

Feedback Welcome About

- data you'd like to be able to view this way.
- better ways of looking at it than this.
- ways of avoiding visualization entirely.

Alva L. Couch

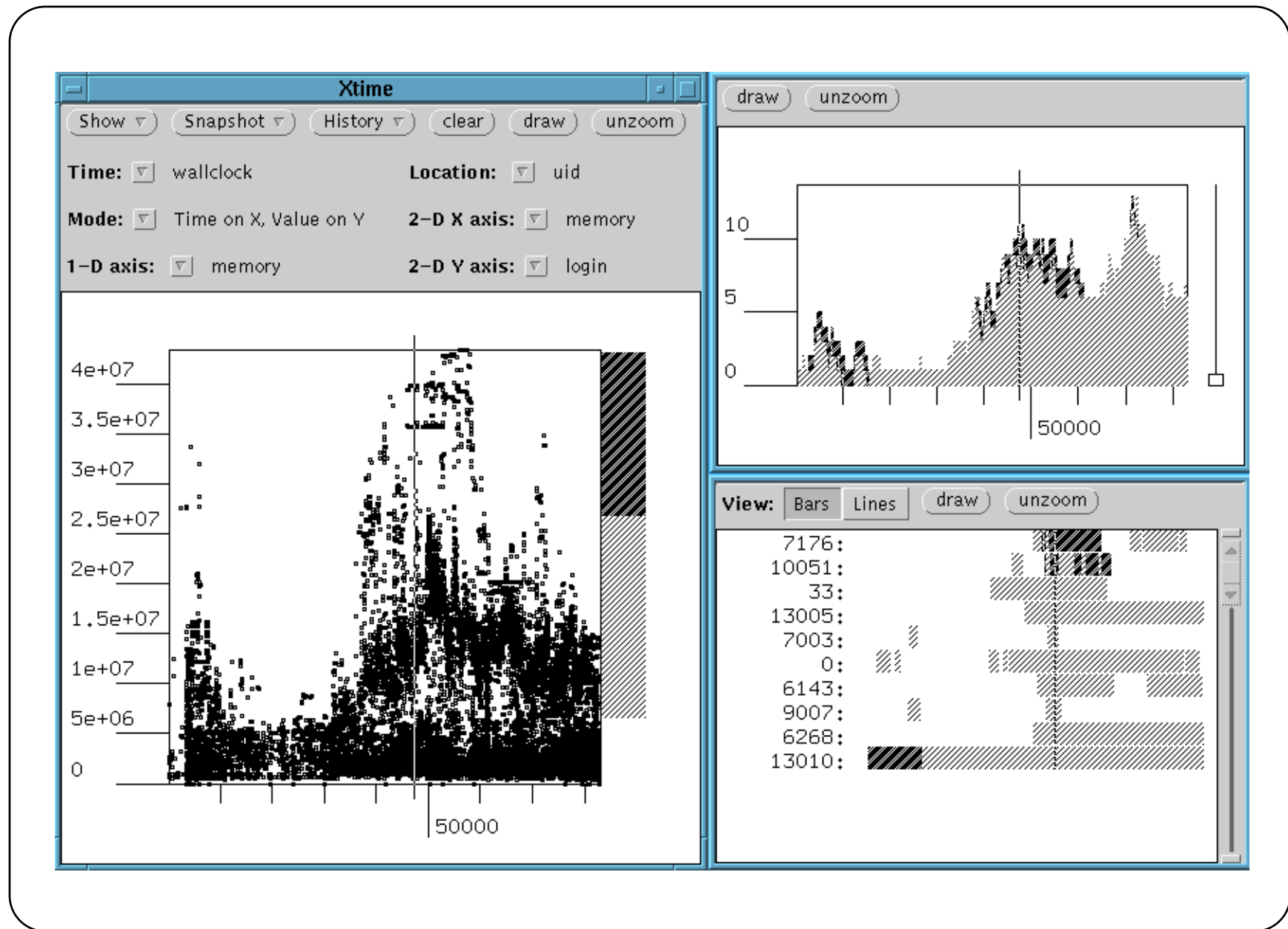
**Electrical Engineering and Computer
Science, Tufts University**

