Show Data
Highlight Month:
J F M A M J J A S O N D
Pan And Zoom

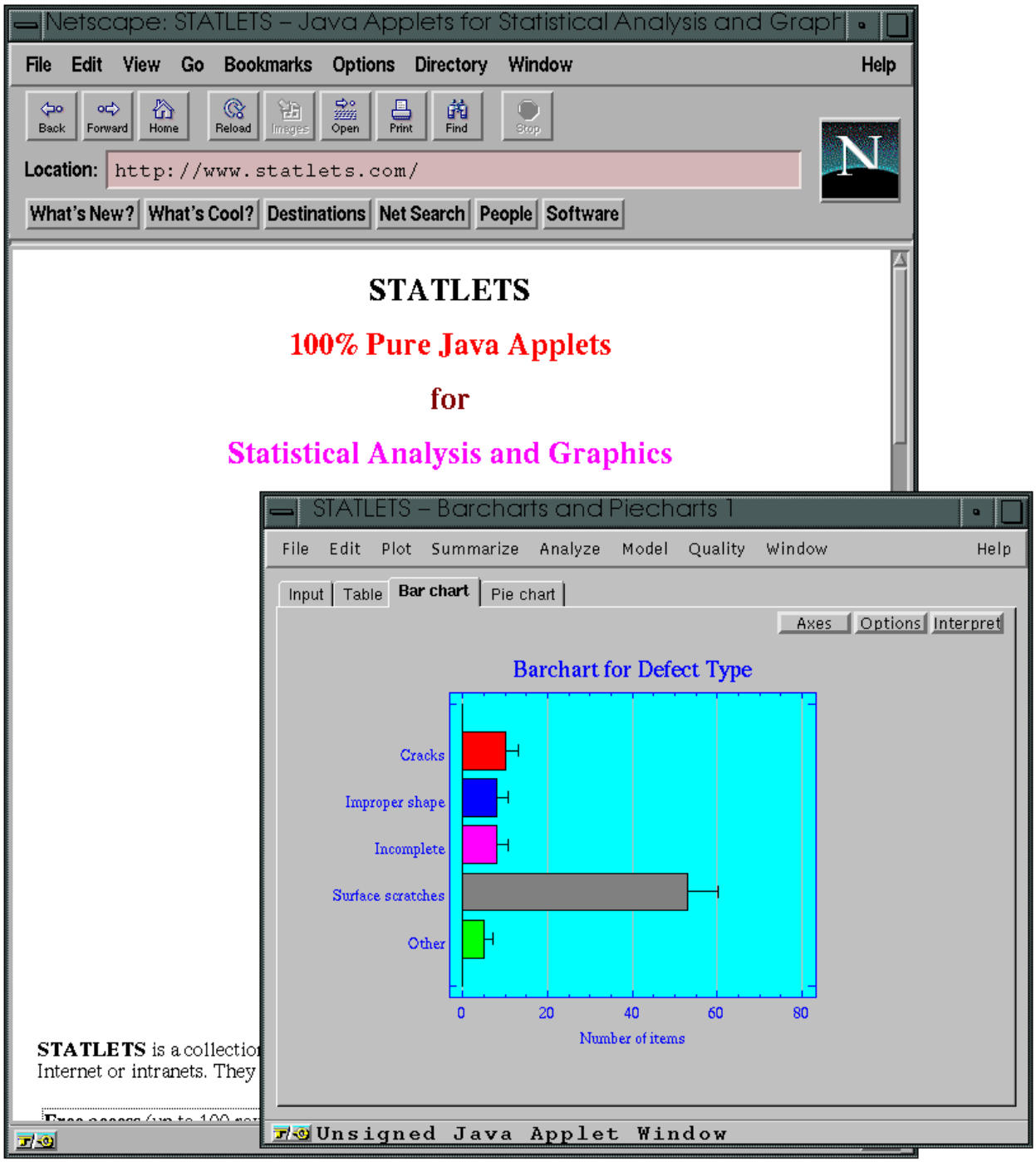

Netscape: DataVista Pro
File Edit View Go Bookmarks Options Directory Window Help
Back Forward Home Reload Images Open Print Find Stop
Netsite: http://www.webcom.com/visual/datavistapro/demos/2dsca1
What's New? What's Cool? Destinations Net Search People Software
DataVista
PRO 2.0
Two-Dimensional Scatter
Changing the Fonts . . .
Density
0.5
0.4
0.3
0.2
0.1
Mar-96
Jun-95
Feb-95
-0.4 -0.3 -0.2 -0.1 0 0.1 0.2 0.3 0.4
Saturation
<
Scatter plot with different fonts
>
Do the following to interact with the graph:
Left-Click to label a point
Right-Click to select a point

