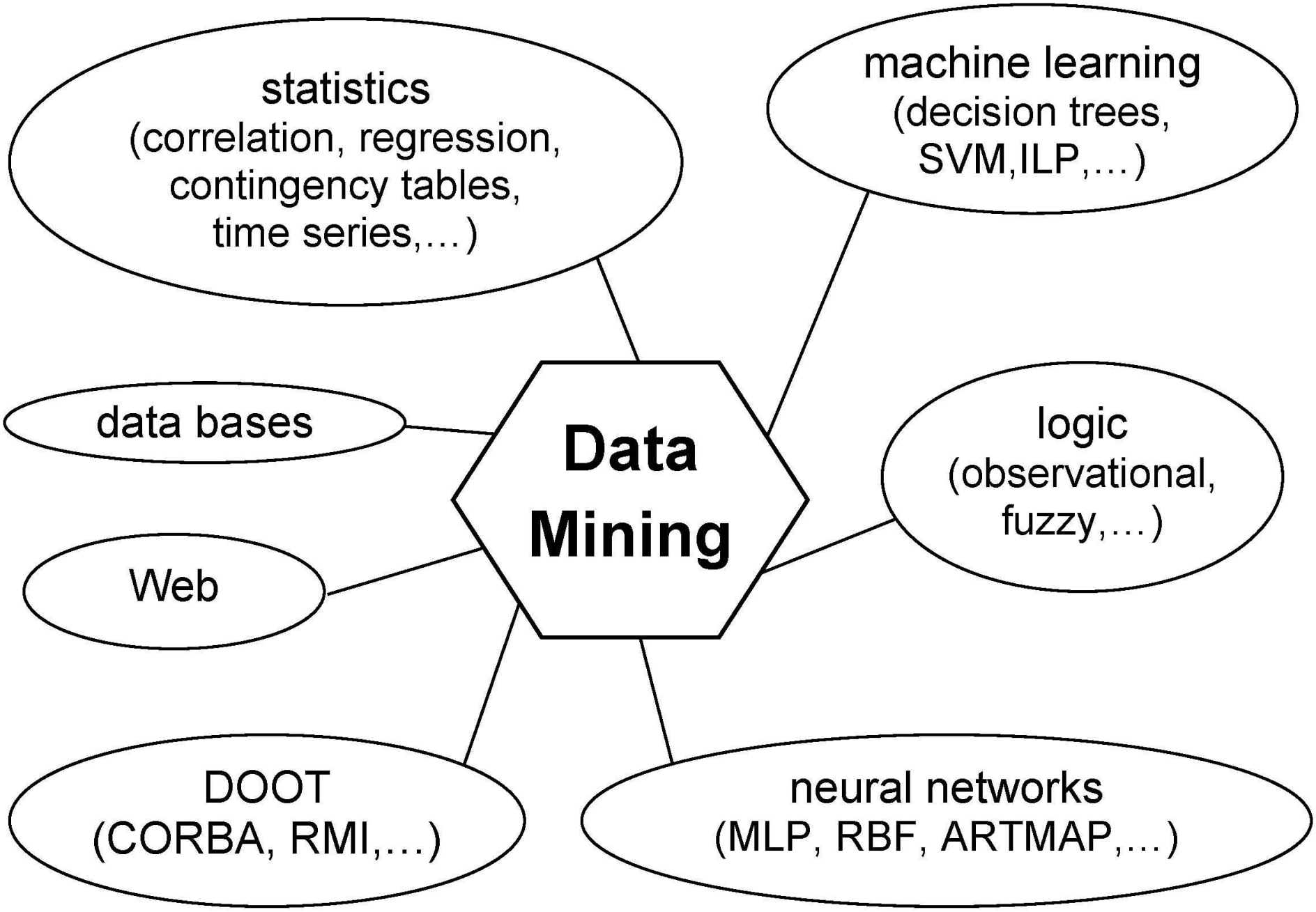
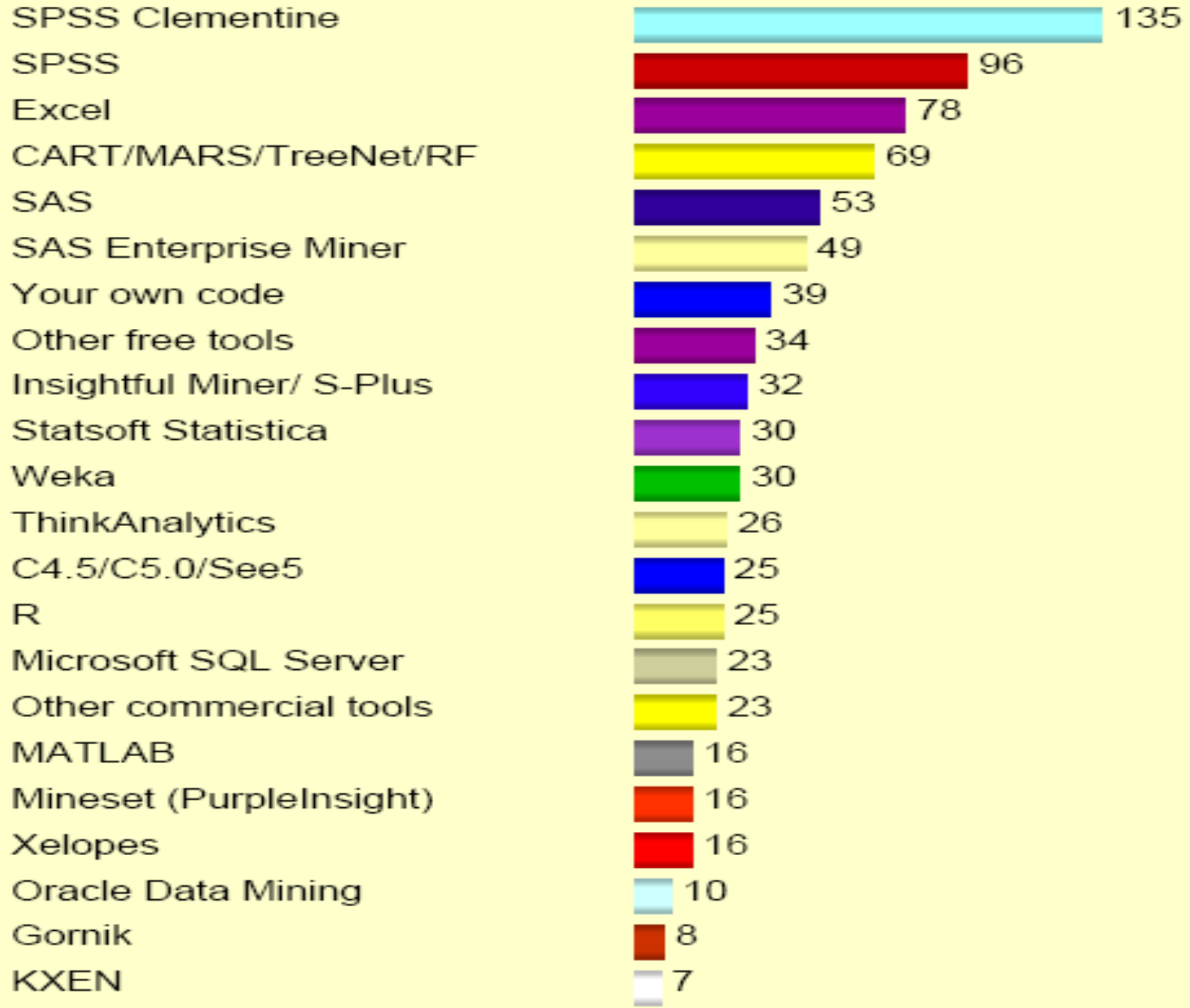
John W. Tukey