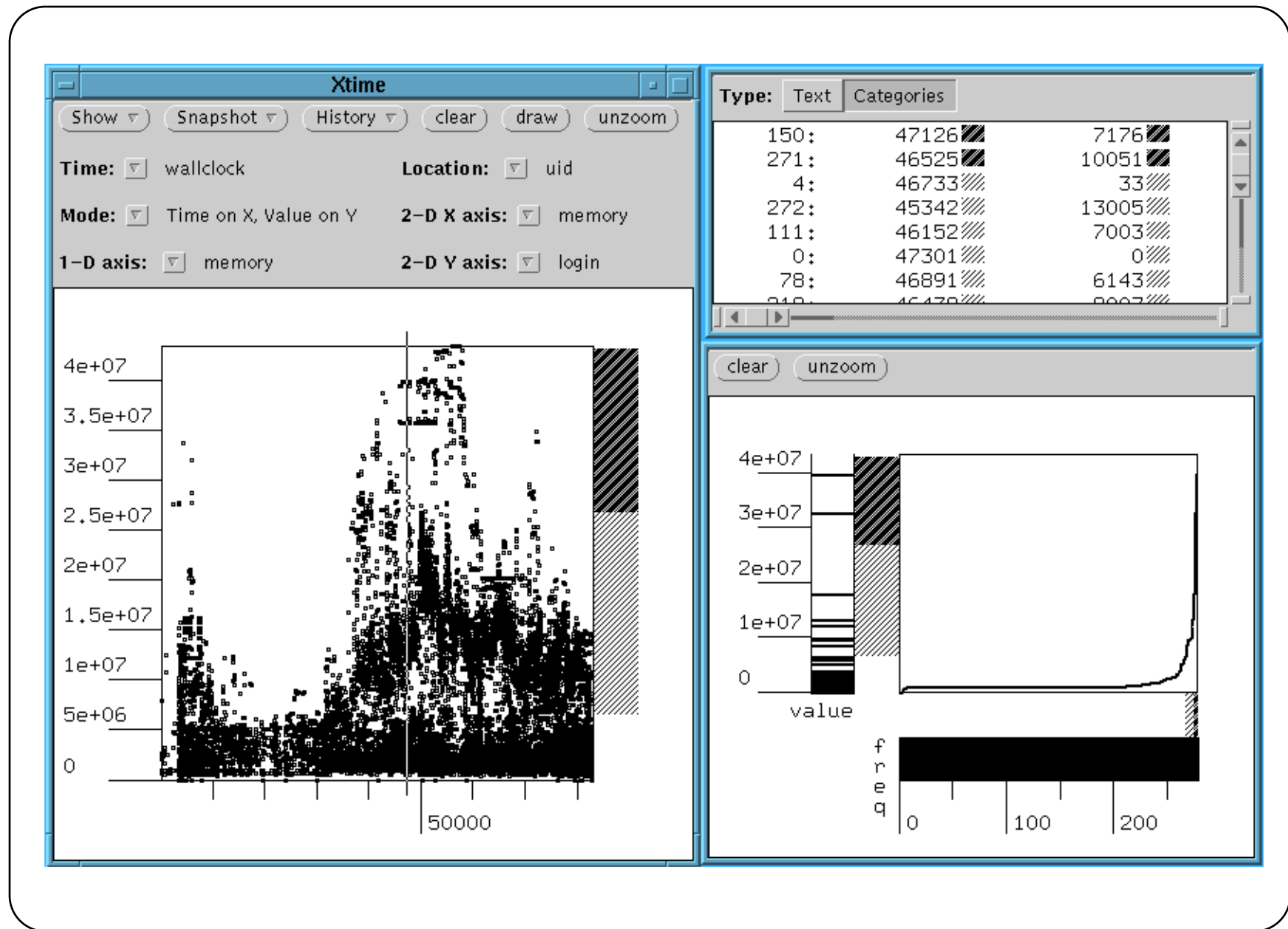
Email: couch@cs.tufts.edu

Web: <http://www.cs.tufts.edu/~couch>

Conclusions

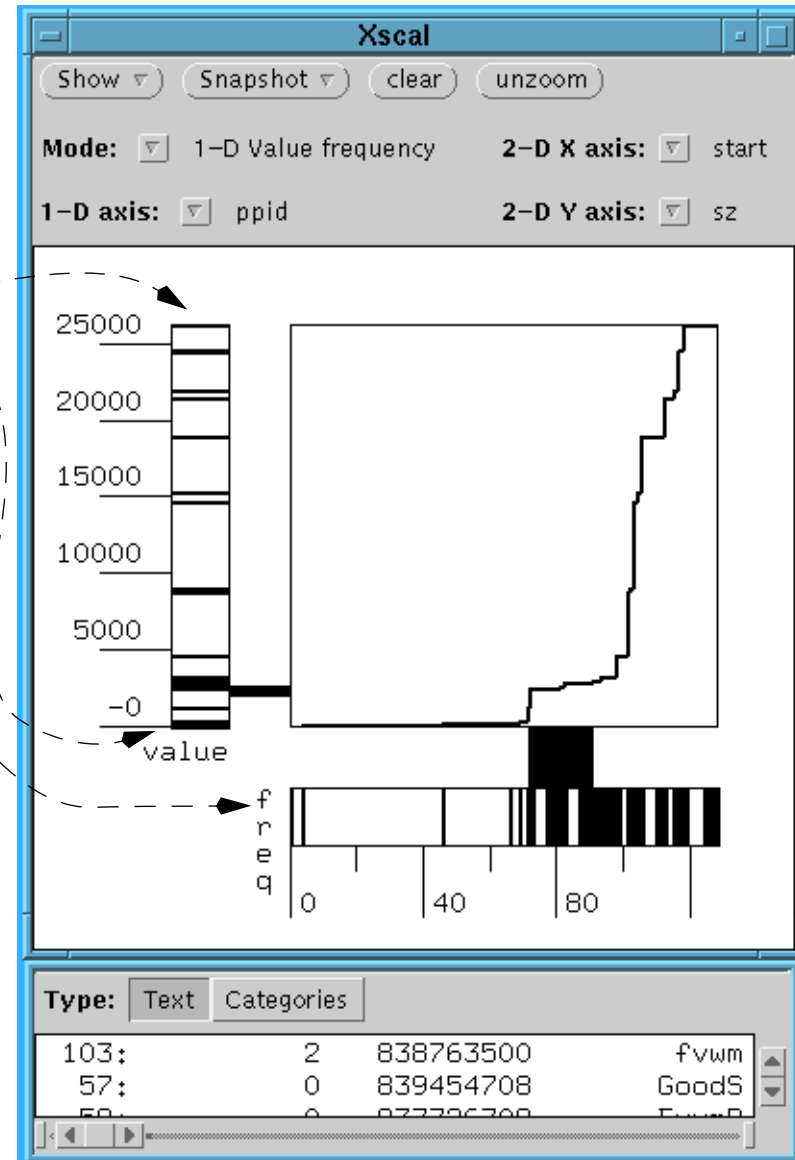
- Xscal is a very powerful tool, but
 - it is tremendously difficult to format data for it
 - ✦ from other formats
 - ✦ over time (see paper)
 - it suffers from the ‘canonical problems’ that plague all visualization tools:
 - ✦ patterns emerge by luck.
 - ✦ must choose the correct view to spot.
- We need a simpler, domain-specific tool.
- I’m willing to create it, but would like feedback on what’s useful to see.