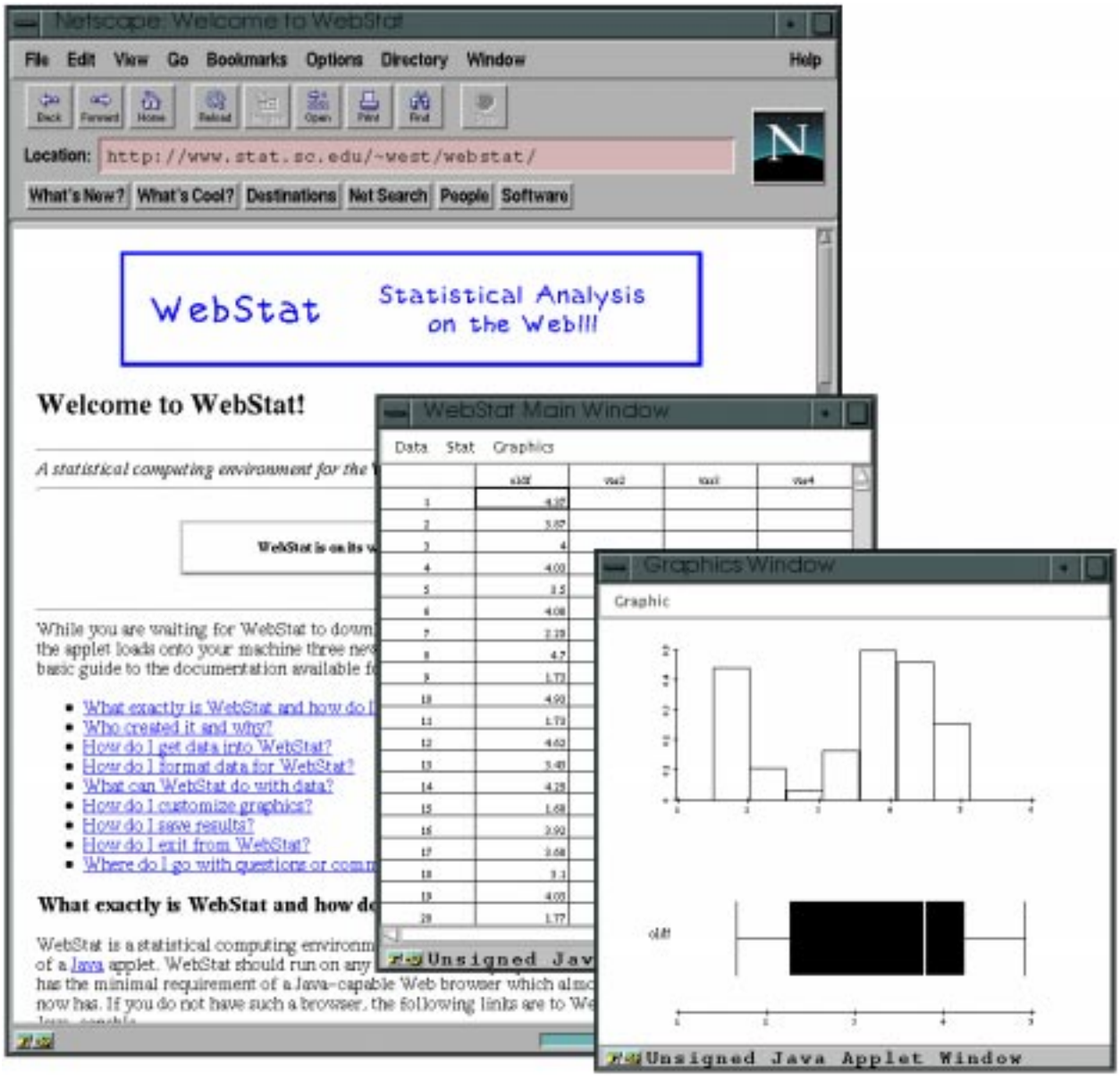

Netscape: Lucent Technologies: SVM demo applet
File Edit View Go Bookmarks Options Directory Window Help
Back Forward Home Reload Images Open Print Find Stop
Location: http://madrigal.ho.lucent.com/SVT/SVMsvt.html
What's New What's Cool Handbook Net Search Net Directory Software
Lucent Technologies
Bell Labs Innovations
Lucent Technologies
Advanced Information Systems Engineering
Demonstration 2-D Pattern Recognition Applet.
To use this applet, click on one of the "+" or "-" buttons to choose a class. You can then click in the light grey applet area to place your points. When you have prepared the data, press "SEND". The solution will be displayed as contours, red for positive, blue for negative points. (NOTE: if the colors look strange, see the note on colormaps below).
You can remove a point by holding down the shift key and clicking over the point, or by clicking twice over the point. You can also "drag" points with the mouse. To see the coordinates of the mouse pointer, press the "p" key. To remove the coordinates, press the "p" key again. To change the default kernel or training parameters, click on the "OPTIONS" button.
The support vectors will appear with circles around them. The training errors will appear with crosses drawn through them.
+ - Options Send Clear

