

# Information Visualization

## Spatial, NecklaceMaps, Myriahedral

### *Ex: Ballotmaps*

**Tamara Munzner**

Department of Computer Science  
University of British Columbia

***Week 7, 20 Oct 2021***

**<https://www.cs.ubc.ca/~tmm/courses/547-21>**

# Plan for today

- last week reading Q&A
  - chap: Networks. papers: Abyss-Explorer, Geneaology
- small group exercises
  - Ballotmaps
- this week reading Q&A
  - chap: Spatial. papers: NecklaceMaps, Myriahedral

# Next time

- reminder: proposals due tomorrow (Thu Oct 21) noon
- next week is unusual!
  - only one reading: Polaris [system] (now known as Tableau!)
  - **no class session**: instead, attend 2 sessions (90 min each) from IEEE VIS conf
    - <http://ieevis.org/>
    - free registration for students <http://ieevis.org/year/2021/info/registration/conference-registration>
    - interactive full program at <https://virtual.ieevis.org/> (6am-1:30pm Sun, 6-11:30am Mon-Fri)
      - serendipitous paper browser is cool!
      - only exclusion: my full-day tutorial Sunday
- **async discussion**
  - usual comment for Polaris paper, due by 1pm Mon Oct 25
  - also comment on what you saw at each VIS session, **due by 5pm Fri Oct 29**
  - response to either Polaris or VIS comment, **due by 5pm Sun Oct 31**

# Ballotmaps: The paper

## BallotMaps: Detecting Name Bias in Alphabetically Ordered Ballot Papers

Jo Wood, Member, IEEE, Doria Badawood, Jason Dykes, and Aidan Slingsby Member, IEEE

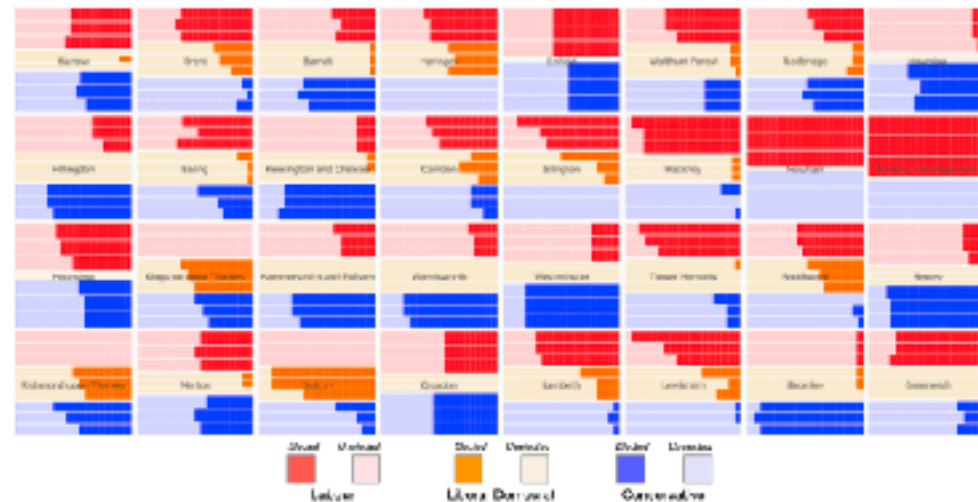


Fig. 1. BallotMap showing electoral success (or otherwise) of each candidate for the three main parties in wards (small rectangles) in each London borough (big squares) in the 2010 local government elections. Vertical ordering of candidates within each borough is by ballot paper position within party (top row first, middle row second, bottom row third). Main parties with three candidates in a ward are shown. If no ballot ordering bias existed there would be no systematic structure to bar lengths. This ballotMap shows that more candidates get elected who are listed first within their party than do candidates who are second or third.