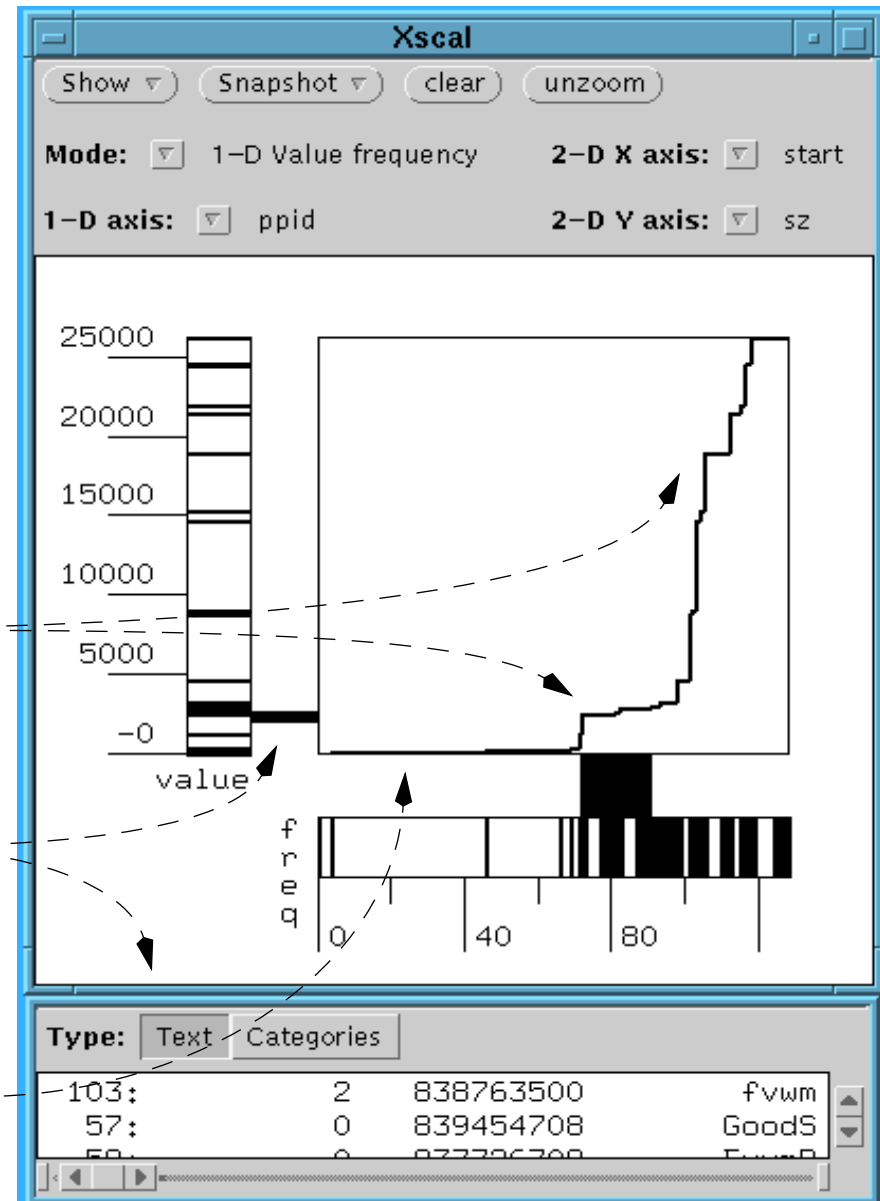
Dealing with Time

- Value scatterplot has a hashmark for each value present
- Frequency scatterplot has a hashmark between each two distinct values
- Plot these against time and present this view for each individual time.



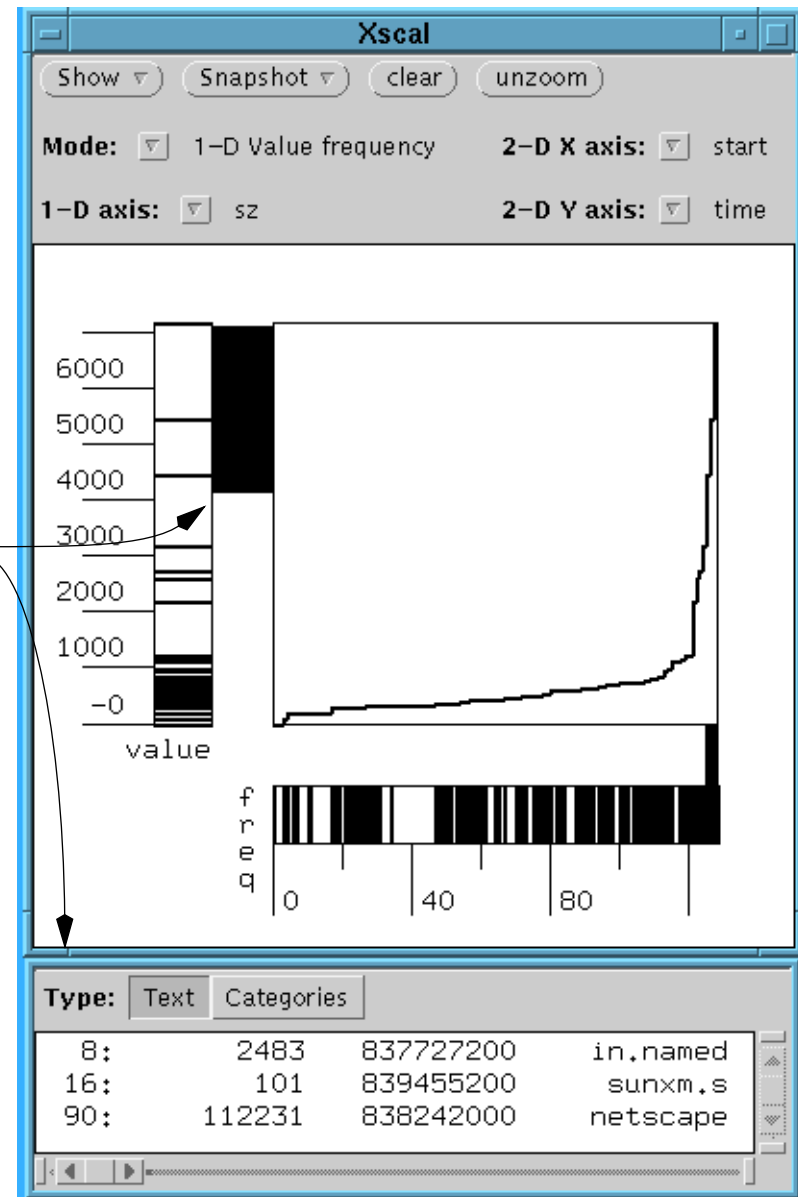
Surprising Information from a mundane source