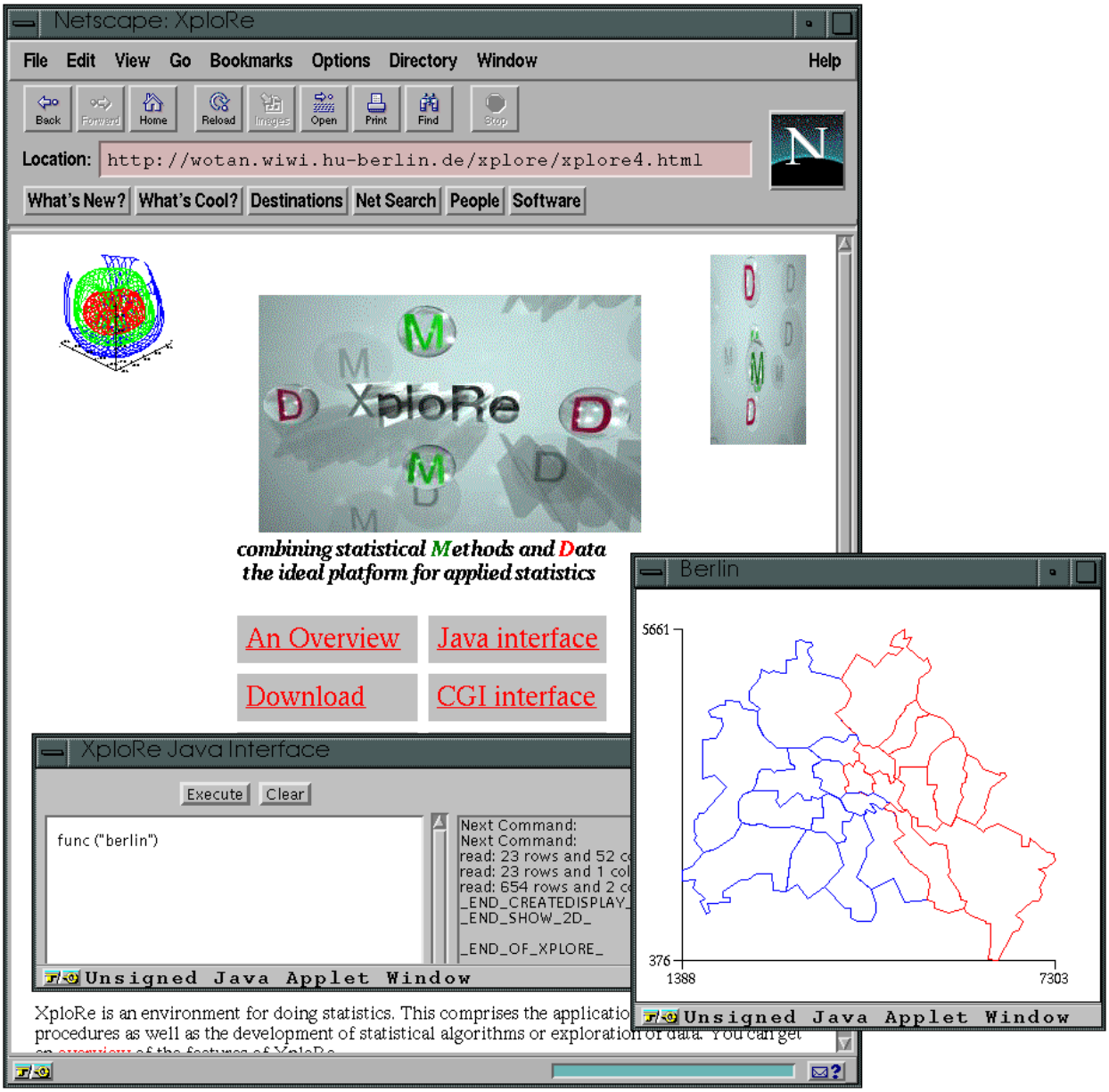

Netscape: Lucent Technologies: SVM demo applet
File Edit View Go Bookmarks Options Directory Window Help
Back Forward Home Reload Images Open Print Find Stop
Location: http://madrigal.ho.lucent.com/SVT/SVMsvt.html
What's New What's Cool Handbook Net Search Net Directory Software
The support vectors will appear with circles around them. The training errors will appear with crosses drawn through them.
+ − Options Send Clear

In terms of general purpose statistical software, there are a number of approaches.

- Statlets
  (NWP Associates, Inc.)
  http://www.statlets.com
- WebStat
  (Webster West, University of South Carolina)
  http://www.stat.sc.edu/~west/webstat/
- XploRe
  (Wolfgang Härdle & Company)
  http://wotan.wiwi.hu-berlin.de/xplore/
- S/Java
  (Bell Labs)
  http://cm.bell-labs.com/stat/project/java


Netscape: STATLETS – Java Applets for Statistical Analysis and Graph
File Edit View Go Bookmarks Options Directory Window Help
Back Forward Home Reload Images Open Print Find Stop
Location: http://www.statlets.com/
What's New? What's Cool? Destinations Net Search People Software
STATLETS
100% Pure Java Applets
for
Statistical Analysis and Graphics
STATLETS – Barcharts and Piecharts 1
File Edit Plot Summarize Analyze Model Quality Window Help
