Visualizing Transaction Data:

A Case Study for the CPI

Edward J. Wegman

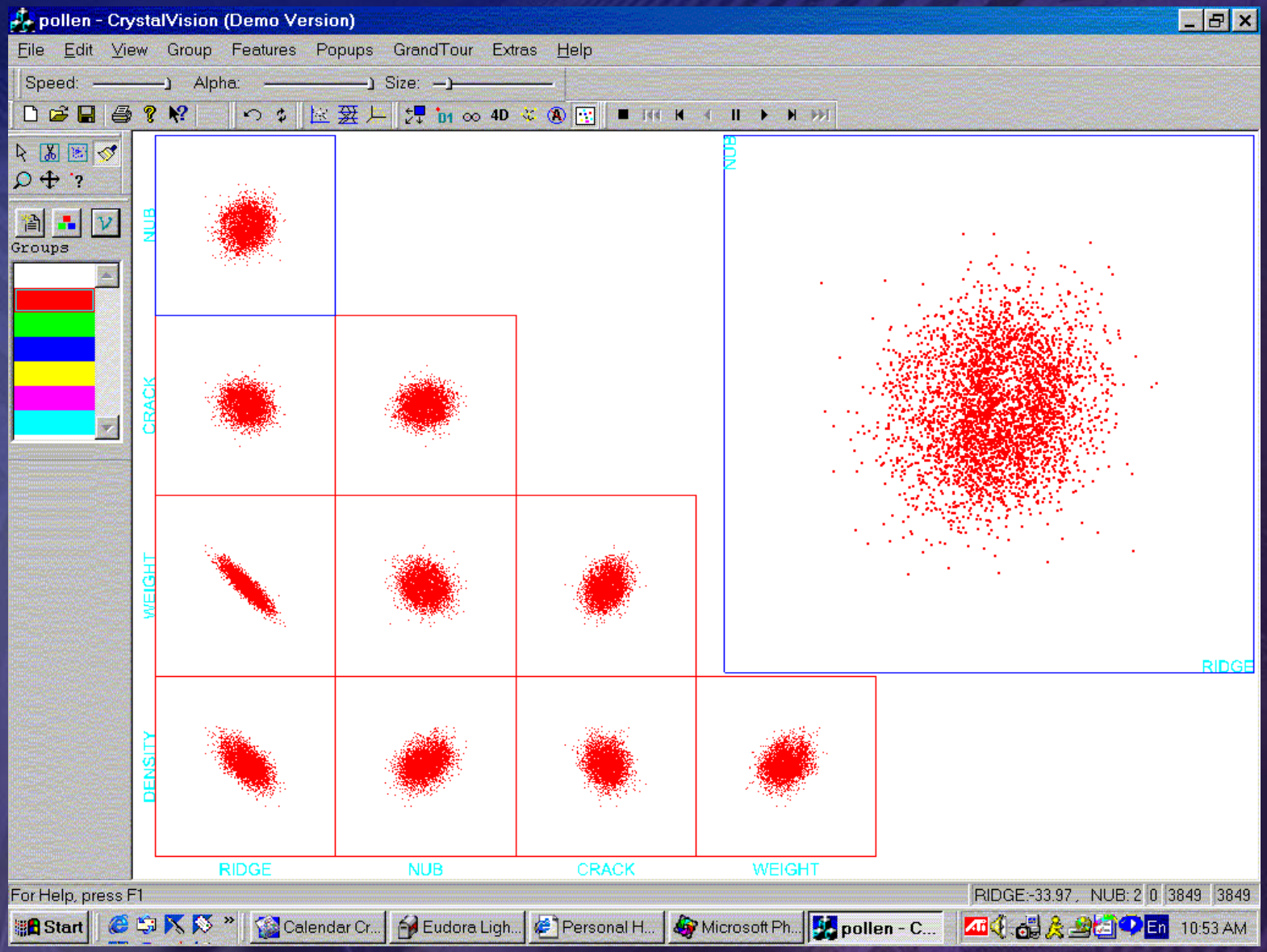
George Mason University

November 13, 2013

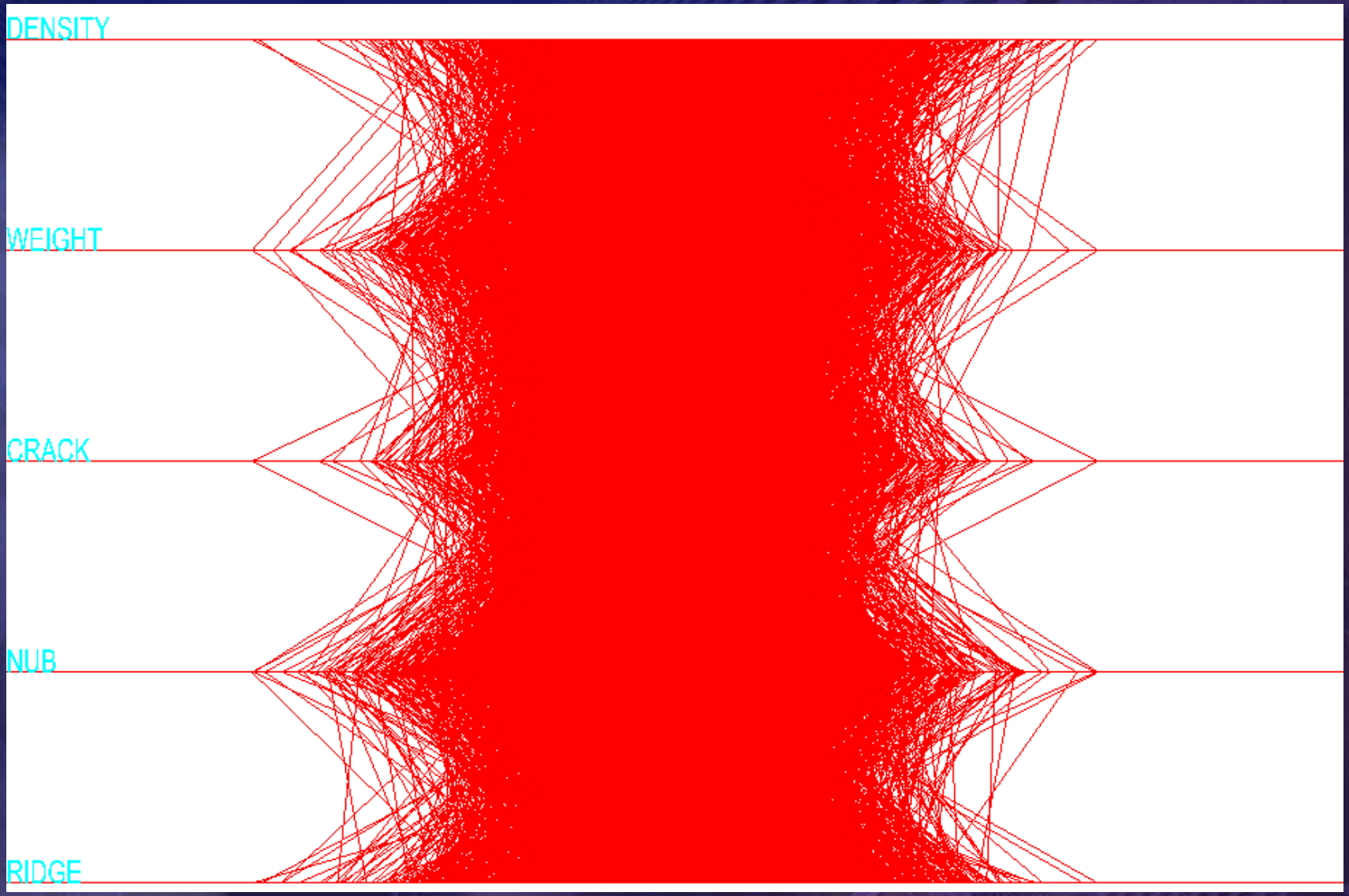
Visual Data Mining Techniques

- Multidimensional Data Visualization
  - Parallel coordinate plots
  - Grand tour
  - Saturation brushing
  - Scatterplot matrix
  - 3-D stereoscopic scatterplots
  - Density plots
  - Linked views
  - Pruning and cropping

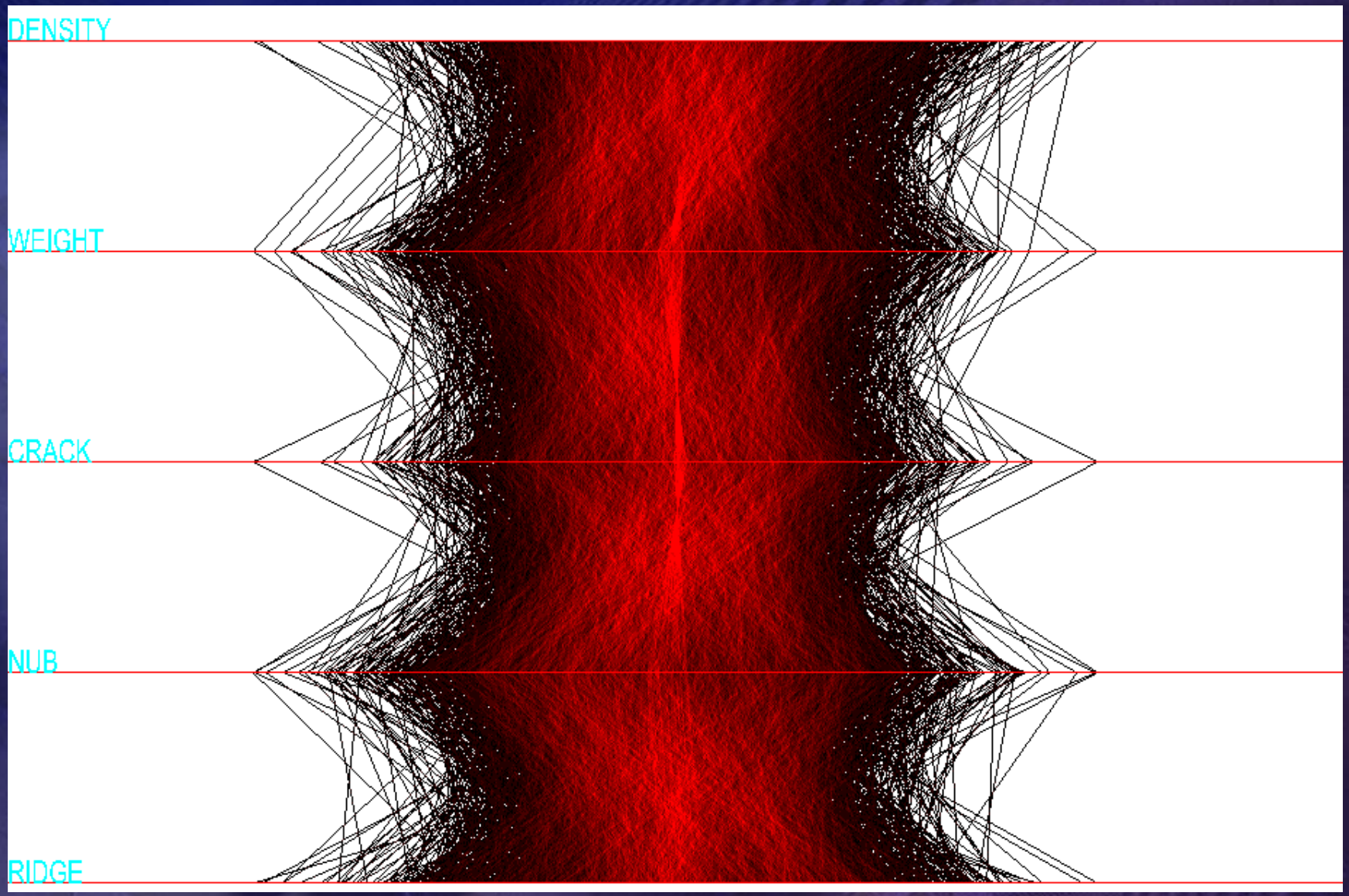
Pollen Data


pollen - CrystalVision (Demo Version)
File Edit View Group Features Popups GrandTour Extras Help
Speed: Alpha: Size:
Groups
NUB
CRACK
WEIGHT
DENSITY
RIDGE
NUB
CRACK
WEIGHT
RIDGE
For Help, press F1
RIDGE:-33.97 , NUB: 2 0 3849 3849
Start
Calendar Cr...
Eudora Ligh...
Personal H...
Microsoft Ph...
pollen - C...
En
10:53 AM