EXPLORATORY DATA ANALYSIS

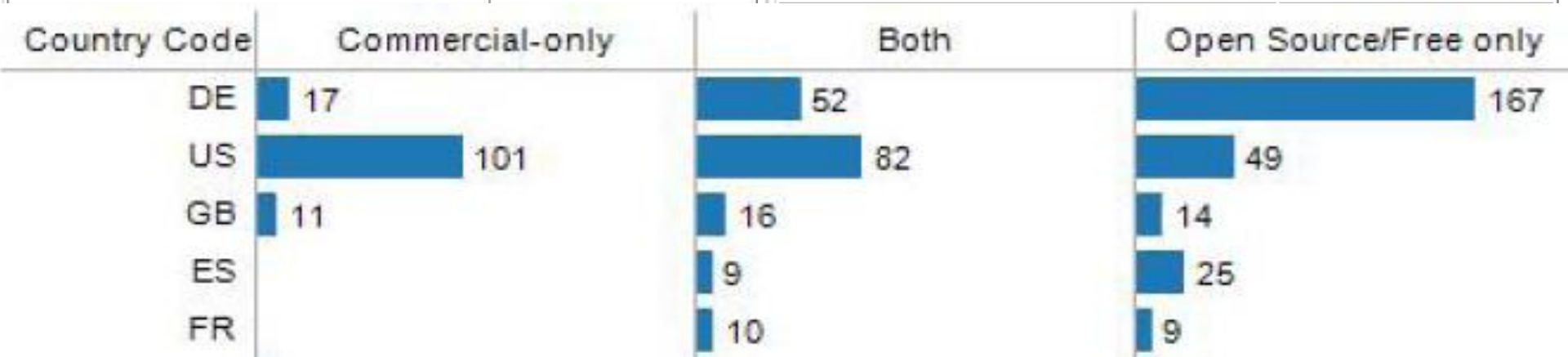

statistics
(correlation, regression,
contingency tables,
time series,…)
machine learning
(decision trees,
SVM,ILP,…)
data bases
Data
Mining
logic
(observational,
fuzzy,…)
Web
DOOT
(CORBA, RMI,…)
neural networks
(MLP, RBF, ARTMAP,…)

Data mining/Analytic tools you used in 2005 [376 voters, 860 votes total]



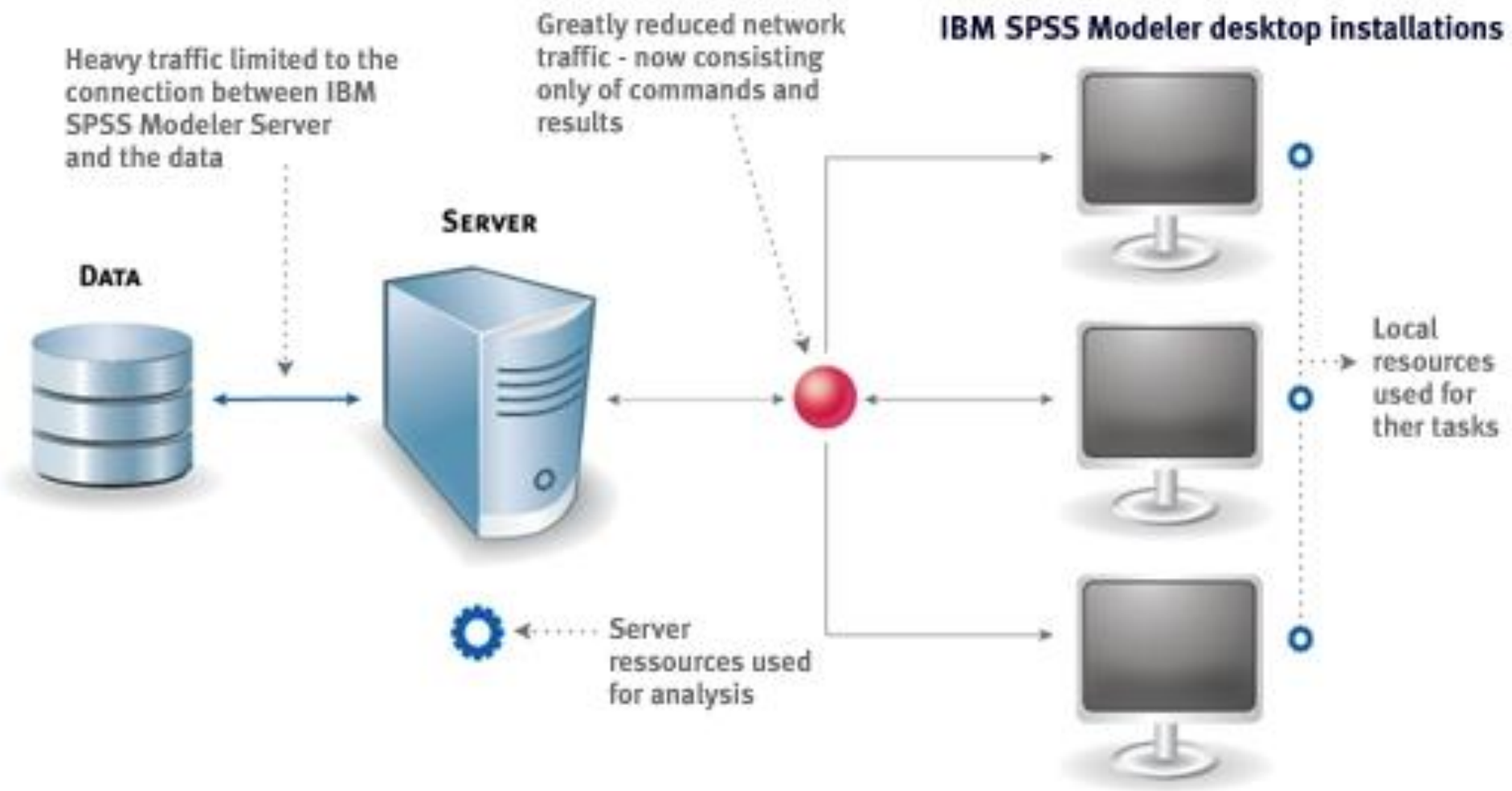
Data Mining Tools Used Poll (May 2009)