**Abstract**—The relationship between candidates' position on a ballot paper and vote rank is explored in the case of 8000 candidates for the UK 2010 local government elections in the Greater London area. This design study uses hierarchical spatially arranged graphics to represent two locations that affect candidates at very different scales: the geographical areas for which they seek election and the spatial location of their names on the ballot paper. This approach allows the effect of position bias to be assessed; that is, the degree to which the position of a candidate's name on the ballot paper influences the number of votes received by the candidate, and whether this varies geographically. Results show that position bias was significant enough to influence rank order of candidates, and in the case of many marginal electoral wards, to influence who was elected to government. Position bias was observed most strongly for Liberal Democrat candidates but present for all major political parties. Visual analysis of classification of candidate names by ethnicity suggests that this too has an effect on votes received by candidates, in some cases overcoming alphabetic name bias. The results found contradict some earlier research suggesting that alphabetic name bias was not sufficiently significant to affect electoral outcome and add new evidence for the geographic and ethnicity influences on voting behaviour. The visual approach proposed here can be applied to a wider range of electoral data and the patterns identified and hypotheses derived from them could have significant implications for the design of ballot papers and the conduct of fair elections.

**Index Terms**—Voting, election, bias, democracy, governance, geomatics, geovisualization, hierarchy, governance.

### 1 INTRODUCTION

There has long been a suspicion from candidates standing for election that the order in which candidate names appear on a ballot paper may,

in part, influence the number of votes received (e.g. [18]). Despite a number of studies investigating the degree of this effect [24, 1, 13, 5, 20, 19], evidence appears inconclusive and sometimes contradictory. This paper considers how information visualization may be designed and applied to investigating the degree to which some form of name bias may exist in influencing votes received by candidates.

The overall aims of the work are twofold: to identify the degree to which the position of candidate name affects numbers of votes received; and to develop a data visualization design appropriate for exploring the spatial and non-spatial influences over candidate votes. Addressing these aims is important because conducting fair and neutral elections is an essential part of the democratic process. The aims

• Jo Wood (jo@city.ac.uk), Doria Badawood (Doria.Badawood@city.ac.uk), Jason Dykes (j.dykes@city.ac.uk), and Aidan Slingsby (aidan.slingsby@city.ac.uk) are at the geCentre (Aldo@gecentre.org) at the School of Geography, City University London. Manuscript received 27 March 2011; accepted 1 August 2011; revised version 27 October 2011; mailed on 14 October 2011. For information on obtaining reprints of this article, please visit <http://www.iee.org>.