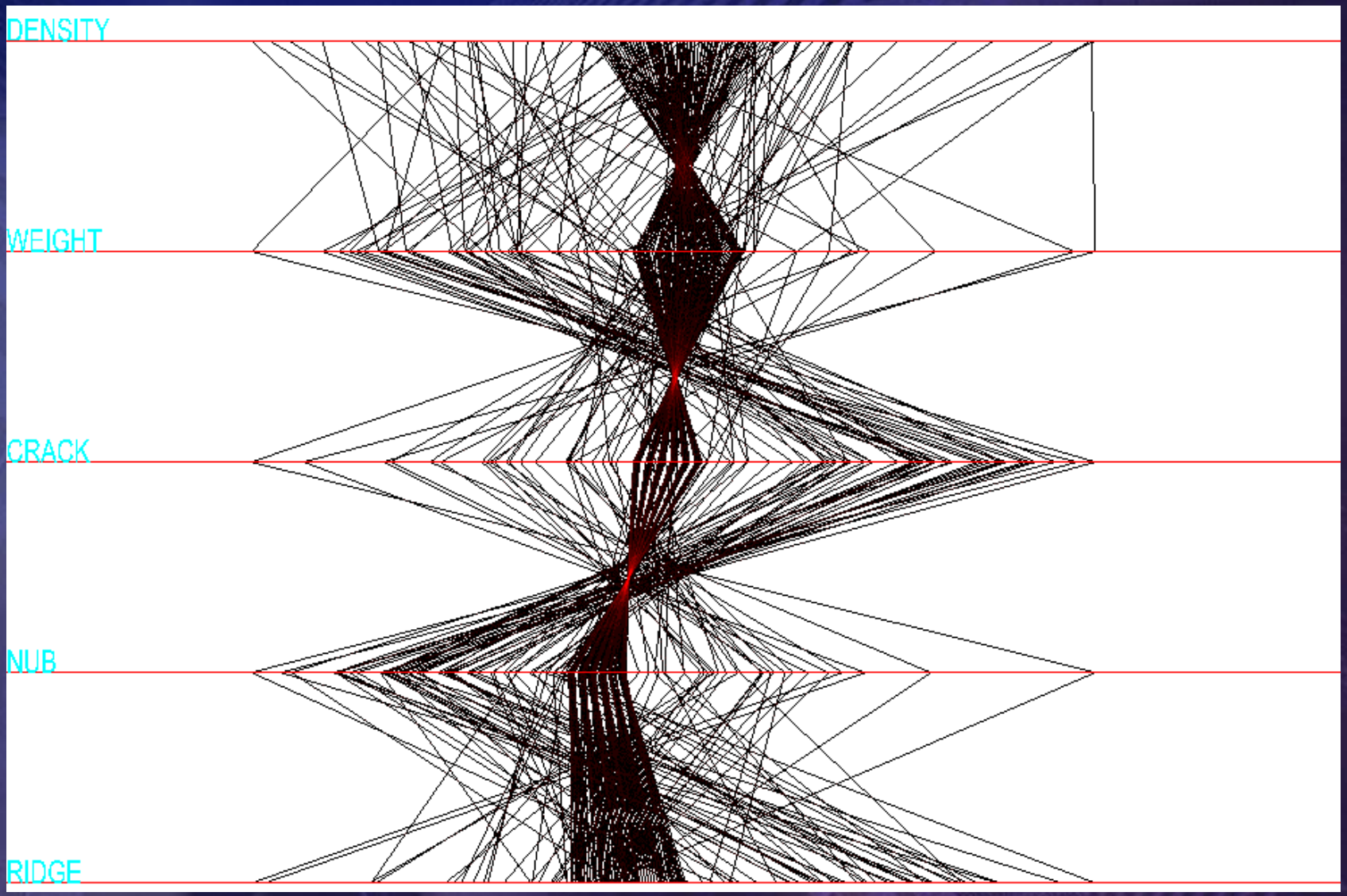
Pollen Data



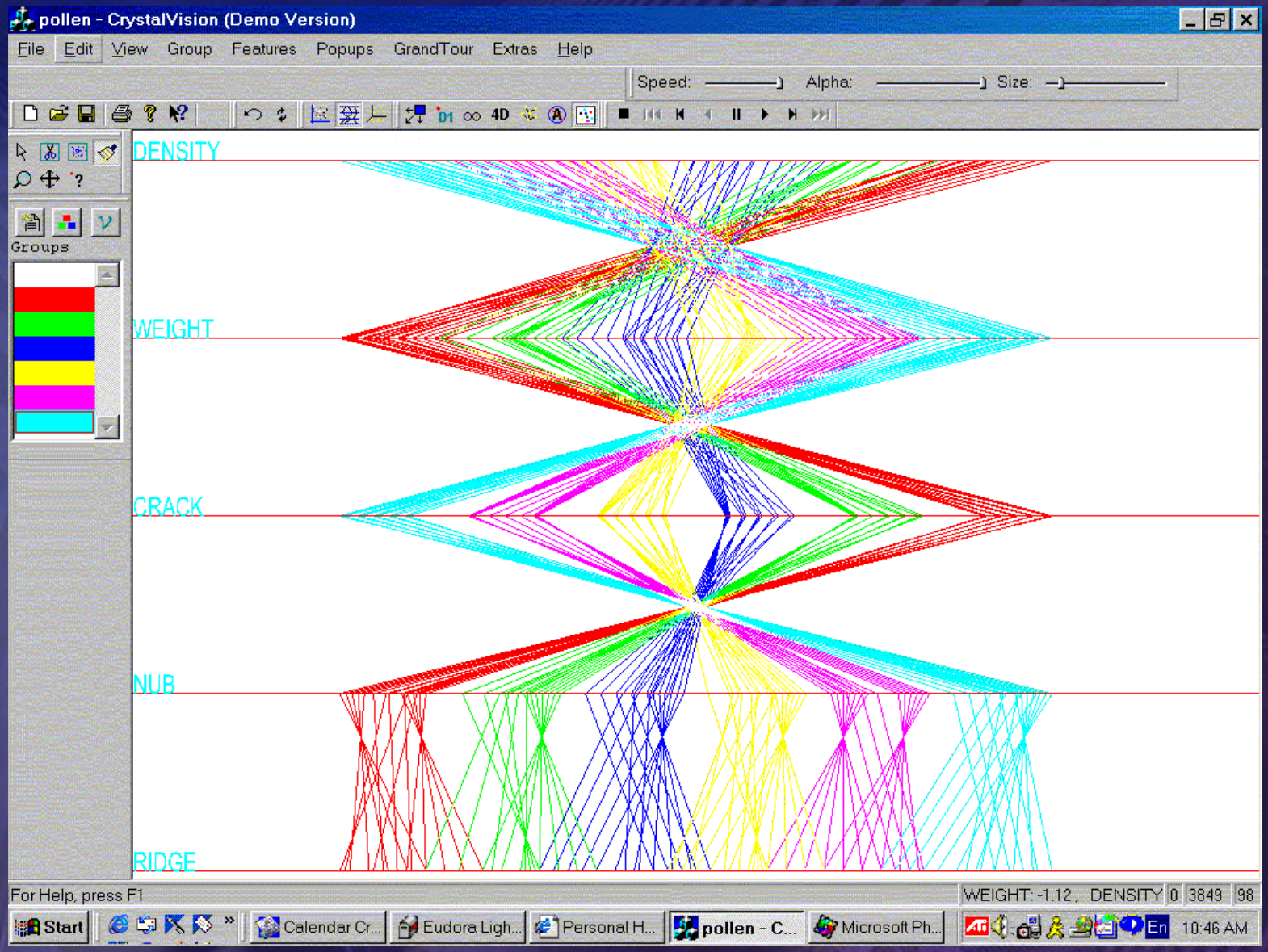
Pollen Data



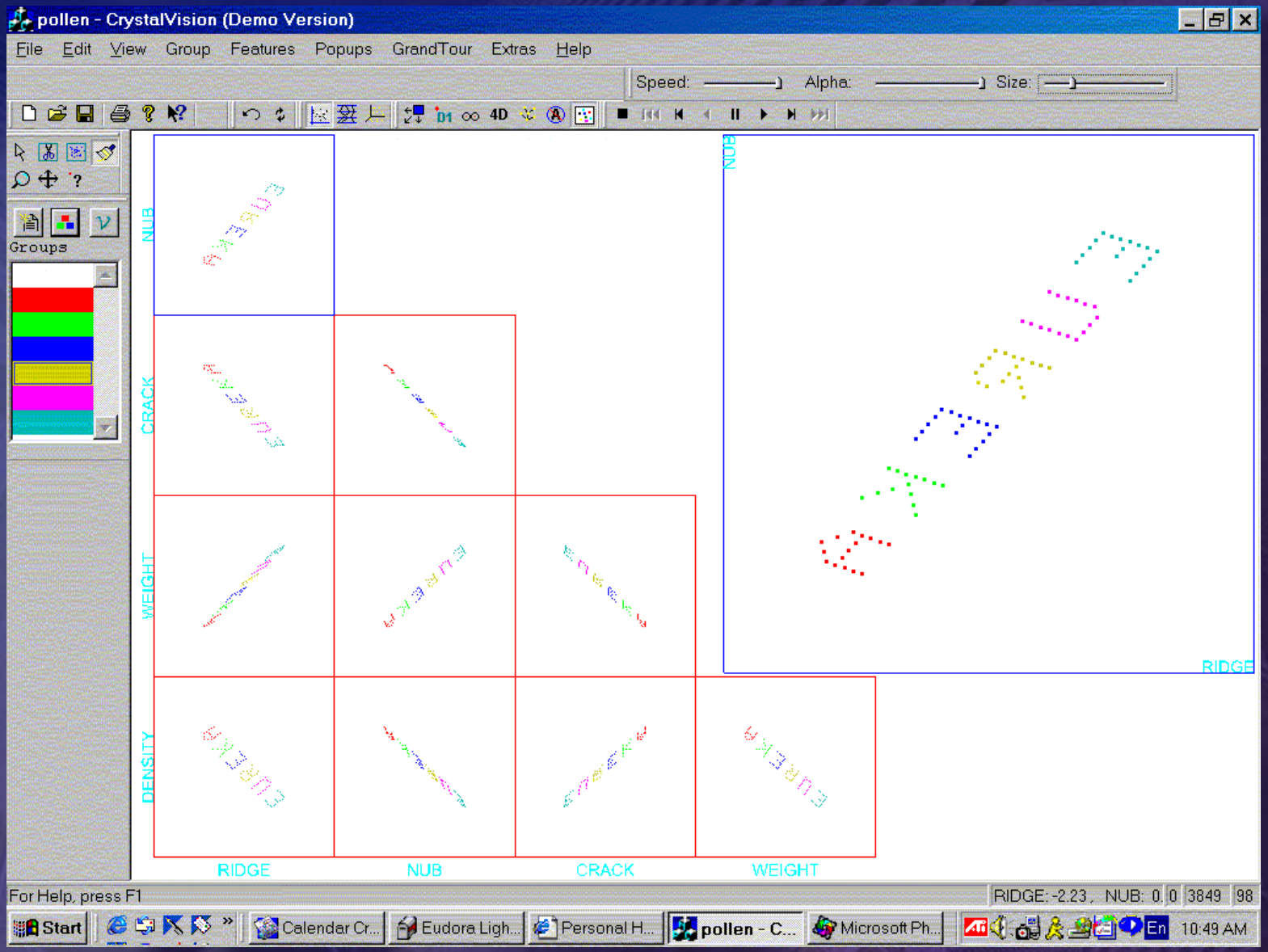
Pollen Data



Pollen Data


pollen - CrystalVision (Demo Version)
File Edit View Group Features Popups GrandTour Extras Help
Speed: Alpha: Size:
Groups
DENSITY
WEIGHT
CRACK
NUB
RIDGE
For Help, press F1
WEIGHT: -1.12 , DENSITY 0 3849 98
Start
Calendar Cr...
Eudora Ligh...
Personal H...
pollen - C...
Microsoft Ph...
10:46 AM