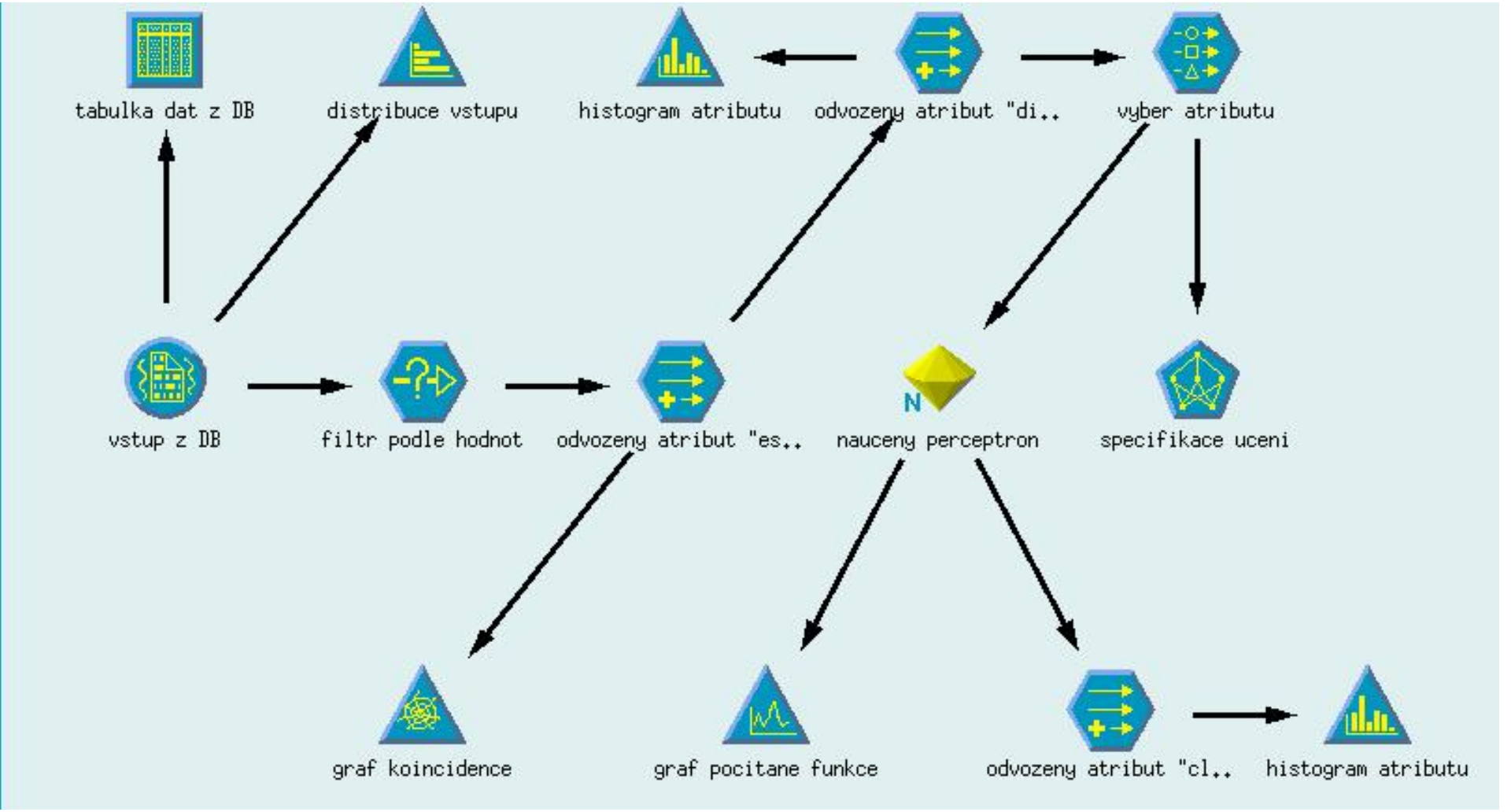
What data mining tools have you used for a real project (not just for evaluation) in the past 6 months? (364 voters)	
SPSS PASW Modeler (formerly Clementine) (68 alone, 52 with other tools, 120 total)	
RapidMiner (36 alone, 41 w. other tools, 77 total)	
SAS (39 alone or with SAS EM; 36 with other tools, 75 total)	
Excel (1 alone, 68 total)	
SAS Enterprise Miner (39 alone or with SAS; 28 w/ other tools; 67 total)	
R (2 alone, 51 total)	
Your own code (3 alone, 44 total)	
KXEN (25 alone, 31 total)	
Weka (now Pentaho) (0 alone, 31 total)	
MATLAB (0 alone, 26 total)	
Other commercial tools (0 alone, 19 total)	
KNIME (1 alone, 18 total)	
Other free tools (0 alone, 15 total)	
Microsoft SQL Server (1 alone, 15 total)	
Zementis (5 alone, 13 total)	
Oracle DM (0 alone, 9 total)	
Statsoft Statistica (0 alone, 8 total)	
Orange (0 alone, 5 total)	
Salford CART, Mars, other (1 alone, 5 total)	
C4.5/C5.0 (0 alone, 4 total)	
Angoss (0 alone, 4 total)	