*BallotMaps: Detecting name bias in alphabetically ordered ballot papers.*

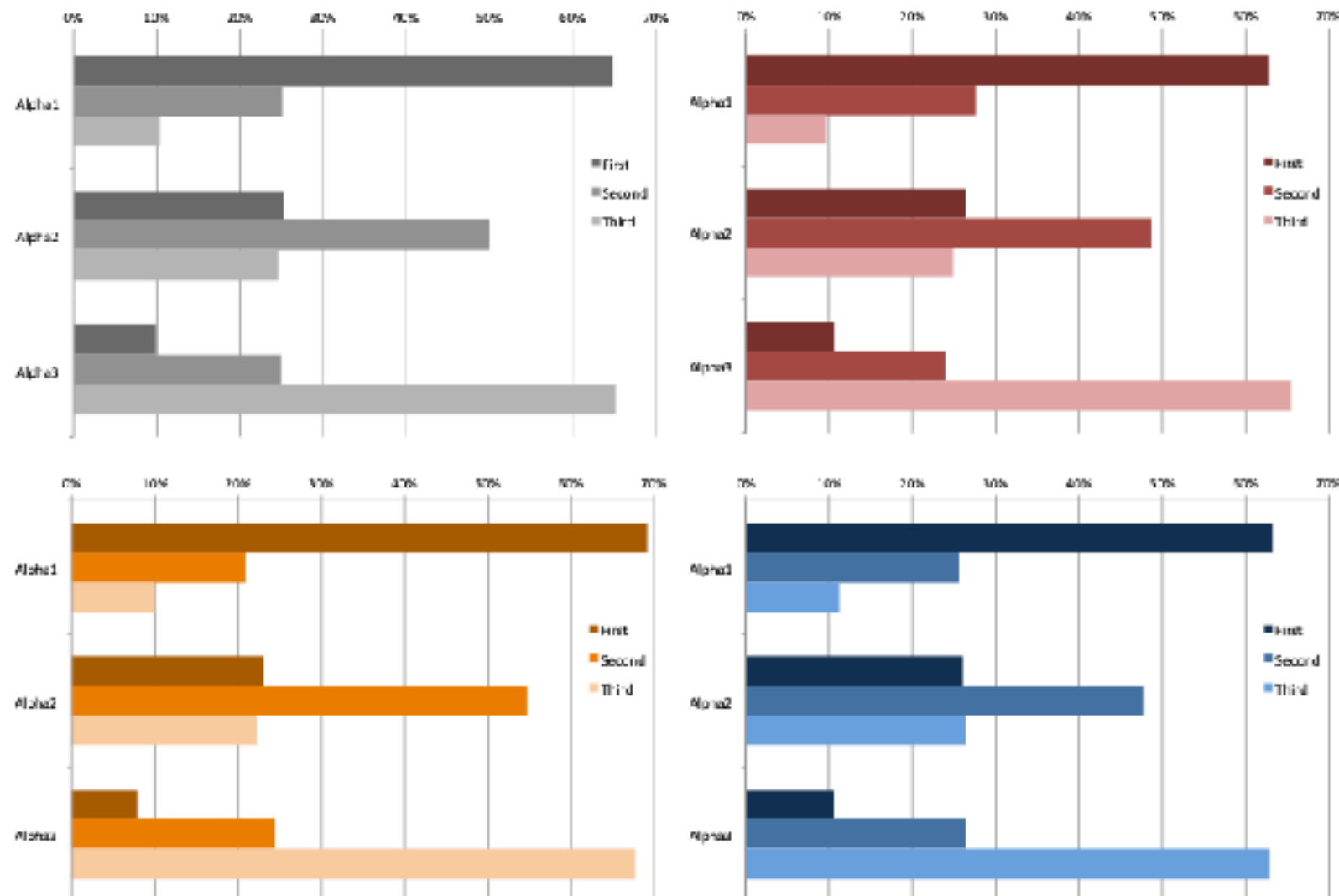
*Wood, Badawood, Dykes, and Slingsby.*

*IEEE TVCG (Proc. InfoVis) 17(12):2384-2391, 2011.*

[http://openaccess.city.ac.uk/436/1/wood\\_ballotmaps\\_2011.pdf](http://openaccess.city.ac.uk/436/1/wood_ballotmaps_2011.pdf)

# PosAcross

- *To what extent does the position in the ballot affect the number of votes received by a candidate, overall? Is there variation across political party?*



