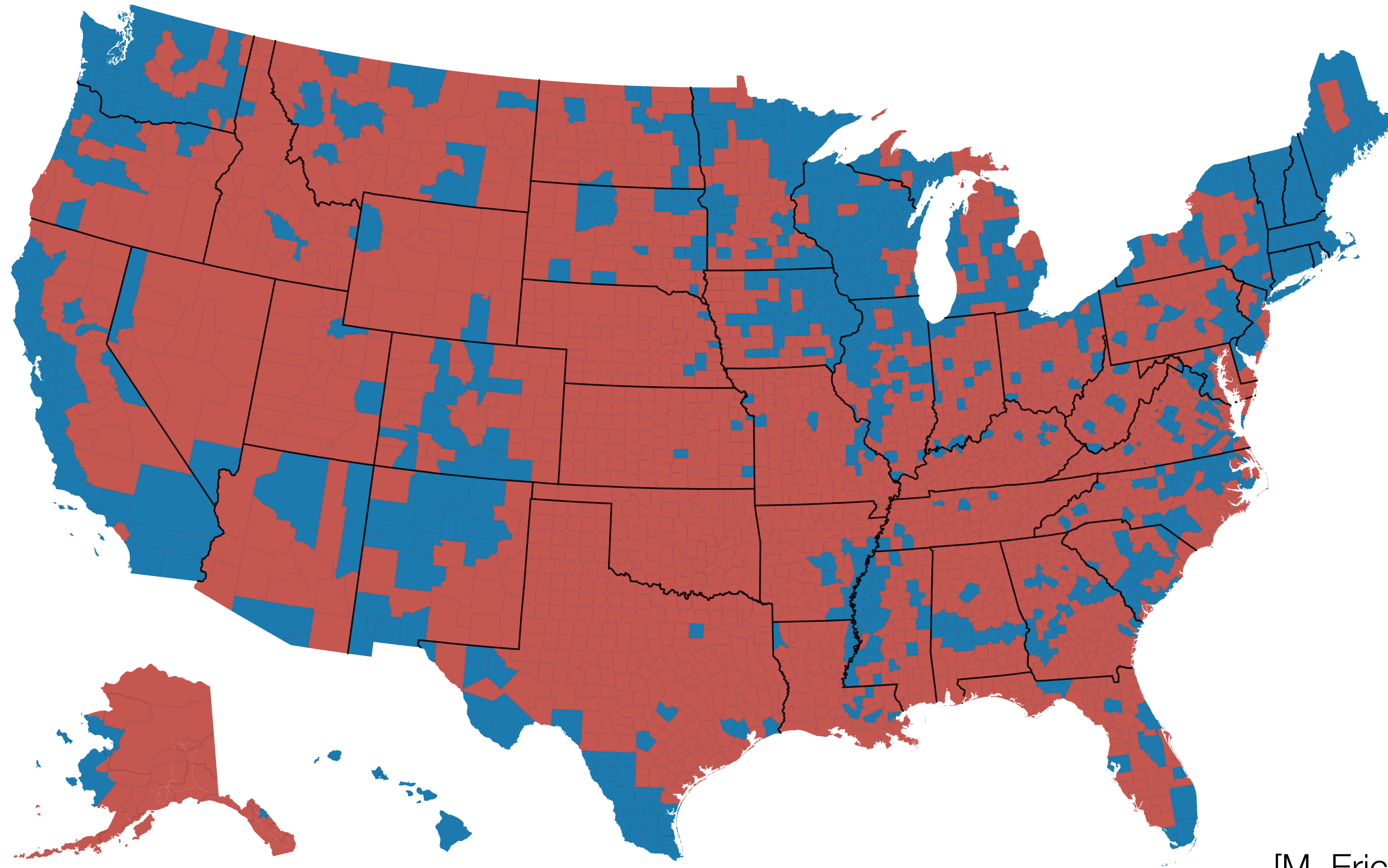


Data Visualization (CSCI 627/490)

Trees

Dr. David Koop

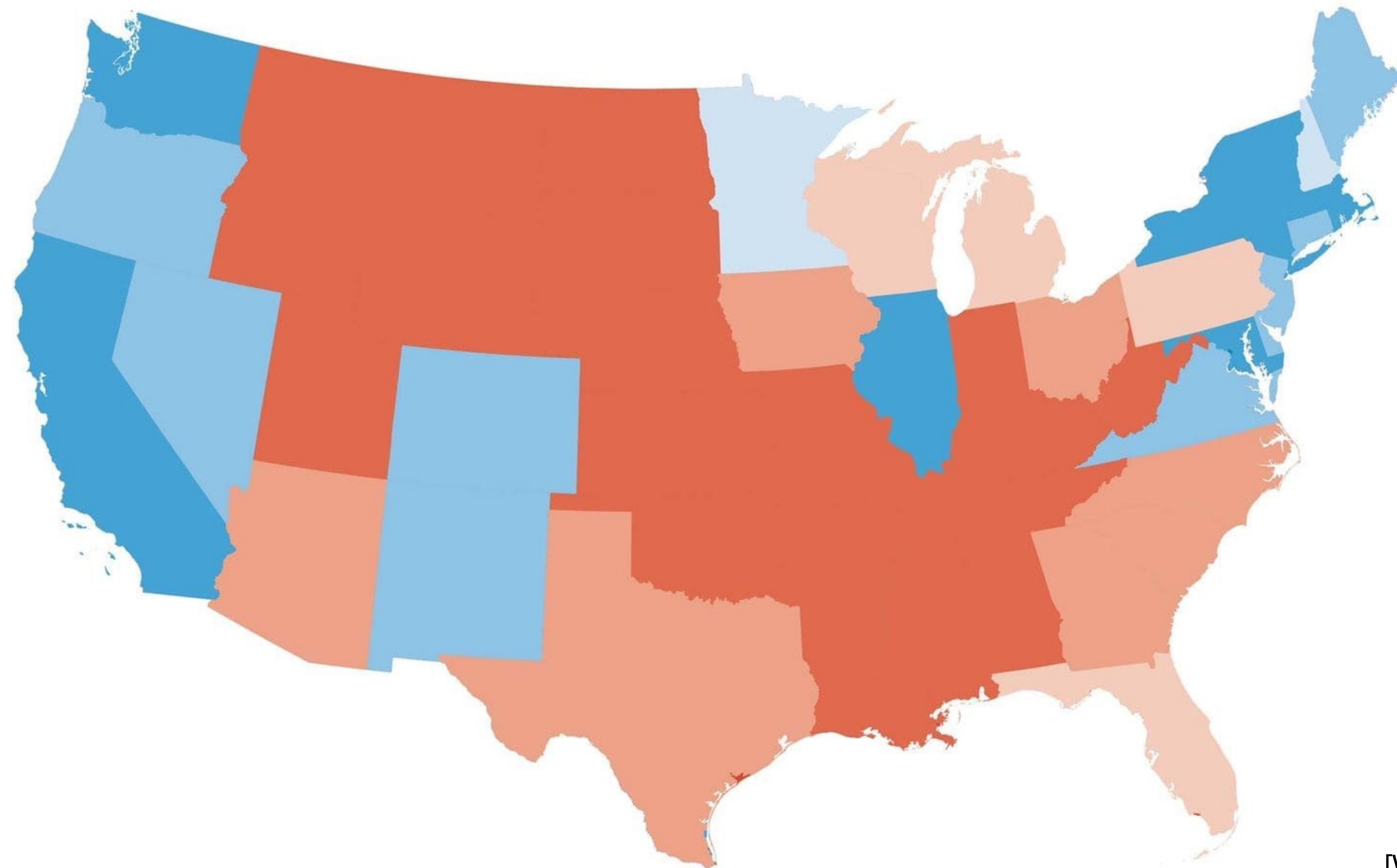
Choropleth (Two Hues)



[M. Ericson, New York Times]

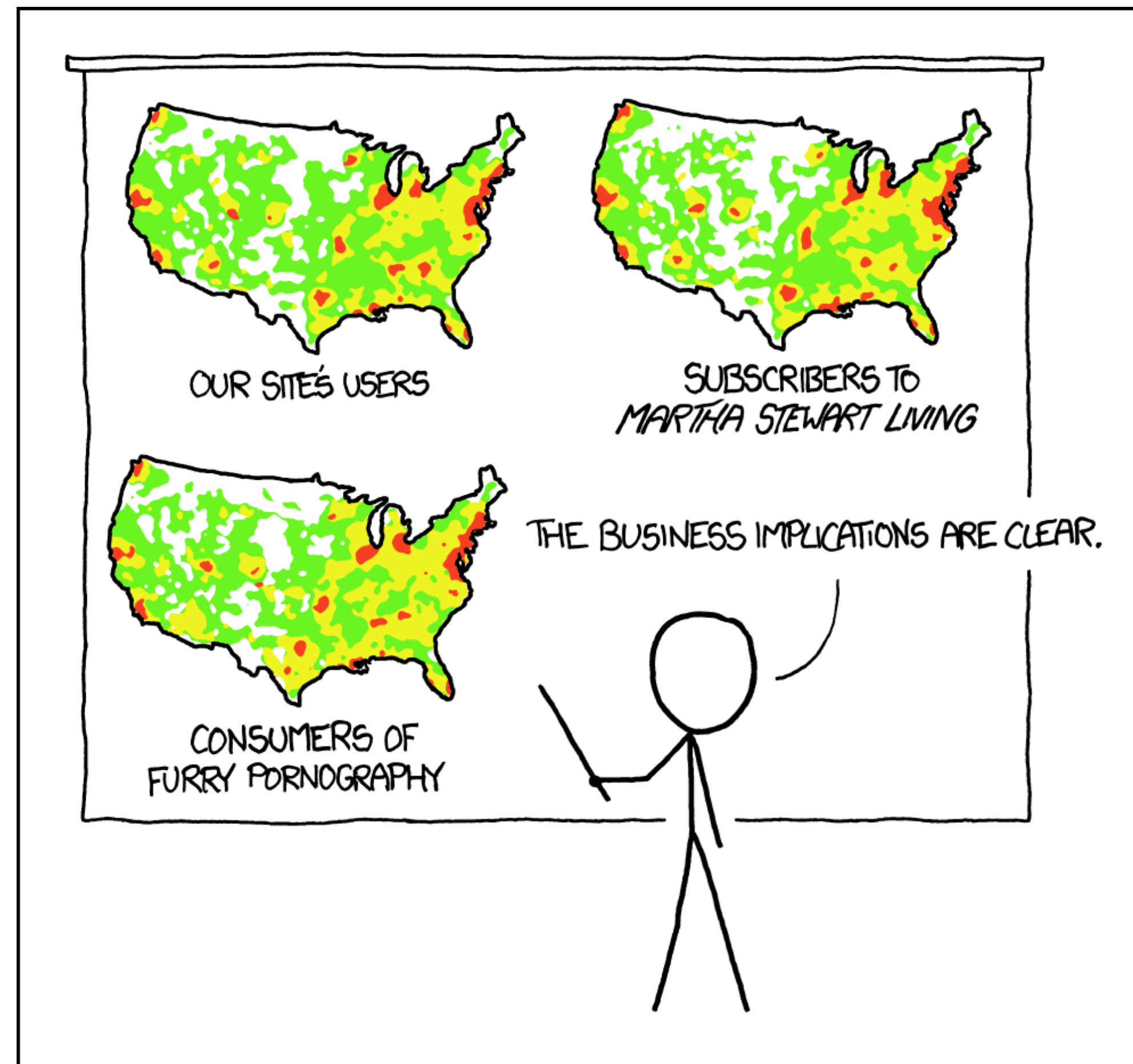
Choropleth (Diverging Attribute)

Clinton +50-100 +15-50 +2.1-15 +0-2.1 Trump +0-2.1 +2.1-15 +15-50 +50-100



[Washington Post, 2018]

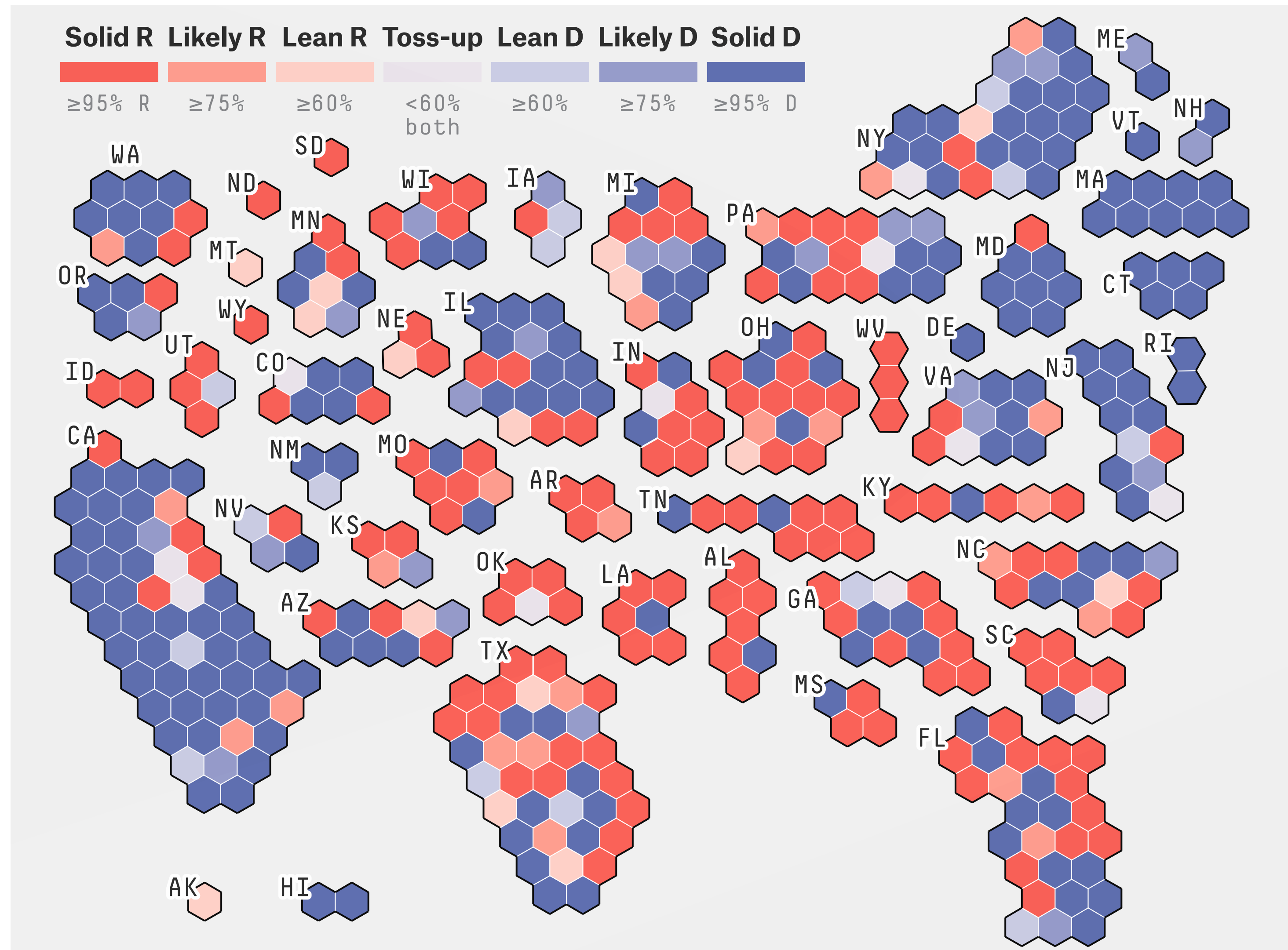
Don't Just Create Population Maps!



PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE
BASICALLY JUST POPULATION MAPS

[xkcd]

House Races: Non-Contiguous "Cartogram"



[FiveThirtyEight, 2020]

Maps Aren't Always Best: Close House Races

12 Lean Democratic

- AZ-02 Open (McSally)
- CA-49 Open (Issa)
- CO-06 Coffman
- IA-01 Blum
- KS-03 Yoder
- MI-11 Open (Trott)
- MN-02 Lewis
- MN-03 Paulsen
- NV-03 Open (Rosen)
- NJ-11 Open (Frelinghuysen)
- PA-07 Vacant (formerly Dent)
- VA-10 Comstock

31 Tossups

- CA-10 Denham
- CA-25 Knight
- CA-39 Open (Royce)
- CA-45 Walters
- CA-48 Rohrabacher
- FL-26 Curbelo
- FL-27 Open (Ros-Lehtinen)
- IL-06 Roskam
- IL-12 Bost
- IA-03 Young
- KS-02 Open (Jenkins)
- KY-06 Barr

25 Lean Republican

- AR-02 Hill
- CA-50 Hunter
- FL-15 Open (Ross)
- FL-16 Buchanan
- GA-06 Handel
- GA-07 Woodall
- IL-13 Davis
- IL-14 Hultgren
- MO-02 Wagner
- MT-AL Gianforte
- NE-02 Bacon
- NY-24 Katko

[New York Times, 2018]

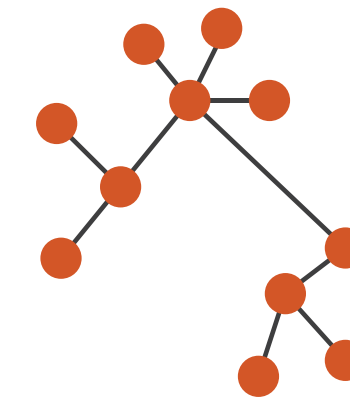
Networks

- Network: nodes and edges connecting the nodes
- Formally, $G = (V, E)$ is a set of nodes V and a set of edges E where each edge connects two nodes.
- Nodes == items, edges connect items
- **Both** nodes and edges may have **attributes**

Arrange Networks and Trees

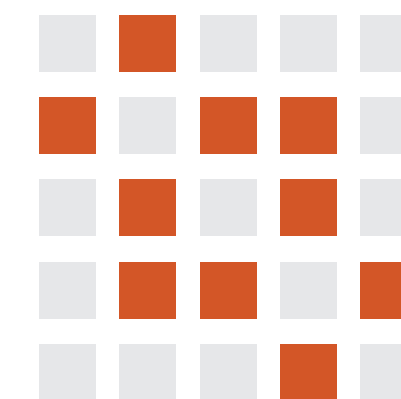
➔ Node–Link Diagrams Connection Marks

✓ NETWORKS ✓ TREES



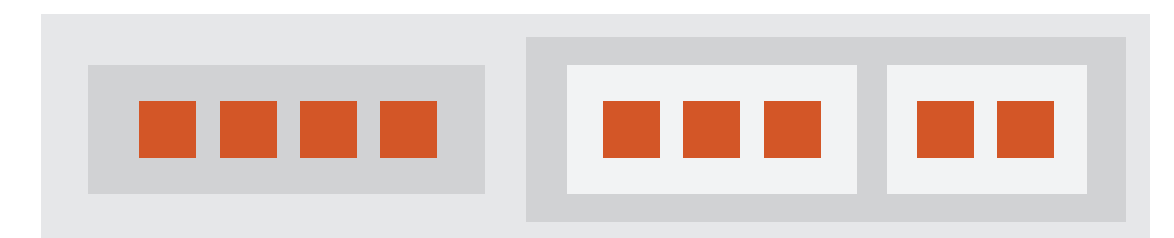
➔ Adjacency Matrix Derived Table

✓ NETWORKS ✓ TREES



➔ Enclosure Containment Marks

✗ NETWORKS ✓ TREES



[Munzner (ill. Maguire), 2014]

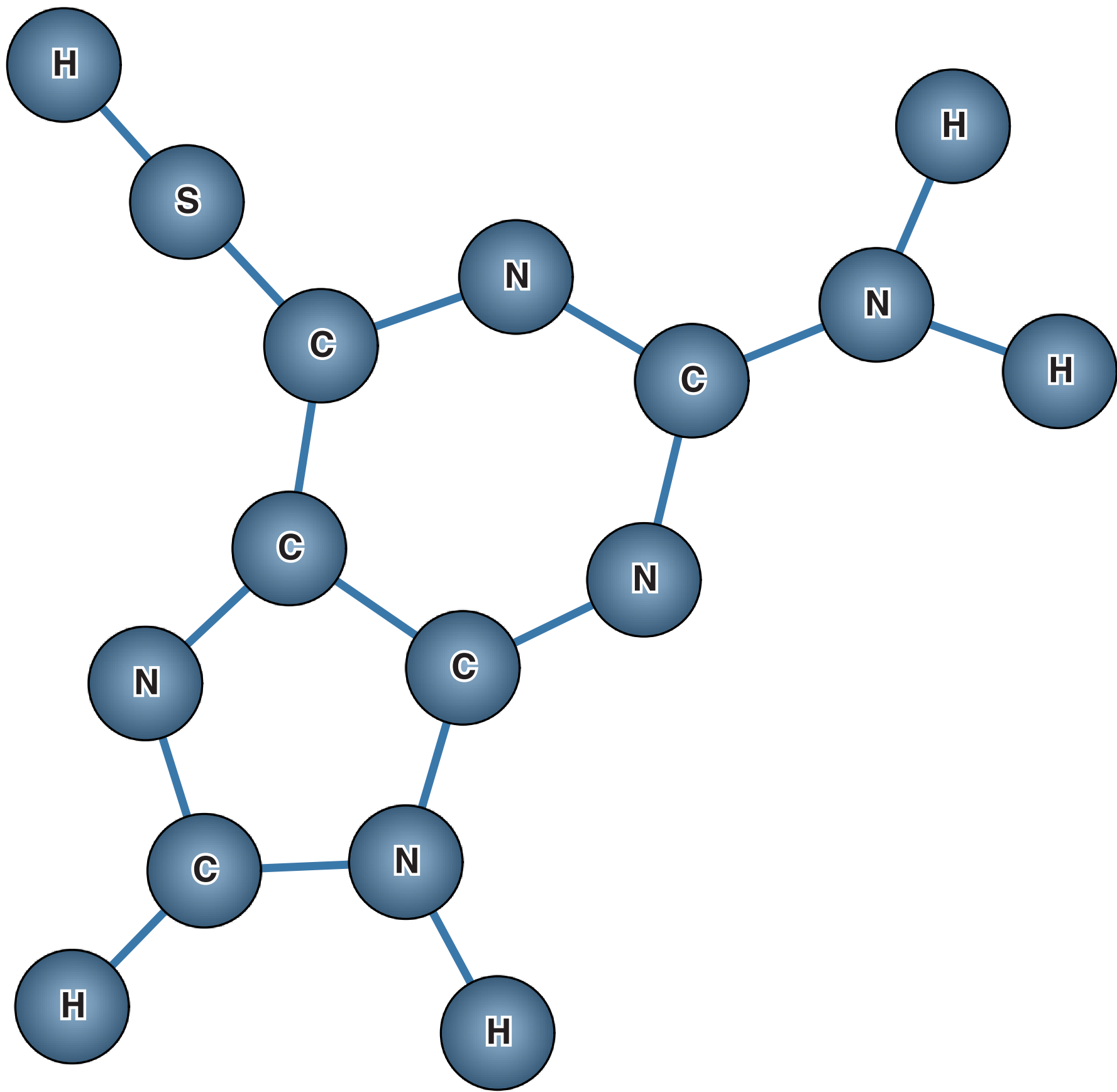
Network Data Represented in Tables

Nodes

ID	Atom	Electrons	Protons
0	N	7	7
1	C	6	6
2	S	16	16
3	C	6	6
4	N	7	7

Edges

ID1	ID2	Bonds
0	1	1
1	2	1
1	3	2
3	4	1



Project Proposal

- Two Possibilities:
 - Create an interactive visualization
 - Work on a research project
- Dataset Choices
 - US Food Safety Data
 - Illinois Hospital Report Card
 - NFL Data
 - US Register of Introduced and Invasive Species
 - Others?
- Proposal Due Wednesday