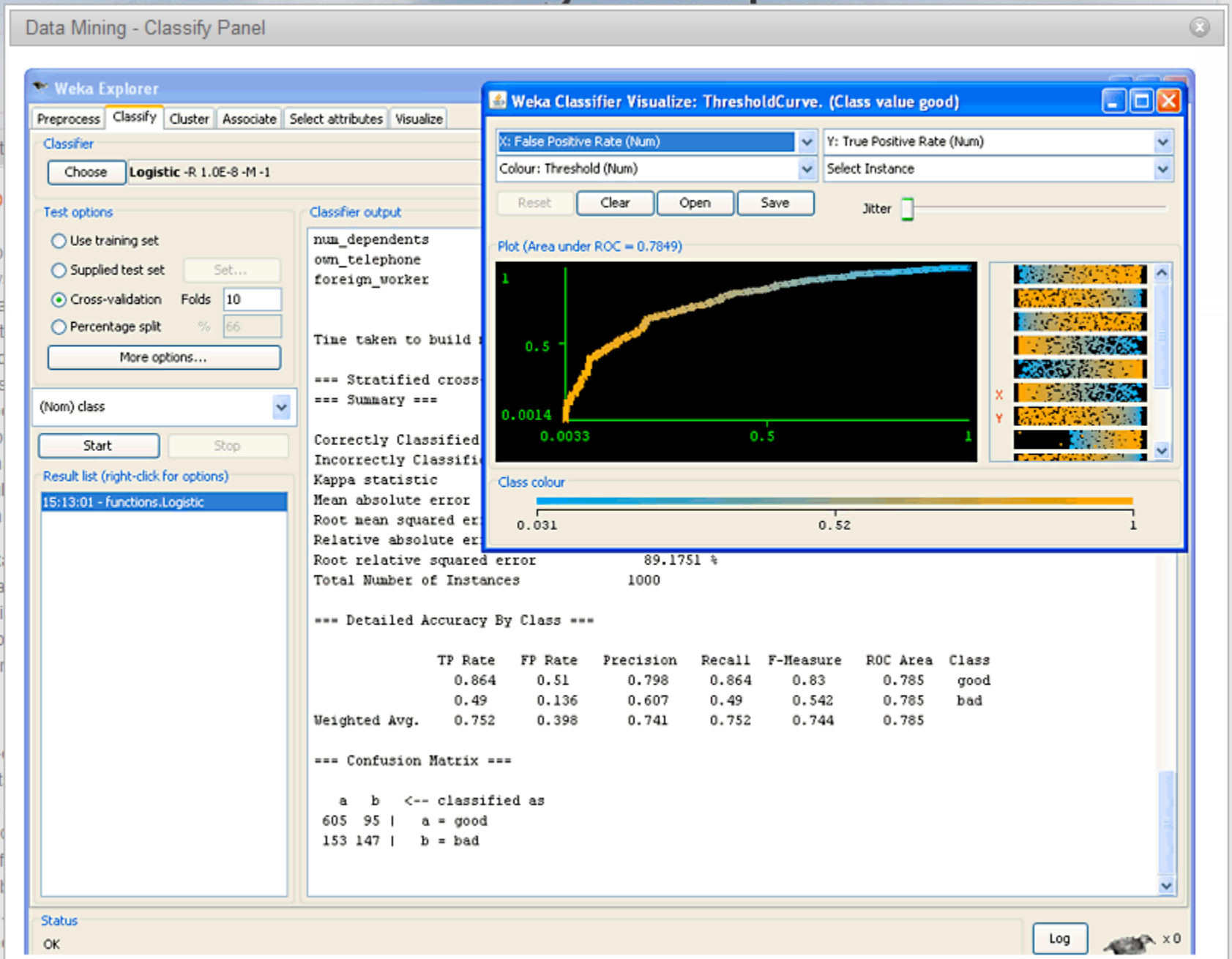
Our customers include:

- All 50 U.S. state governments
- 22 of the top 24 global commercial banks
- 100% of the top U.S. universities
- The 18 top property and casualty insurance companies in the U.S.
- The leading 12 global pharmaceutical companies




tabulka dat z DB
distribuce vstupu
histogram atributu
odvozeny atribut "di..
vyber atributu
vstup z DB
filtr podle hodnot
odvozeny atribut "es..
N
nauceny perceptron
specifikace uceni
graf koincidence
graf pocitane funkce
odvozeny atribut "cl..
histogram atributu

Pentaho Data Mining Enterprise Edition



RapidMiner@RapidNTB02

File Edit Process Tools View Help

Result Overview | ExampleSet (Retrieve)