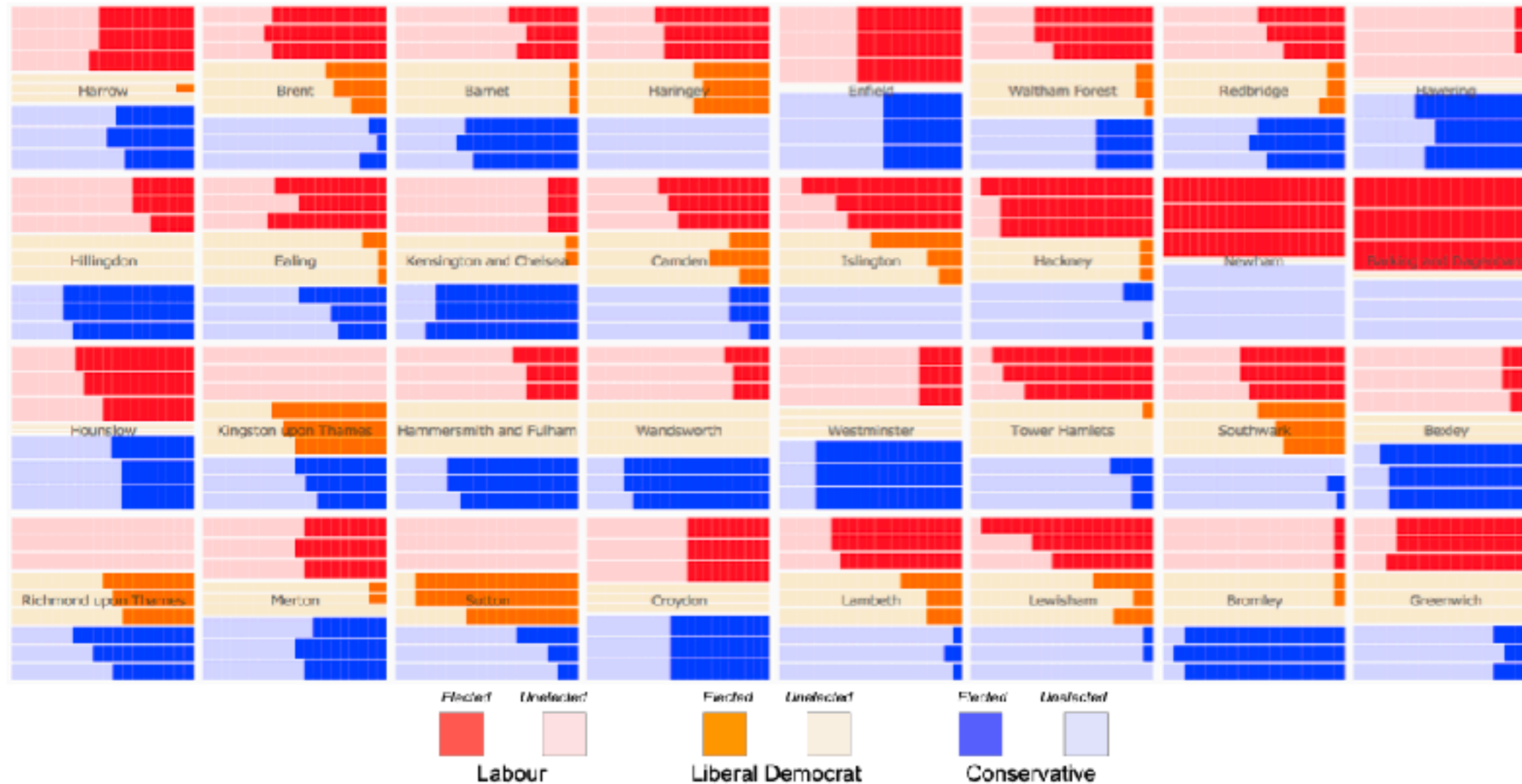
- Votes: 0-N
- Position\_Overall: 1-9
- Party: 1-3 (Lab, LibDem, Cons)
- **derived: Position\_Within: 1-3**

[Fig 5. *BallotMaps: Detecting name bias in alphabetically ordered ballot papers.* Wood, Badawood, Dykes, and Slingsby. *IEEE TVCG (Proc. InfoVis)* 17(12):2384-2391, 2011.]

Fig. 5. Alpha position and vote order for, all candidates (gray); Labour candidates (red); Conservative (blue) and Liberal Democrat (orange). If no name order bias existed, all bars would be about the same length.