Assignment 4

- To be announced soon

Wednesday

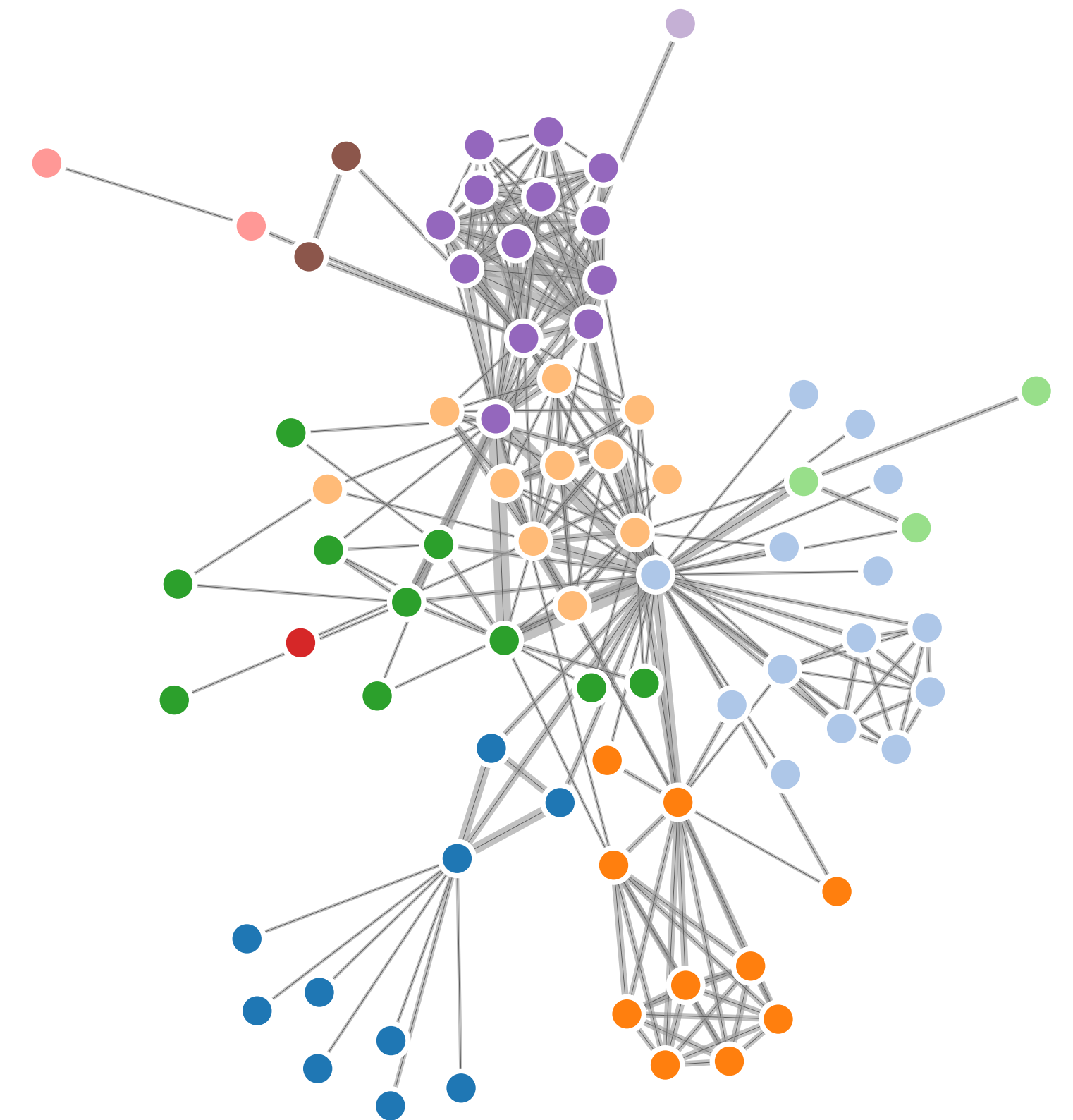
- Online Video Lecture
- Have external meeting
- Was told I was going to have a break during this class time but timing was shifted in current agenda
- Office Hours shifted to the afternoon 2-3:15pm

Networks Need Layouts!

- Need to use spatial position when designing network visualizations
- Otherwise, nodes can **occlude** each other, links hard to distinguish
- How?
 - With bar charts, we could order using an attribute...
 - With networks, we want to be able to see connectivity and topology (not in the data usually)
- Possible metrics:
 - Edge crossings
 - Node overlaps
 - Total area

Force-Directed Layout

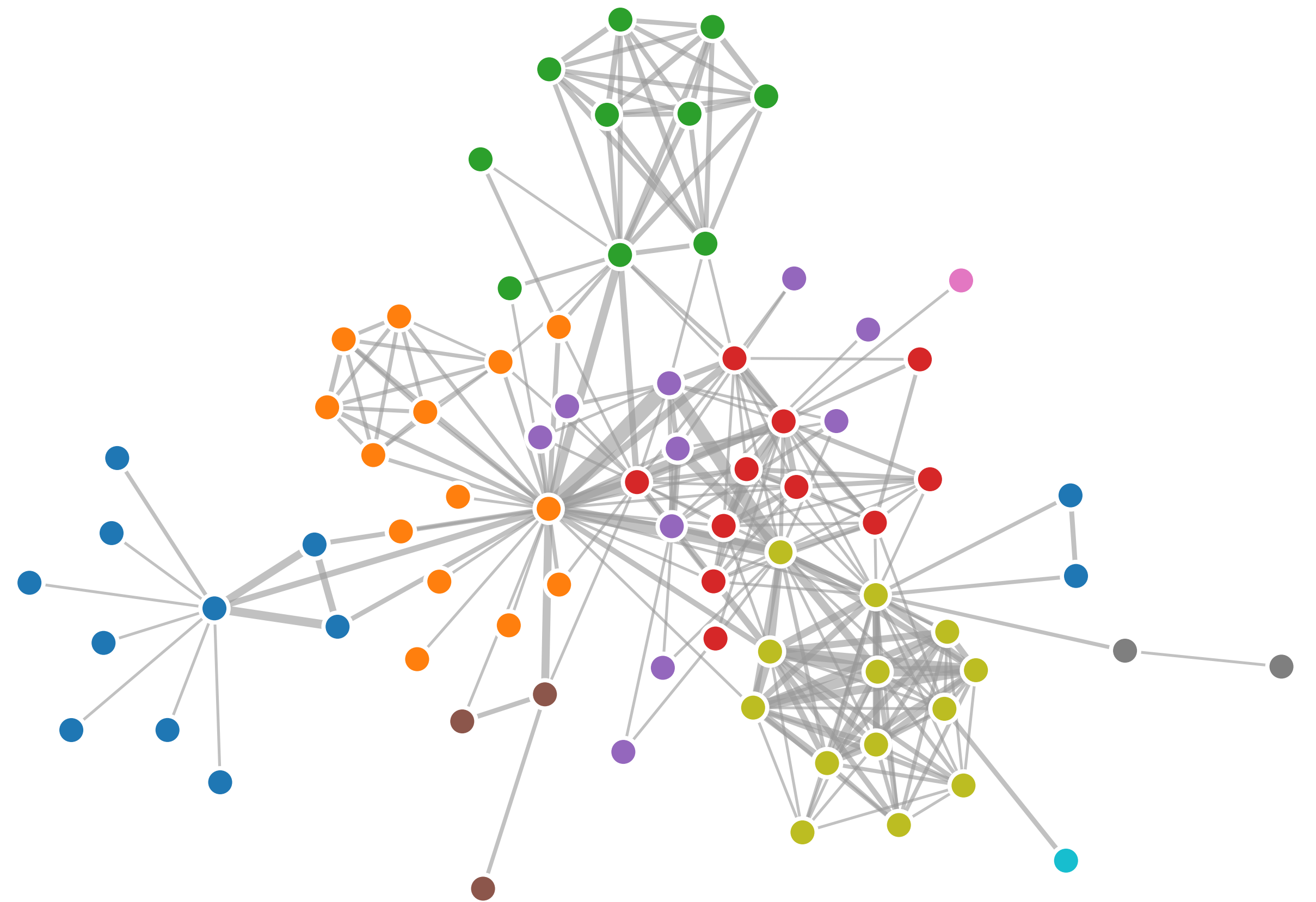
- Nodes push away from each other but edges are springs that pull them together
- Weakness: nondeterminism, algorithm may produce different results each time it runs



[M. Bostock, 2017]

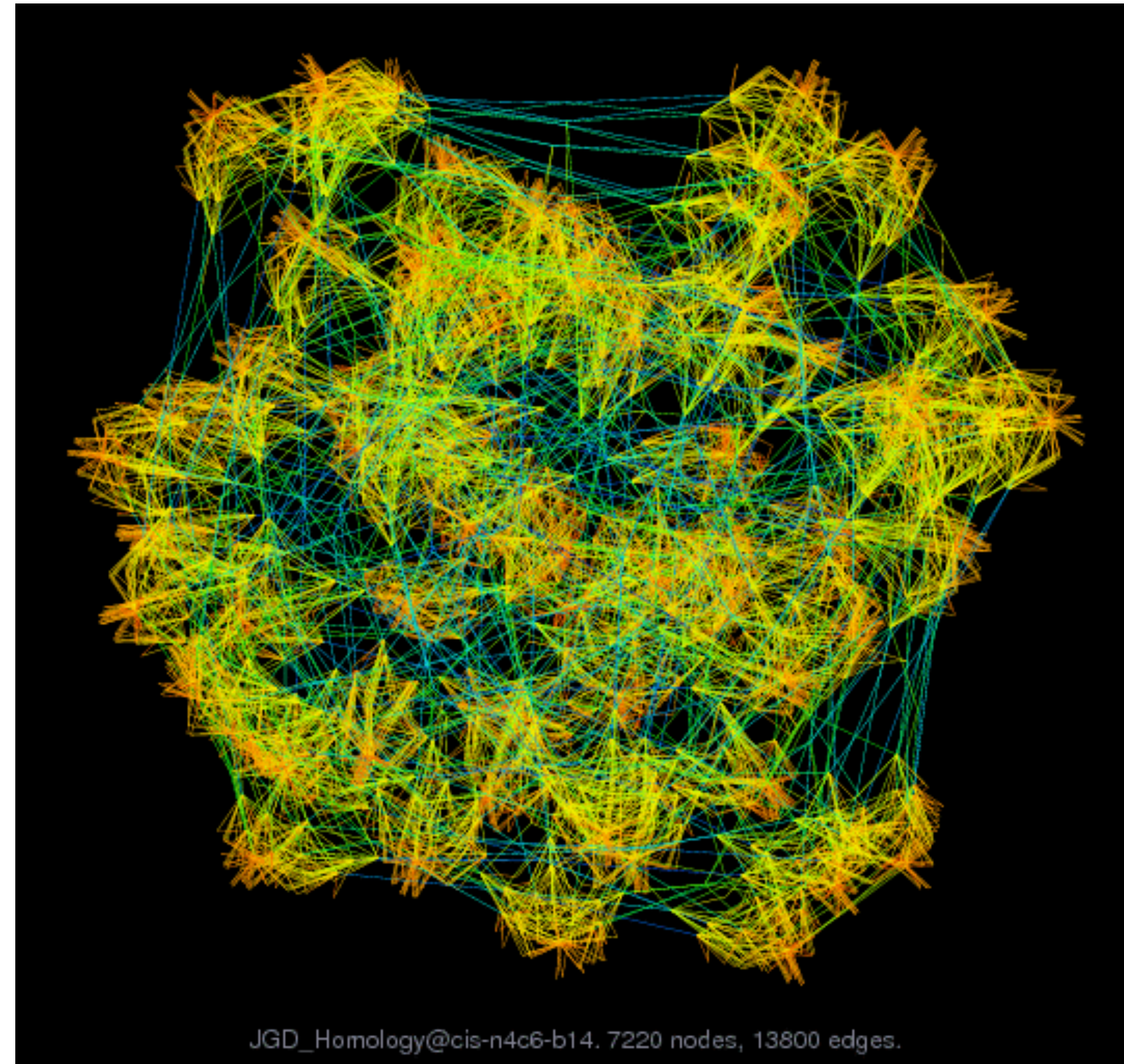
Constraint-Based Optimization (CoLa)

- Higher quality layout
- More **stable** in interactive applications (no "jitter")
- Allows user specified constraints such as alignments and grouping
- Can avoid overlapping nodes
- Provides flow layout for directed graphs
- May be **less scalable** to very large graphs
- Can route edges around nodes



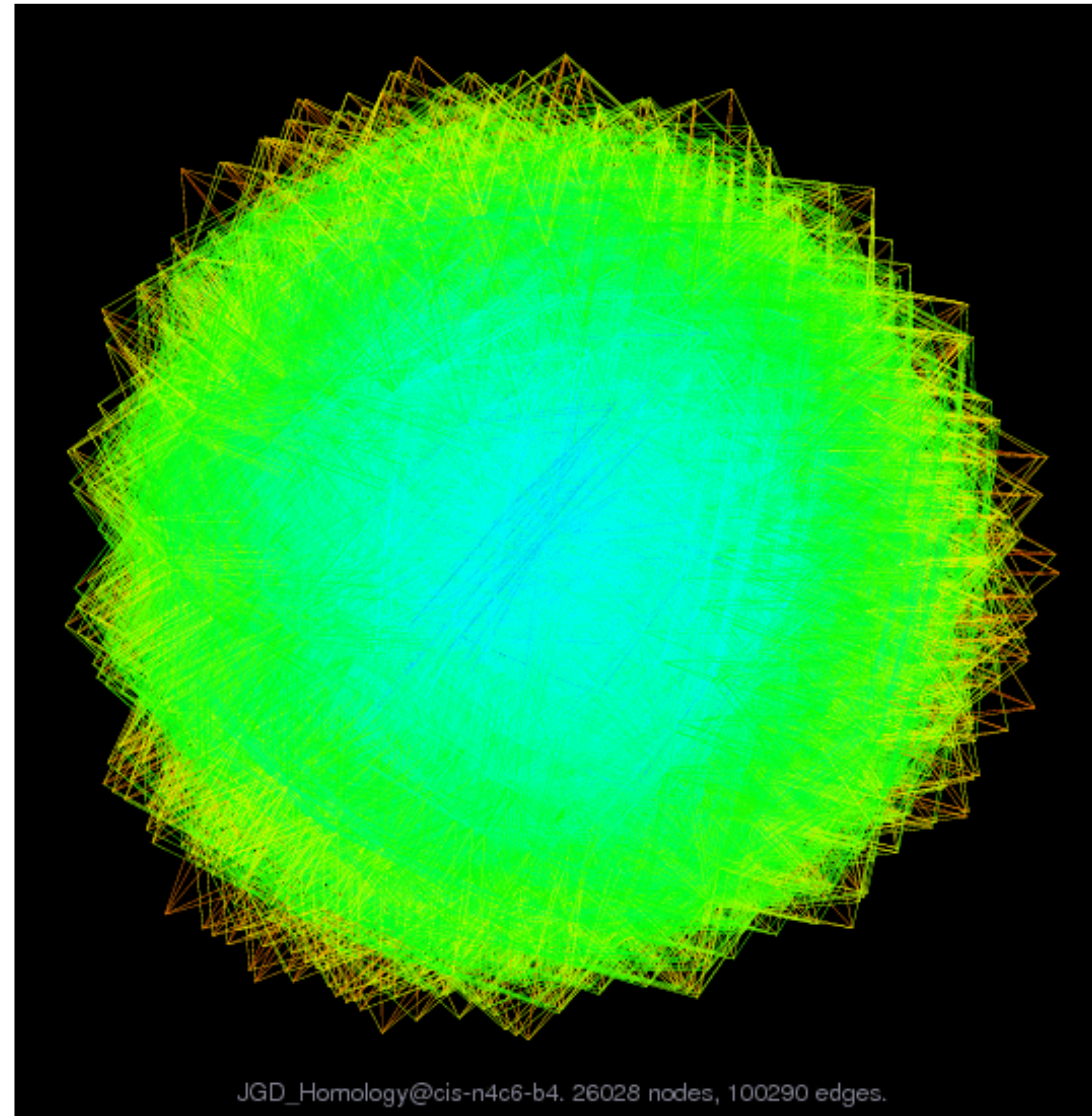
[T. Dwyer et al. (WebCoLa); M. Bostock (Example), 2018]

sfdp



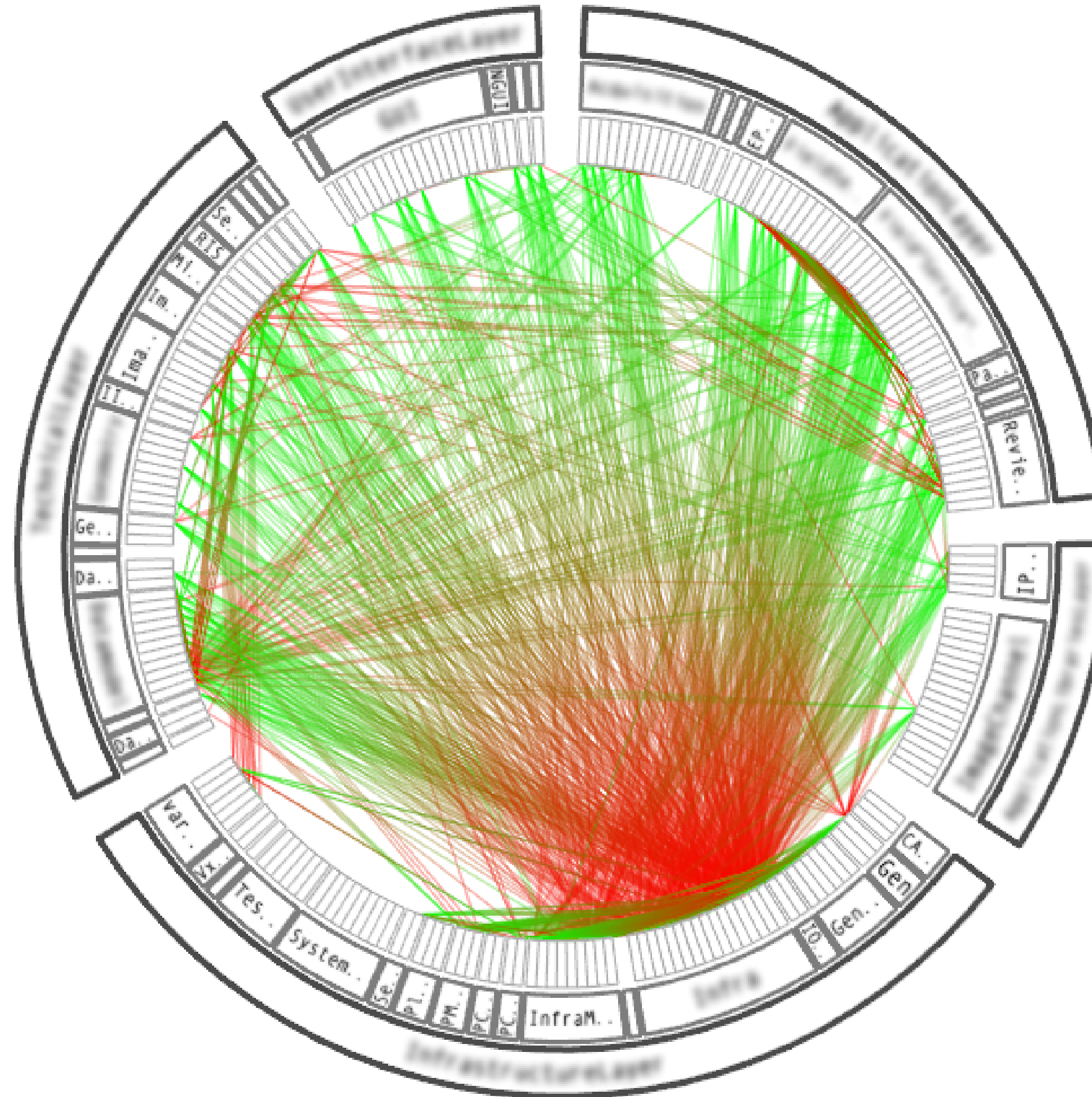
[Hu, 2005]

“Hairball”



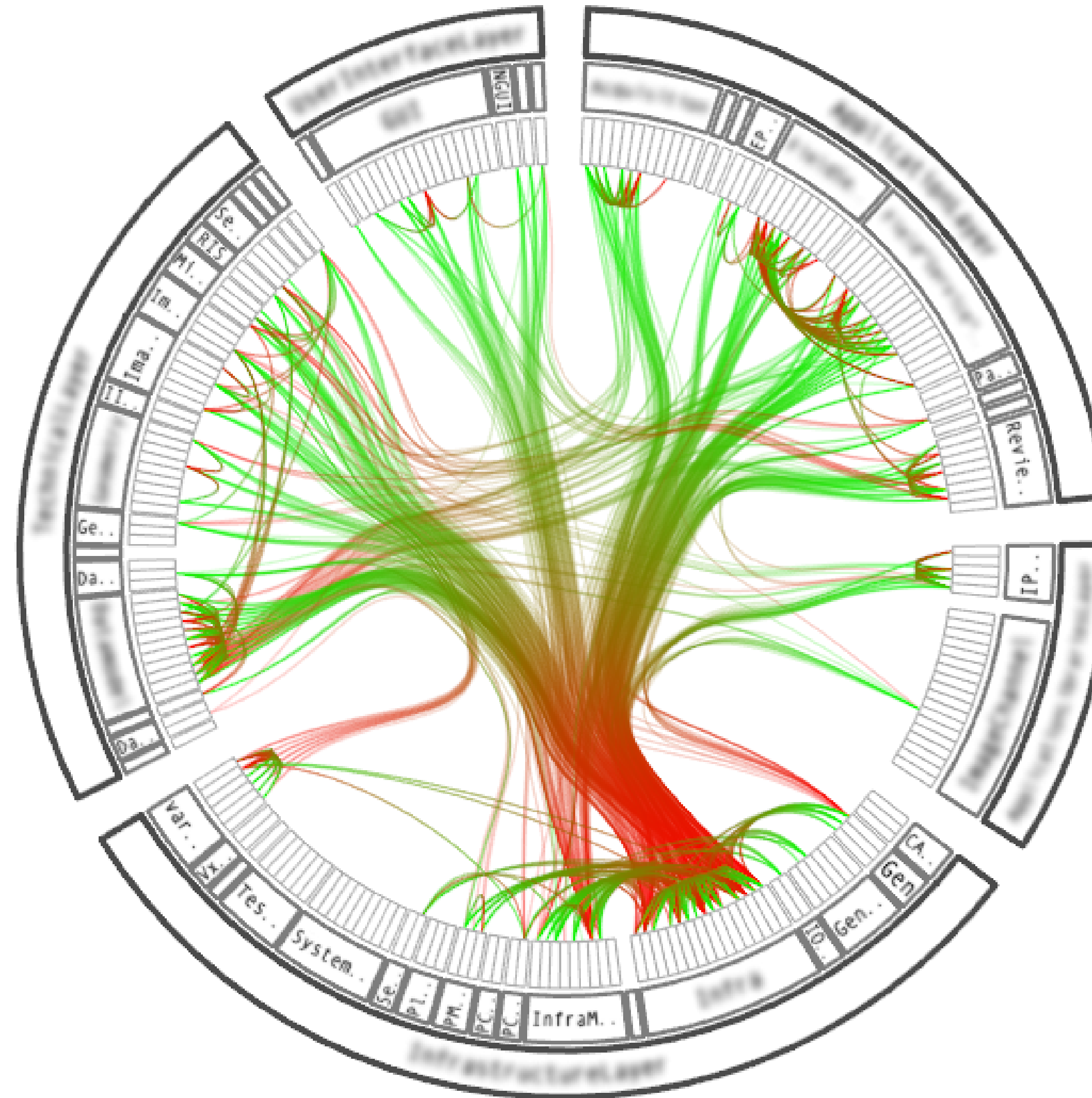
[Hu, 2014]