- Parent Process ID's
- Plateaux show one user session's contribution to total processes
- Zooming in shows that long plateaux are X11 sessions.
- Flat section at bottom represents system processes (from init)



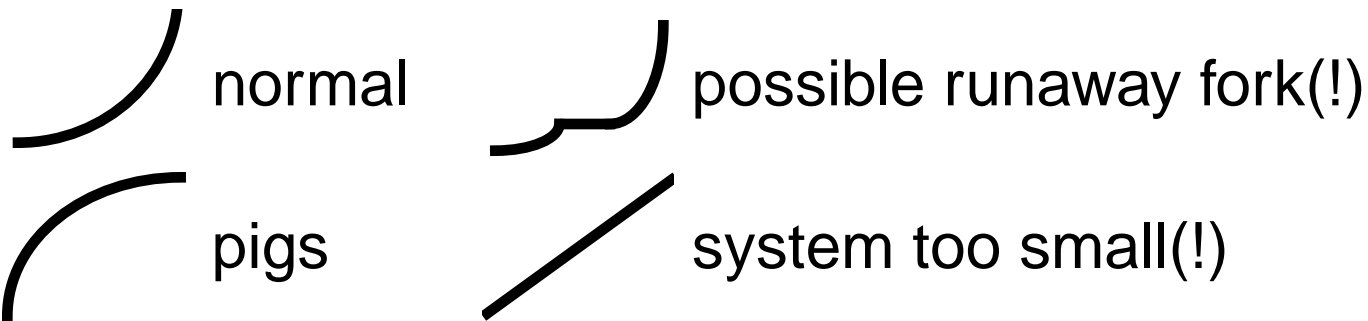
Problem Pursuit

- The value-frequency view hides the identity of participating entities.
- Highlighting problem areas displays identities of problem processes.
- Multiple colors for different kinds of processes.

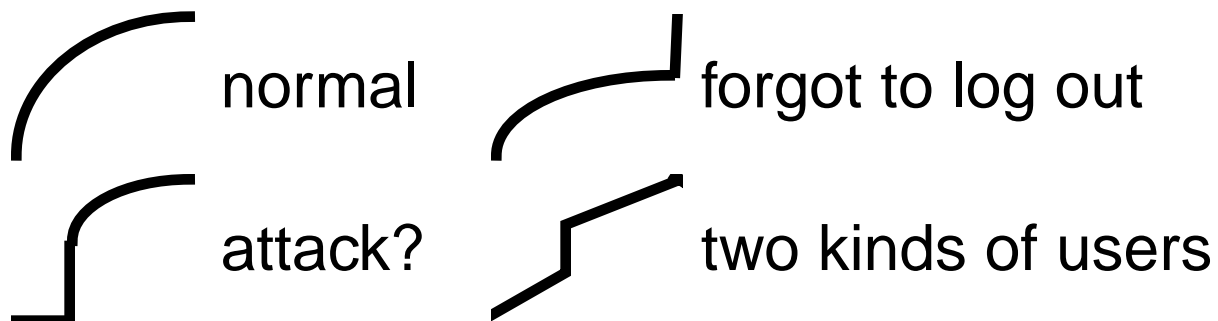


What's Normal?