Input Table Bar chart Pie chart
Axes Options Interpret
Barchart for Defect Type
Cracks
Improper shape
Incomplete
Surface scratches
Other
0 20 40 60 80
Number of items
Unsigned Java Applet Window
STATLETS is a collectio
Internet or intranets. They


Netscape: Welcome to WebStat
File Edit View Go Bookmarks Options Directory Window Help
Location: http://www.stat.sc.edu/~west/webstat/
What's New? What's Cool? Destinations Net Search People Software
WebStat
Statistical Analysis on the Web!!!
Welcome to WebStat!
A statistical computing environment for the
While you are waiting for WebStat to down
the applet loads onto your machine three new
basic guide to the documentation available f
What exactly is WebStat and how do I
Who created it and why?
How do I get data into WebStat?
How do I format data for WebStat?
What can WebStat do with data?
How do I customize graphics?
How do I save results?
How do I exit from WebStat?
Where do I go with questions or comm
What exactly is WebStat and how d
WebStat is a statistical computing environm
of a Java applet. WebStat should run on any
has the minimal requirement of a Java-capable Web browser which almo
now has. If you do not have such a browser, the following links are to We
WebStat Main Window
Data Stat Graphics
Unsigned Jav
Graphics Window
Graphic
Unsigned Java Applet Window