Hierarchical Edge Bundling



[Holten, 2006]

Hierarchical Edge Bundling



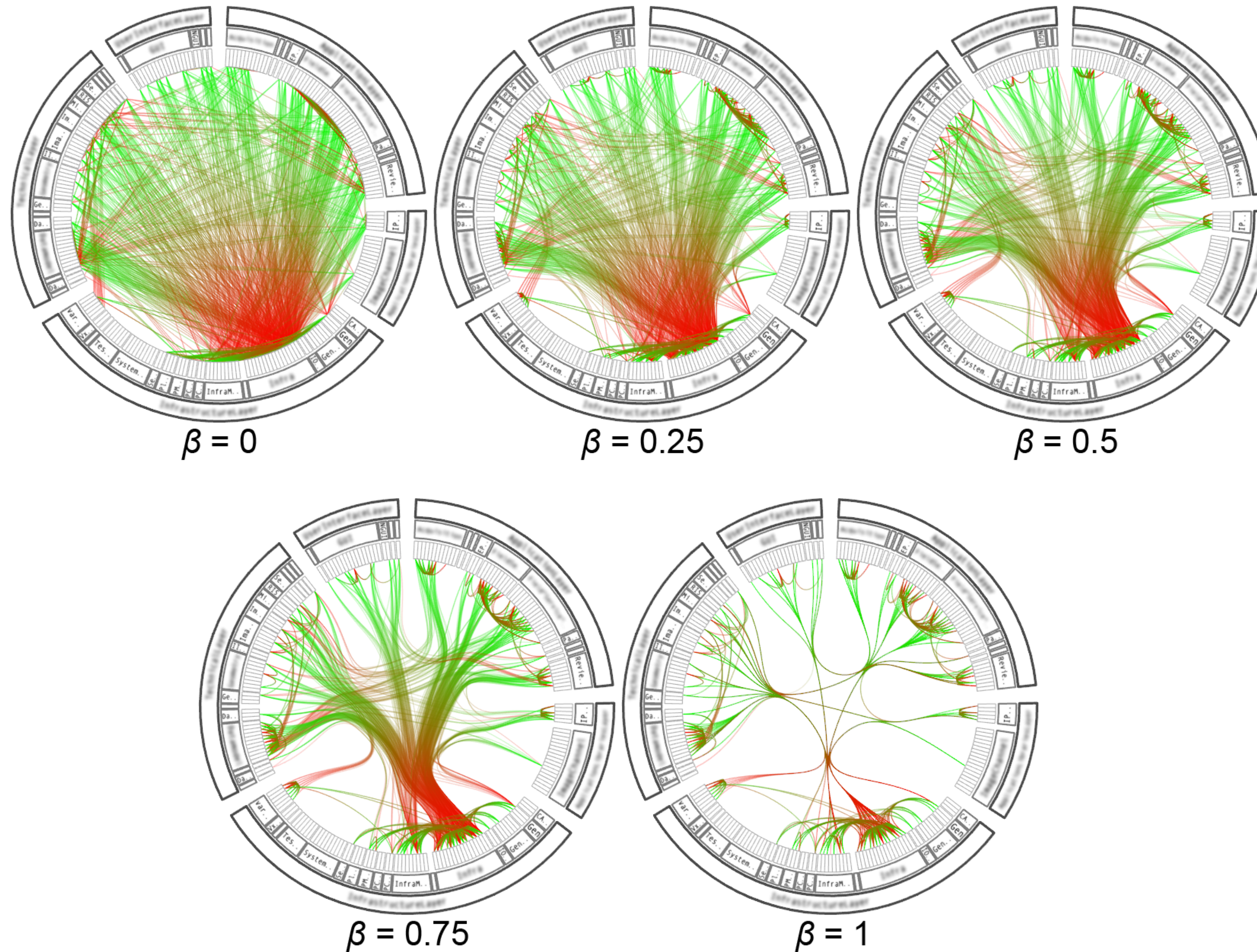
[Holten, 2006]

Hierarchical Edge Bundling

- Flexible and generic method
- Reduces visual clutter when dealing with large numbers of adjacency edges
- Provides an intuitive and continuous way to control the strength of bundling.
 - Low bundling strength mainly provides low-level, node-to-node connectivity information
 - High bundling strength provides high-level information as well by implicit visualization of adjacency edges between parent nodes that are the result of explicit adjacency edges between their respective child nodes

[Holten, 2006]

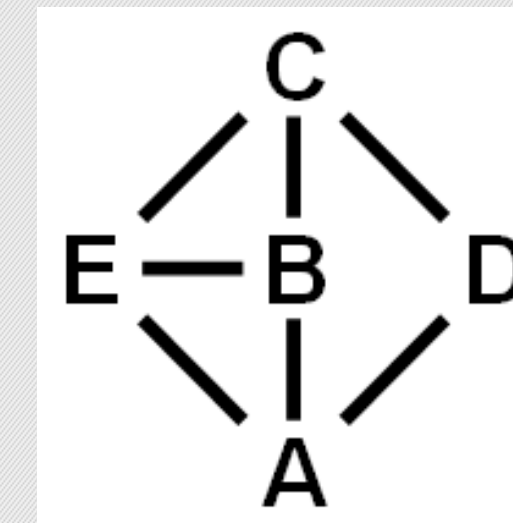
Bundling Strength



[Holten, 2006]

Adjacency Matrix

- Change network to tabular data and use a matrix representation
- Derived data: nodes are keys, edges are boolean values
- Task: lookup connections, find well-connected clusters
- Scalability: millions of edges
- Can encode **edge weight**, too

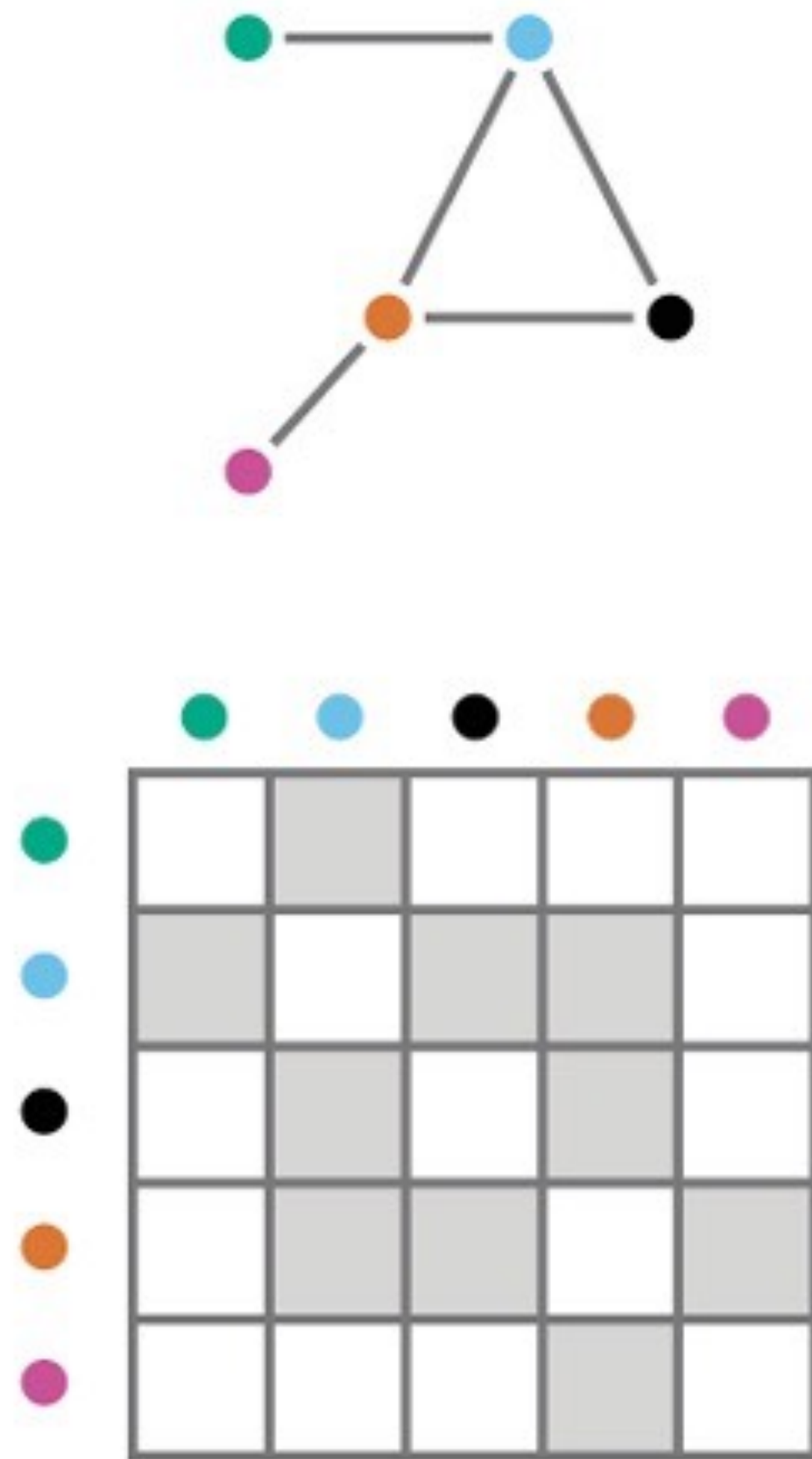


	A	B	C	D	E
A	A				
B		B			
C			C		
D				D	
E					E

[Henry et al., 2007]

Cliques in Adjacency Matrices

a

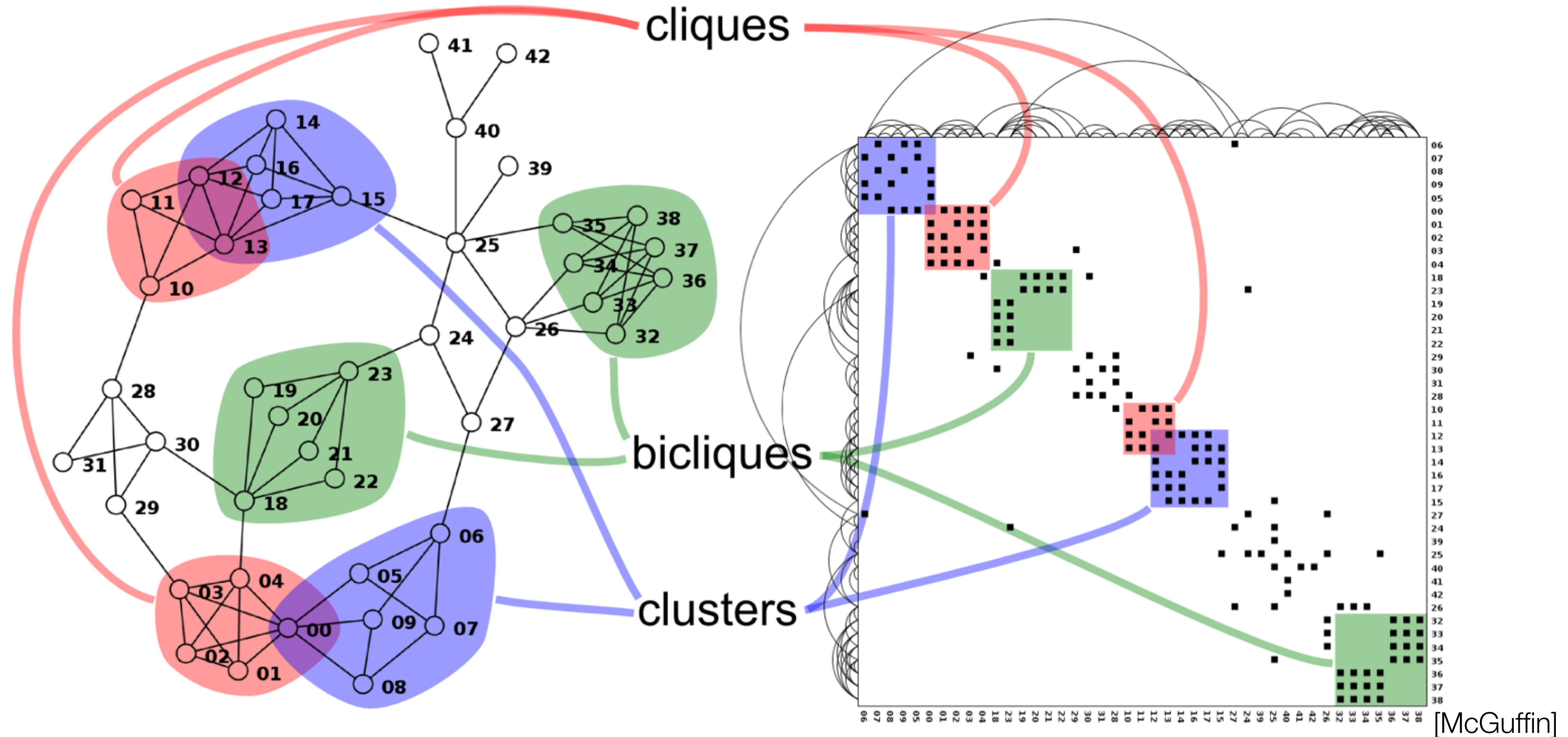


b



[Gehlenborg and Wong]

Structures from Adjacency Matrices



[McGuffin]

