Dense Pixel Displays

Forest Gagnon and Jack Israel

What is a dense pixel display?

- A dense pixel display is a method of visualizing every single data value in a data set at the same time
- A dataset with n records and q dimensions has n x q data values. A dataset as small as 300 x 8 already has 2400 data values. This number gets big fast, but the dense pixel display method accommodates great volume by:
  - Wasting no space: each data value is confined to a single pixel, and no white space is necessary
  - Ensuring no overlapping occurs: each data value occupies its own space; there is no danger that one data value will obscure another
- Each dimension uses the same color scale, with the minimum data value in the dimension being black and the maximum being some bright color
- Allow for visualization of more dimensions that mosaic plots
- Allow for ‘cleaner’ visualization than radar charts

Example: Call Detail Data

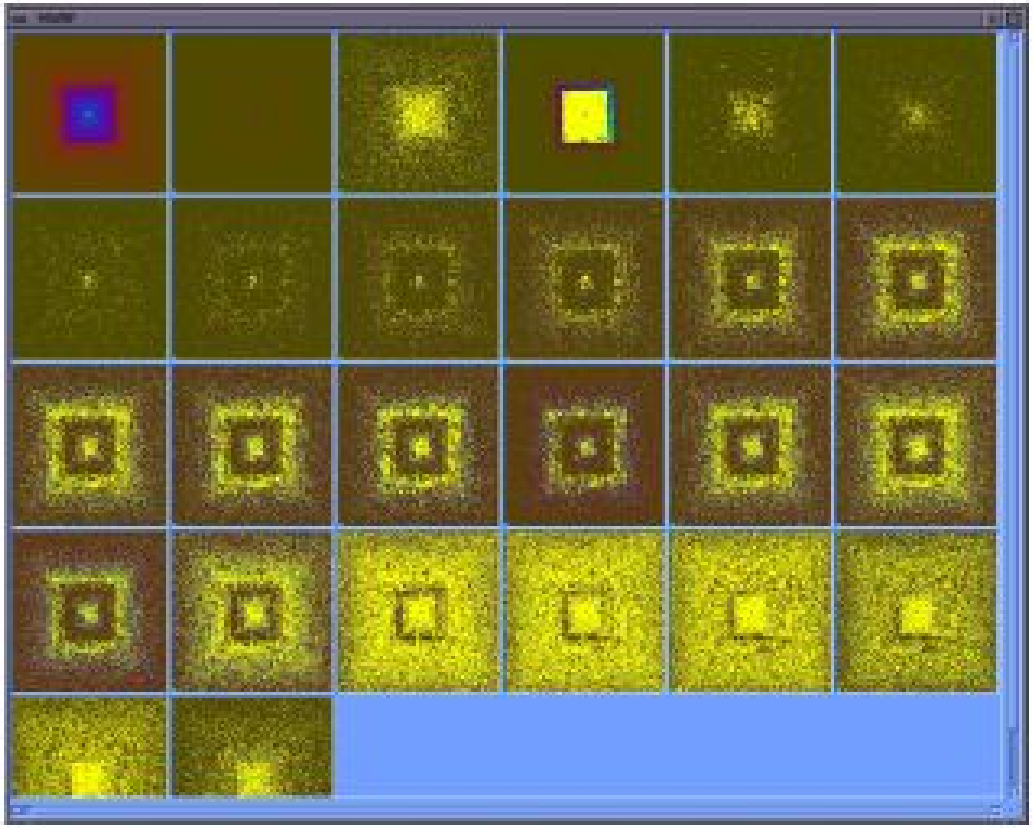


Figure 5: Query-dependent Visualization of Call Detail Data

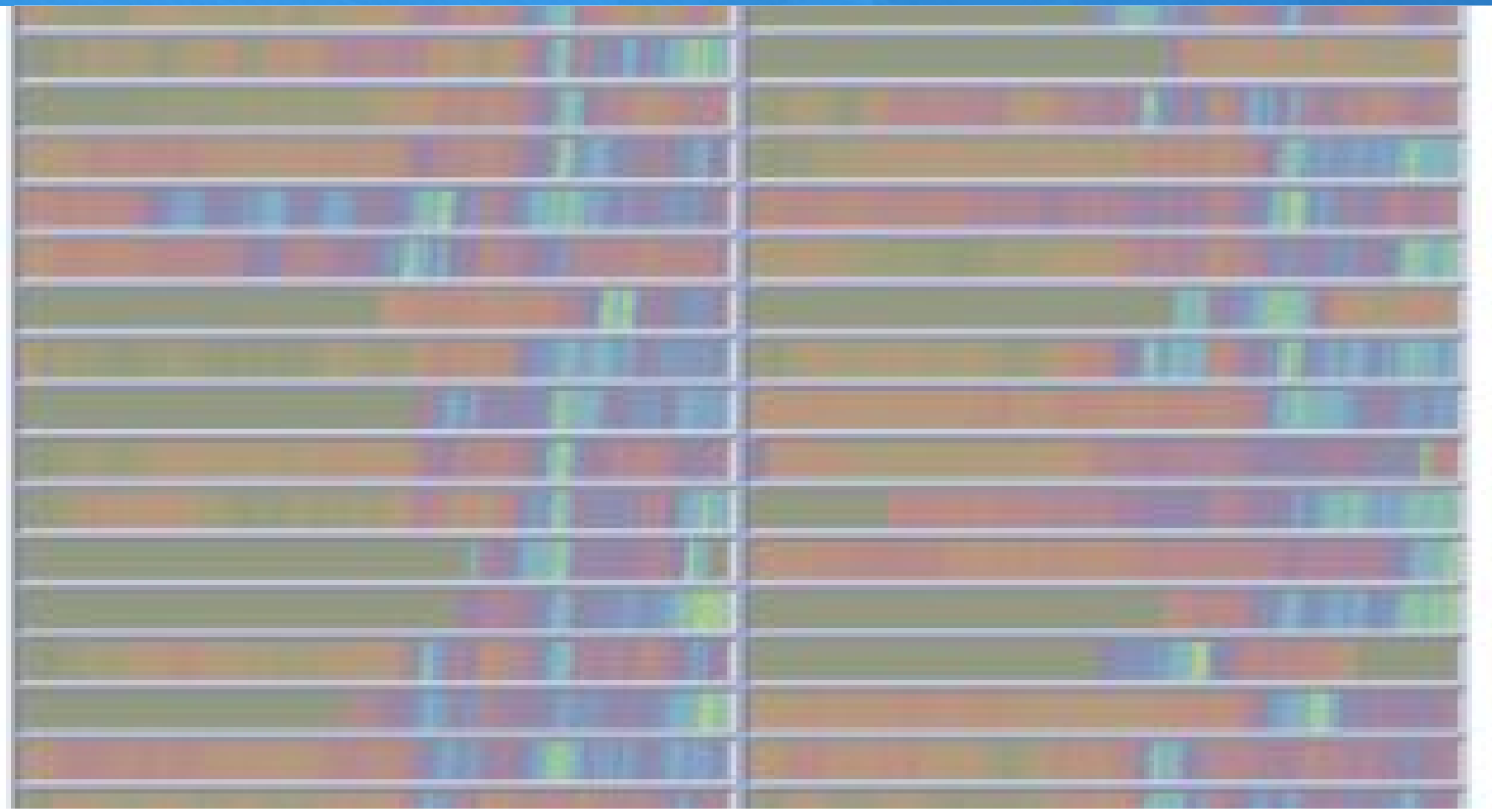
First Implementation

- There are 22,500 records and 3 variables: 67,500 data values
- r is a random floating point number between 1 and 2
- Column 1: $[1*r_1, 2*r_2, 3*r_3, 4*r_4, \ldots, 22500*r_{22500}]$
- Column 2: $[1*r_{22501}, 2*r_{22502}, 3*r_{22503}, 4*r_{22504}, \ldots, 22500*r_{45000}]$