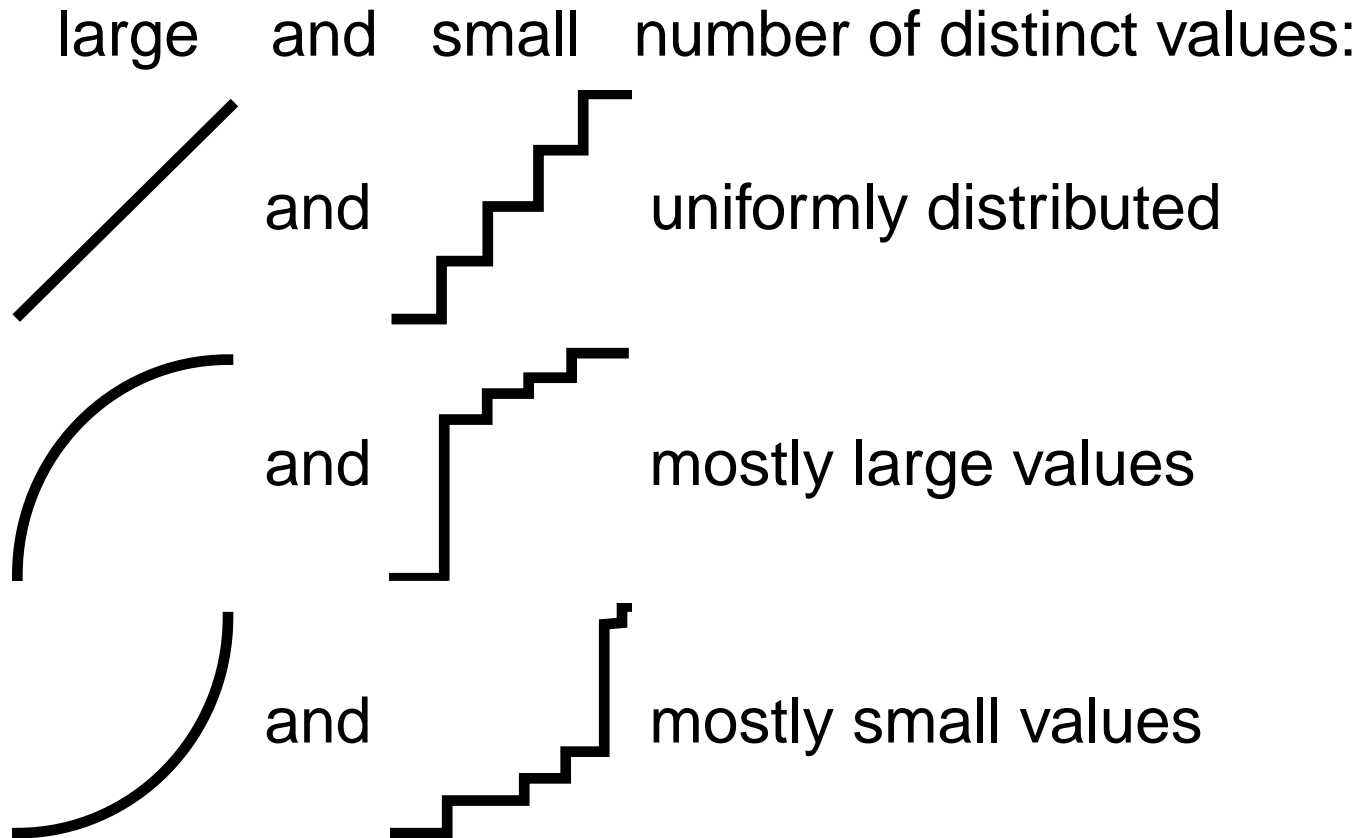
○ Process memory size:



○ Session time:

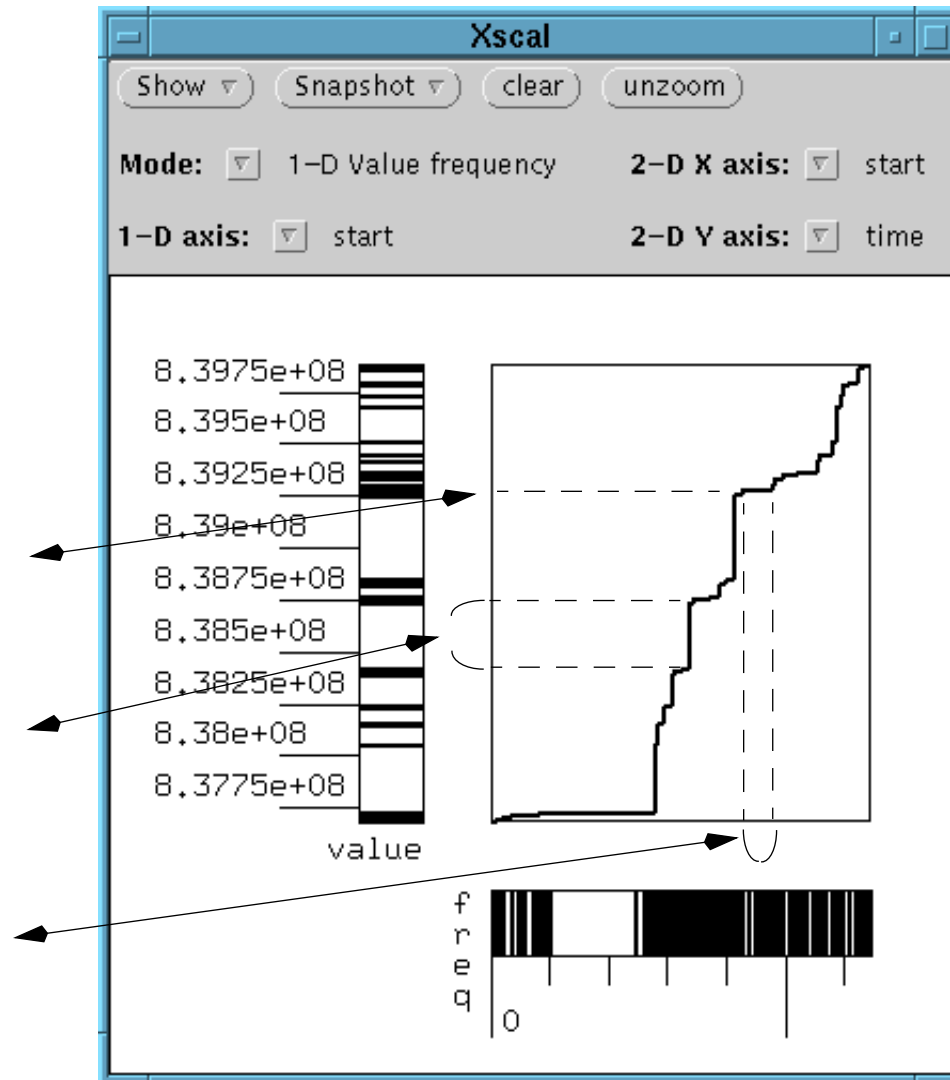


Characteristic Curves



Xscal's Value-Frequency View

- value on Y axis
- frequency of values on X axis.
- height of each horizontal line indicates value
- verticals show value gaps
- width of each horizontal line shows frequency



XSCAL

- Comparison views for thousands of variables
- Scale-invariant representation of data, equally usable when there are
 - ☐ two or millions of measurements.
 - ☐ two or millions of distinct values measured.
- Utilizes value-frequency duality
 - ☐ frequency is required to make sense of a few distinct values
 - ☐ If there are many distinct values, frequency is relatively unimportant.
- Solution: combine value and frequency in one view.

Overview

- The administrator of a large system or network is constantly exposed to vast amounts of performance data
 - snapshots: details for one instant of time.
 - traces: details over a time period.
- To use this data to detect abnormal conditions, it is first necessary to **define normality**, and then to **detect abnormal conditions** based upon this definition.
- Visualization tools can help, if there is a view in which these behaviors look different.

VISUALIZING HUGE TRACEFILES WITH XSCAL

Alva L. Couch

Assoc. Prof. of Electrical

Engineering and Computer Science

Tufts University

Email: *couch@cs.tufts.edu*

Web: *http://www.cs.tufts.edu/~couch*