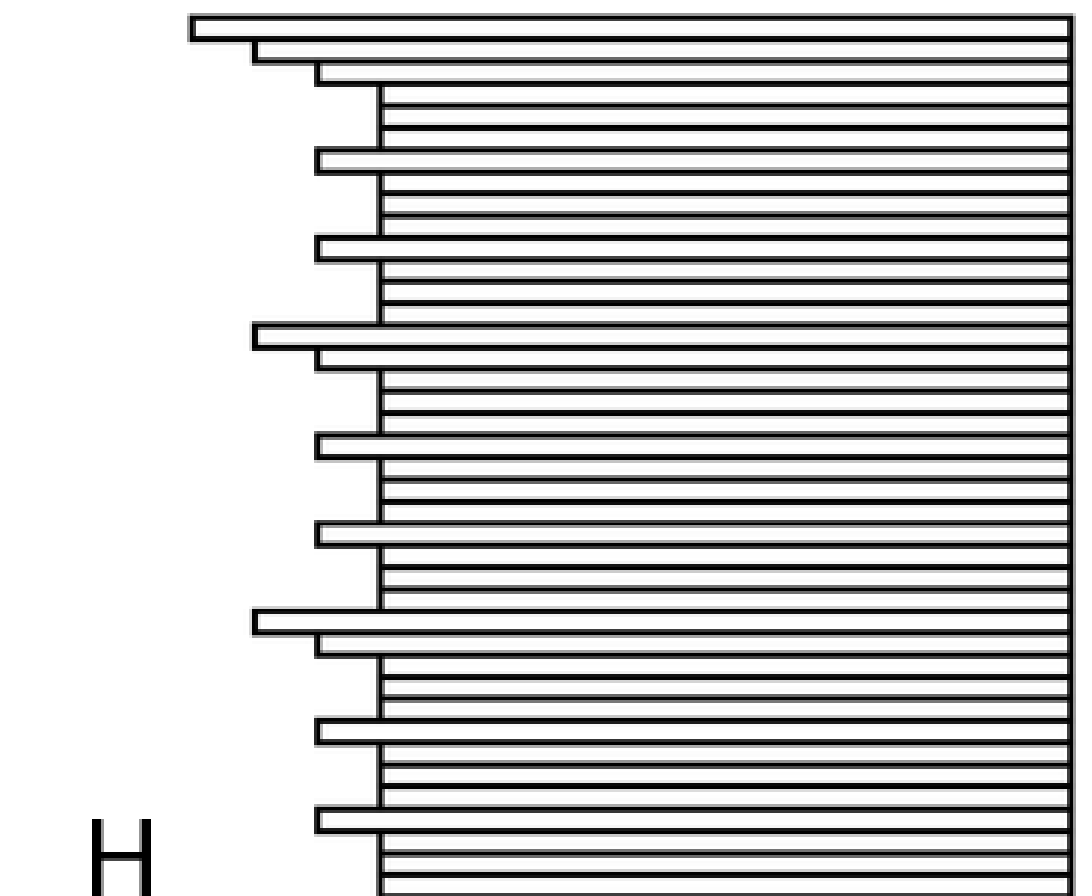
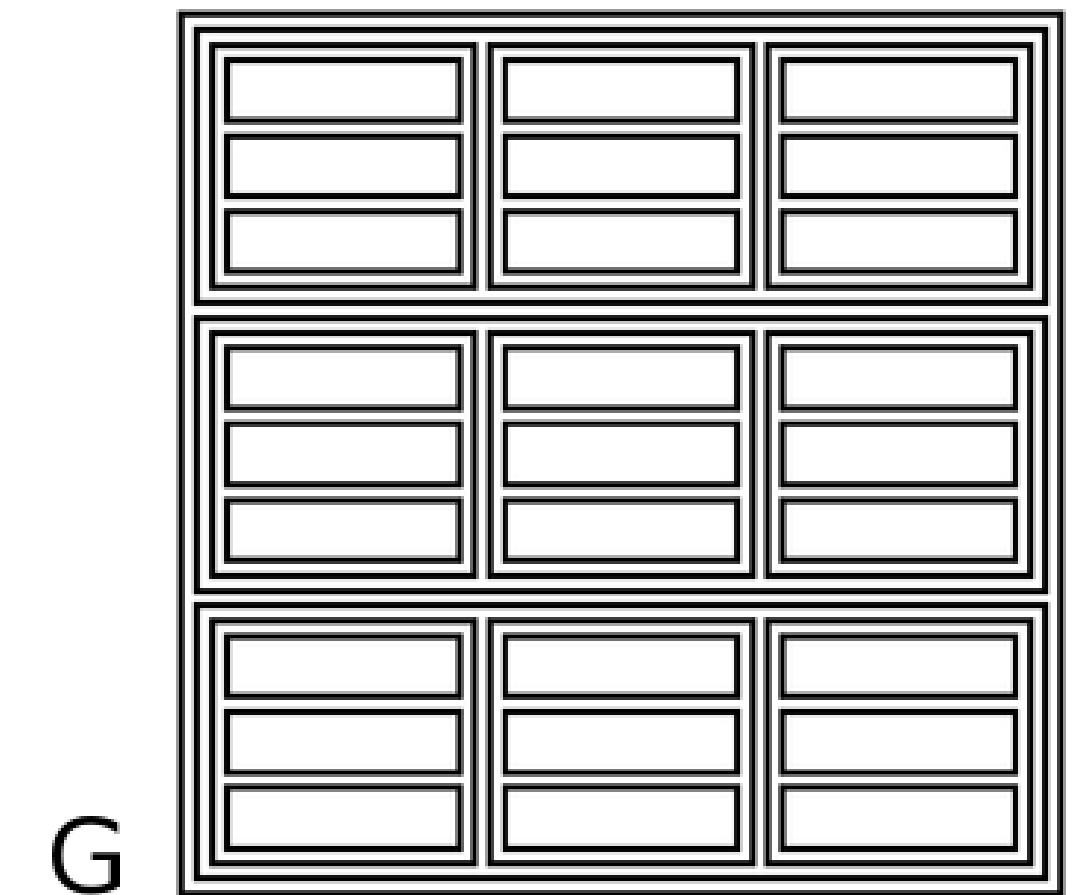
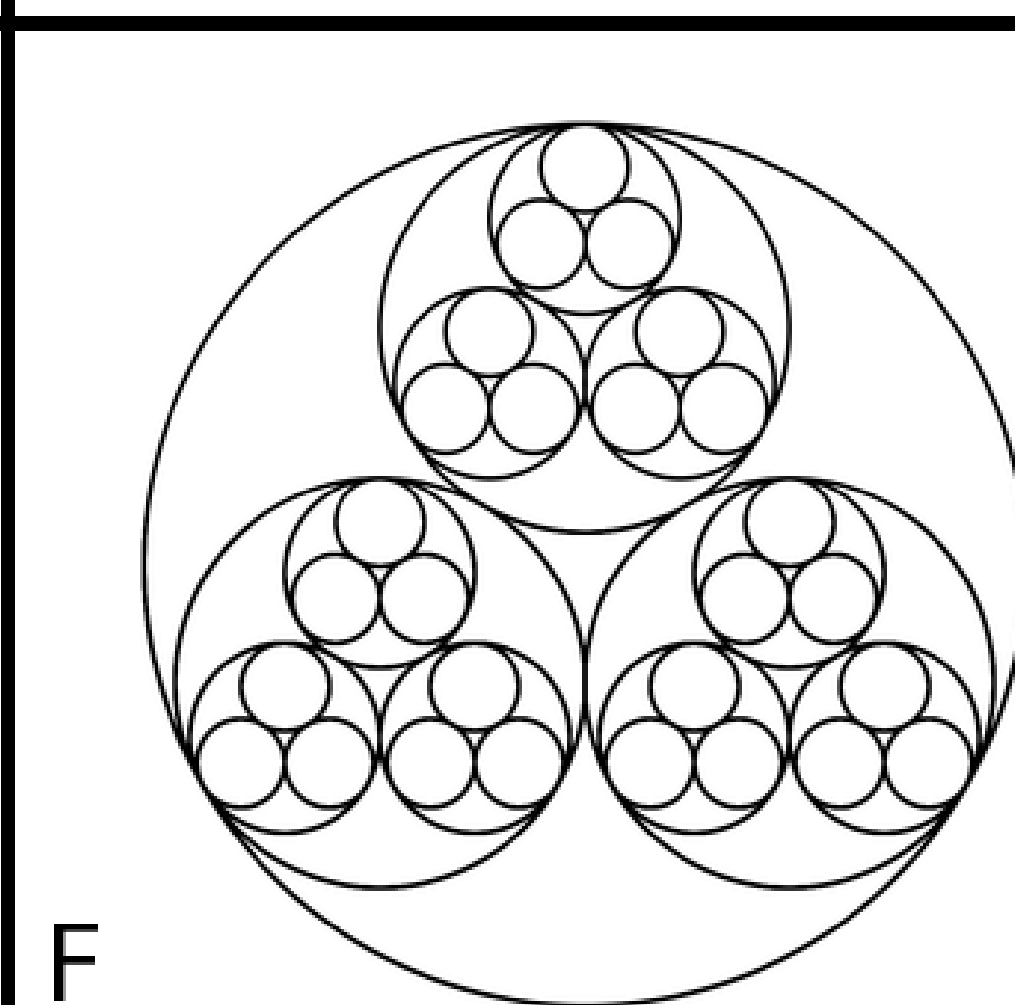
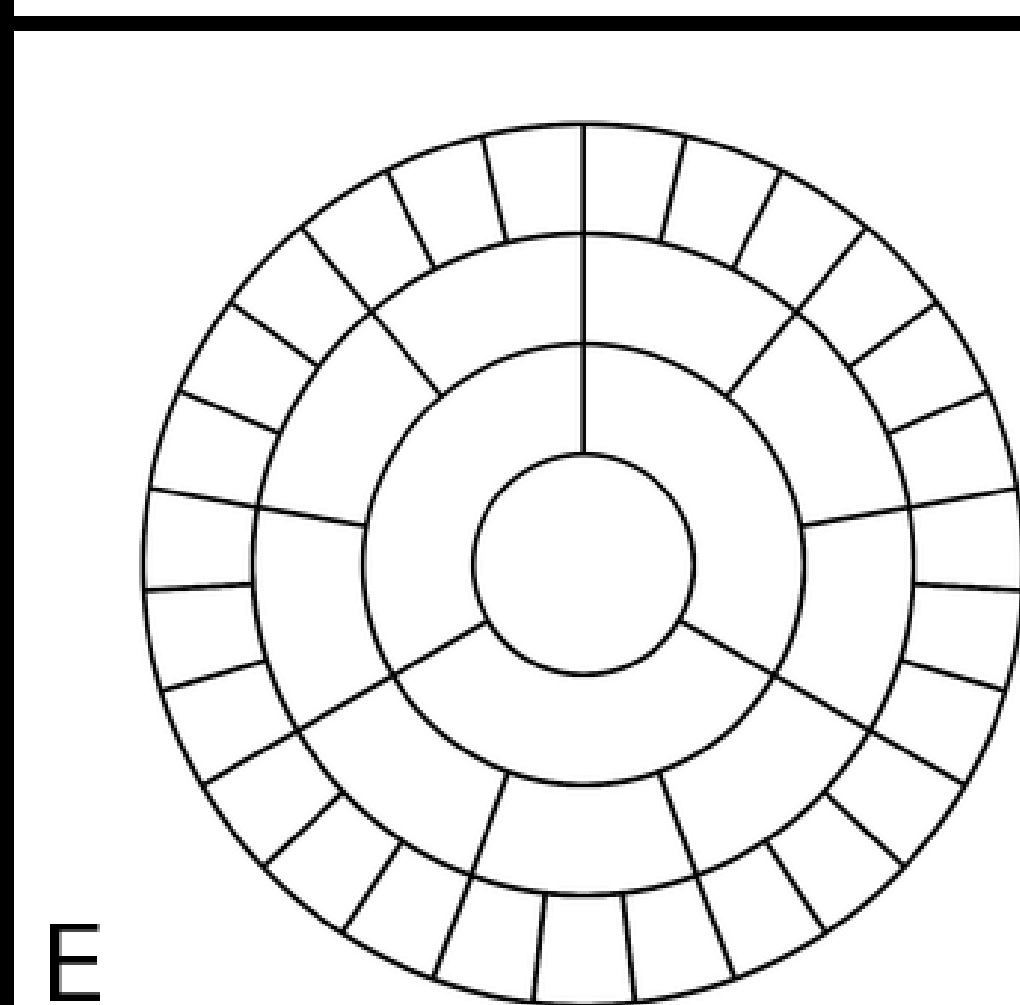
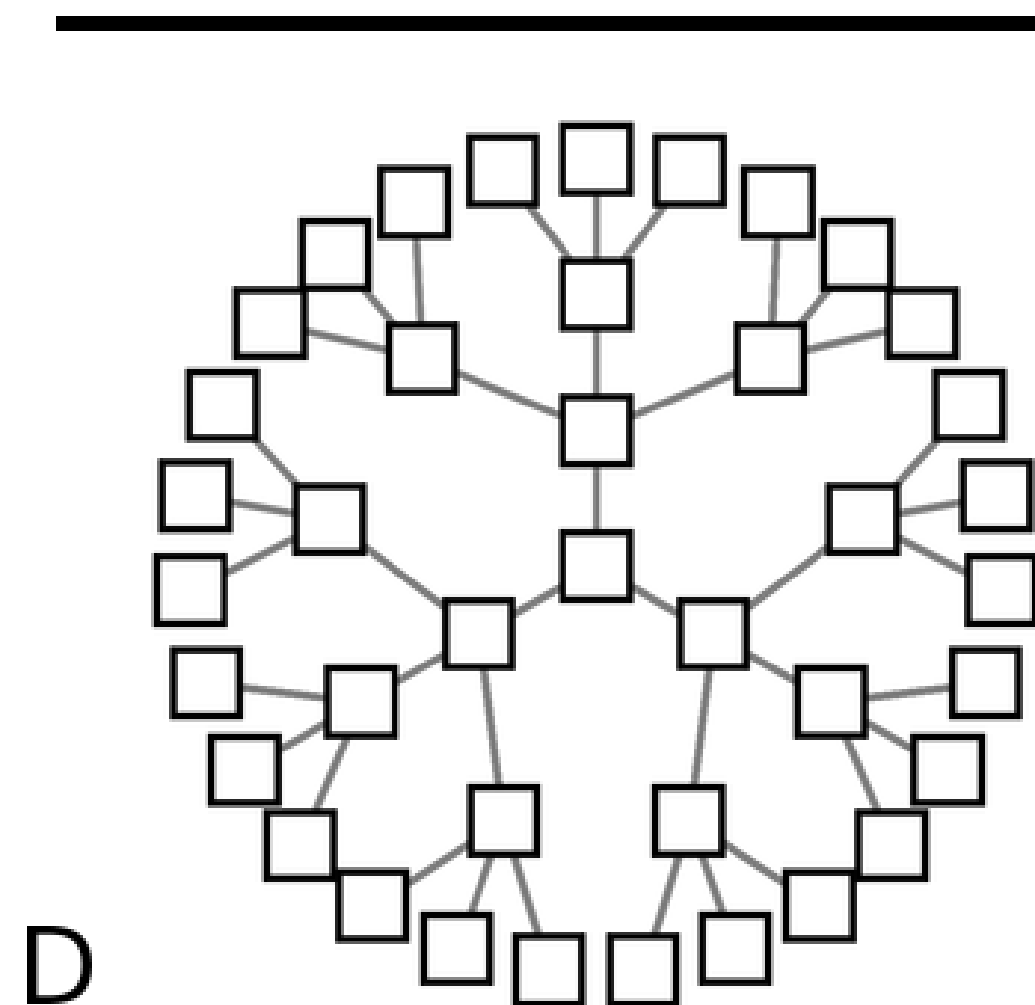
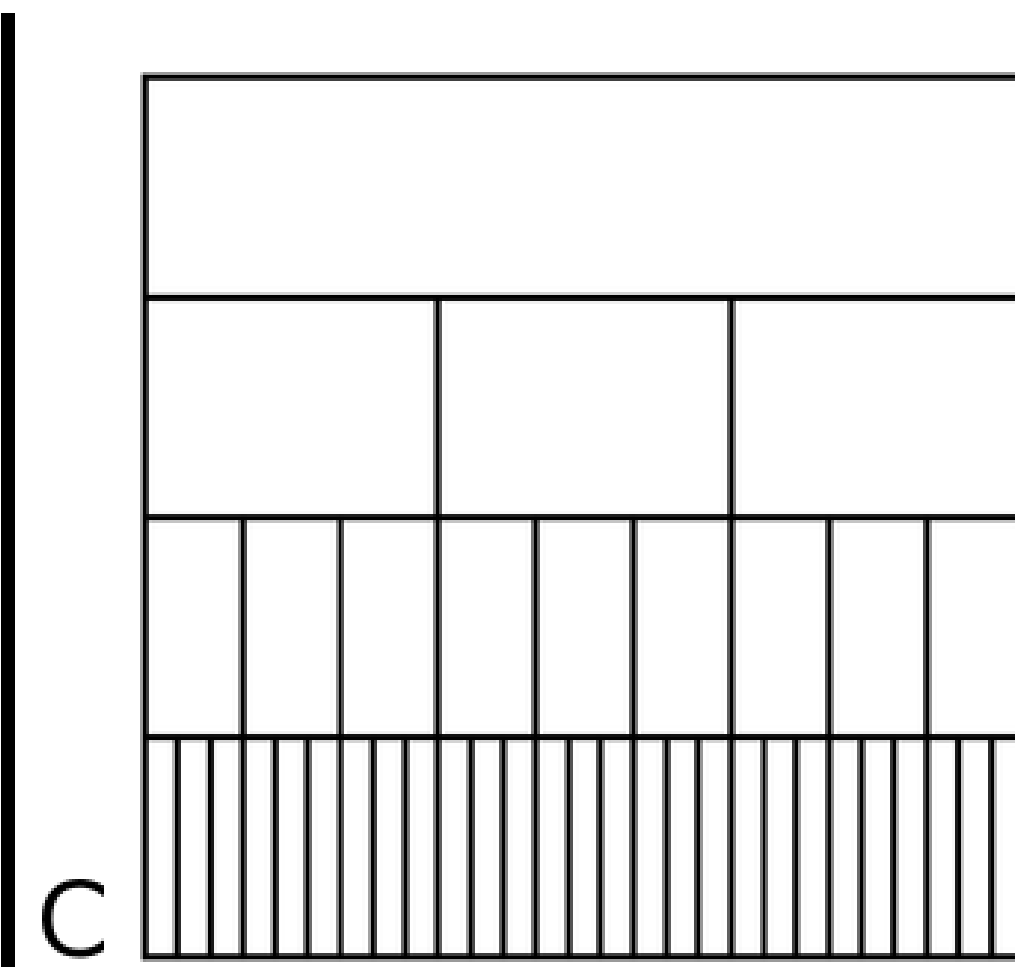
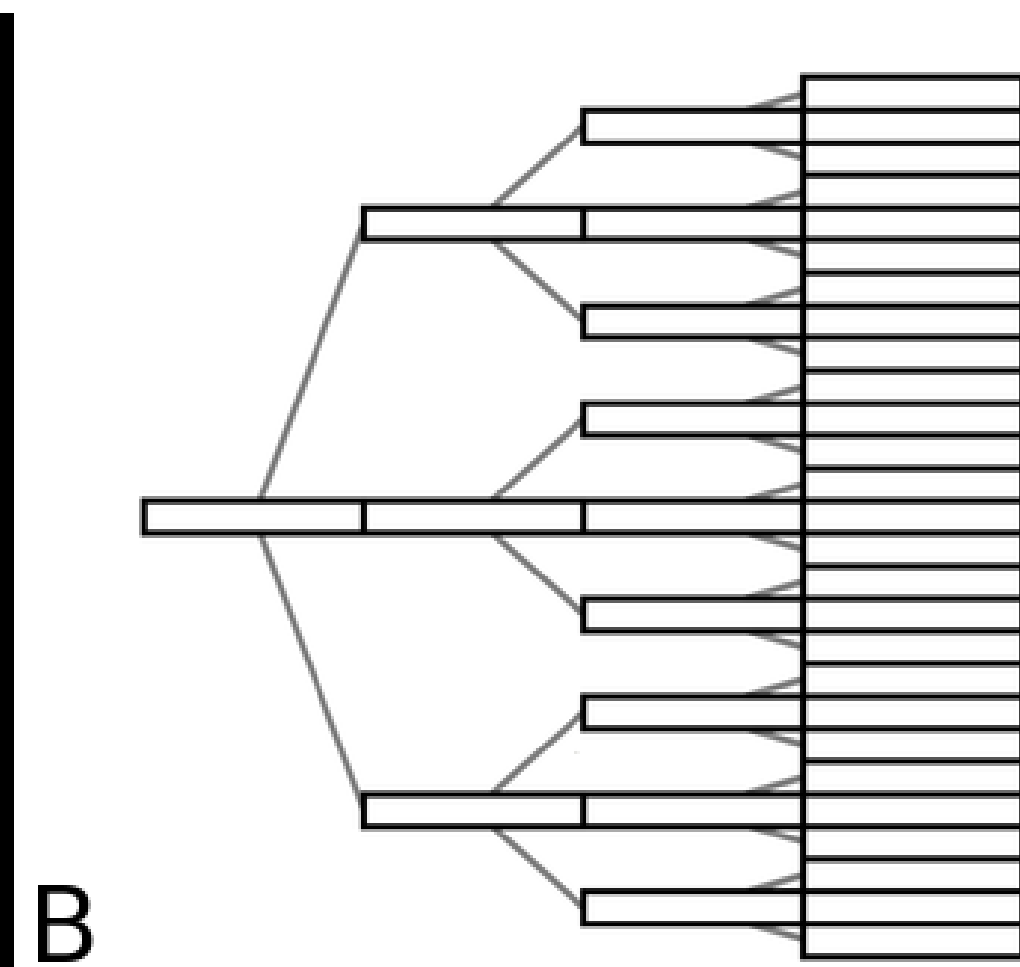
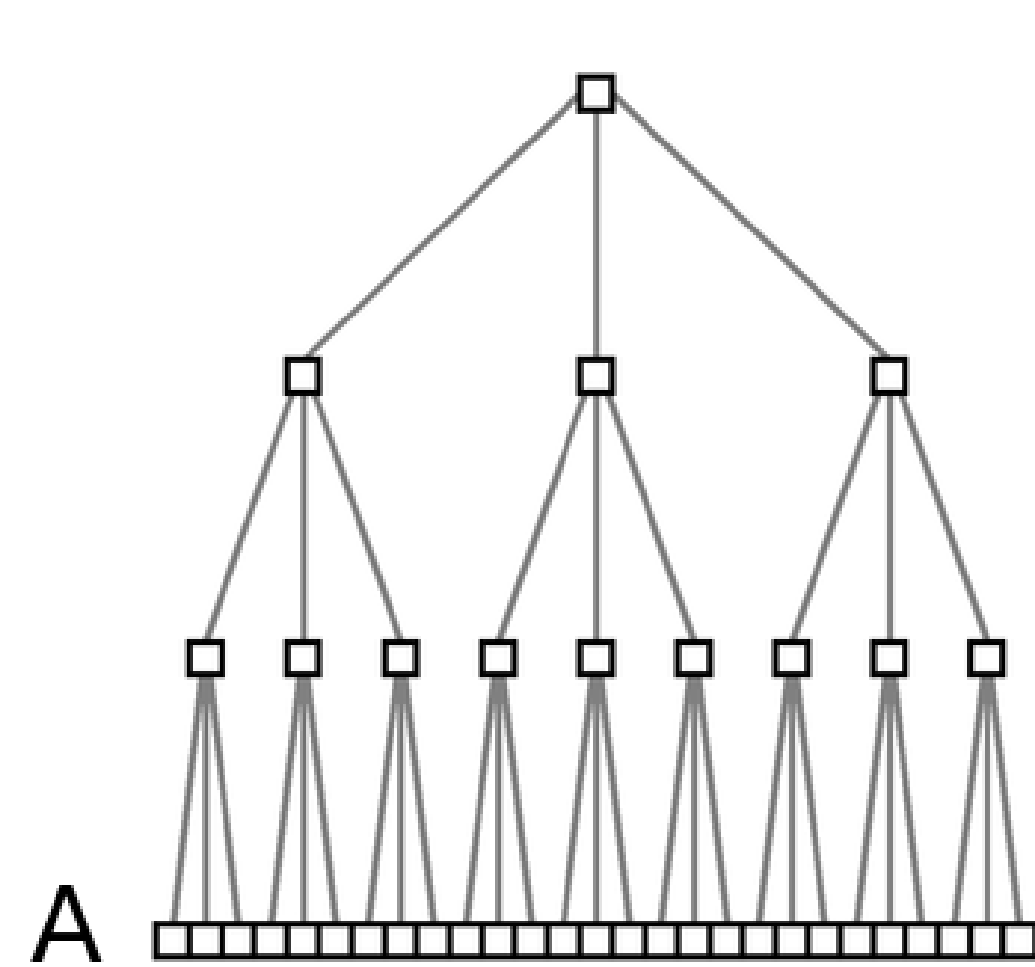
Node-Link or Adjacency Matrix?

- Empirical study: For most tasks, node-link is better for small graphs and adjacency better for large graphs
- Multi-link paths are hard with adjacency matrices
- Immediate connectivity or neighbors are ok, estimating size (nodes & edges also ok)
- People tend to be more familiar with node-link diagrams
- Link density is a problem with node-link but not with adjacency matrices

Trees

- Trees are directed acyclic **networks**
 - each edge has a direction: the origin is the parent, the destination is the child
 - cannot get back to a node after leaving it
- ...plus each node has **at most one** parent node
- A tree has a **root** (every other node hangs off it)
- Can consider enclosure in trees using parent-child relationships

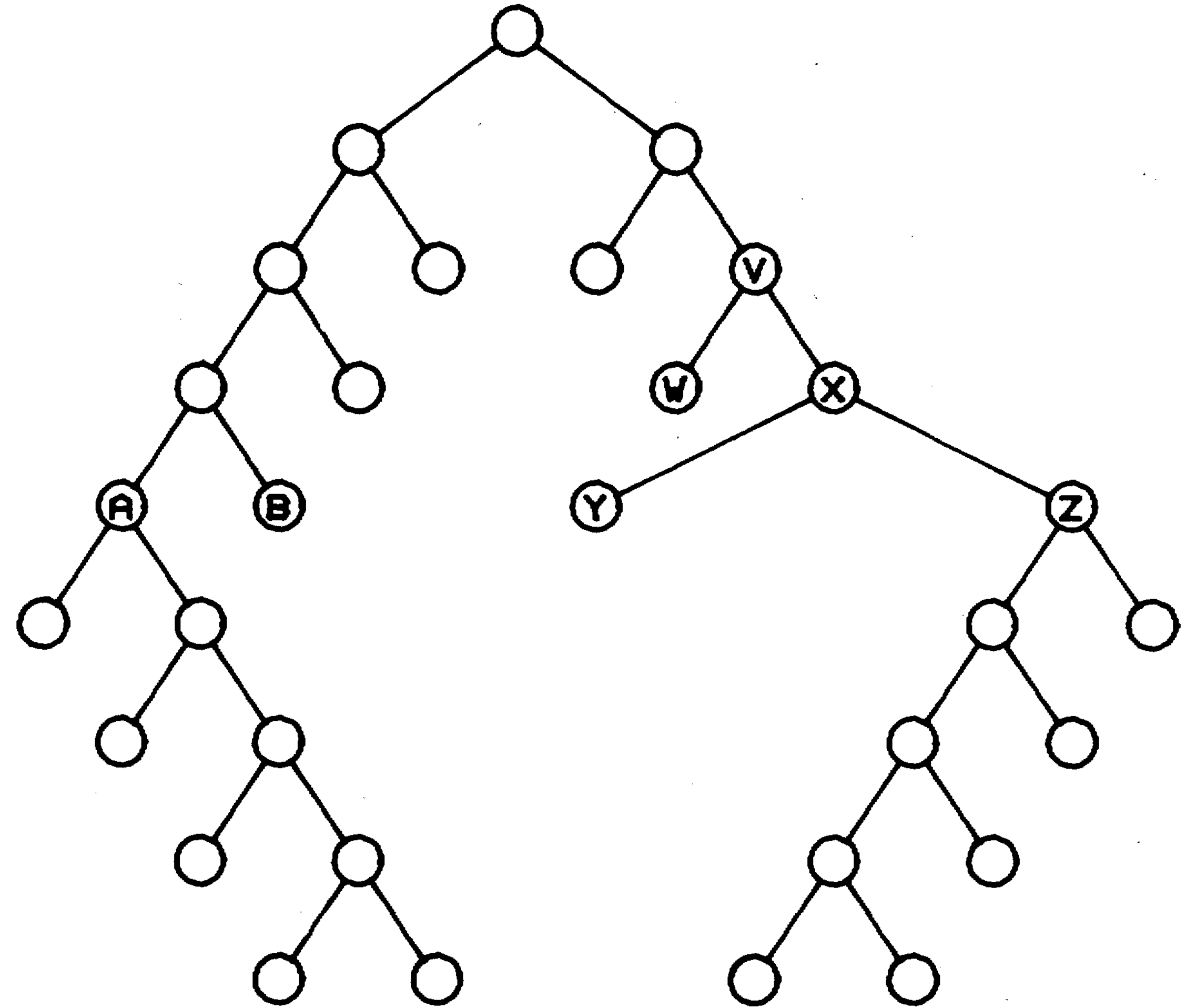
Tree Visualizations



[McGuffin and Robert, 2010]