Meta Data View | Data View | Plot View

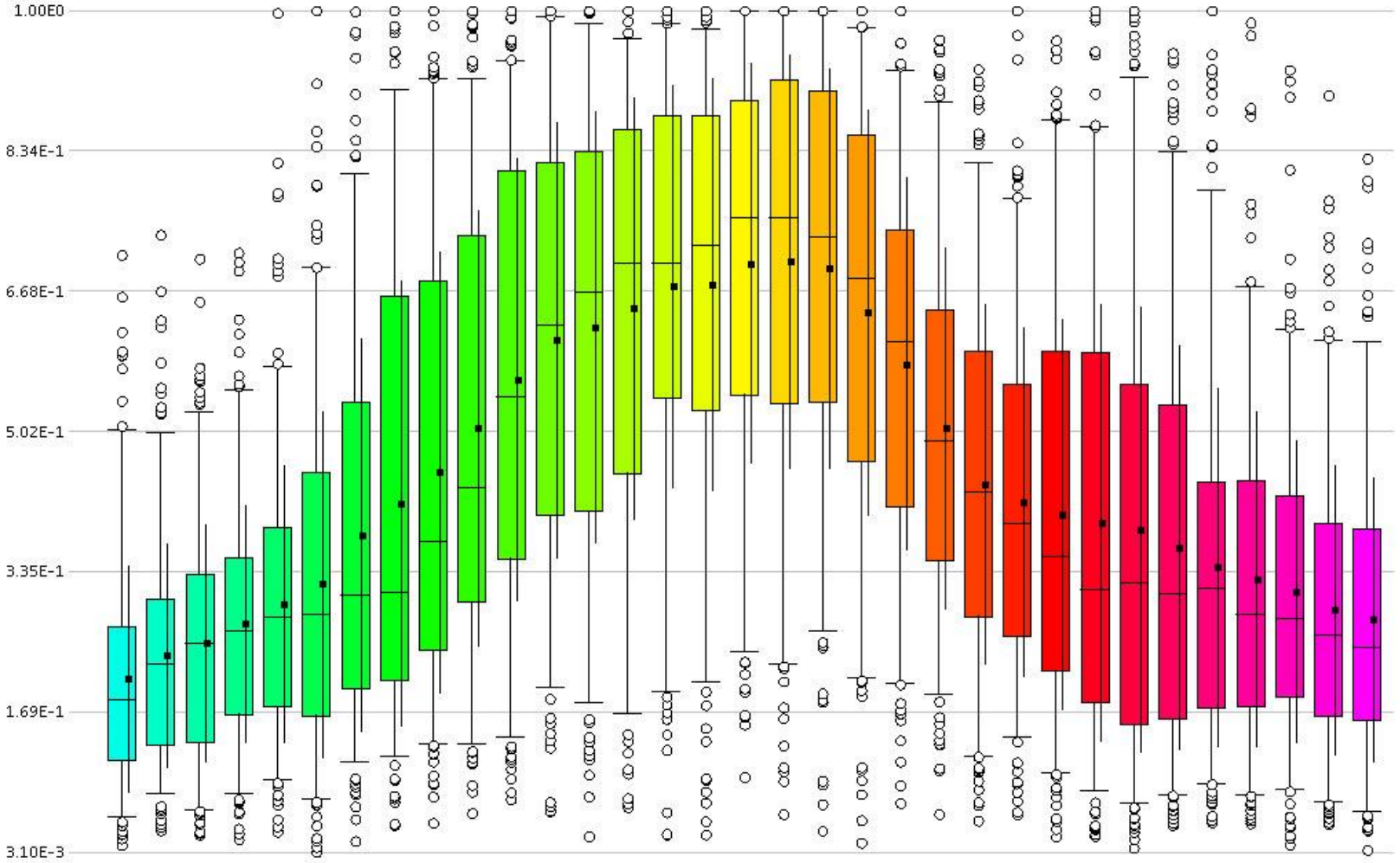
Plotter

Quartile

Dimensions

attribute_8
attribute_9
attribute_10
attribute_11
attribute_12
attribute_13
attribute_14
attribute_15
attribute_16
attribute_17
attribute_18
attribute_19
attribute_20
attribute_21
attribute_22
attribute_23
attribute_24
attribute_25
attribute_26
attribute_27
attribute_28
attribute_29
attribute_30
attribute_31
attribute_32
attribute_33
attribute_34
attribute_35
attribute_36
attribute_37
attribute_38
attribute_39
attribute_40
attribute_41

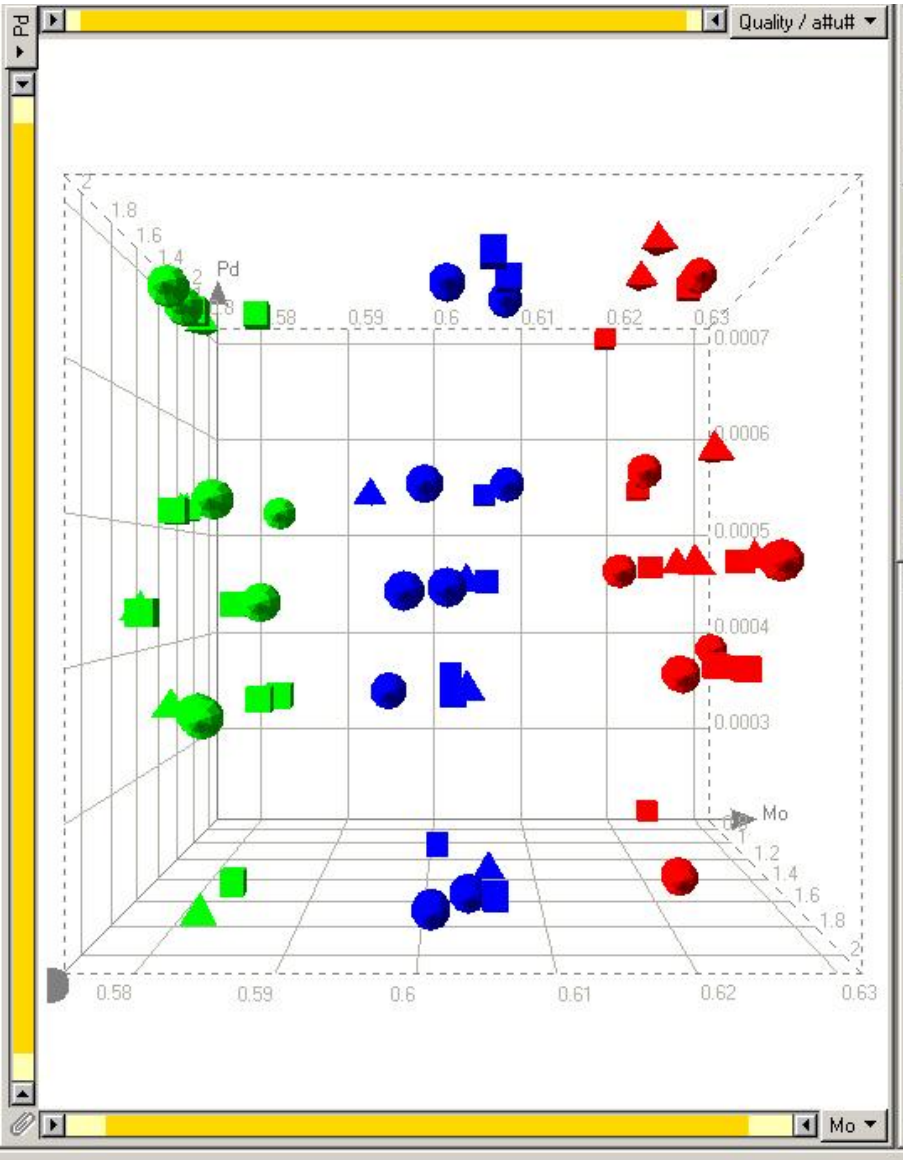
Save Image...