Pollen Data


pollen - CrystalVision (Demo Version)
File Edit View Group Features Popups GrandTour Extras Help
Speed: Alpha: Size:
Groups
NUB
CRACK
WEIGHT
DENSITY
RIDGE
NUB
CRACK
WEIGHT
RIDGE
For Help, press F1
RIDGE: -2.23, NUB: 0.0 3849 98
Start
Calendar Cr...
Eudora Ligh...
Personal H...
pollen - C...
Microsoft Ph...
En
10:49 AM

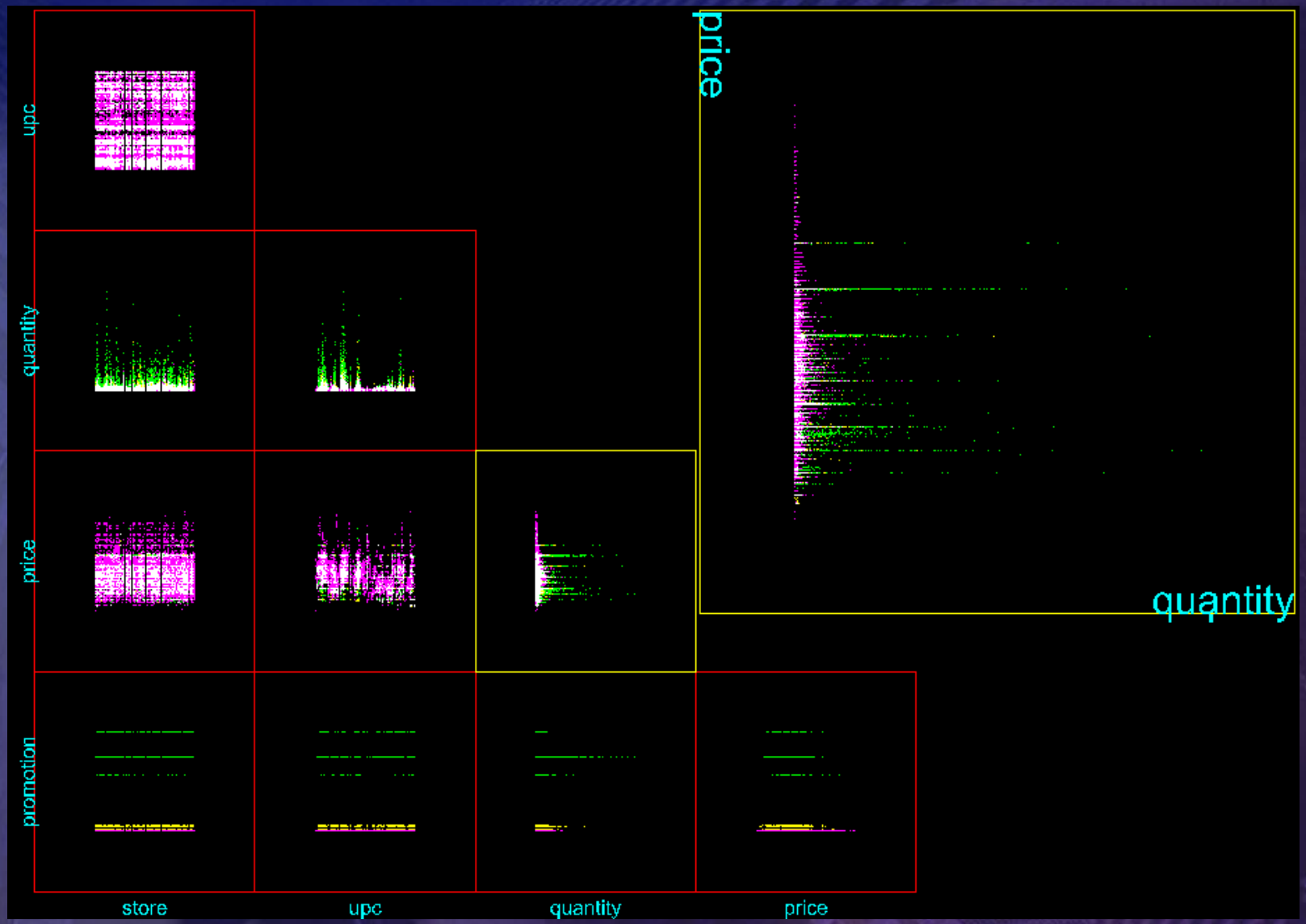
Scanner Data for Breakfast Cereals

- 5.5 gigabytes of scanner data in relational database
  - Price, sales volume, promotion, store, chain, PSU, UPC
  - Work done at BLS