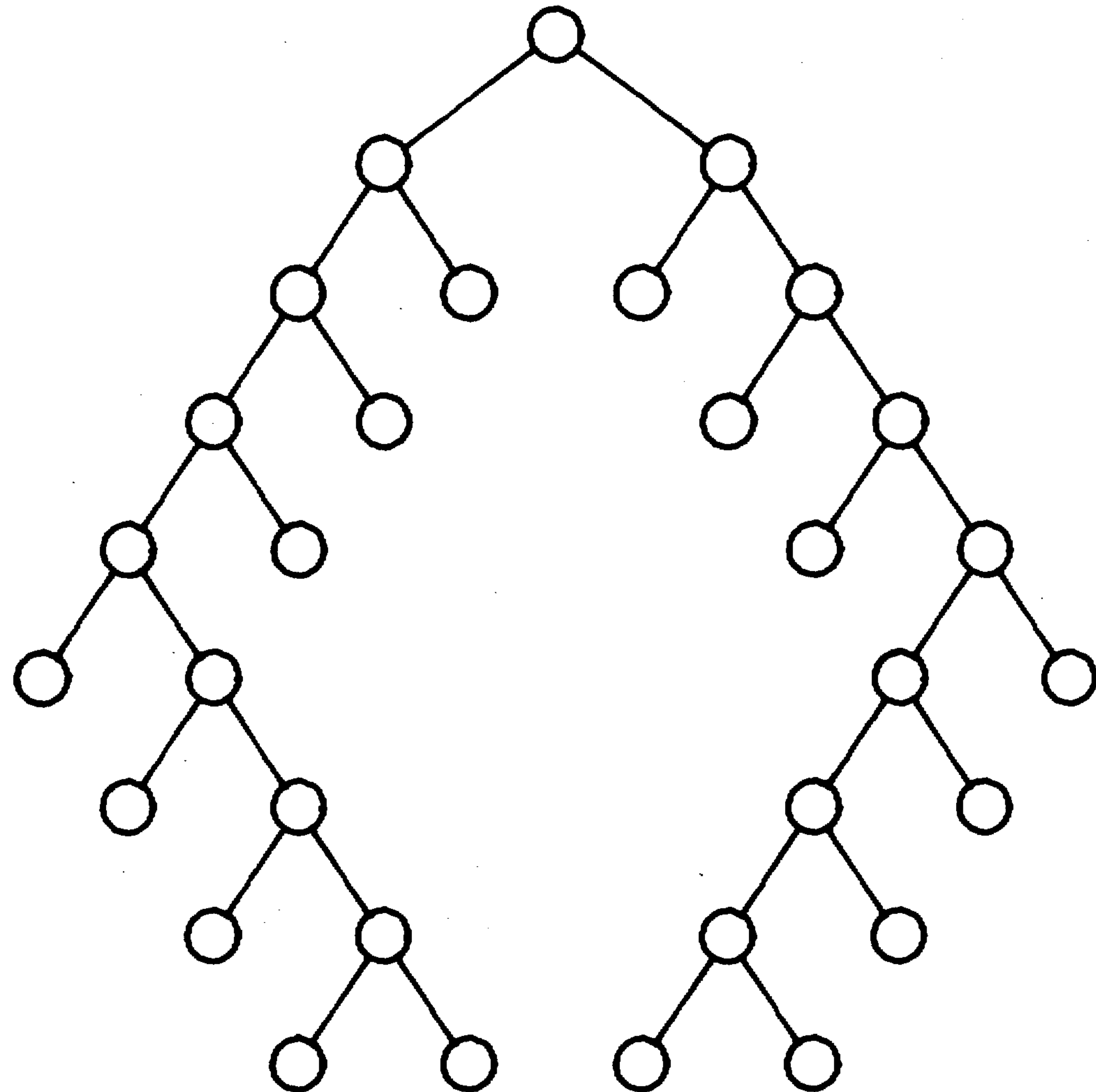
Node-Link Diagram

- Trees are graphs
- ...but we have more structure
- Horizontal or vertical
- Idea 1: partition space for each node via recursion
- Idea 2: “Tidy” Drawing
 - Wetherell & Shannon: Don’t waste space (overlapping parent nodes is ok)
 - Reingold and Tilford: Keep symmetry, subtrees look similar



[WS Alg., Reingold and Tilford, 1981]

Reingold-Tilford Algorithm

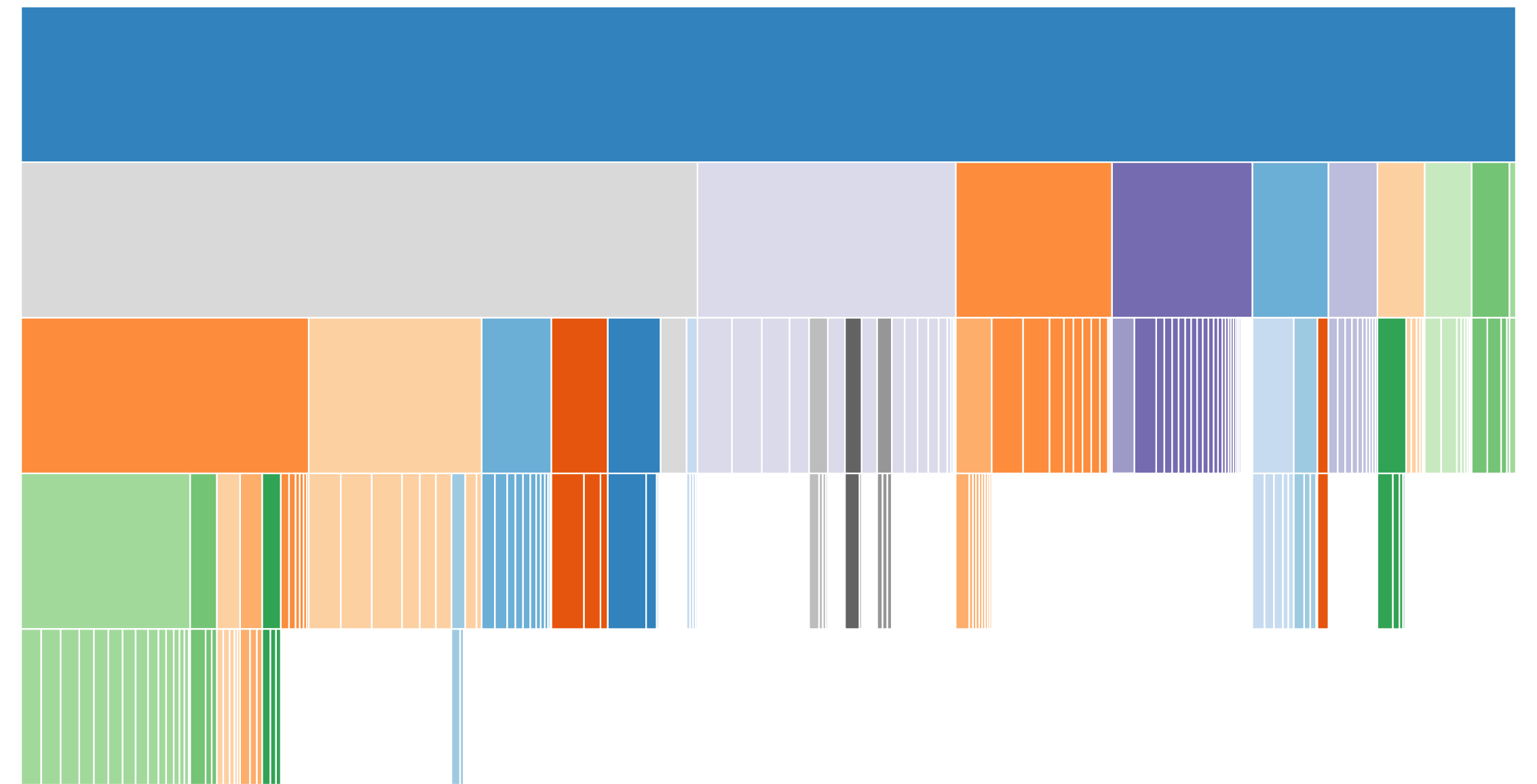


- Recurse on left and right subtrees
- Shift subtree over as long as it doesn't overlap
- Place parent centered above the subtrees
- Originally, only binary trees, extended by Walker

[Reingold and Tilford, 1981]

Icicle Plot

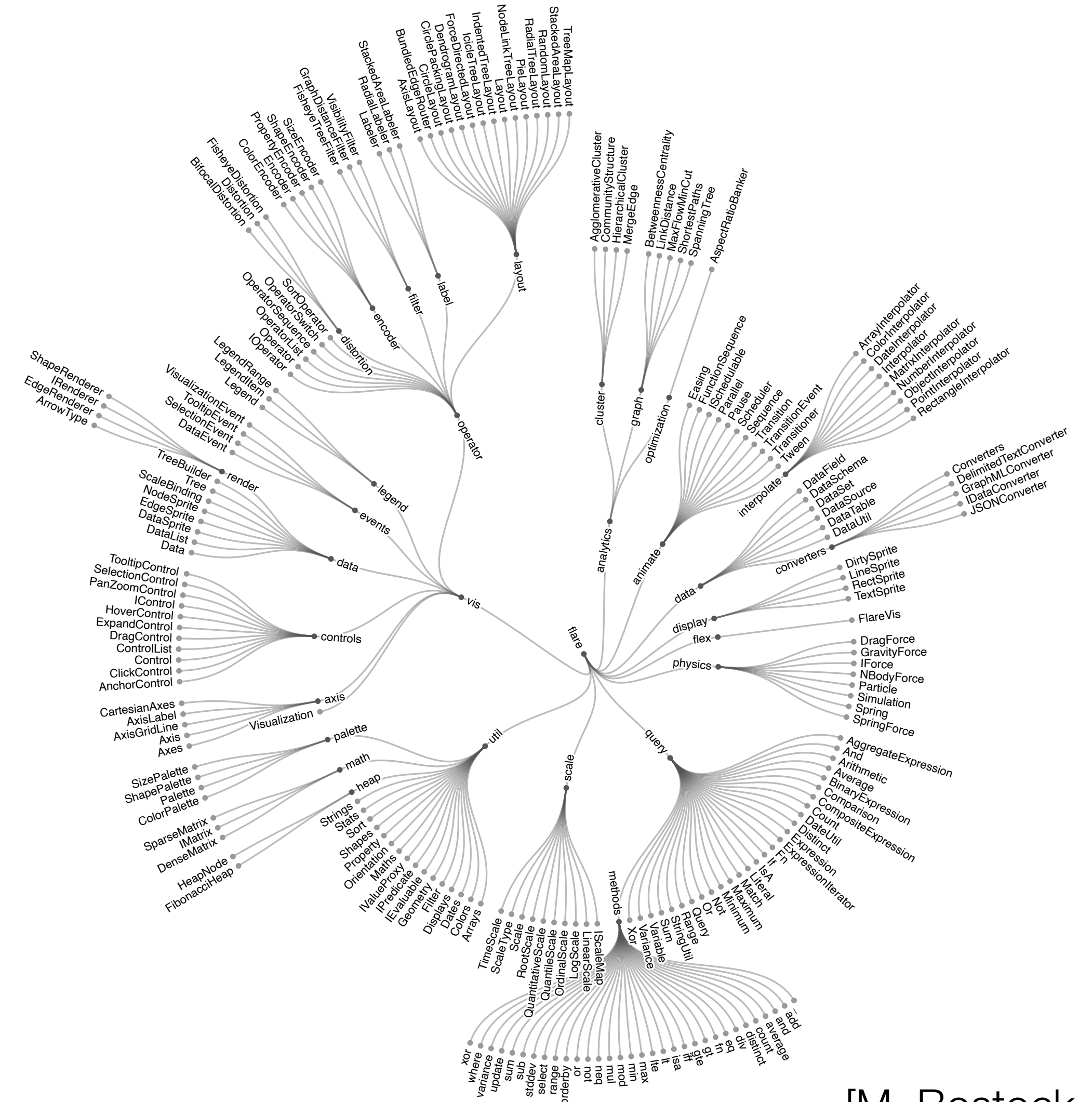
- Line marks
- Vertical position shows depth
- Horizontal position shows links and sibling order
- Scalability: 1 pixel leaves, but harder to label



[Bostock, 2011]

Radial Node-Link

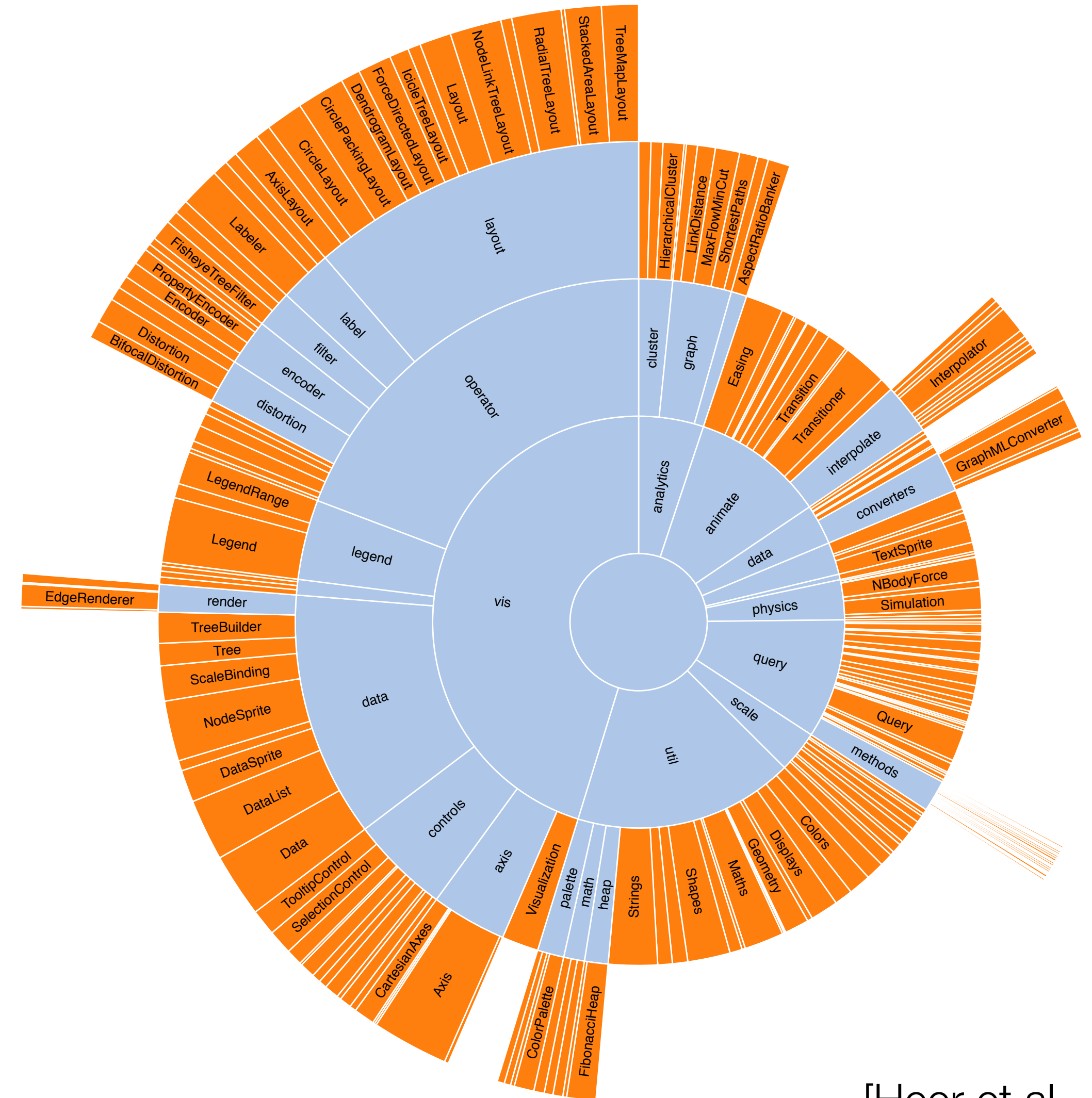
- Use polar coordinates instead of rectilinear
- Same layout algorithms work (e.g. Reingold-Tilford)
- Benefit: space usage, labels



[M. Bostock, 2017]

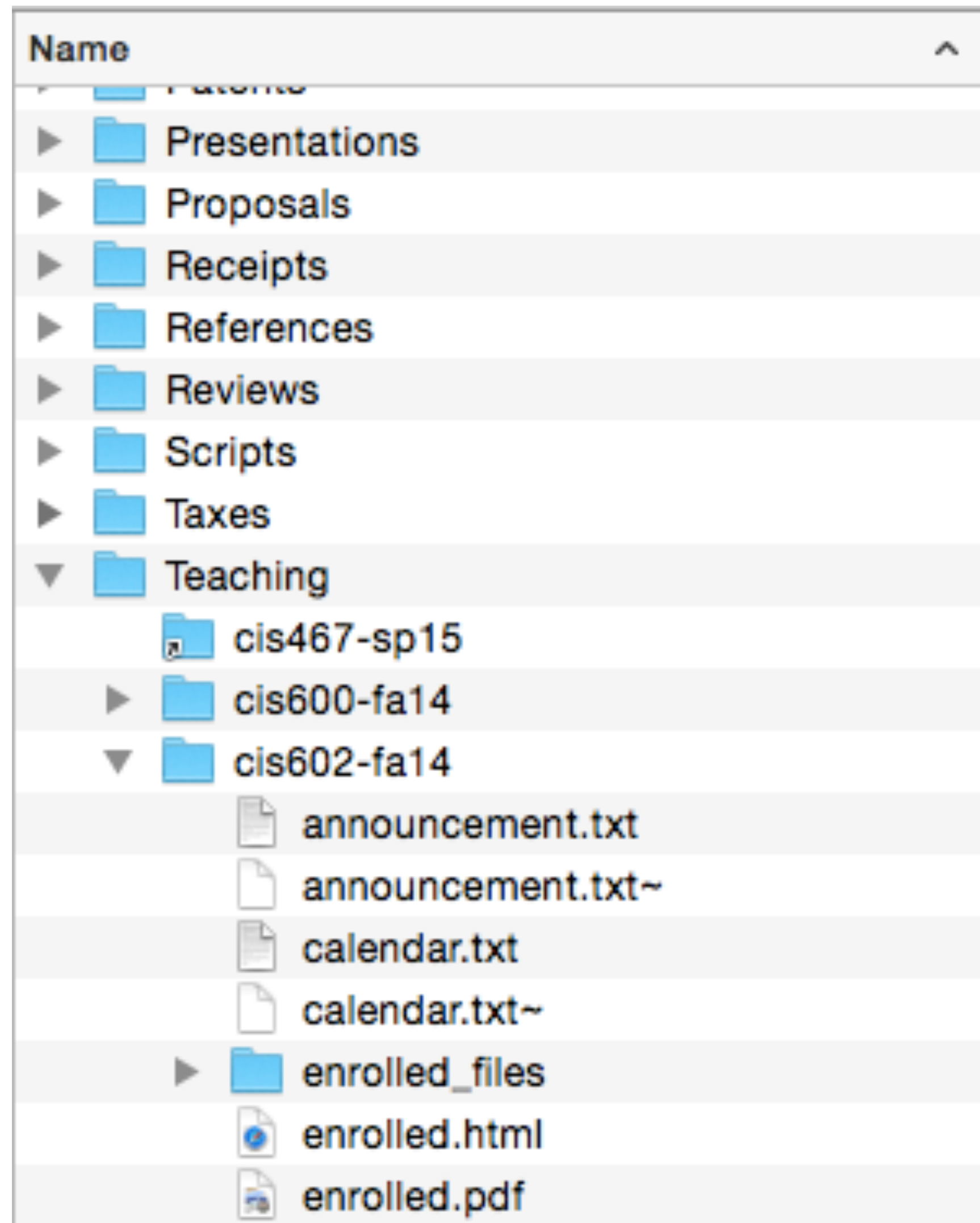
Sunburst

- Icicle plot in a radial layout
- Reading labels?
- Intuitive navigation



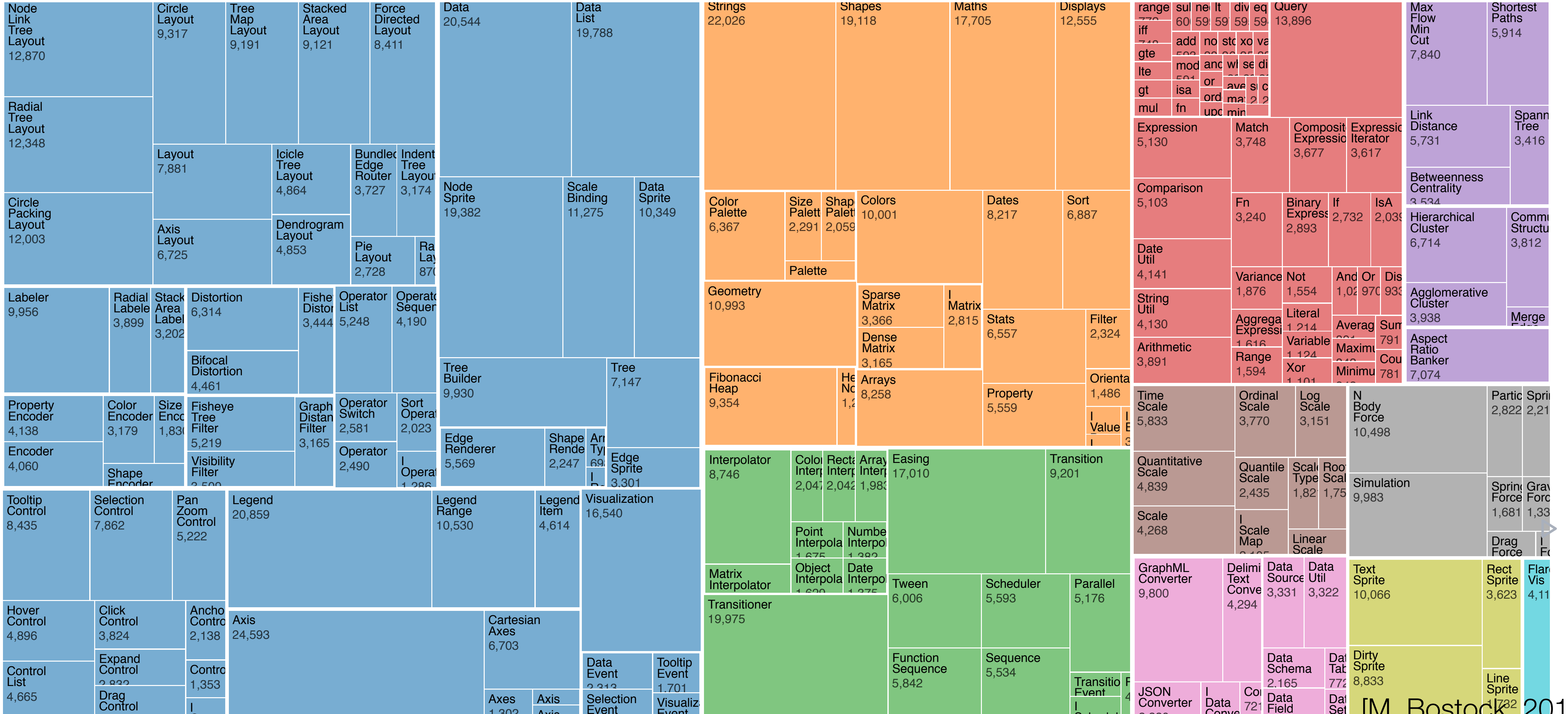
[Heer et al., 2012]

Indented Outline



- Like a filesystem tree
- Use horizontal position to show depth, vertical positions show sibling/order

Treemap



[M. Bostock, 2017]

Car/Truck Treemap

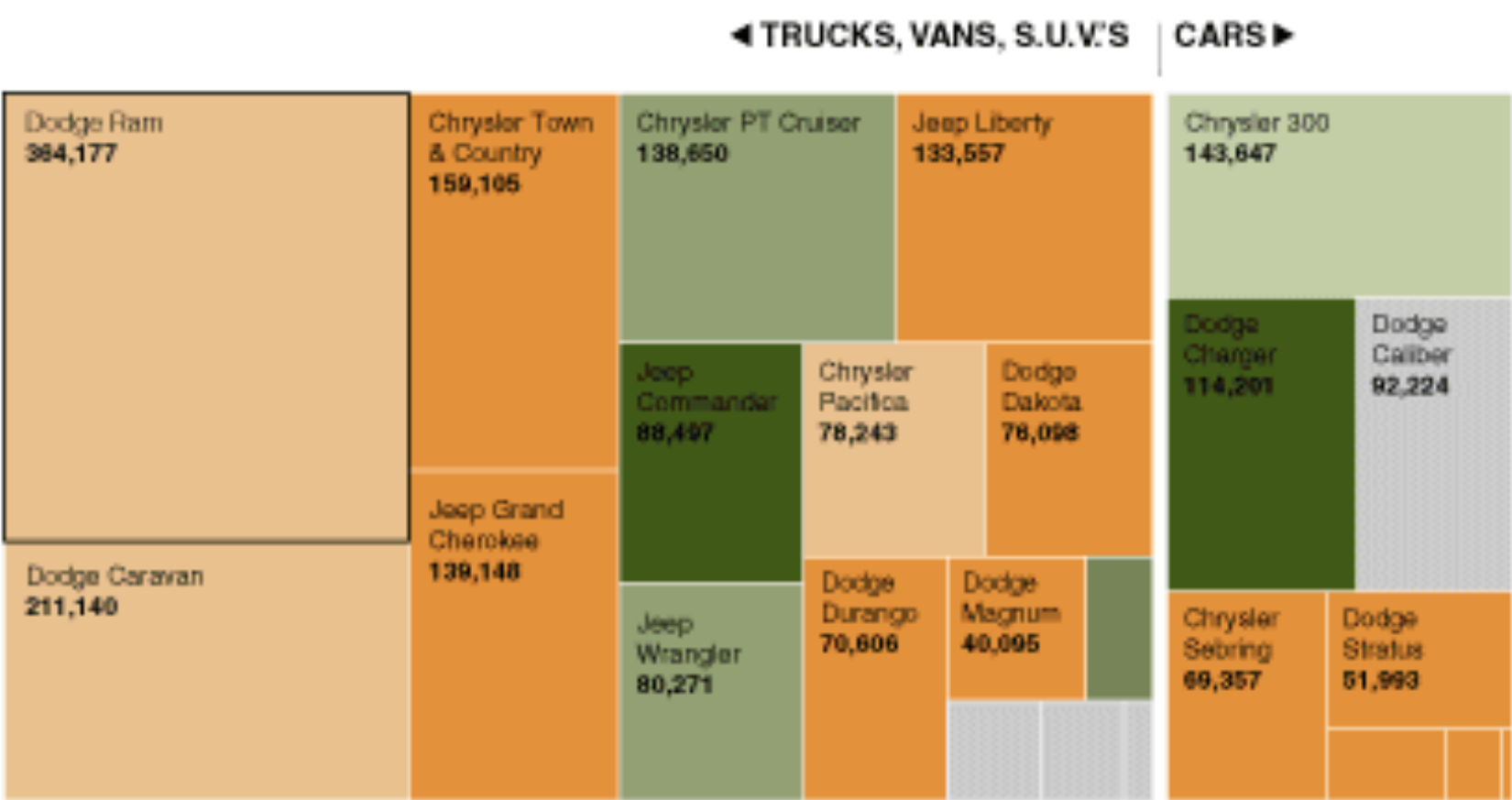
Truck Sales Slip, Tripping Up Chrysler

Over the past few years, Chrysler executives said they were following the lead of Toyota and Honda, focusing on vehicles that met the needs of their customers. But as American consumers turned away from large trucks and S.U.V.'s in 2006, Chrysler continued to churn out big vehicles, which are now sitting unsold at dealerships across the country.