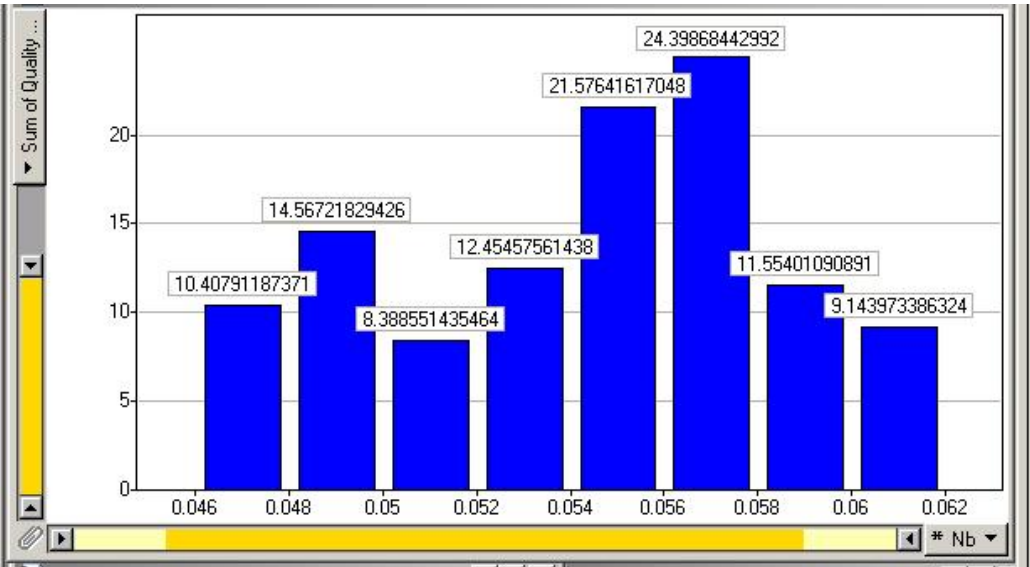


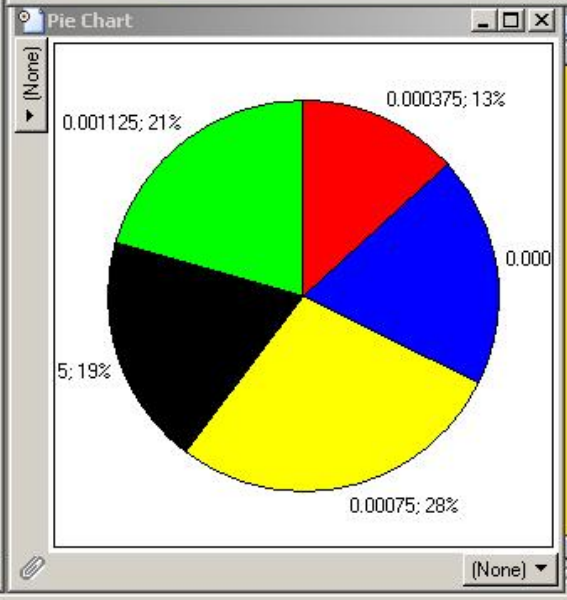
Log

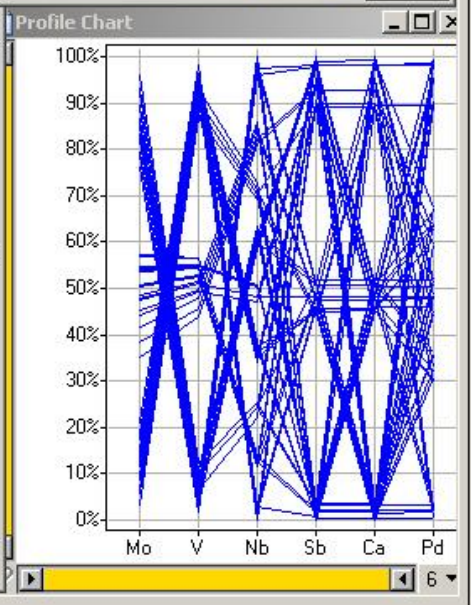
Jan 20, 2010 10:18:24 AM INFO: No filename given for result file, using stdout for logging results!
Jan 20, 2010 10:18:24 AM INFO: Process starts
Jan 20, 2010 10:18:24 AM INFO: Loading initial data