Scanner Data for Breakfast Cereals

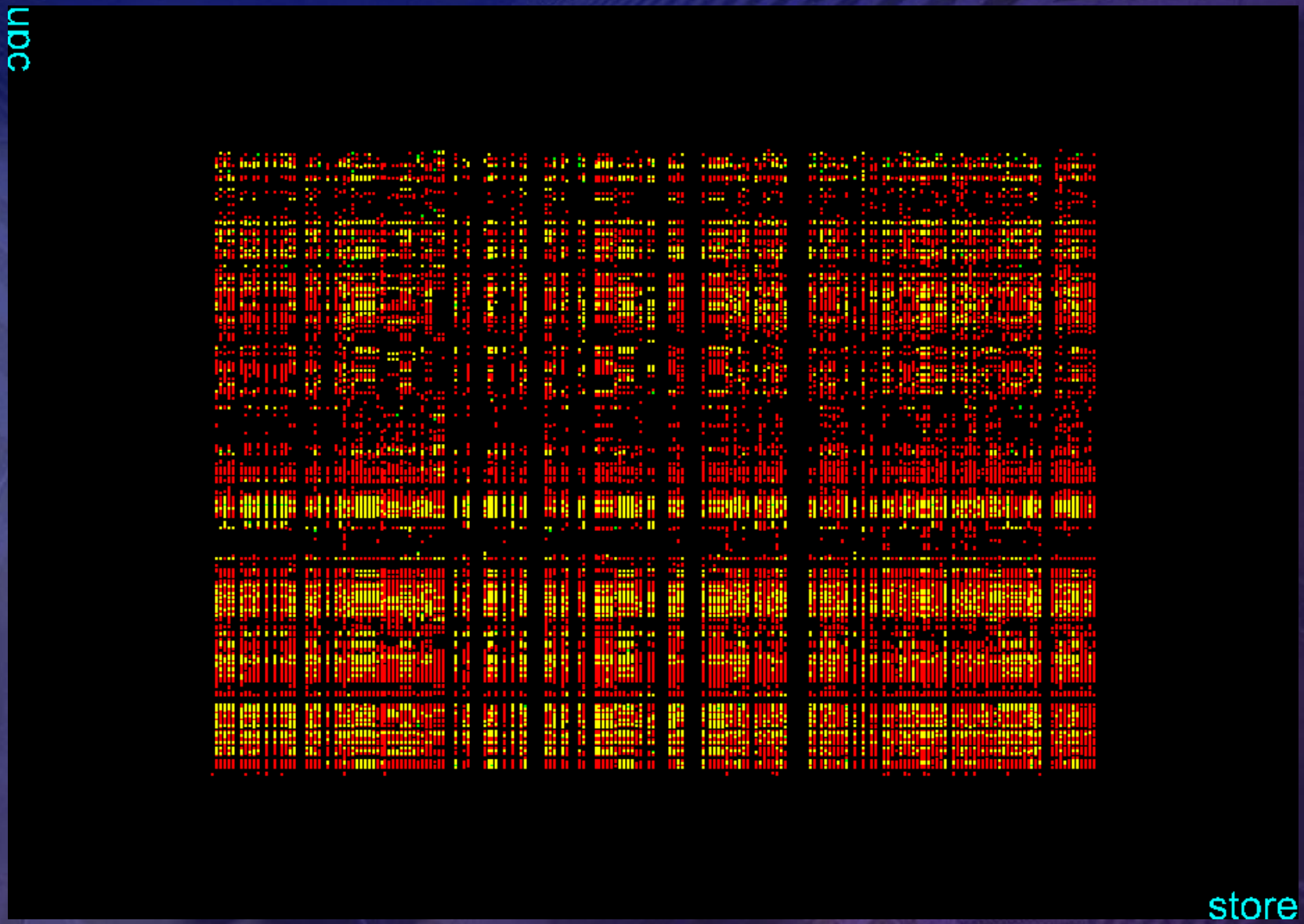
Phase 1

Scanner Data for Breakfast Cereals



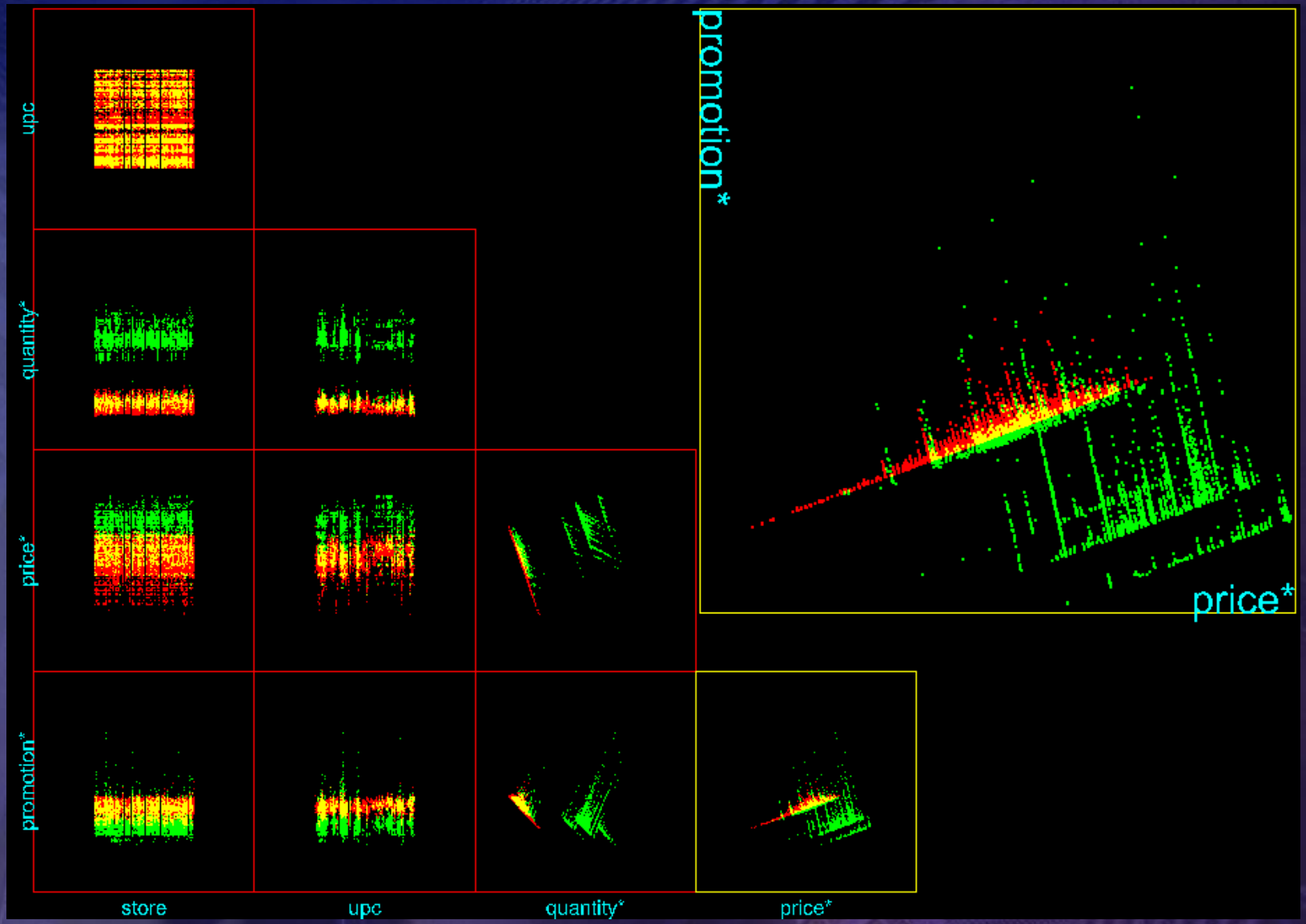
Promotion has huge impact on sales volume