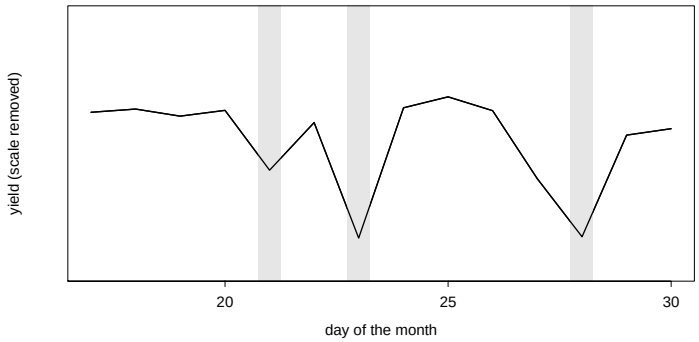

Netscape: XploRe
File Edit View Go Bookmarks Options Directory Window Help
Back Forward Home Reload Images Open Print Find Stop
Location: http://wotan.wiwi.hu-berlin.de/xplore/xplore4.html
What's New? What's Cool? Destinations Net Search People Software
XploRe
combining statistical Methods and Data
the ideal platform for applied statistics
An Overview
Java interface
Download
CGI interface
XploRe Java Interface
Execute Clear
func ("berlin")
Next Command:
Next Command:
_END_CREATEDISPLAY_
_END_SHOW_2D_
_END_OF_XPLORE_
Unsigned Java Applet Window
Berlin
5661
376
1388
7303
Unsigned Java Applet Window
XploRe is an environment for doing statistics. This comprises the applicatio
procedures as well as the development of statistical algorithms or explorati

Some Background

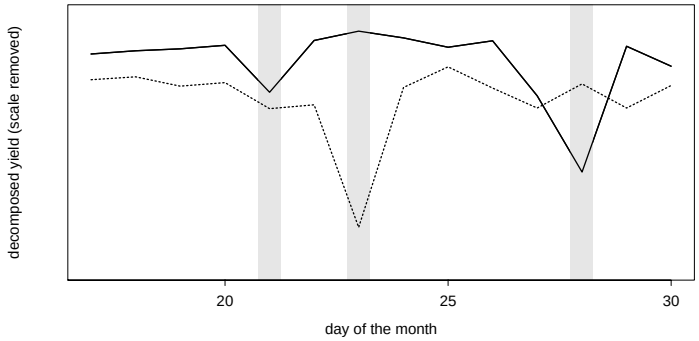
- Research $\rightarrow$ S Software $\rightarrow$ BU process & decision-making
- The WWW as a medium for sharing the results

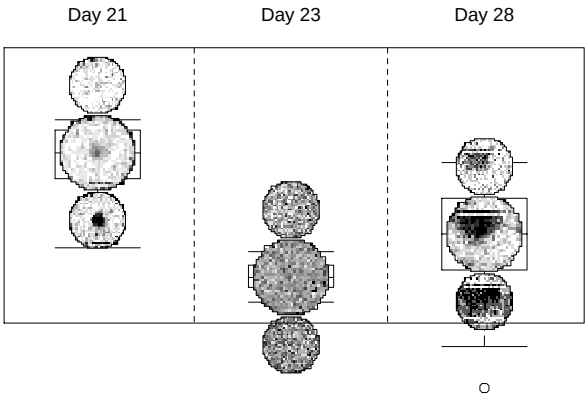
Example: Semiconductor manufacturing

Wafer Yield by Day



Decomposing Yield into Clustered and Random Effects





The Advantages to Users

- Quick transfer of new technology
- More informative, interesting displays
- Sense of community, regardless of geography

Stages of Interaction

1. Graphical & Textual Displays
   (distributing PostScript or PDF)

2. Live Interaction with Displays
   (Java or VRML managing interactions)

3. Specialized Data Analysis
   (Java, CGI, specialized clients and servers)

4. General Use of S
   (serving objects → links to other statistical services)

Examples

1. Wafer Plots
2. Triogram Interactive Plots
3. Animation
4. S GUI

Sample it on-line, starting at

http://cm.bell-labs.com/cm/ms/who/cocteau

Comments on Examples

- Java provides dynamic graphics, interaction
- The data involved is non-trivial
- BUT: the programming effort in such examples need not be large

Java and S
Made for each other?

- Both object oriented

  Every object in S has a Java representation in the interface!

- Both handle communication & events
- Both use a general, high-level programming model
- But they have complementary strengths

Complementary Strengths

- Java is strong in detailed user interface, low-level graphics, and window management
- S is strong in data structures, higher-level graphics (data visualization), computing with data

How It Works
Starting An Application

1. An S Server, for example, a Linux box, provides access to S and some cgi-bin scripts

2. A user on the Web invokes a script (URL) on the server

3. The script returns an (applet, S process)-pair

4. Under the user's control, the applet generates tasks and/or data for the S server, receives data from the server, and displays information with which the user can interact

How It Works

Incorporating New Java Clients

An existing Java system implementing some interactive display or task can often be adapted to use the S/Java interface with minimal effort

(Examples include dynamic graphical displays and spreadsheet editors)

How It Works
Incorporating New Java Clients

1. The data input method is overwritten to make use of a Java S-object from the interface

2. The remaining Java code is incorporated as is

3. On the S side, the programmer needs to write a function to generate the data (e.g. a matrix). Methods in the interface handle data transmission transparently.