- f is a random integer between 1 and (2*22,500) (45,000 chosen because it is max possible value of columns 1 and 2)
- Column 3: $[f_1, f_2, f_3, f_4, \ldots, f_{22500}]$

Separated Groupings of Values Dilute the Display

- Column 4 added: [-22,500*r_{45001}, -22499*r_{45002}, -22498*r_{45003}, … , 0*r]
- Value range used to be 0:45000; now: -45000:45000
- Usefulness of color interpolation is halved
- Solution: apply interpolation to each sub-window separately instead of to the entire dataset at once
- Some functionality is lost if each sub-window represent the same variable under some modification (eg. call volume based on time of day)
- No functionality is lost if each sub-window represents a totally different variable and all we care about is comparing ratios

Example: Comparing Ratios



Evaluation

Advantages of DPDs:

Excellent information density that takes advantage of limited display space

They make it easy to identify patterns, especially correlations between different variables.

Disadvantages:

For some data with non-natural ordering, complicated sorting/clustering algorithms are needed to make the pixel display useful for comparing trends.

It is not suited for use in cases where you want to see clear individual values (it is more useful for pattern finding / as an overview)

Bibliography

- http://www.sdml.info/library/Keim00.pdf
- http://bib.dbvis.de/uploadedFiles/172.pdf