Scanner Data for Breakfast Cereals



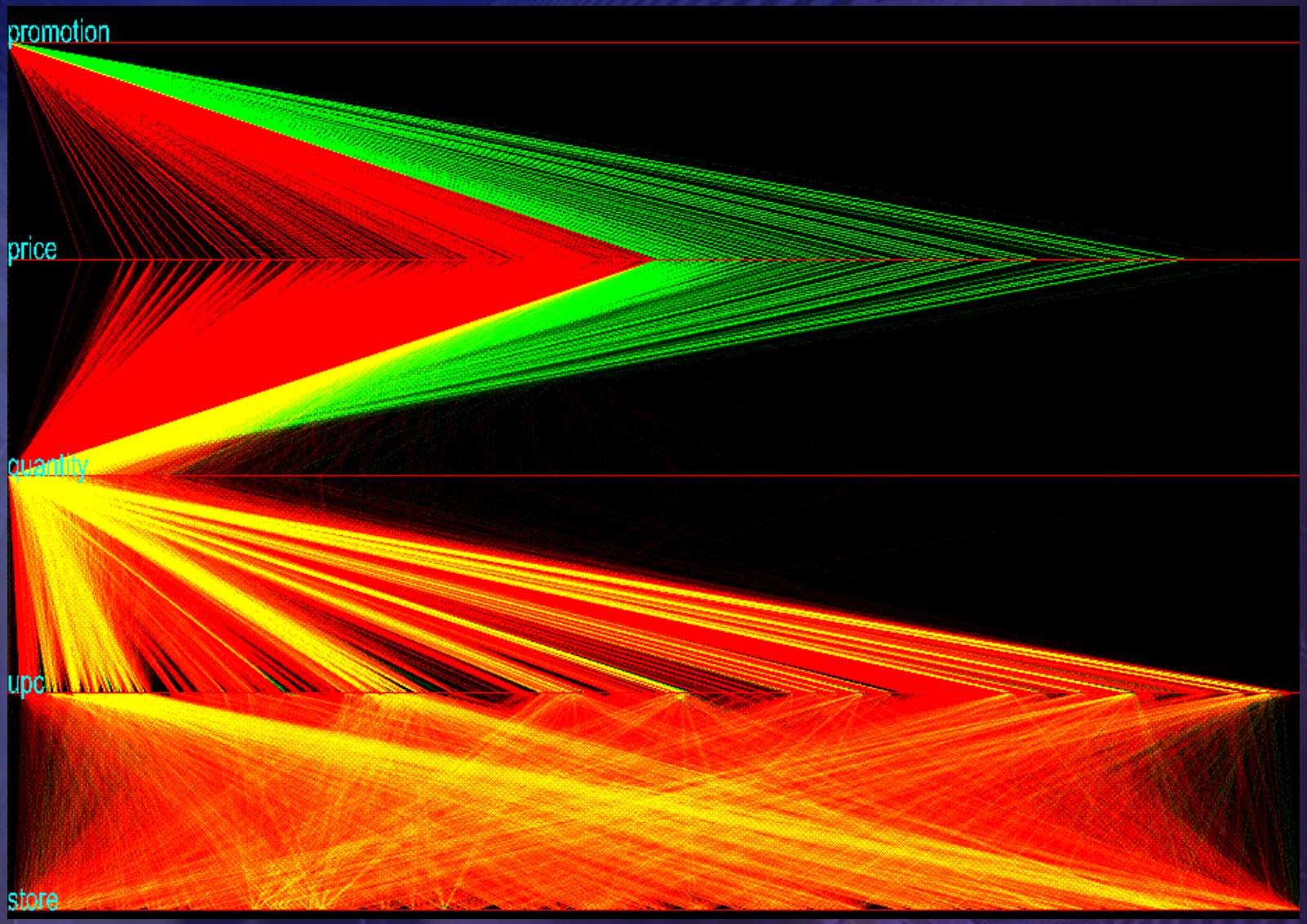
Stores not randomized

Scanner Data for Breakfast Cereals



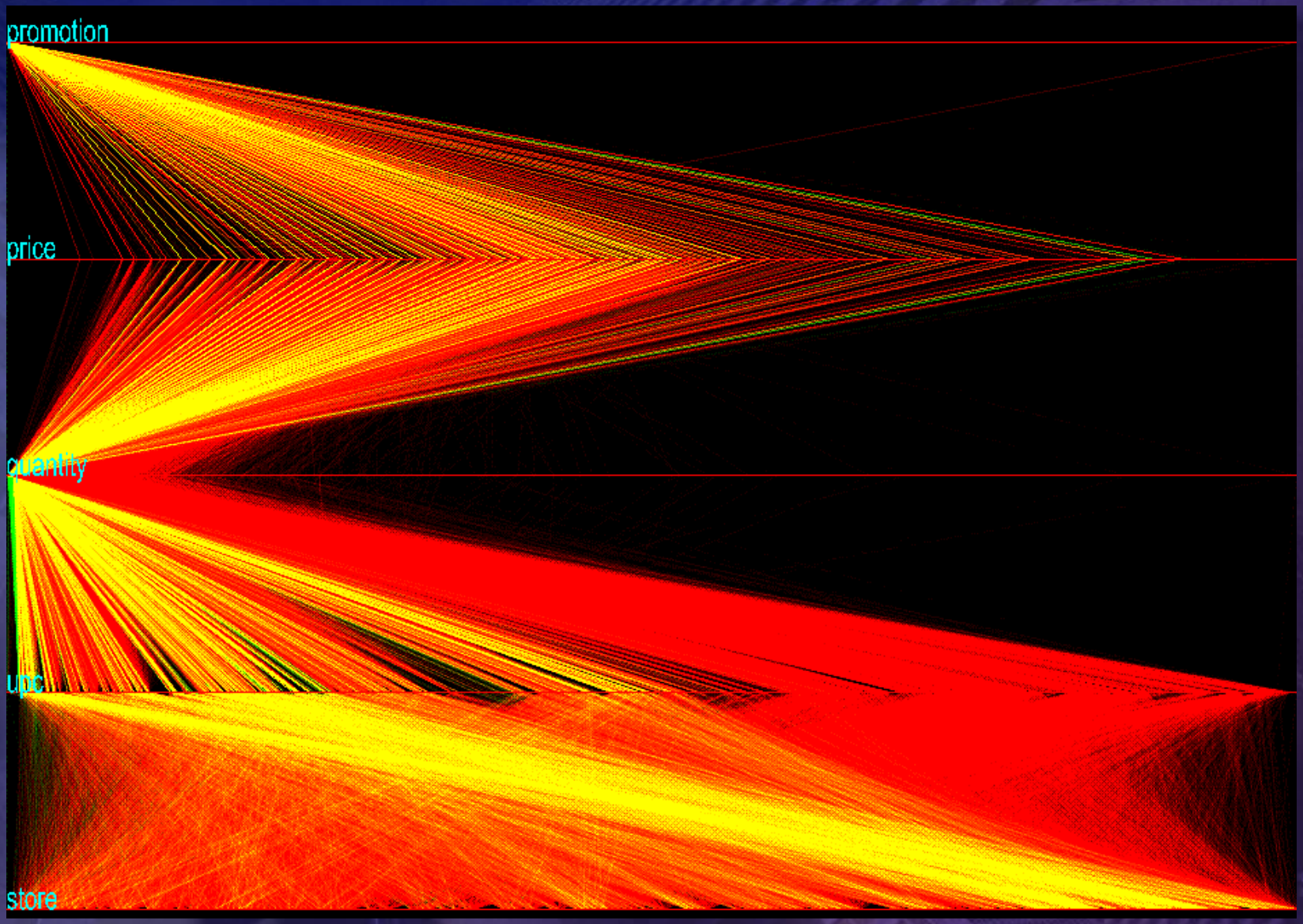
Aggressive promotion pays

Scanner Data for Breakfast Cereals



High Priced Cereals in Preferred Markets

Scanner Data for Breakfast Cereals

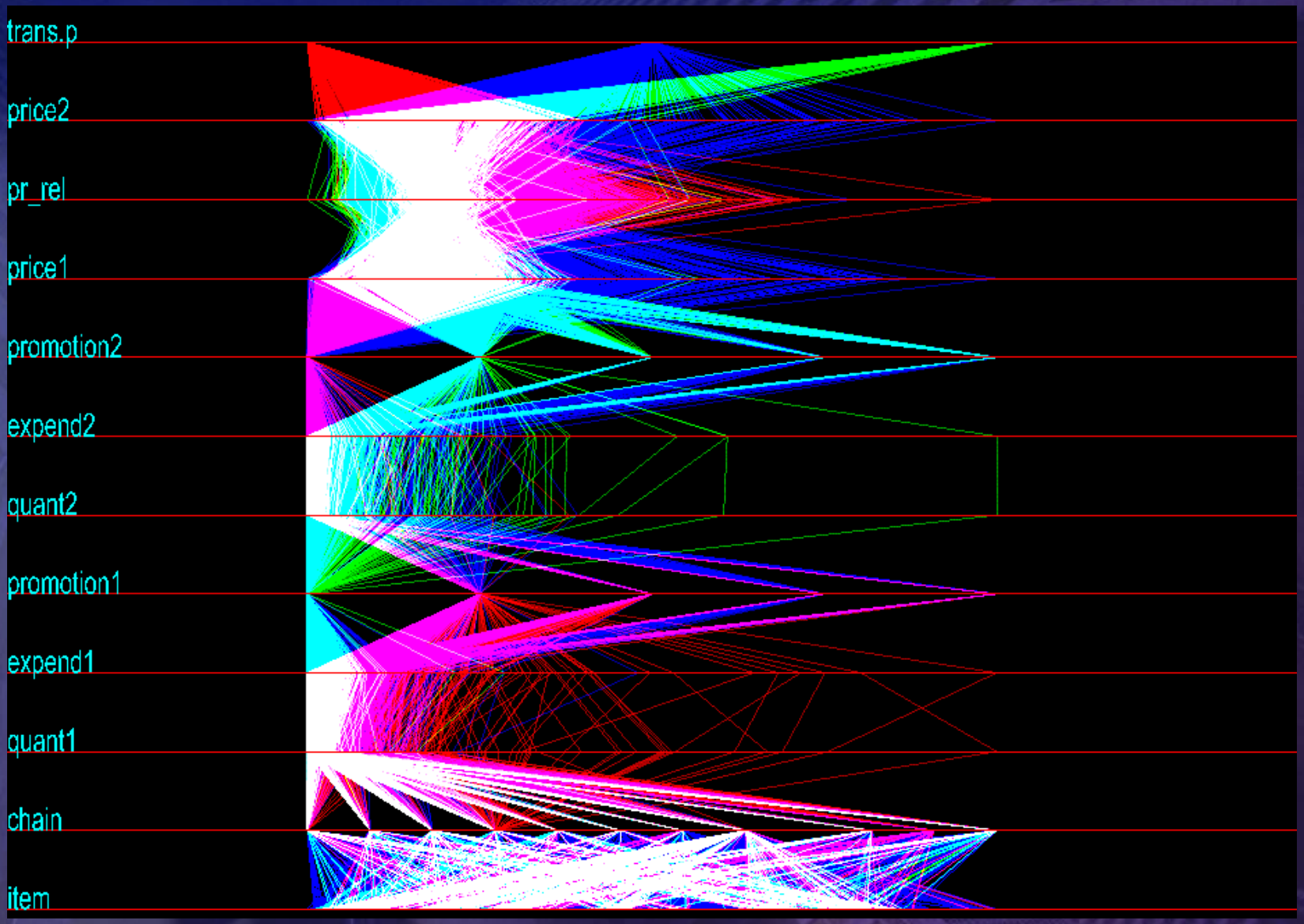


Low Volume Cereals in Similar Markets

Scanner Data for Breakfast Cereals