Off-Line / On-Line

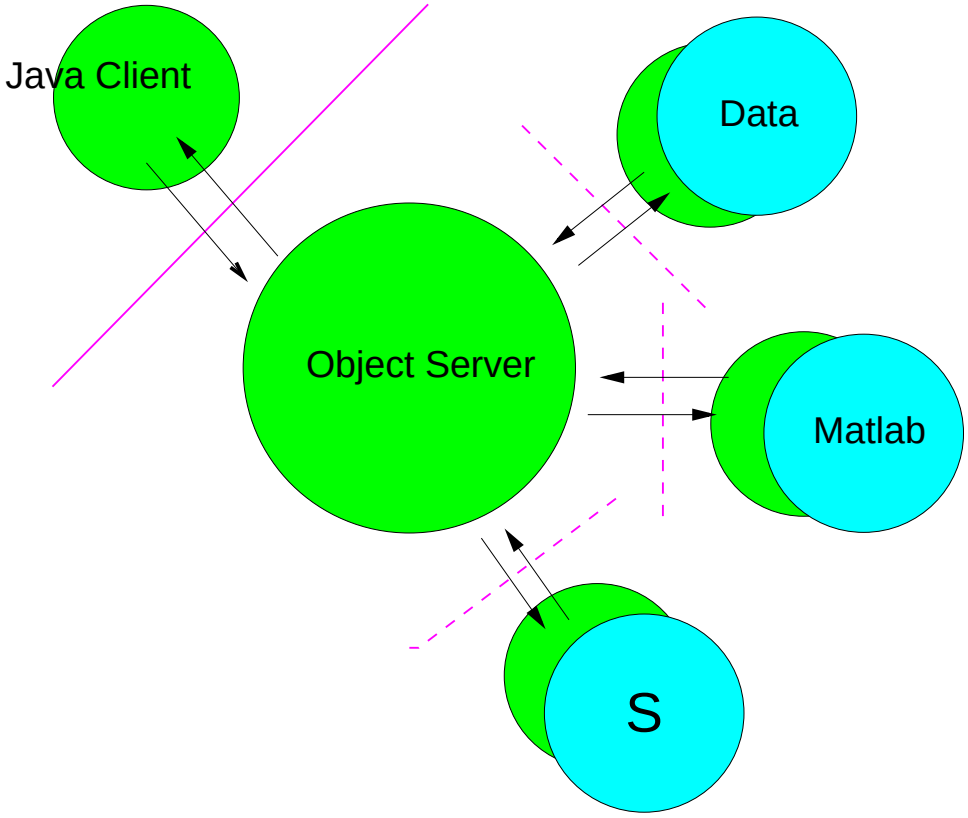
Java interaction can be on-line or off-line; that is, with S running during the interaction, or by running S off-line to generate and save data that is to be examined later

- Off-line eliminates inter-process hassles and is faster,
- BUT limits the user's interactive options

On-line use doesn't imply that the user is conscious of using S. Depending on the application, the actual S tasks may be generated with little or no "programming."

S ↔ Java Communication

- Data transmission is handled by a Java application known as an Object Server
- Currently, all objects in the interface are of class ANY_S
- The class of the object 'dispatches' a corresponding method to do whatever was requested
- This simple framework allows us to link easily into the data base capabilities of Java. We are also incorporating servers other than S, Matlab being our first project.


Java Client
Data
Object Server
Matlab
S

Conclusion

- There is a tremendous rush to (re)produce statistical platforms in languages like Java
  (portability, interactivity, network computing)

- Our S/Java interface allows us to distribute new statistical methods easily, making clear the separation between client and server
  (not all clients are created equal)

- Through the Object Server, we are now able to link in other statistical platforms
  (what is the language on the wire?)

PC WEEK
JUNE 30, 1997

JOHN DODGE: THIS PC WEEK

STANDARDS CAN'T KEEP PACE WITH NET GROWTH



FOR ALL ITS BENEFITS, STANDARDS-BASED COMPUTING has some major failings.

By definition, standards are compromises. The bodies that create them are analogous to Congress, which sets a dubious standard for inaction, bickering and watered-down thinking.