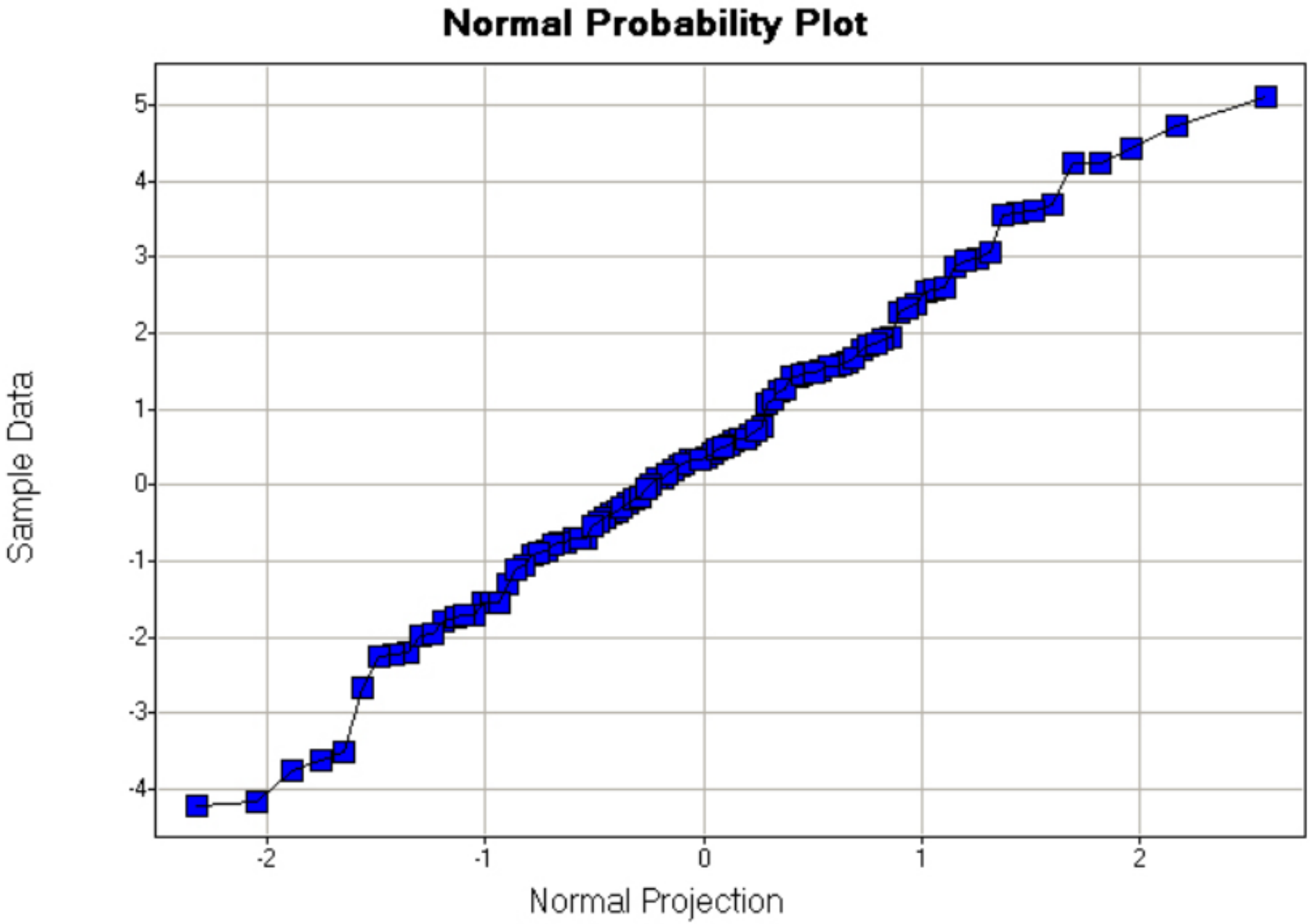
System Monitor

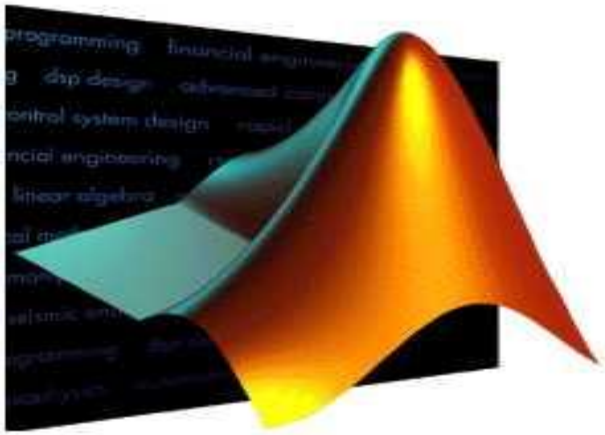

Pd
Quality / a#u#
Pd
Mo
0.58
0.59
0.6
0.61
0.62
0.63
0.0007
0.0006
0.0005
0.0004
0.0003
Mo


Sum of Quality ...
24.39868442992
21.57641617048
14.56721829426
12.45457561438
11.55401090891
10.40791187371
9.143973386324
8.388551435464
0
5
10
15
20
0.046
0.048
0.05
0.052
0.054
0.056
0.058
0.06
0.062
* Nb


Pie Chart
(None)
0.000375; 13%
0.001125; 21%
0.000
5; 19%
0.00075; 28%
(None)


Profile Chart
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Mo
V
Nb
Sb
Ca
Pd
6


Normal Probability Plot
Sample Data
5
4
3
2
1
0
-1
-2
-3
-4
-2
-1
0
1
2
Normal Projection



MATLAB®

The Language of Technical Computing

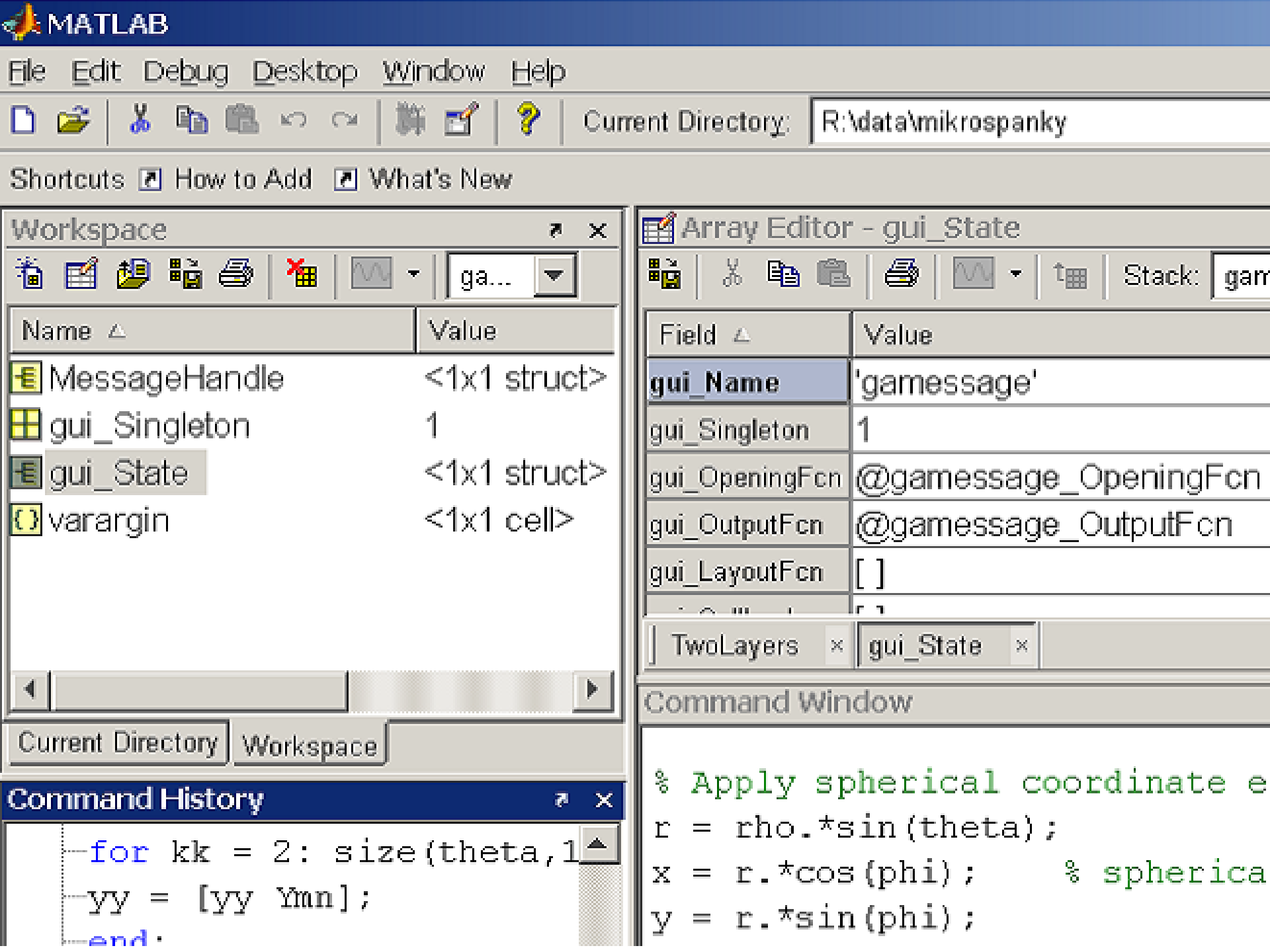
The MATH WORKS Inc.