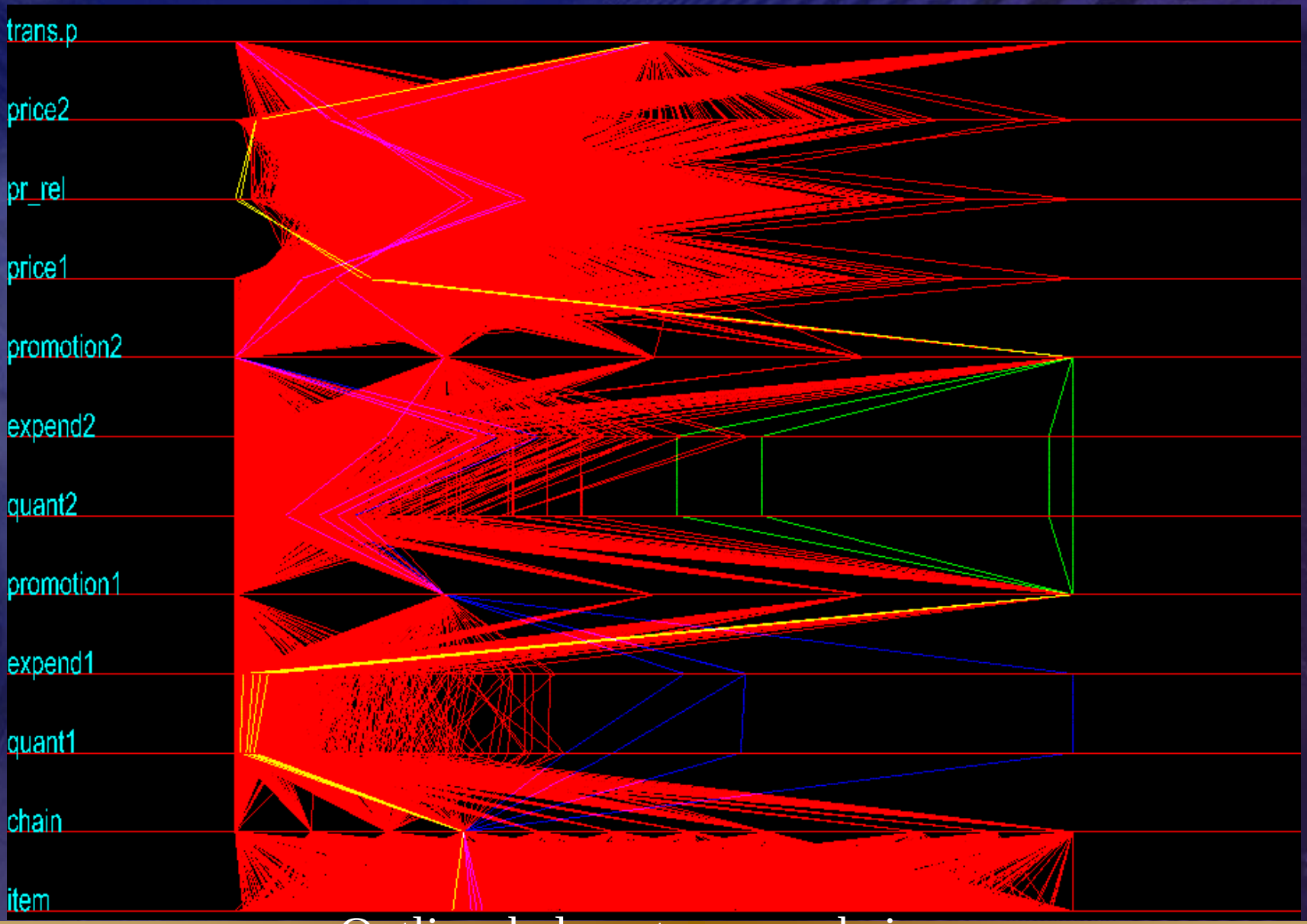
Phase 2

Scanner Data for Breakfast Cereals



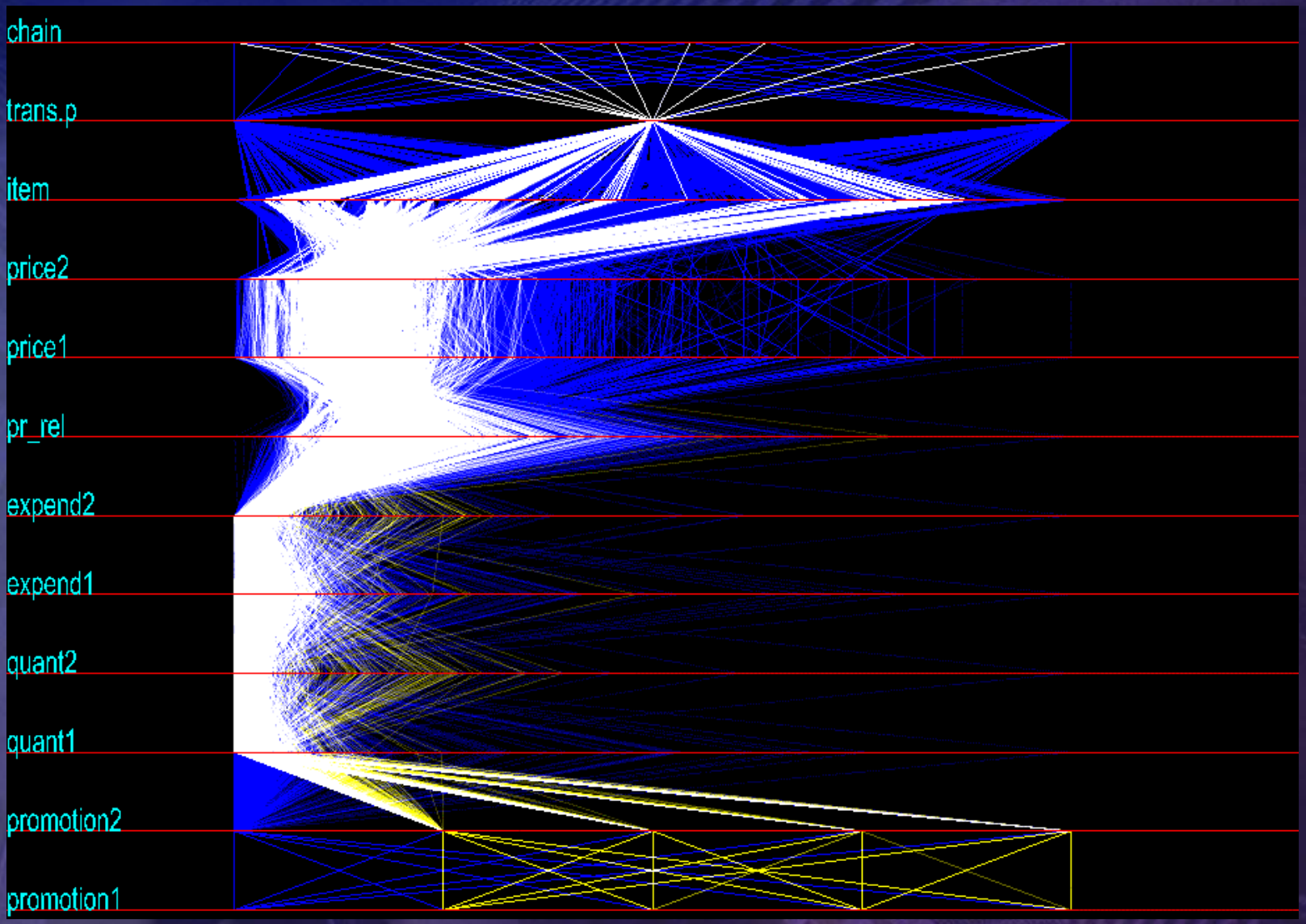
Promotion of Cereals Creates Outliers

Scanner Data for Breakfast Cereals



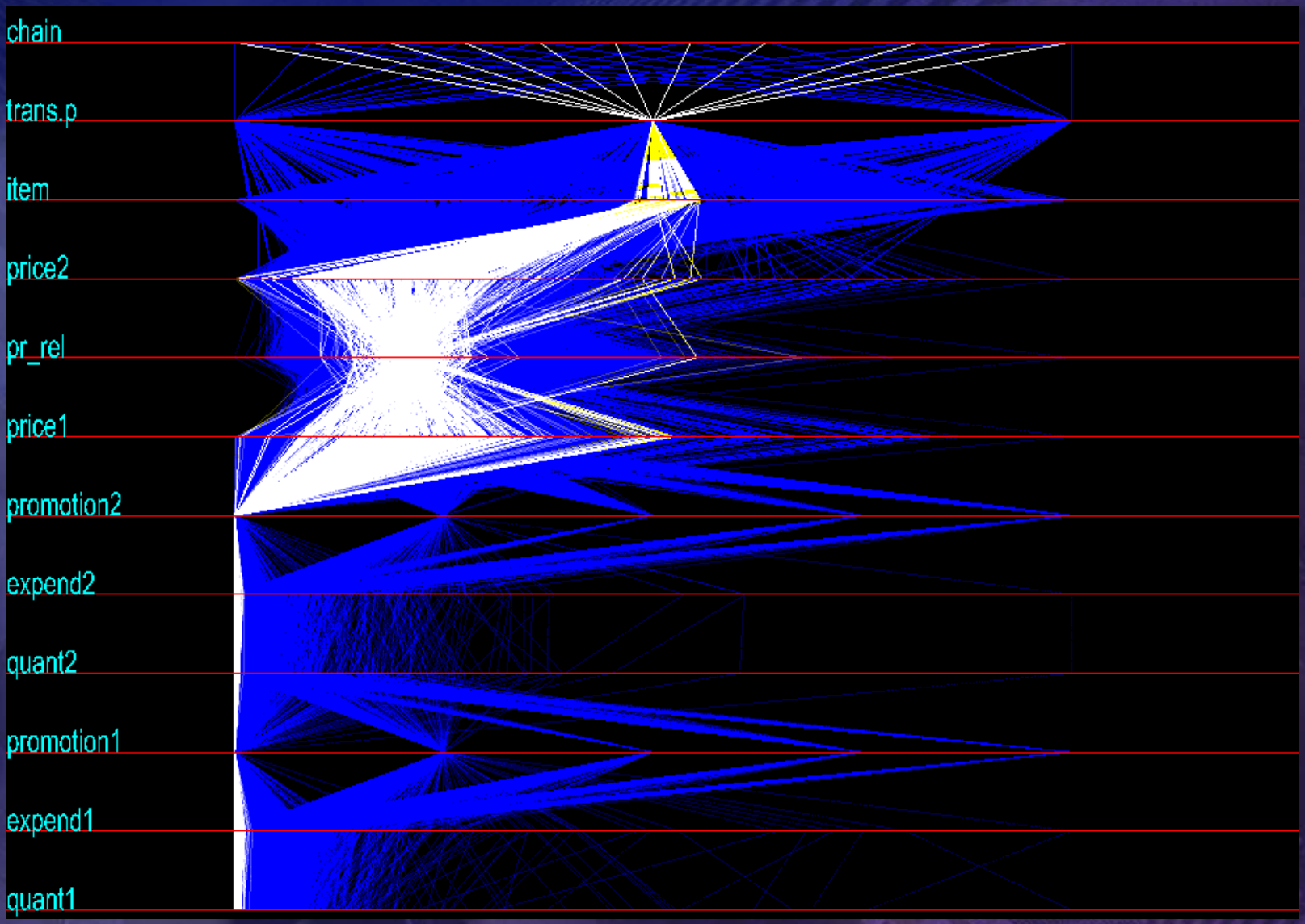
Outliers belong to same chain

Scanner Data for Breakfast Cereals



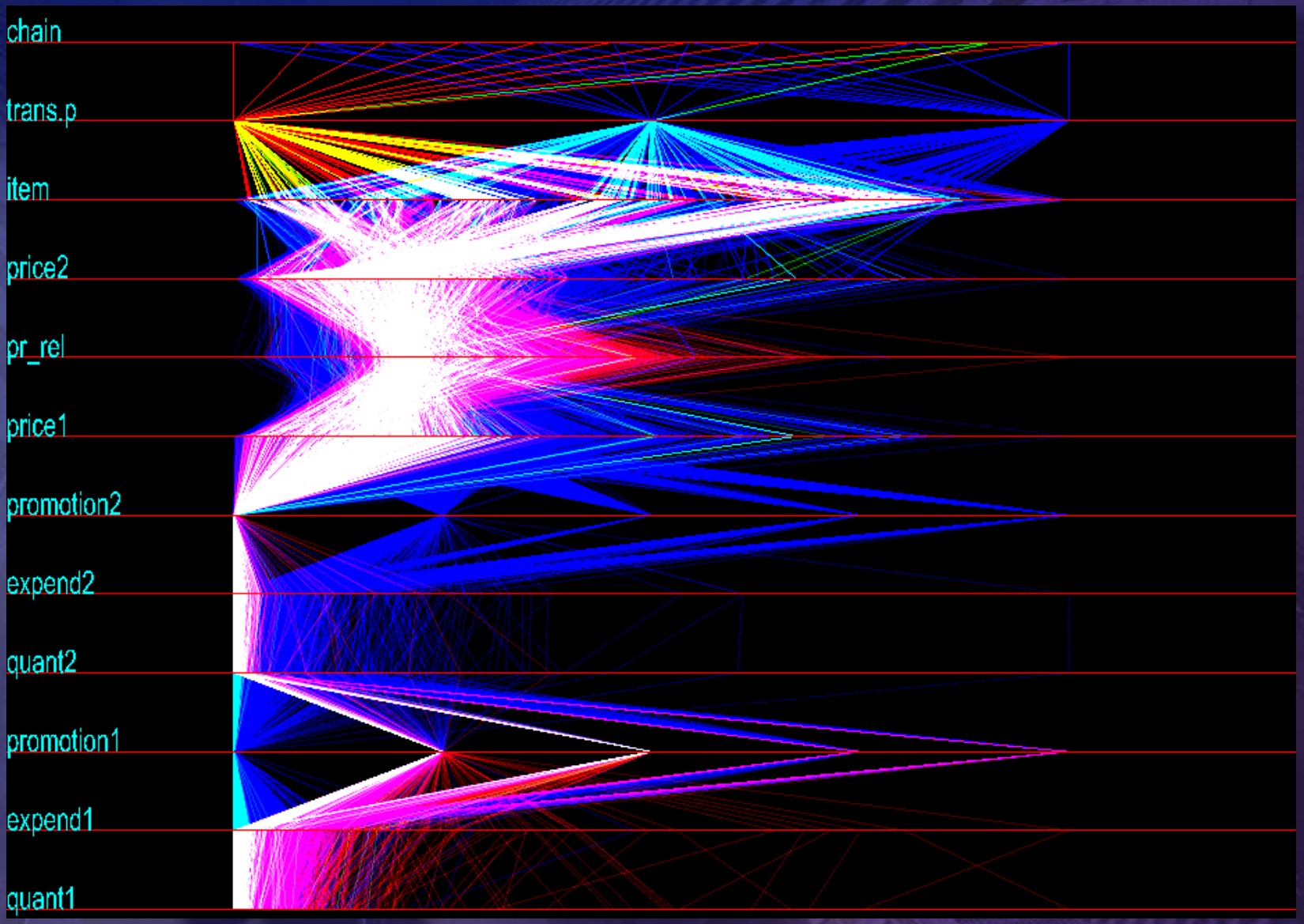
Promotion both years

Scanner Data for Breakfast Cereals



Range of items with no promotion

