Hierarchical Aggregation for Information Visualization: Overview, Techniques, and Design Guidelines

Niklas Elmqvist, Member, IEEE, and Jean-Daniel Fekete, Member, IEEE

Maithili Mishra

Motivation

Most analysts start with an overview of the data before gradually refining their view to be more focused and detailed

However, OVERVIEW is becoming increasingly difficult to effectively achieve with the ever-increasing size of real-world datasets.

The motivation of the work is to make visual representations more understandable, visually scalable and less cluttered.

HIERARCHICAL AGGREGATION FOR VISUALIZATION

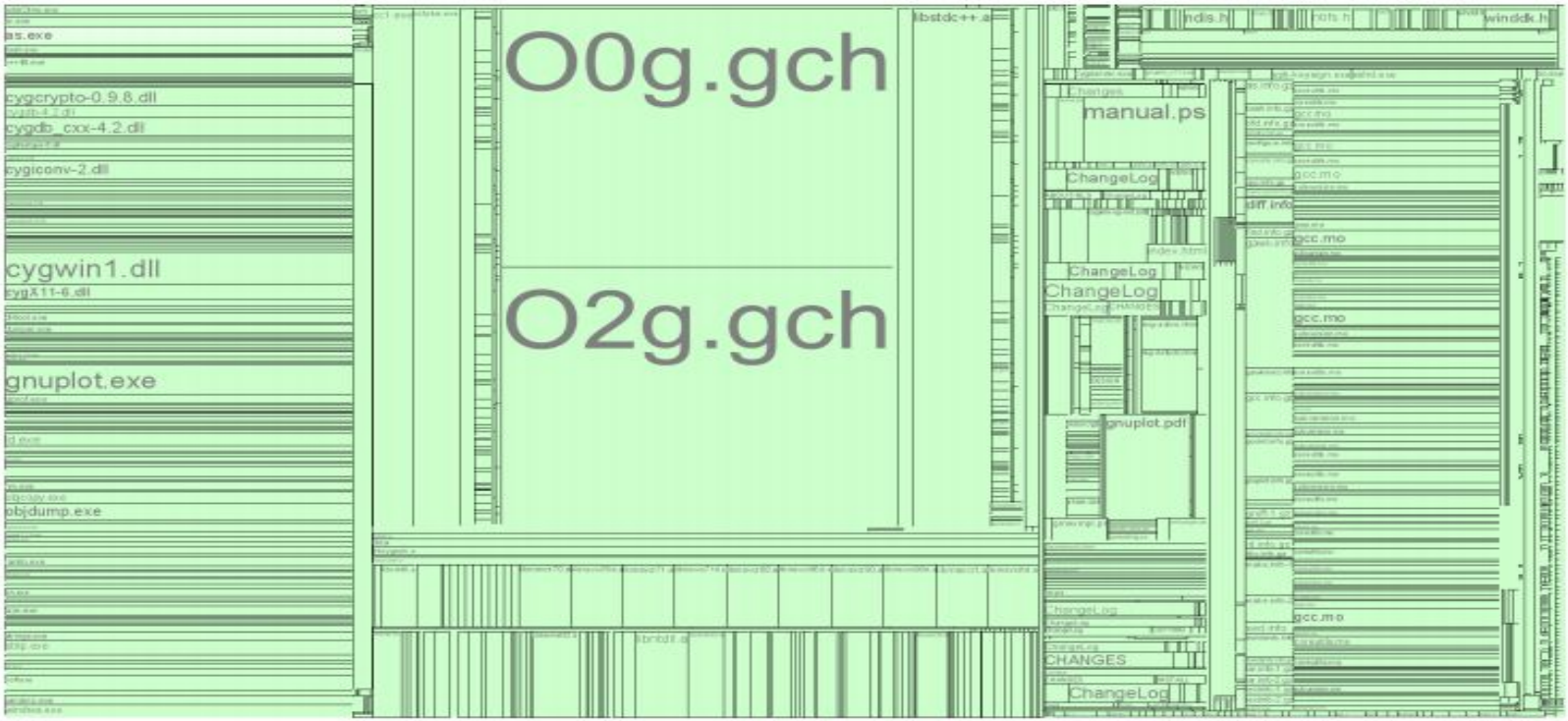
- Turns any visualization into a multiscale structure that can be rendered at any desired level-of-detail
- Provides the user with a manageable overview
- Hides any clutter arising from details
- Gives a reasonable indication of the data size, extents, or distribution

Follows the visual information seeking mantra: “overview first, zoom and filter, then details on demand.”