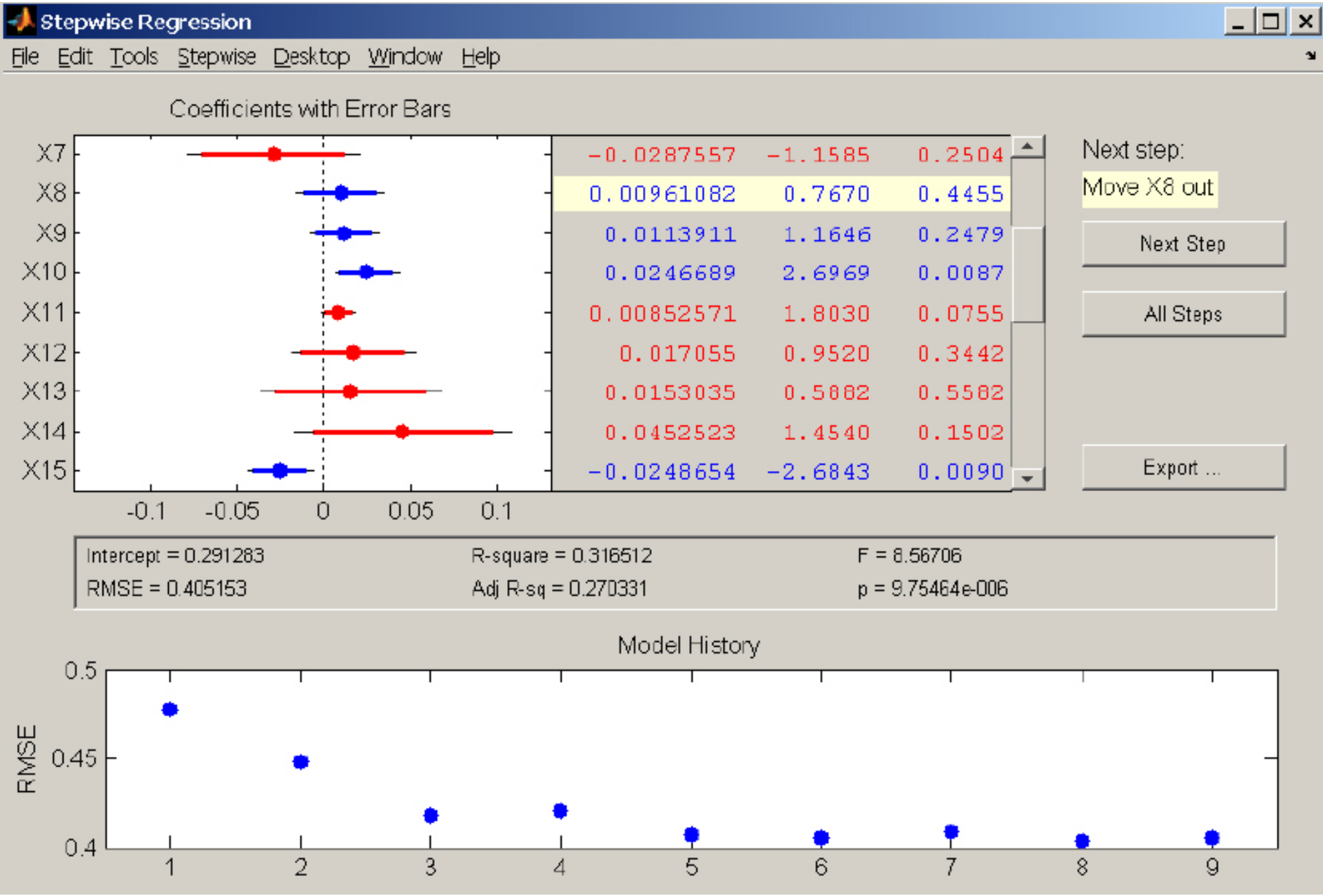
REVENUE

- $500 million in 2009, with more than 50% from outside the United States
- Profitable every year since its founding

FAST FACTS

- Founded in 1984
- Privately held
- There are more than:
  - » 1 million users of MATLAB worldwide
  - » 350 third-party solutions that build on MATLAB and Simulink
  - » 1200 MATLAB based books in 26 languages


MATLAB
File Edit Debug Desktop Window Help
Current Directory: R:\data\mikrospanky
Shortcuts How to Add What's New
Workspace
ga...
Name
Value
MessageHandle
<1x1 struct>
gui_Singleton
1
gui_State
<1x1 struct>
varargin
<1x1 cell>
Current Directory
Workspace
Command History
for kk = 2: size(theta,1
yy = [yy Ymn];
end
Array Editor - gui_State
Stack:
Field
Value
gui_Name
'gamessage'
gui_Singleton
1
gui_OpeningFcn
@gamessage_OpeningFcn
gui_OutputFcn
@gamessage_OutputFcn
gui_LayoutFcn
[ ]
TwoLayers
gui_State
Command Window
% Apply spherical coordinate e
r = rho.*sin(theta);
x = r.*cos(phi);
% spherica
y = r.*sin(phi);


Stepwise Regression
File Edit Tools Stepwise Desktop Window Help
Coefficients with Error Bars
X7
X8
X9
X10
X11
X12
X13
X14
X15
-0.1 -0.05 0 0.05 0.1
-0.0287557 -1.1585 0.2504
0.00961082 0.7670 0.4455
0.0113911 1.1646 0.2479
0.0246689 2.6969 0.0087
0.00852571 1.8030 0.0755
0.017055 0.9520 0.3442
0.0153035 0.5682 0.5582
0.0452523 1.4540 0.1502
-0.0248654 -2.6843 0.0090
Next step:
Move X8 out
Next Step
All Steps
Export ...
Intercept = 0.291283
RMSE = 0.405153
R-square = 0.316512
Adj R-sq = 0.270331
F = 8.56706
p = 9.75464e-006
Model History
RMSE
0.5
0.45
0.4
1 2 3 4 5 6 7 8 9