Scanner Data for Breakfast Cereals

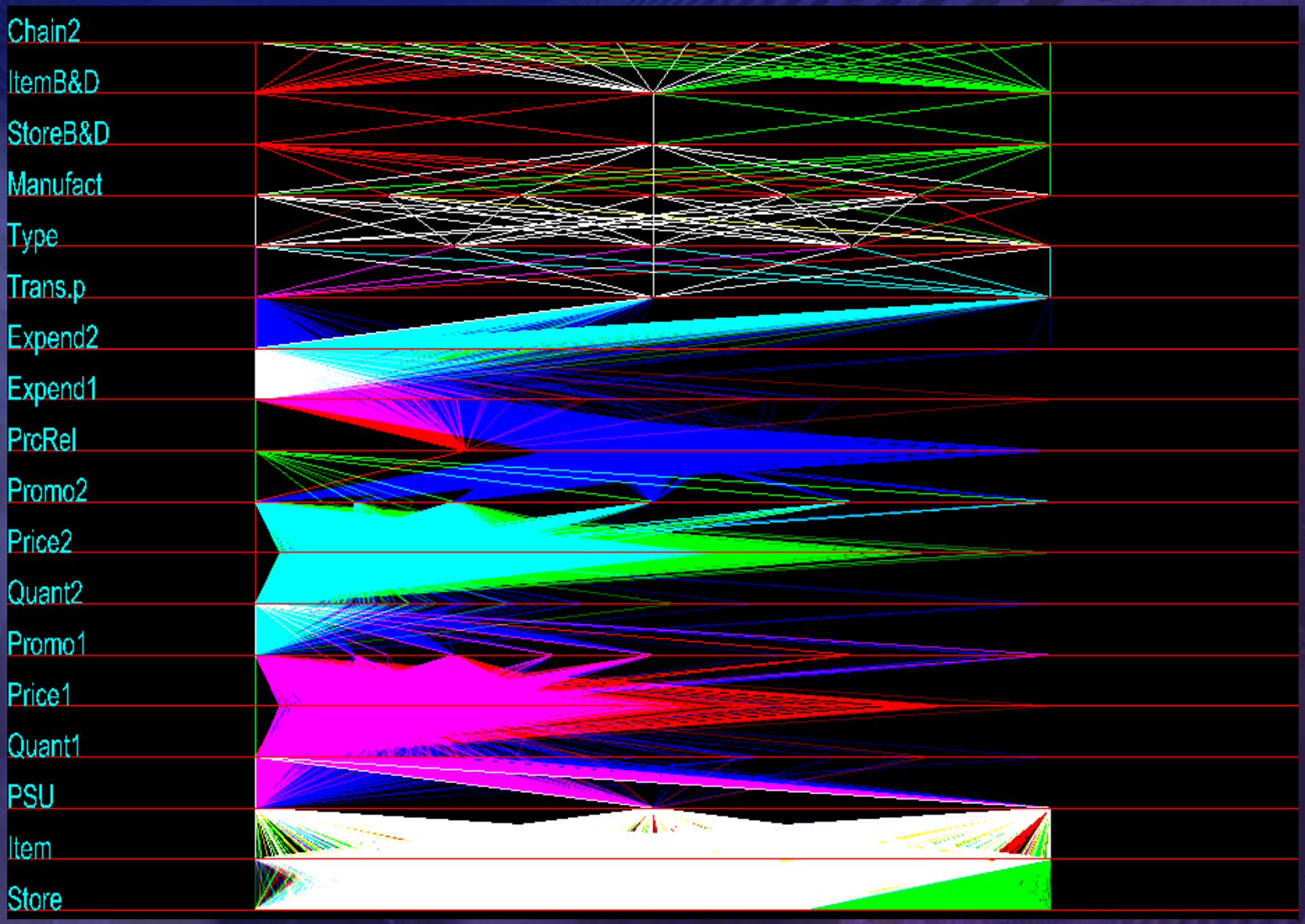


One chain ceased promotions

Scanner Data for Breakfast Cereals

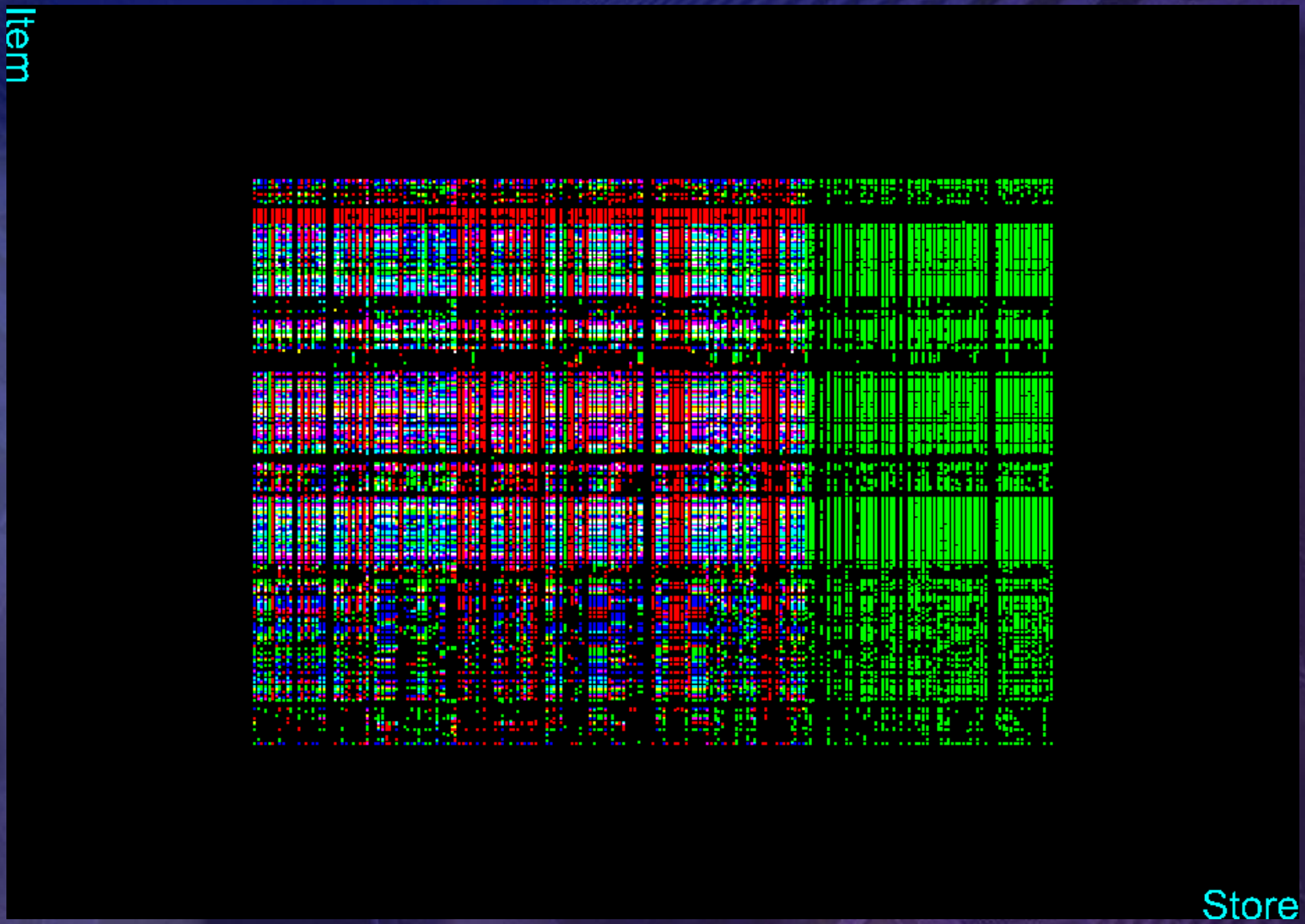
Phase 3

Scanner Data for Breakfast Cereals



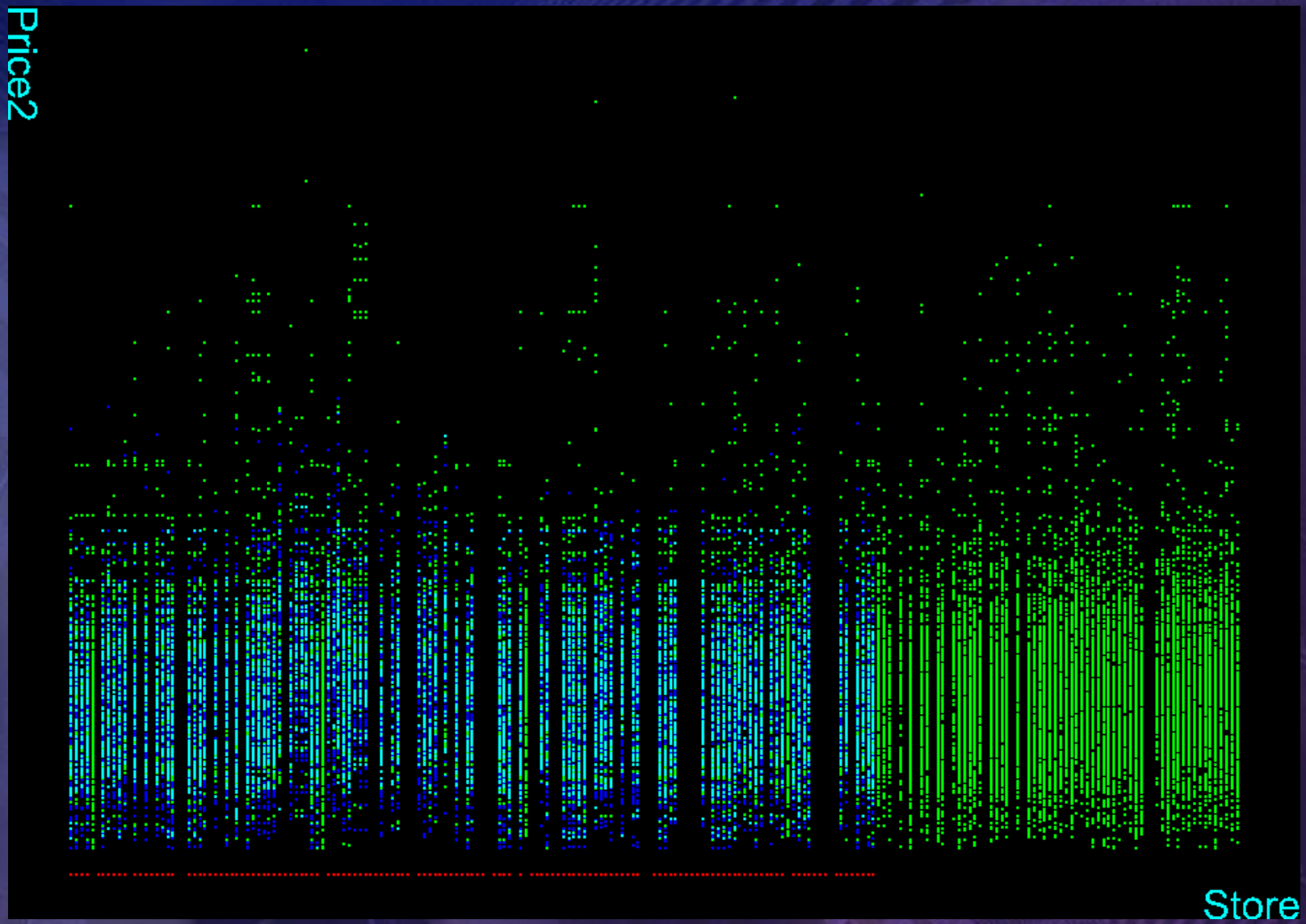
Churning comes from both new items and new stores

Scanner Data for Breakfast Cereals



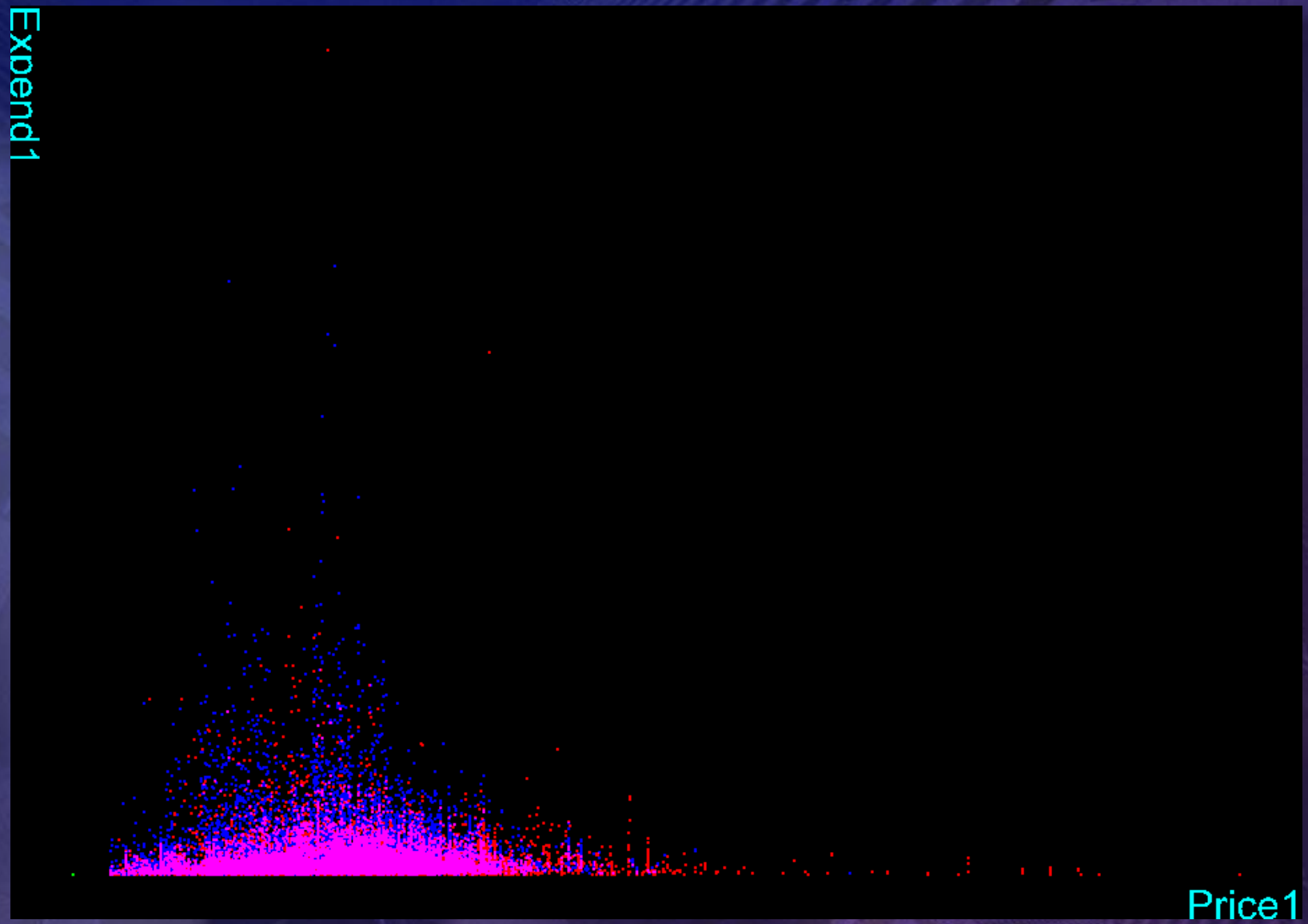
Churning Effects: Red: PR=0, Blue: PR>0, Green PR=infinity