DESIGN GUIDELINES

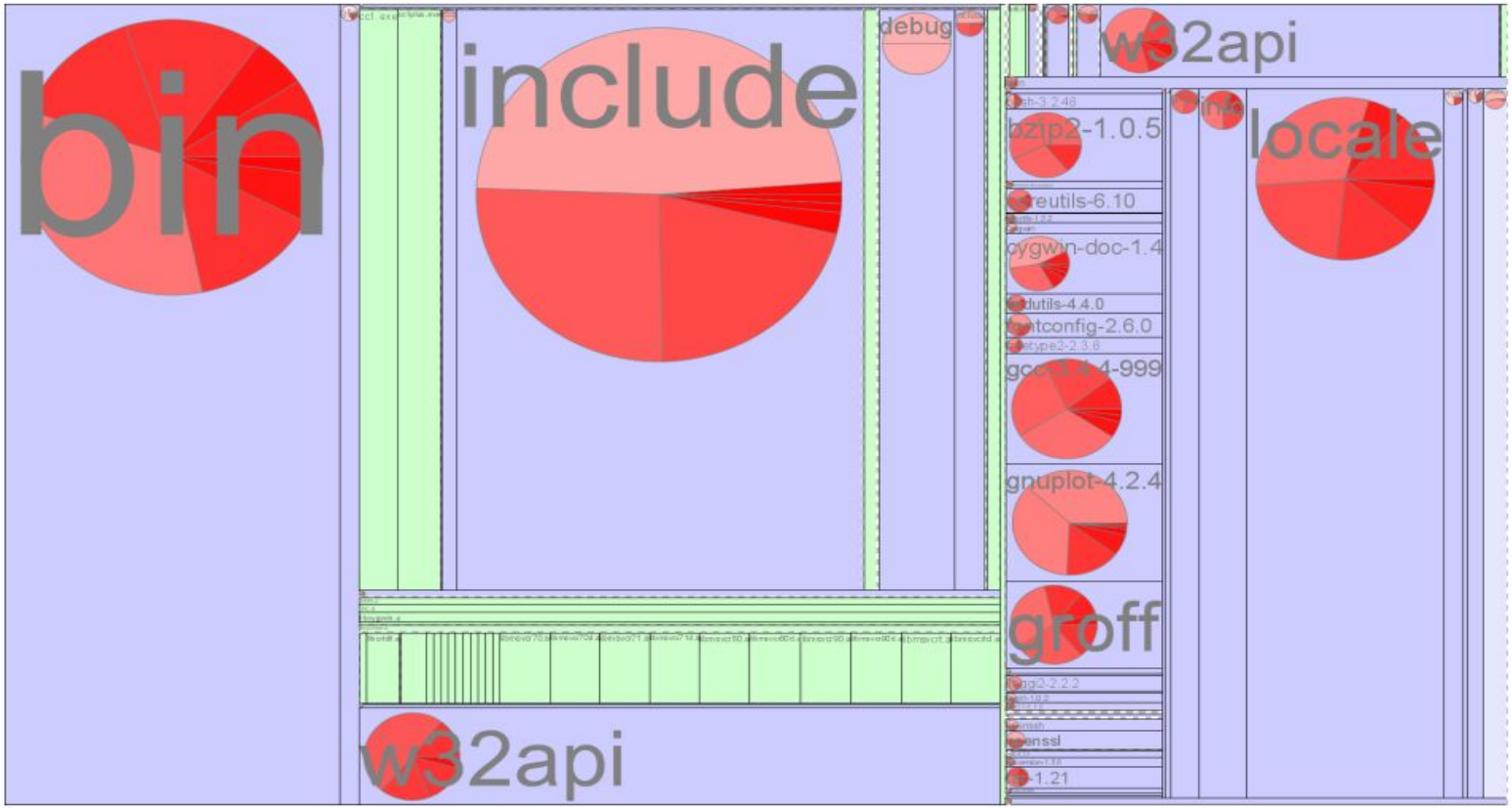
G1 Entity budget. Maintain an entity budget;
G2 Visual summary. Aggregates should convey information about underlying data;
G3 Visual simplicity. Aggregates should be clean and simple;
G4 Discriminability. Aggregates should be distinguishable from data items;
G5 Fidelity. Beware that abstractions may lie; and
G6 Interpretability. Aggregate items only so much so that the aggregation is still correctly interpretable within the visual mapping.