Chrysler Group **-7.0%**
Trucks/vans/S.U.V.'s 1.6 million
Cars 0.5 million

Pickups, minivans and S.U.V.'s made up 76 percent of Chrysler's sales, which left it vulnerable when consumers shifted to cars.

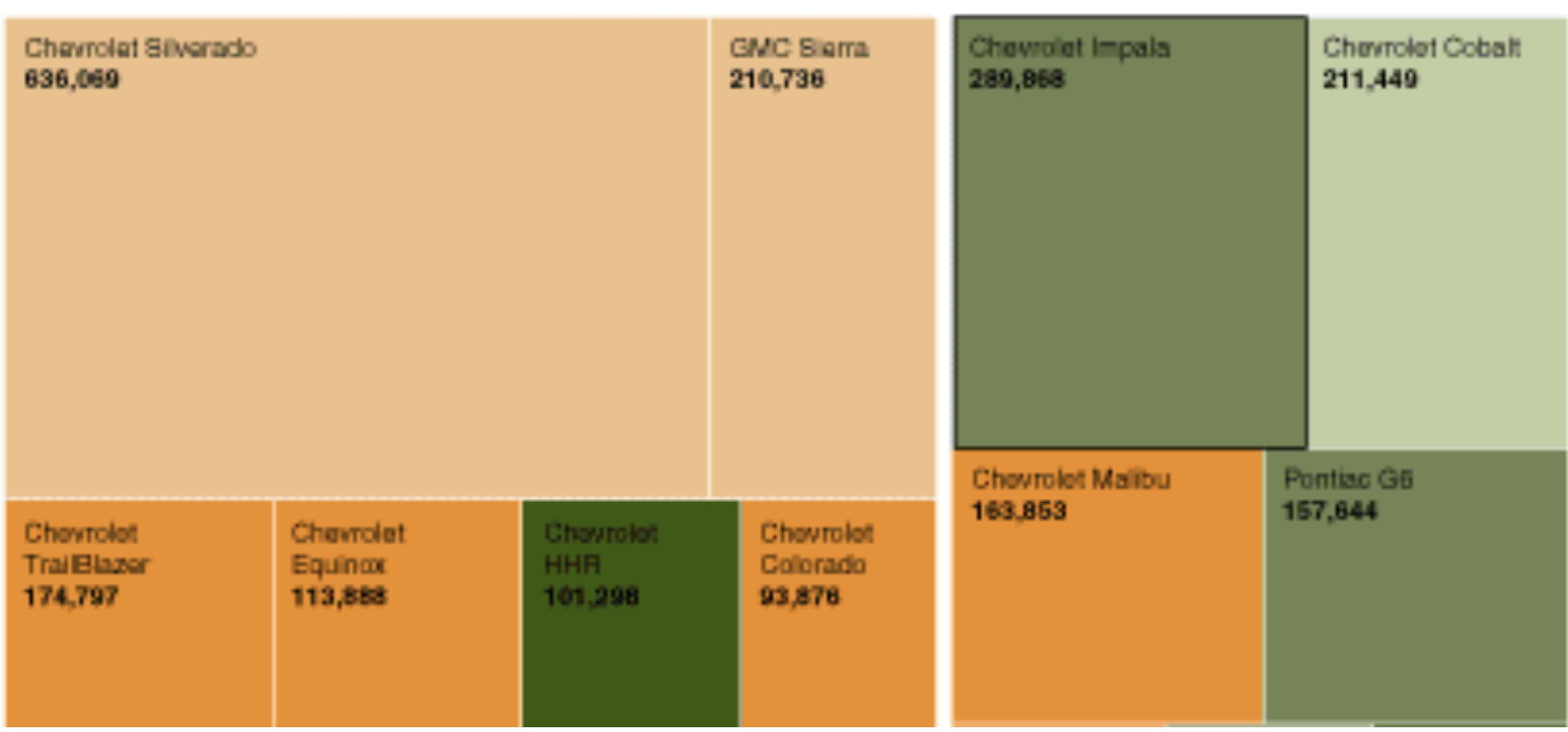
Dodge Ram



General Motors **-8.7%**
Trucks/vans/S.U.V.'s 2.5 million
Cars 1.6 million

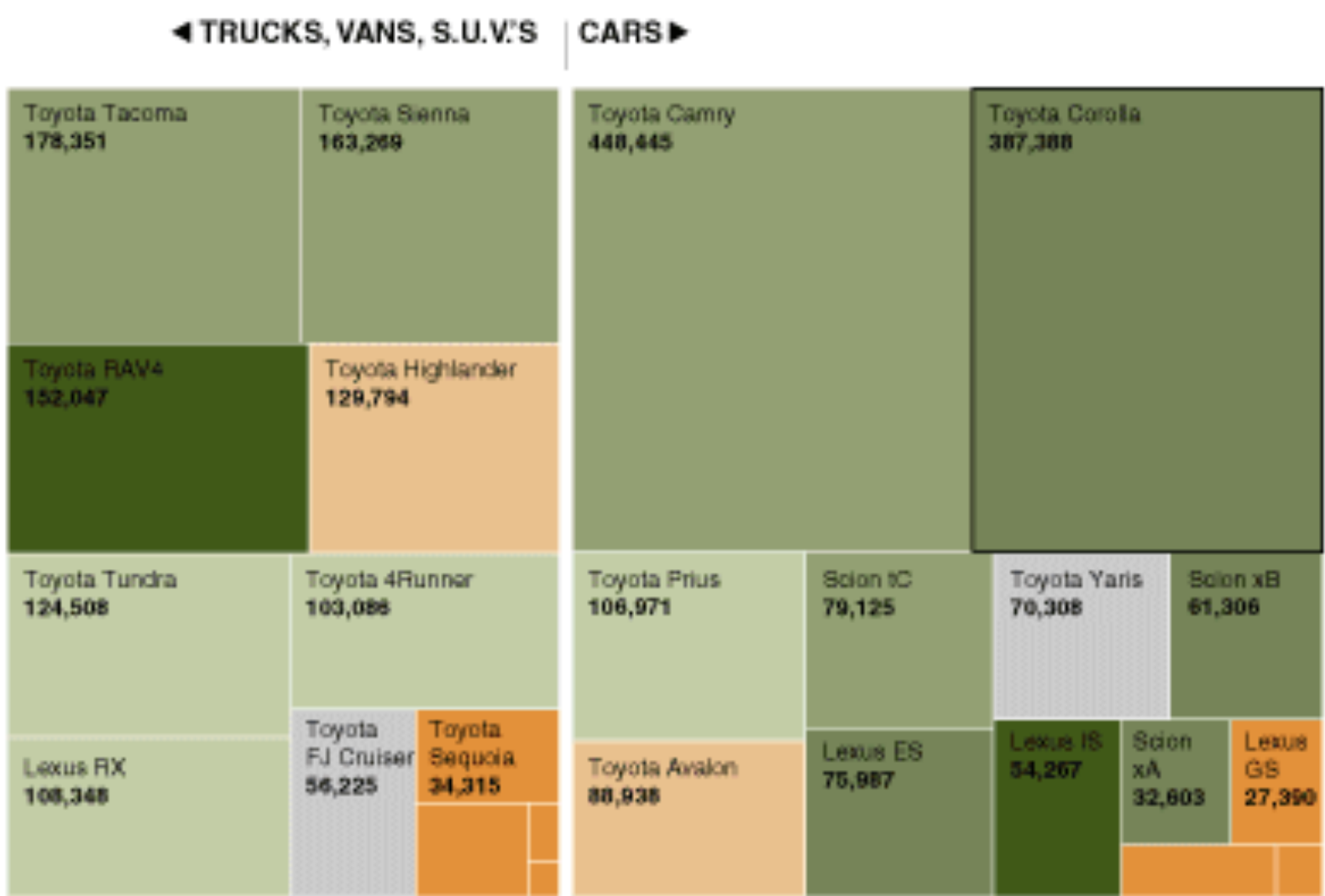
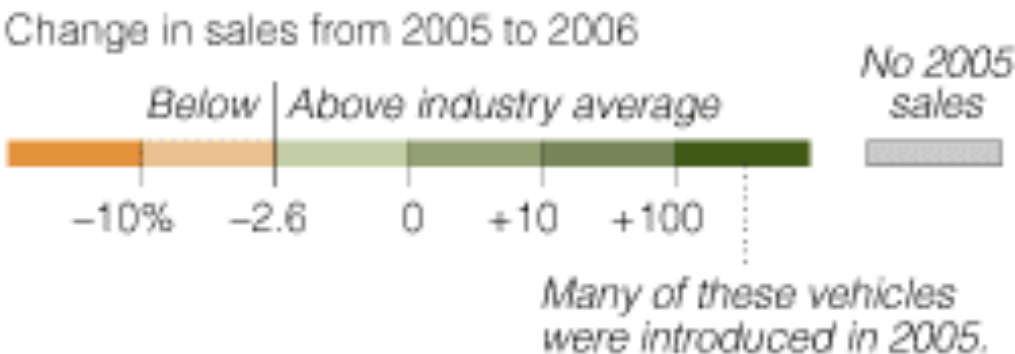
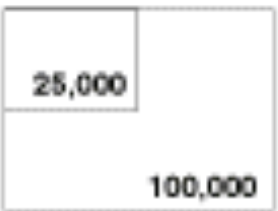
G.M. introduced new versions of its large S.U.V.'s in late 2005, hoping they would bolster sales. Instead, sales of big vehicles were hurt when gas prices climbed. One of the few standouts was the Chevrolet HHR, new in 2005.

Chevrolet



READING THE CHART

Boxes are scaled proportionally according to number of cars sold in 2006



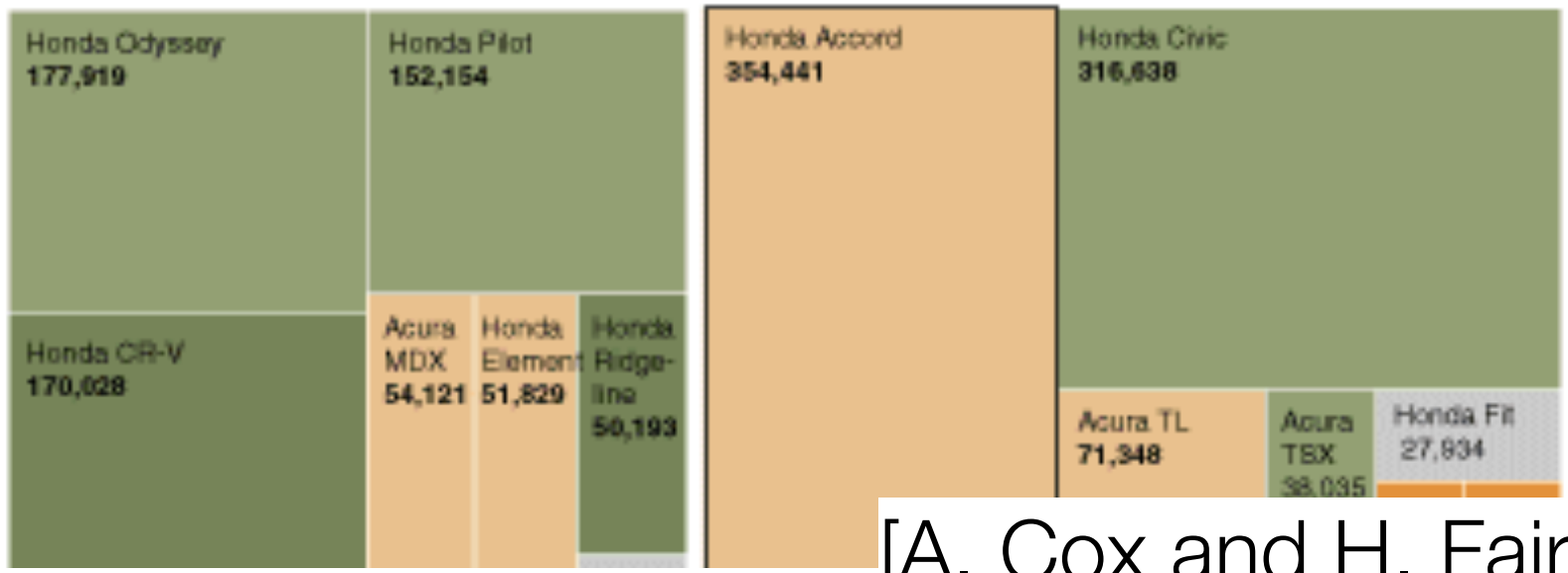
Toyota **+12.5%**
Trucks/vans/S.U.V.'s 1.1 million
Cars 1.5 million

Toyota rolled out a new version of the Camry, and once again it was the country's best-selling car.

Toyota Corolla



Corolla sales also jumped, along with gas prices. Toyota could not escape the decline in sales of supersized S.U.V.'s like its Sequoia.

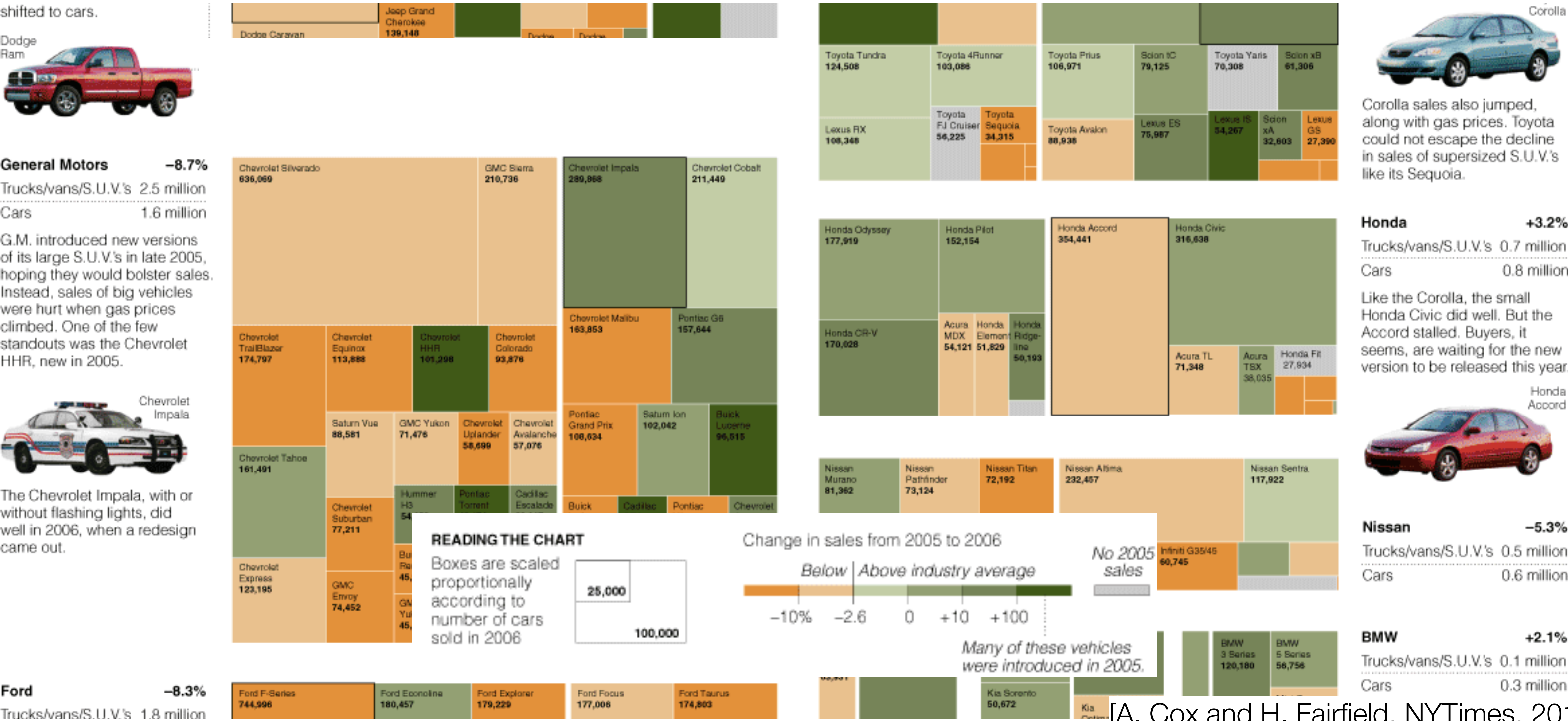


Honda **+3.2%**
Trucks/vans/S.U.V.'s 0.7 million
Cars 0.8 million

Like the Corolla, the small Honda Civic did well. But the Accord stalled. Buyers, it seems, are waiting for the new version to be released this year.

[A. Cox and H. Fairfield, NYTimes, 2012]

Car/Truck Treemap



[A. Cox and H. Fairfield, NYTimes, 2012]

Treemap

- Containment marks instead of connection marks
- Encodes some attribute of the items as the **size** of the rectangles
- Not as easy to see the intermediate rectangles
- Scalability: millions of leaf nodes and links possible

- Need a layout algorithm!

Layout Algorithms

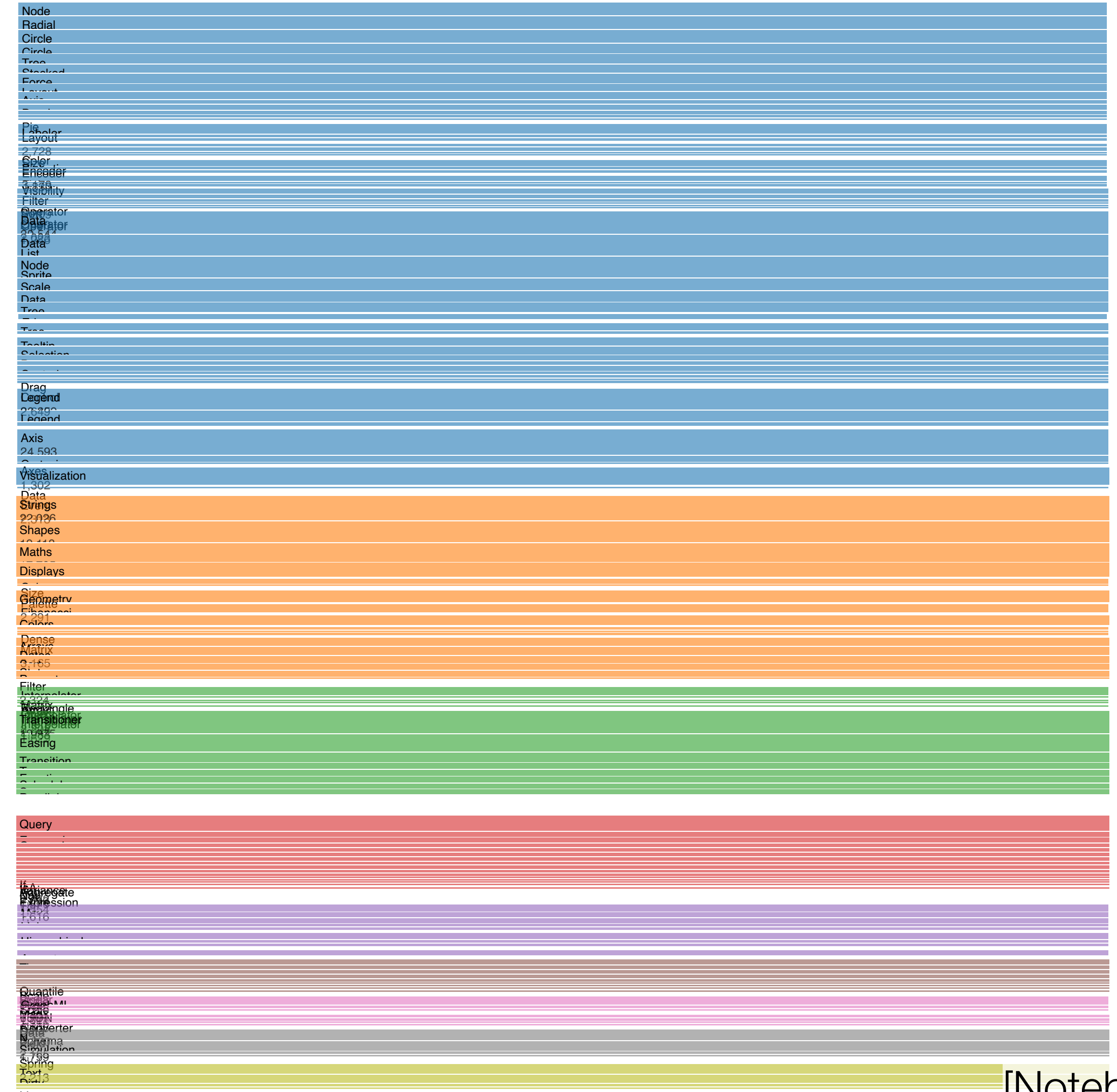
- How do we generate the area marks?
- What considerations should we try to keep in mind?