Scanner Data for Breakfast Cereals



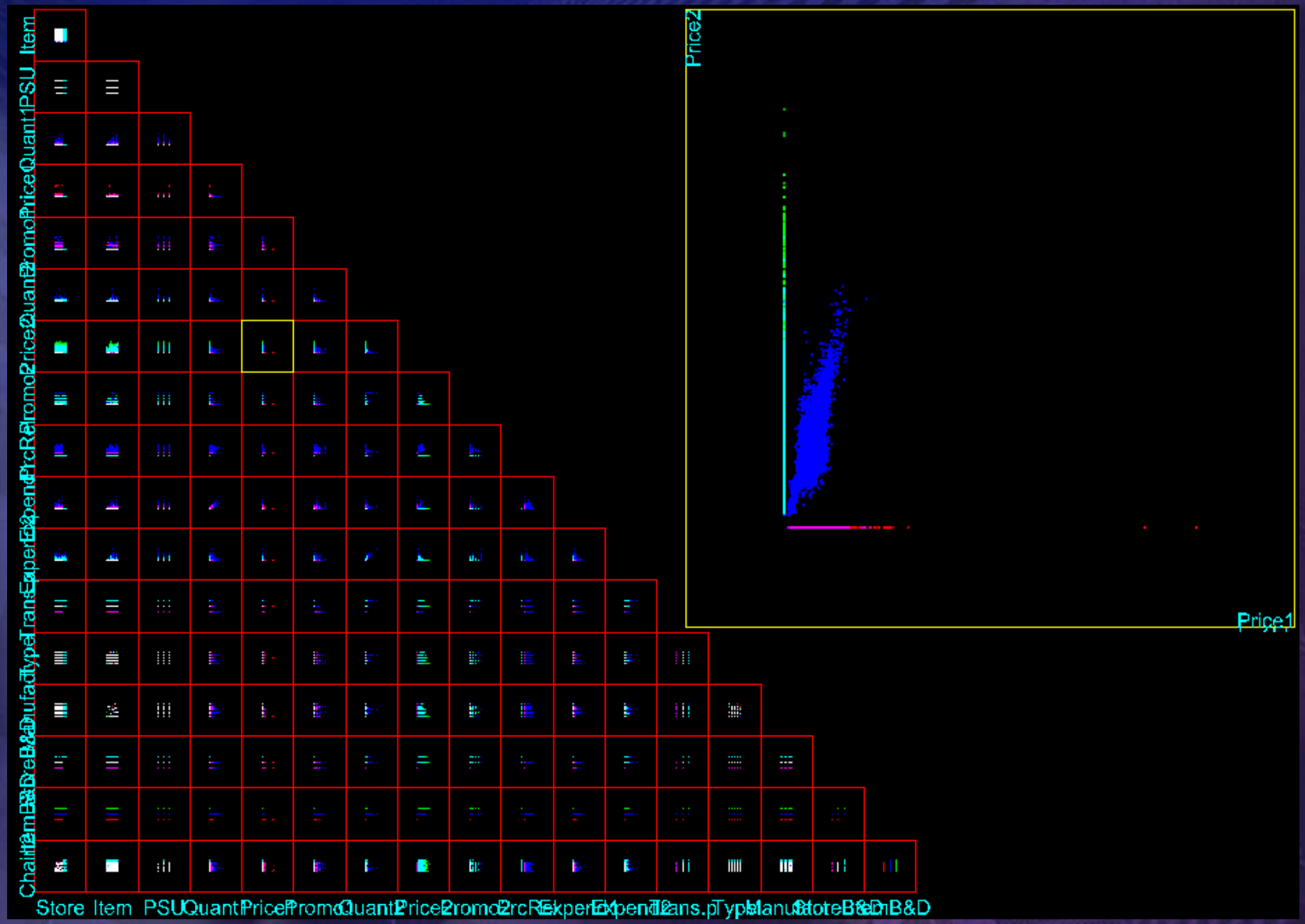
New items tend to have higher prices

Scanner Data for Breakfast Cereals



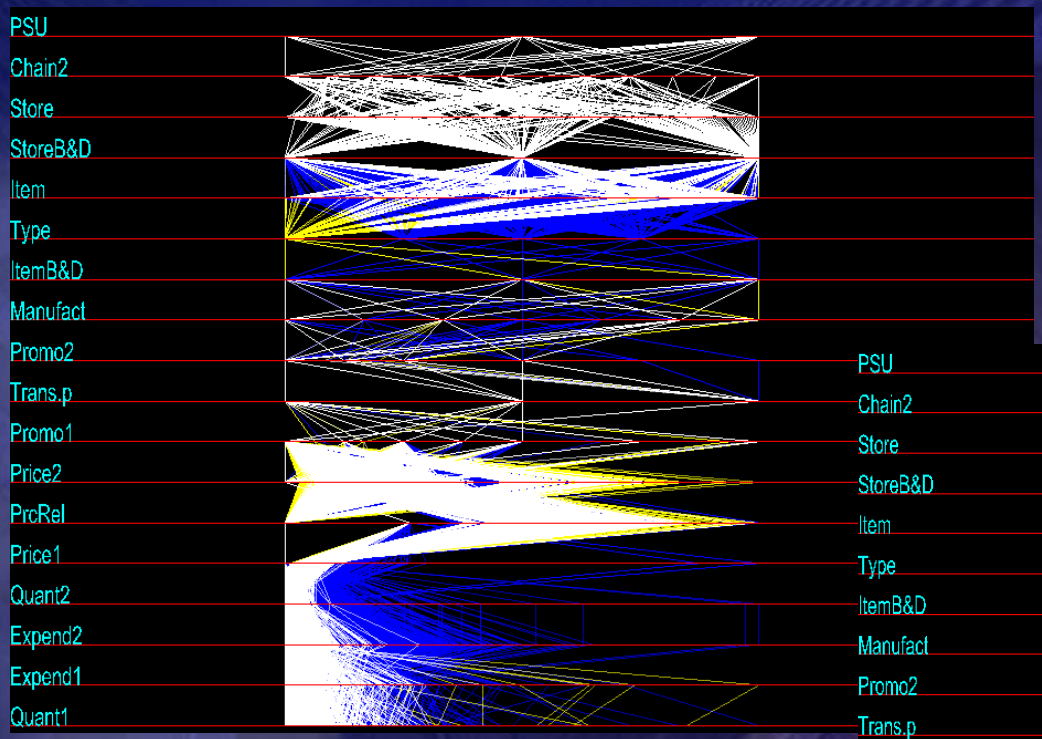
Many discontinued items have high expenditures

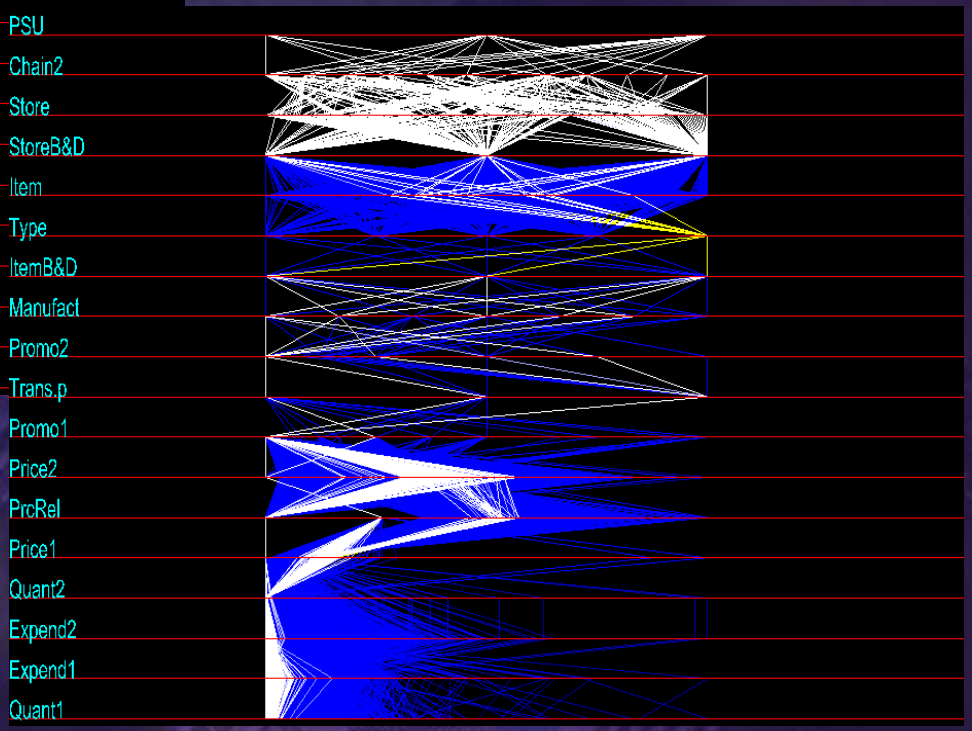
Scanner Data for Breakfast Cereals



Outlier due to price coding error

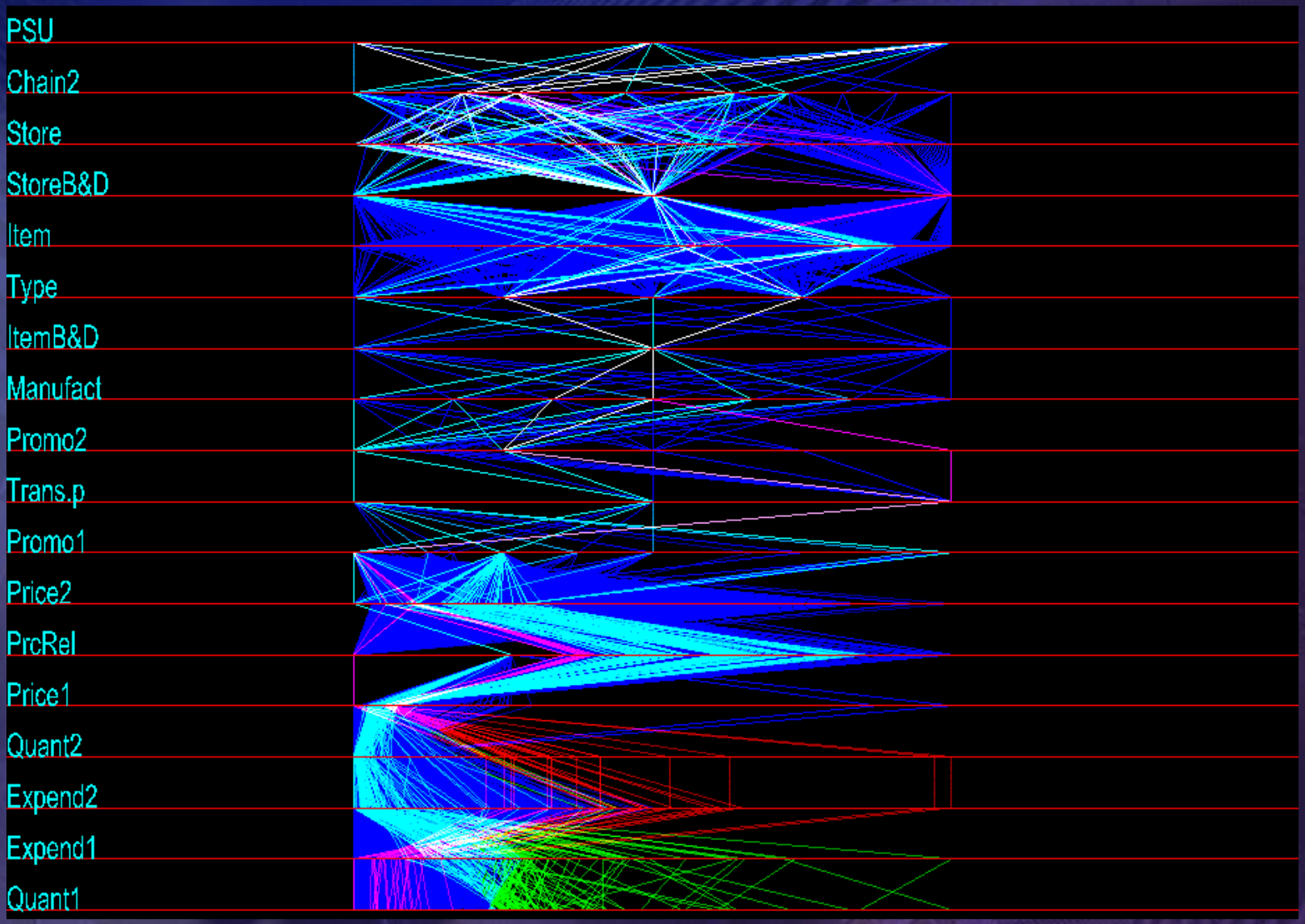
Scanner Data for Breakfast Cereals





Effects of Cereal Types

Scanner Data for Breakfast Cereals



Quantity Effects

Visualizing Transaction Data

Contact Information:

Edward J. Wegman
George Mason University, MS 3F3
Center for Computational Data Sciences
Fairfax, VA 22030-4444

Email: ewegman@gmail.com