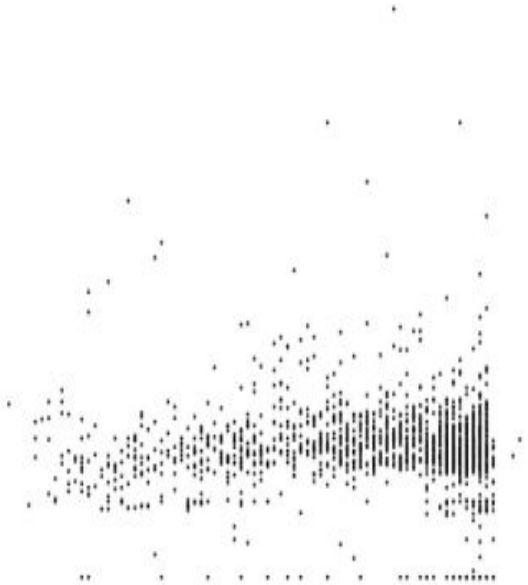
Slice-and-dice treemap visualization showing a typical Cygwin distribution for Windows (648 directories, 6552 files, 181 MB).



Aggregate treemap visualization of file structure

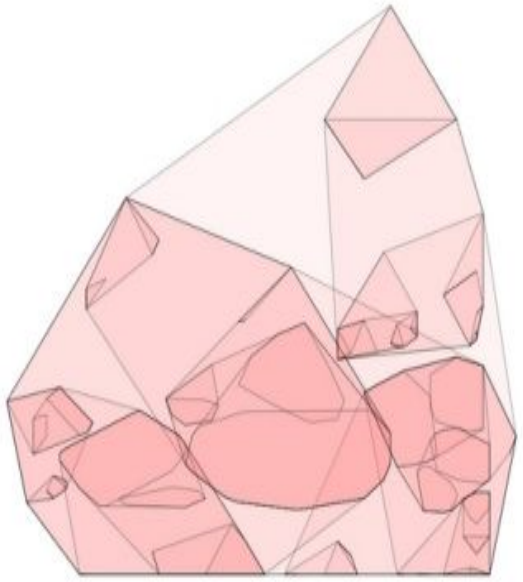
Aggregate treemap visualization of cygwin file distribution





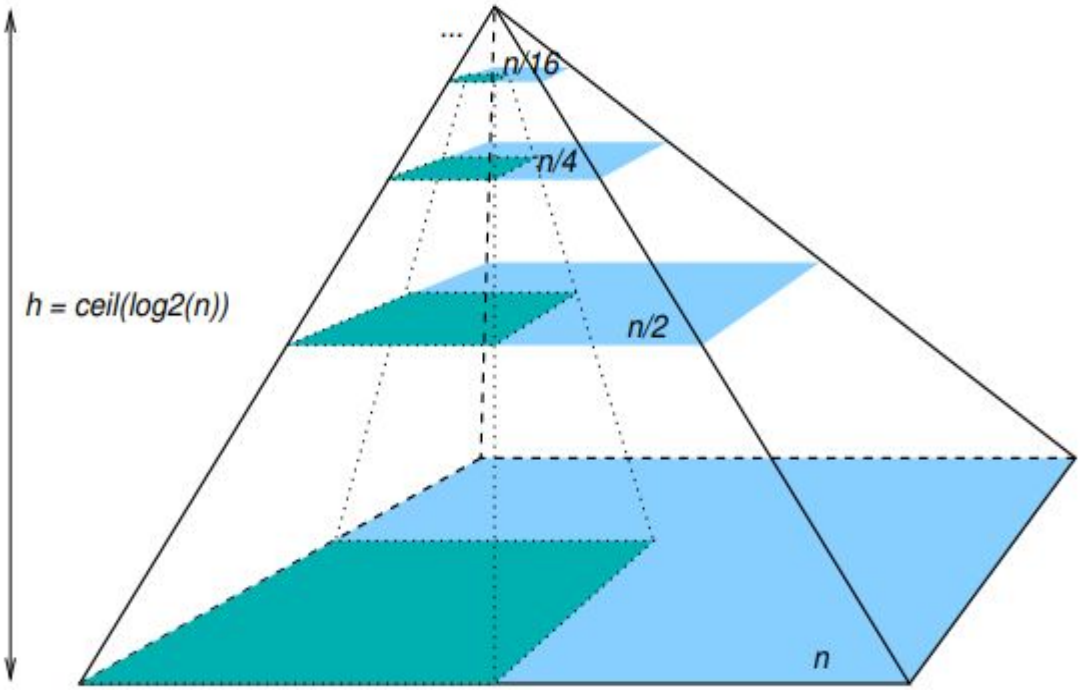
(a) 2D scatterplot visualization.