Layout Algorithms

- How do we generate the area marks?
- What considerations should we try to keep in mind?
 - area true to quantitative value
 - show hierarchy
 - aspect ratio
- Also...
 - ordering
 - stability

Treemap Layouts: Slice

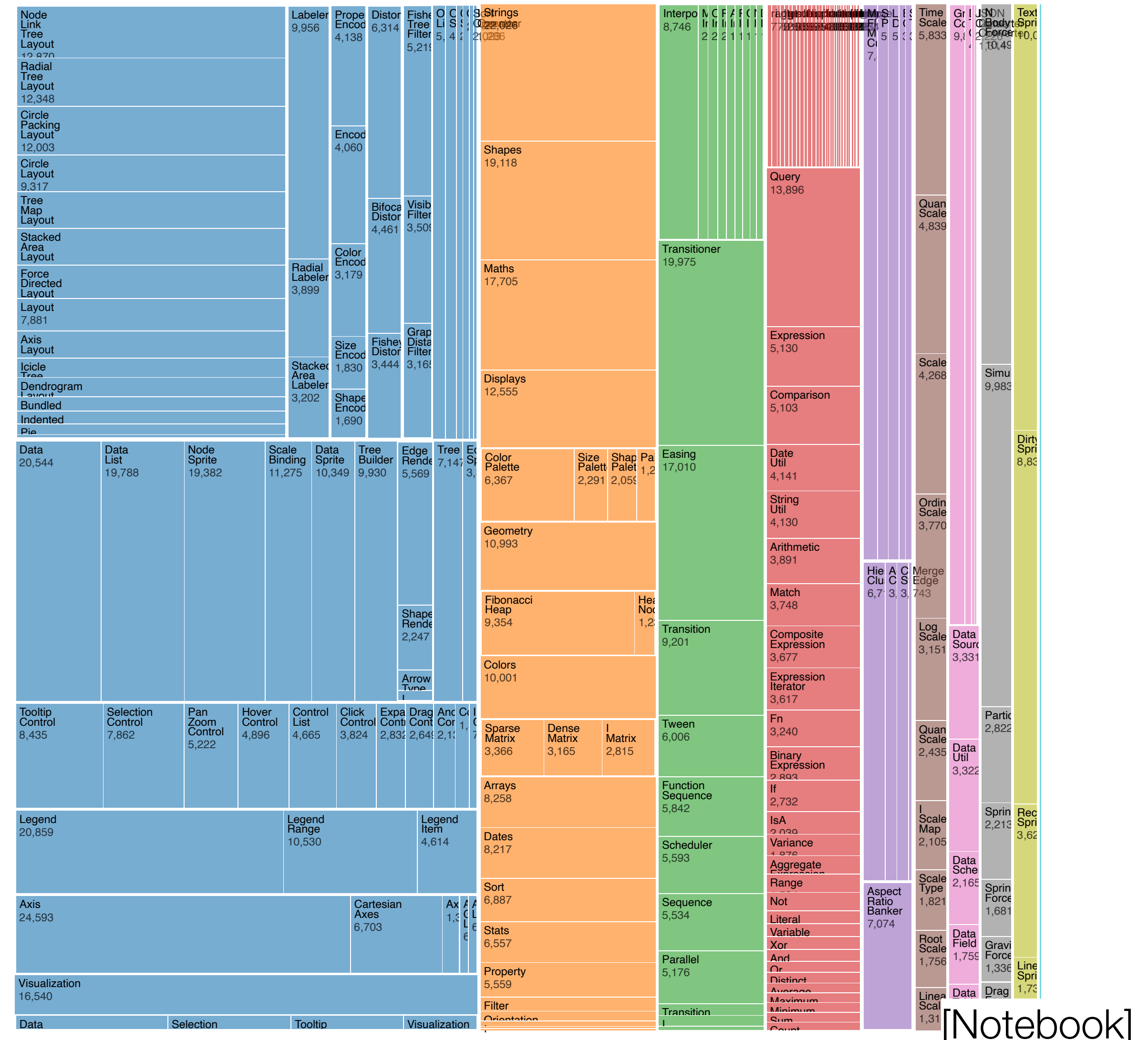
- Just divide horizontally
- Dice is similar, just vertical
- Problem: Bad aspect ratio!
 - Very skinny rectangles
 - Makes it harder to compare sizes, see labels, select rectangles
 - Want rectangles that are closer to squares
 - Aspect ratio = width/height



[Notebook]

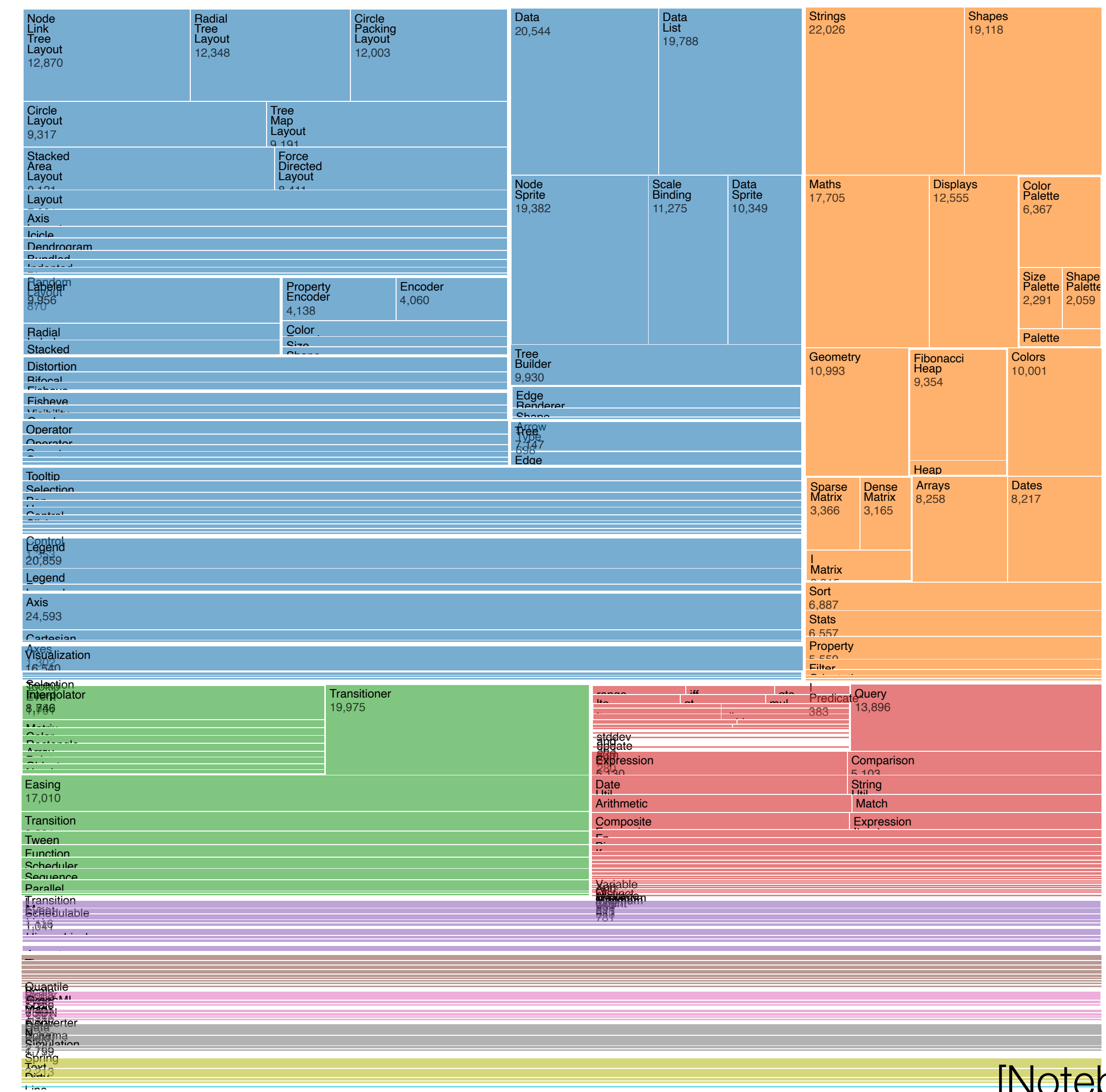
Treemap Layouts: Slice & Dice

- Split at each level into strips
- At each step, orientation of division (horizontal/vertical) changes
- Better, but some rectangles still have bad aspect ratio



Treemap Layouts: Strip

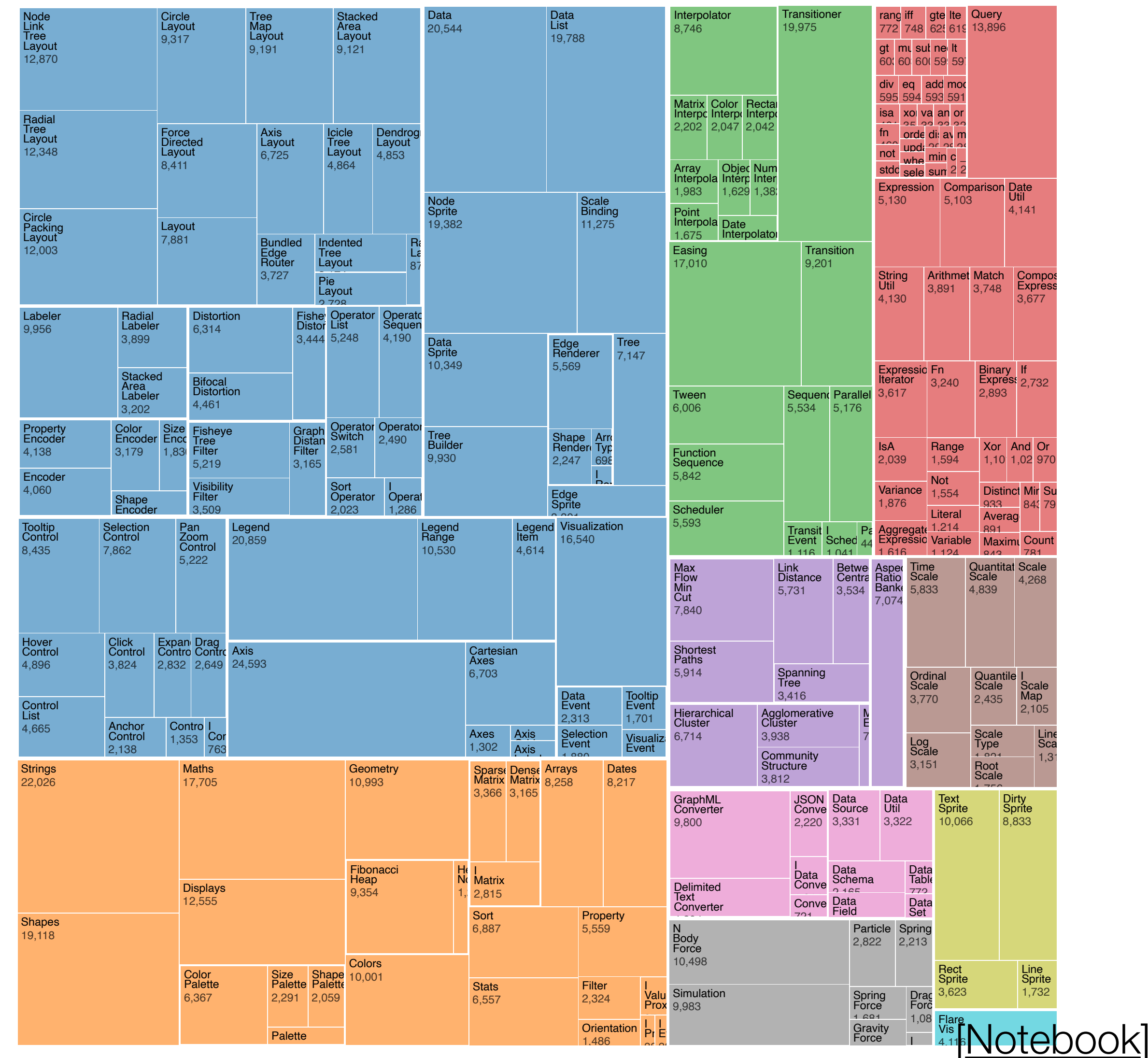
- Consider aspect ratio when adding rectangles
- Do one row at a time by processing rectangles in sorted order by size
 - Check if adding the next rectangle to the row improves aspect ratio
 - When it doesn't, go to next row
- Problem: Last rectangles have bad aspect ratios
- Solution: Look ahead to decide if would be better to add to previous row



[Notebook]

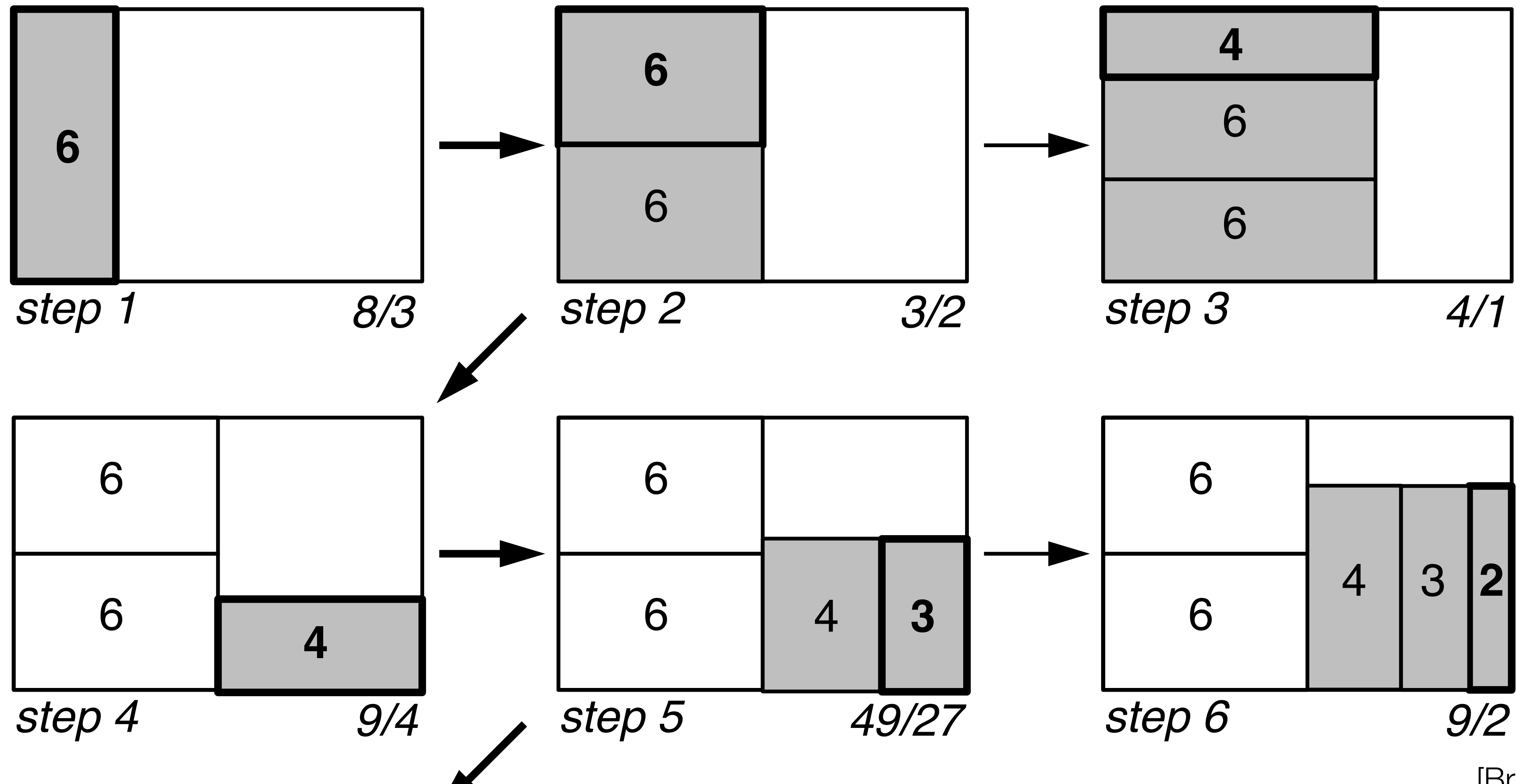
Treemap Layouts: Squarify

- Slice & Dice and Strip can lead to bad **aspect ratios**
- Solution: Strip only uses rows, allow columns to be used, too
- Choose divisions (x/y) based on the width/height of region in order to maintain good aspect ratios
 - Use left and right side
 - Process large rectangles first
- Ordering not preserved which may cause issues if the data is updated



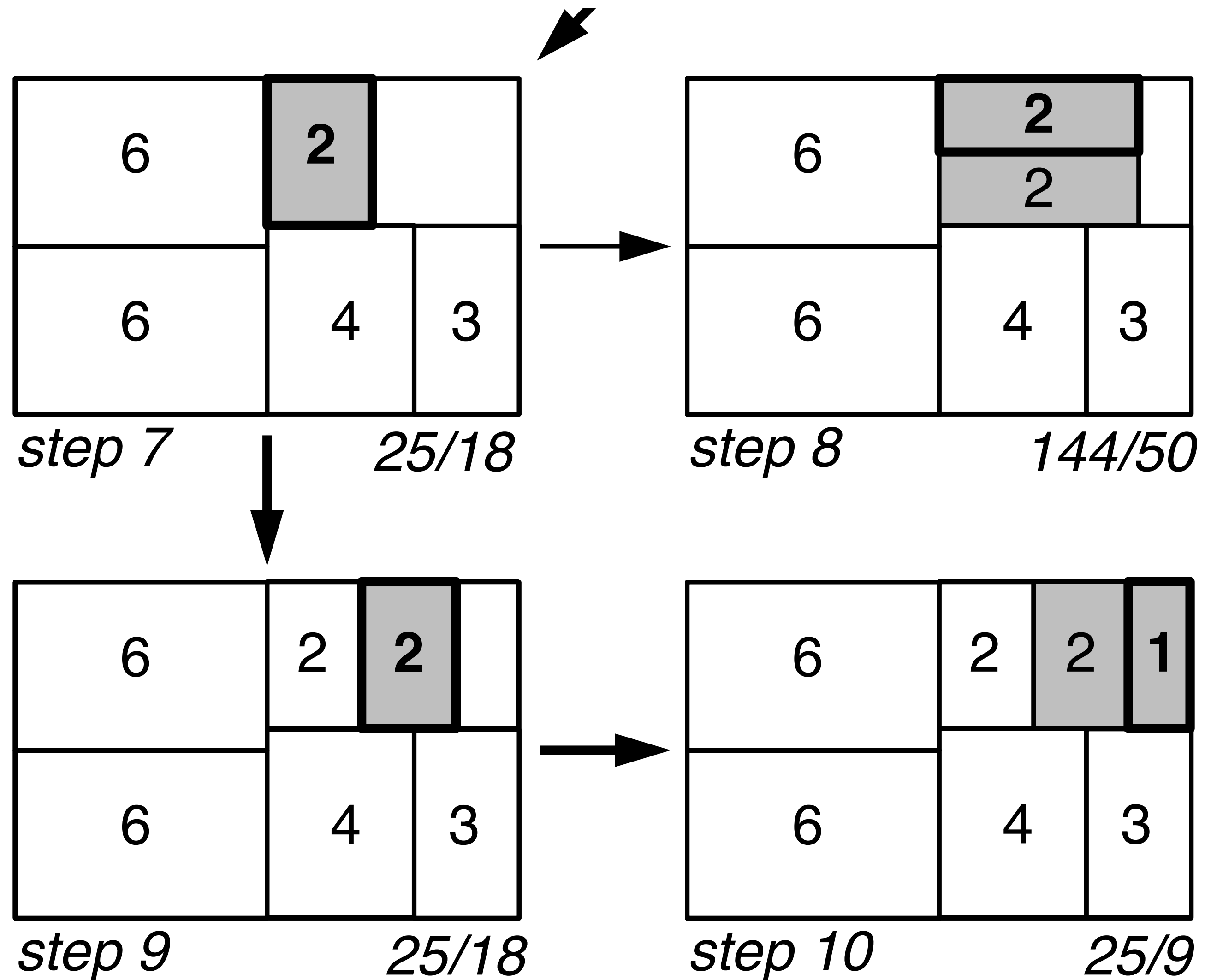
[Notebook]

Squarification Algorithm



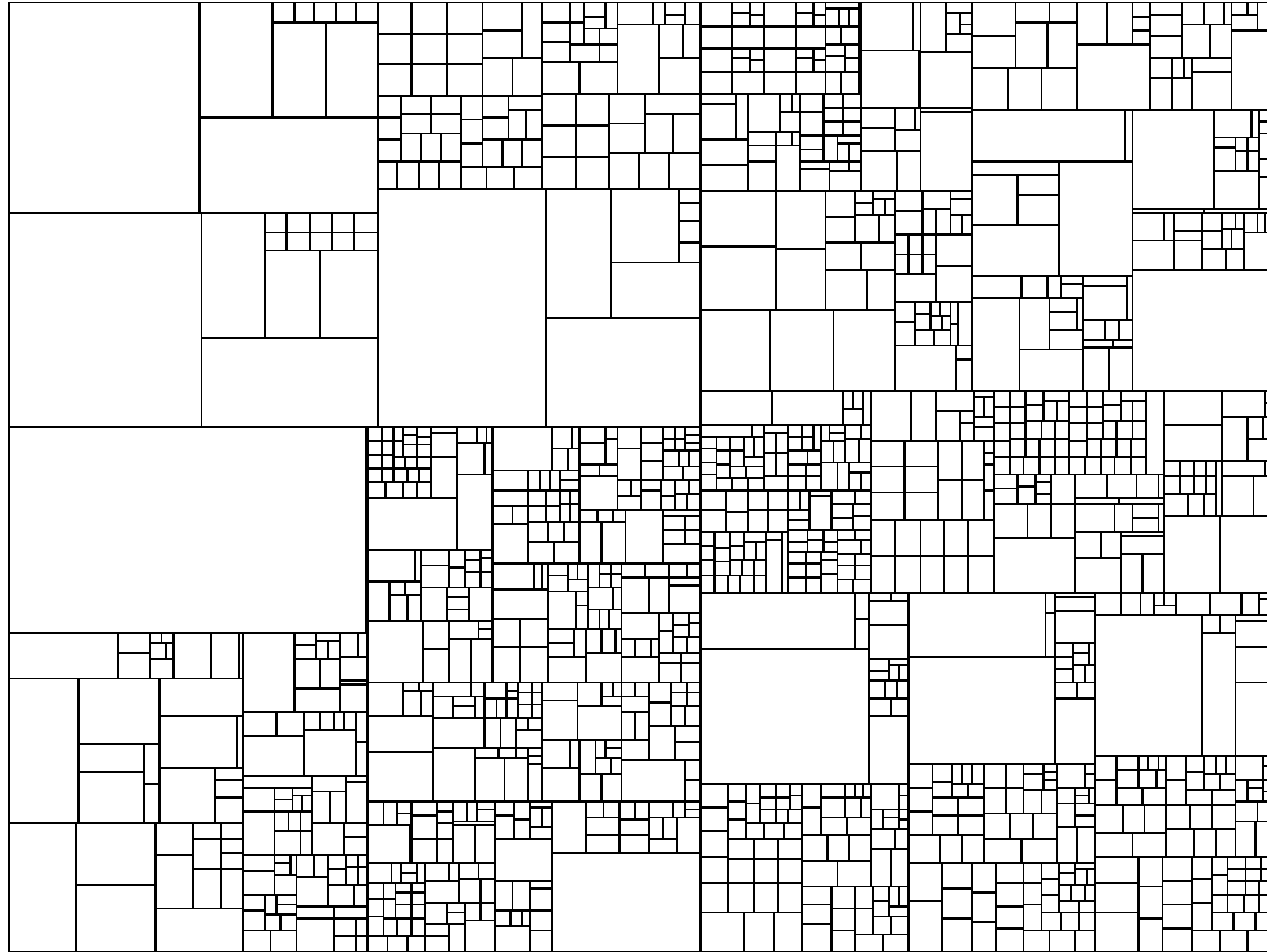
[Brus et al., 1999]