(b) 2D bounding box aggregation.

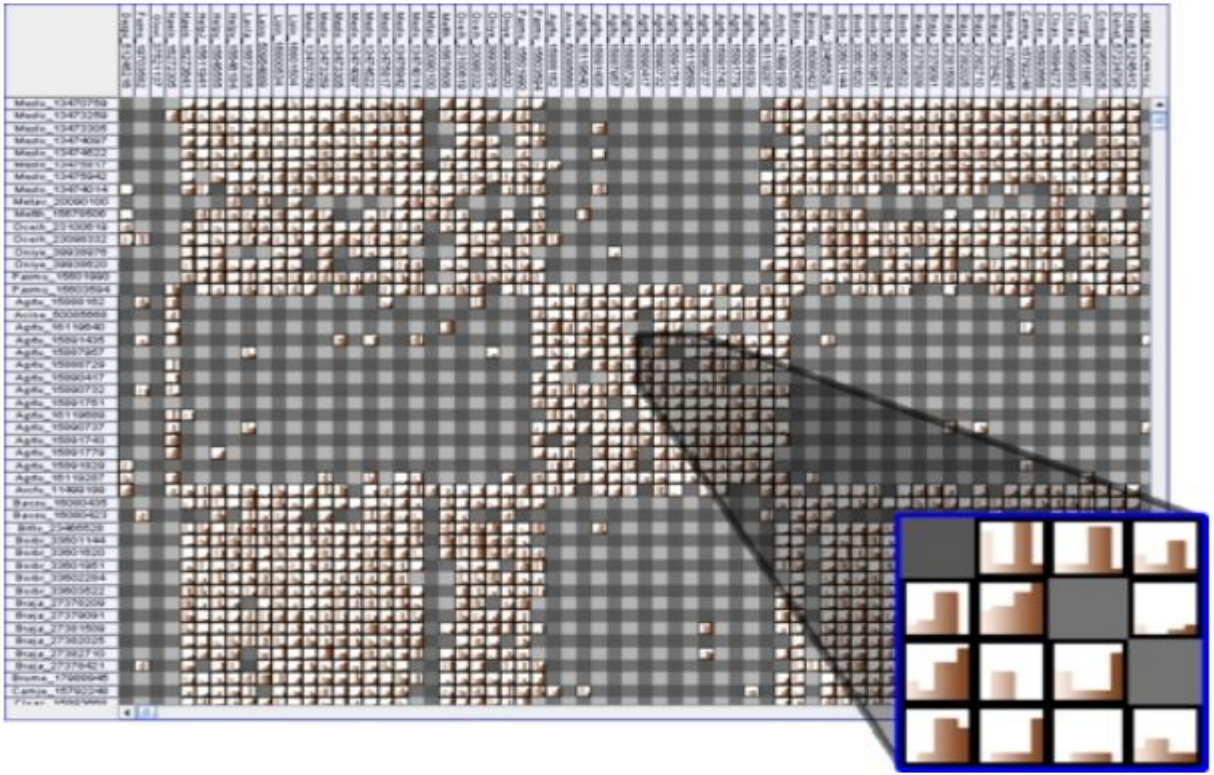


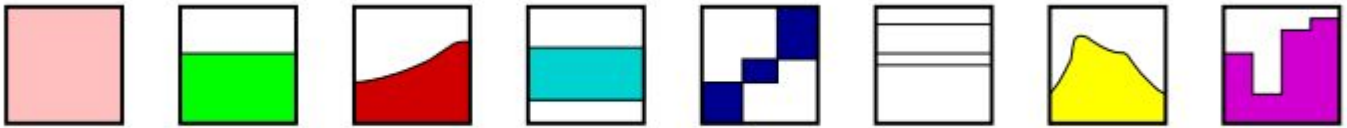
(c) 2D convex hull aggregation.

Pyramid aggregation hierarchy for the ZAME system. Each aggregate node represents four nodes in the level below



ZAME visualizing a protein-protein interaction dataset of 100,000 nodes and 1,000,000 edges. Inset shows a magnified view of step histogram edge aggregates in the matrix.





Visual edge aggregate representations for ZAME (average color shade, average, min/max curve, min/max range, min/max tribox, Tukey box, smooth histogram, step histogram)

Interaction

- Zoom and Pan
- Drill-down and Roll-up
- Local Aggregation Control
- Flipping
- Coupled Zooming and Drilling

Thank you