Squarification Algorithm

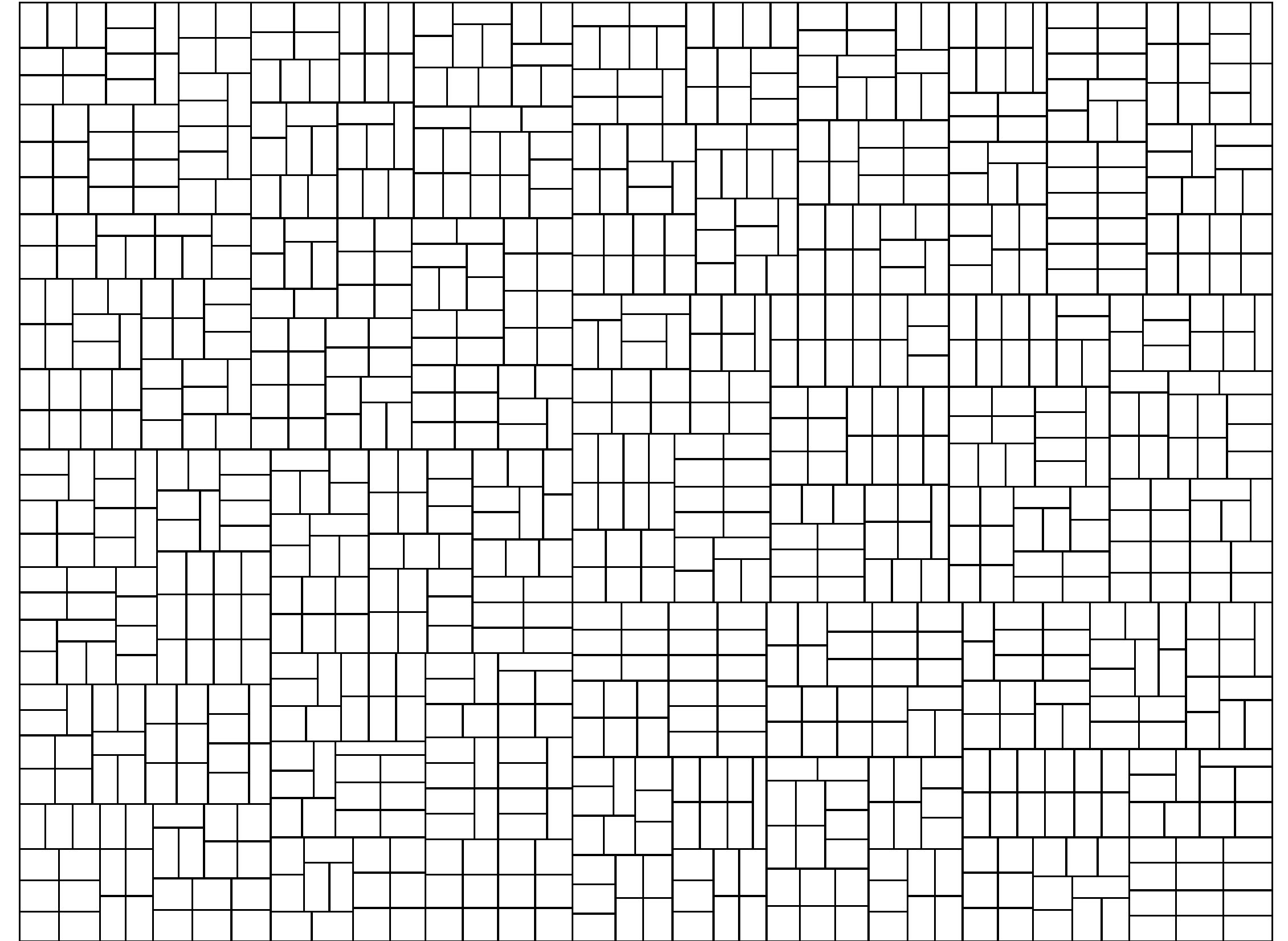


[Brus et al., 1999]

Squarified Treemaps



(a) File system



(b) Organization

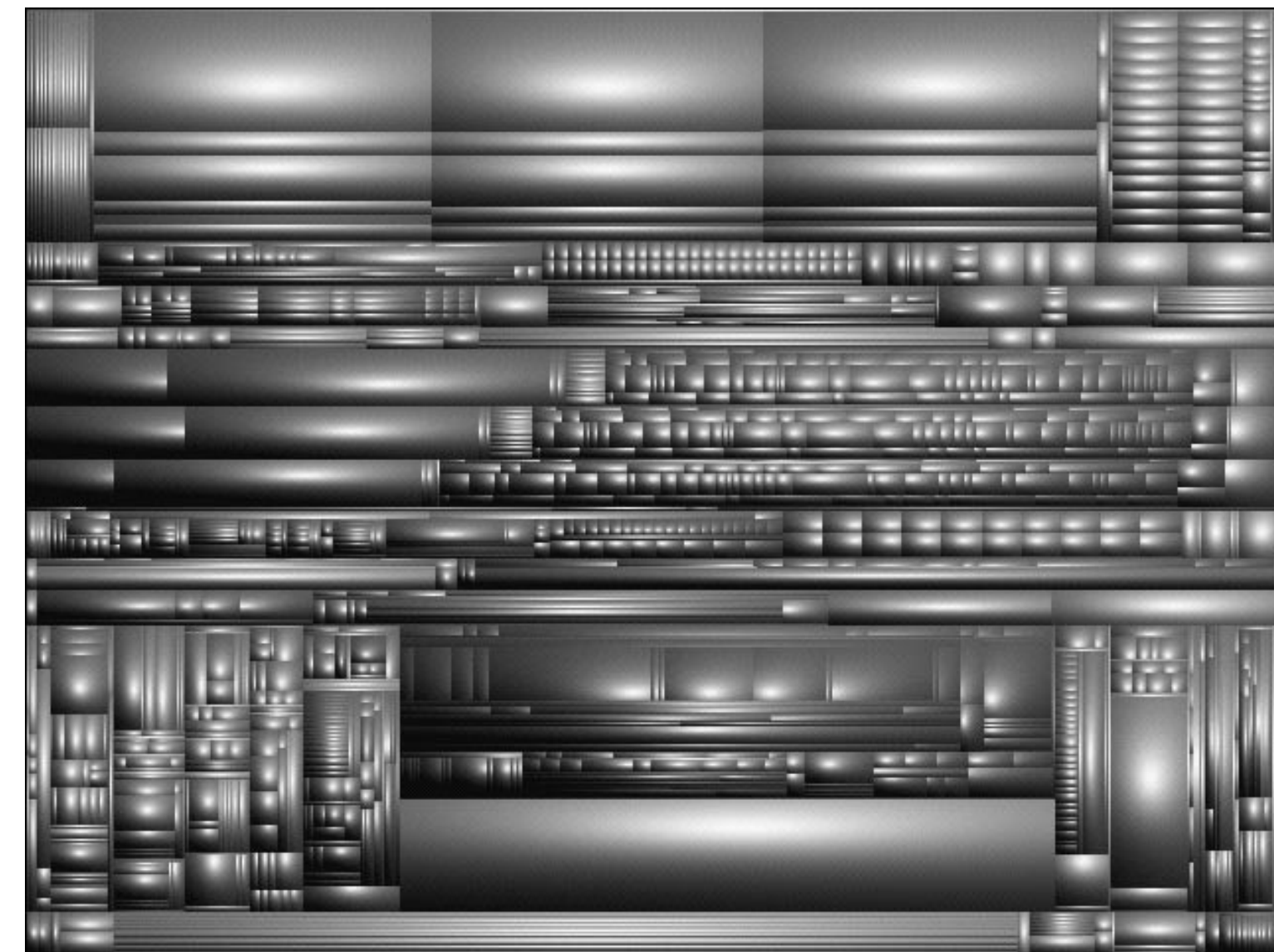
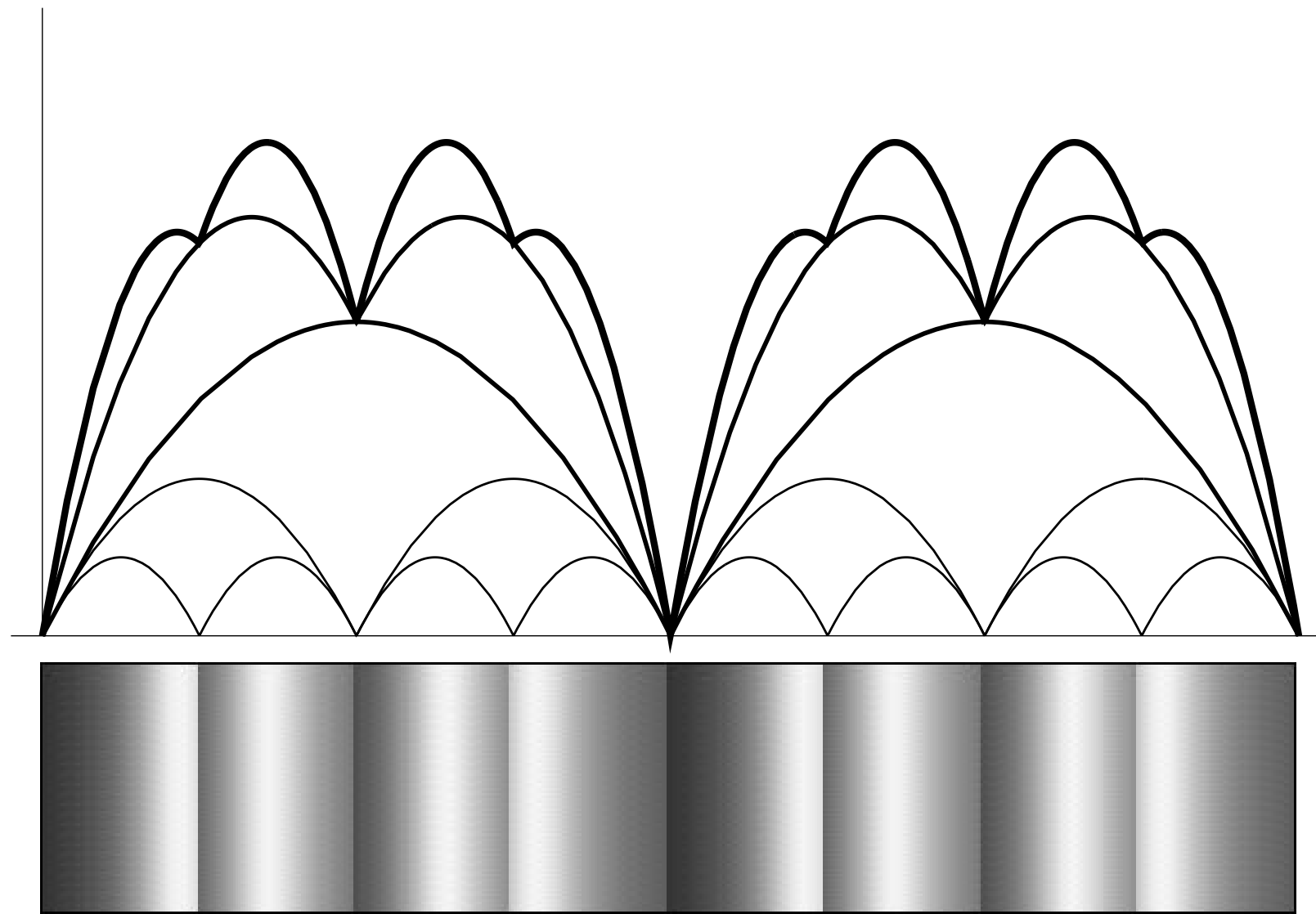
[Brus et al., 1999]

Squarified Layout

- Sort values
- Switch orientation whenever necessary to obtain best aspect ratios

Improving Treemaps (Cushion)

- Leaves are ok, but it can be difficult to find the hierarchy
- Encode this as shading information
- More effective to understand hierarchy



[van Wijk and van de Wetering, 1999]

Disk Inventory

iPhoto (31,3 MB)

+

-

Zoom In Zoom Out

Move To Trash Show Package Contents

NameSize

▼ Contents31,3 MB

▶ Resources21,5 MB

▶ NetServices5,7 MB

▼ MacOS2,2 MB

iPhoto1,9 MB

iPhotoDPA...273 kB

photocd70 kB

▶ Plugins1,7 MB

.DS_Store6 kB

▶ Plugins Disabl6 kB

Info.plist1 kB

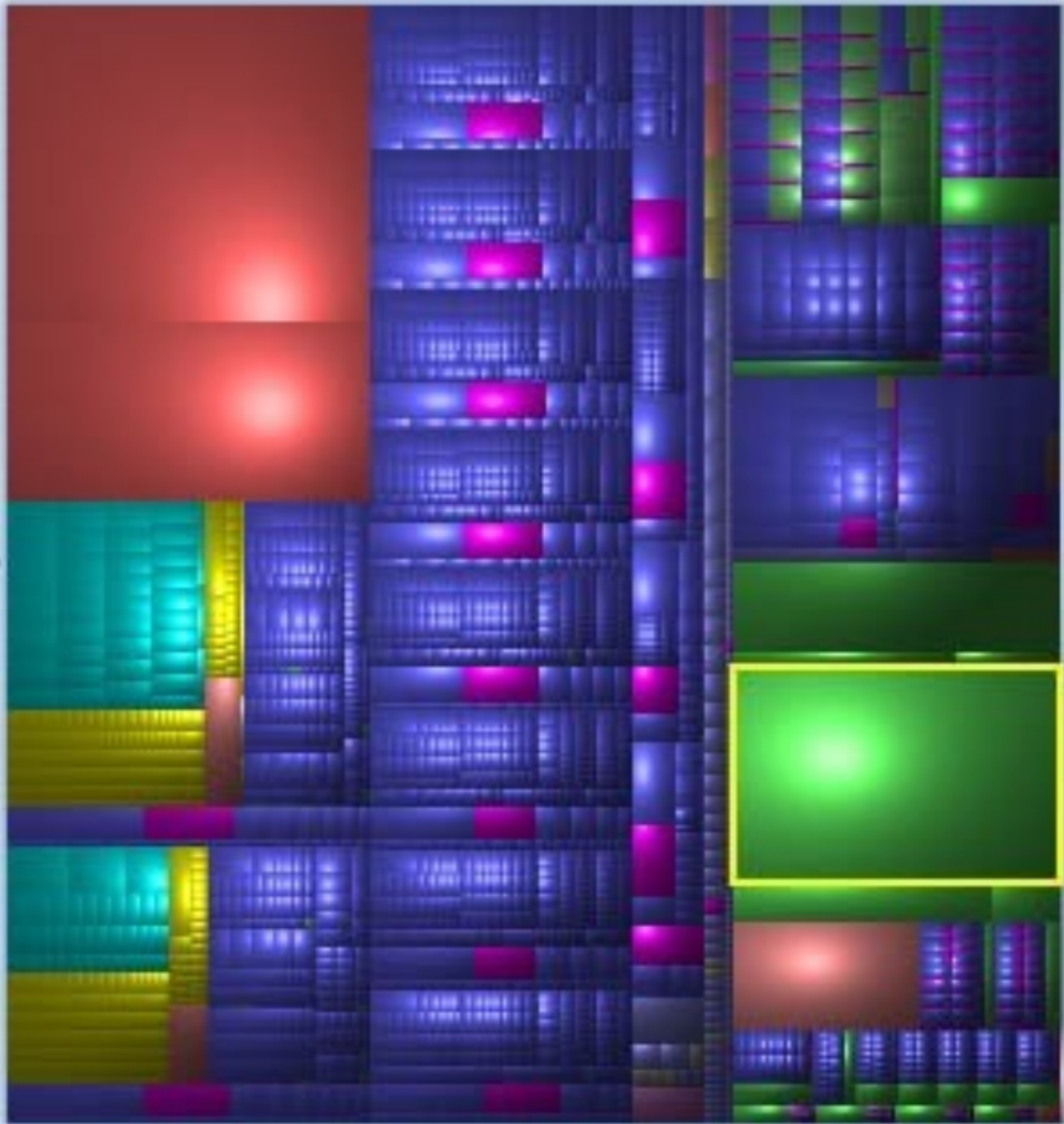
version.plist463 Byte

PkgInfo8 Bytes

▶ Resources Dis0 Bytes













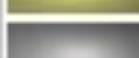







Icon65 kB

.DS_Store6 kB



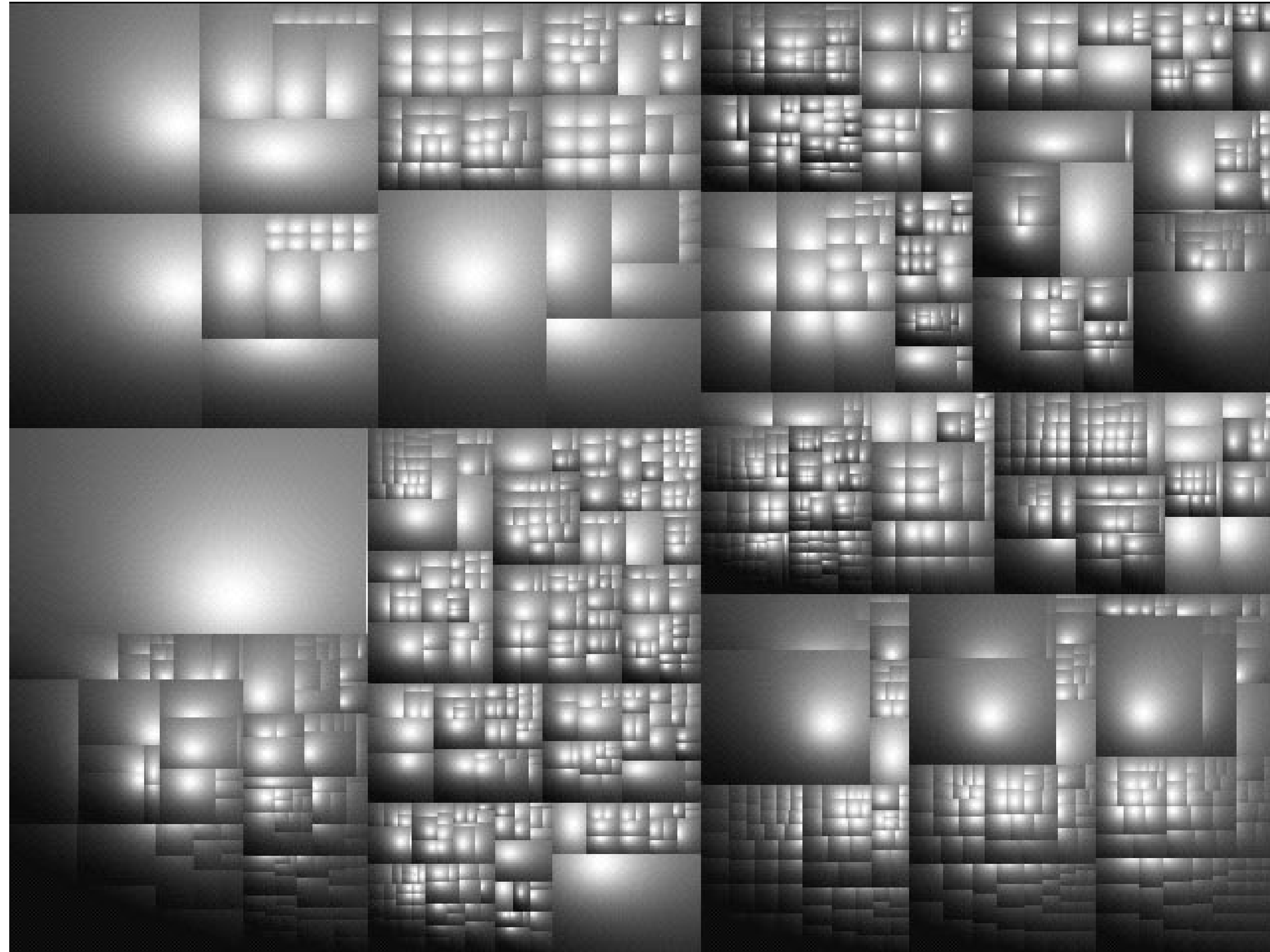
iPhoto (iPhoto/Contents/MacOS)

Unix Executable File, 1,9 MB

Color	Kind	Size	Files
	Interface Builder Document	15,4 MB	2104
	MP3 Audio File	4,8 MB	2
	Unix Executable File	3,8 MB	23
	JPEG Image	1,6 MB	74
	Strings File	1,4 MB	348
	HTML document	1,3 MB	333
	TIFF Document	1,0 MB	310
	Document	886 kB	16
	Portable Network Graphi	635 kB	21
	XML Property List File	183 kB	332
	Apple Icon Image	109 kB	2
	AIFF Audio	67 kB	2
	Finder Document	65 kB	1
	Script	35 kB	5
	Rich Text Format (RTF) d	30 kB	2
	AppleScript Suite Definit	7 kB	1
	AppleScript Suite Termin	6 kB	1
	Graphics Interchange Fo	5 kB	12
	Cascading Style Sheet (C	4 kB	4
	Symbolic Link	164 Byte	9

[Disk Inventory X]

Squarified + Cushioned Treemaps



(a) File system



(b) Organization

[Brus et al., 1999]