# AcrossDistrib

- *Does the ballot-position influence vary geographically in different boroughs?*



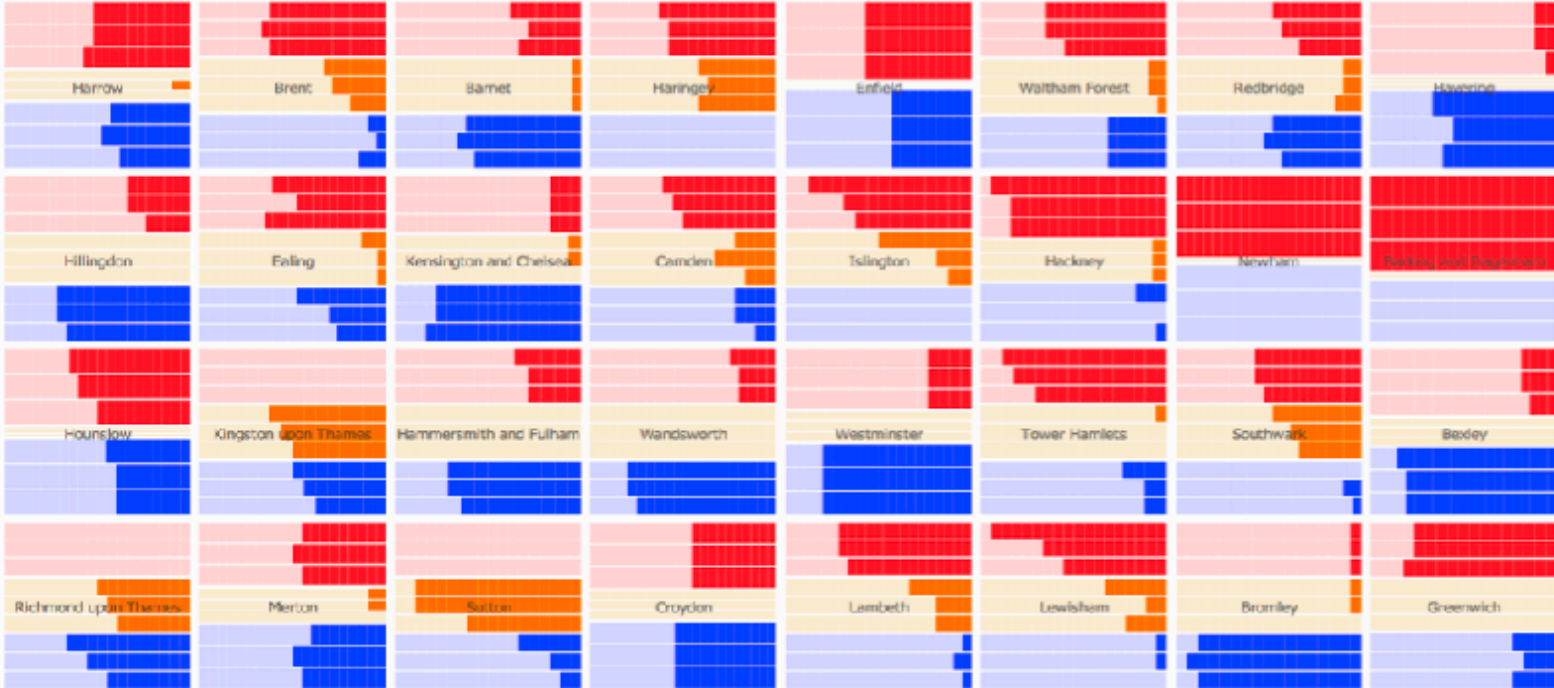
- Borough: count of 32
- Ward: count of 614

[Fig 1. *BallotMaps: Detecting name bias in alphabetically ordered ballot papers.* Wood, Badawood, Dykes, and Slingsby. *IEEE TVCG (Proc. InfoVis)* 17(12):2384-2391, 2011.]

Fig. 1. BallotMap showing electoral success (or otherwise) of each candidate for the three main parties in wards (small rectangles) in each London borough (grid squares) in the 2010 local government elections. Vertical ordering of candidates within each borough is by ballot paper position within party (top row first, middle row second, bottom row third). Main parties with three candidates in a ward are shown. If no ballot ordering bias existed there would be no systematic structure to bar lengths. This ballotMap shows that more candidates get elected who are listed first within their party than do candidates who are second or third.

# Best of both worlds: quasi-geographic positioning

- choropleth: size issues
- tabular: lose geographic position information



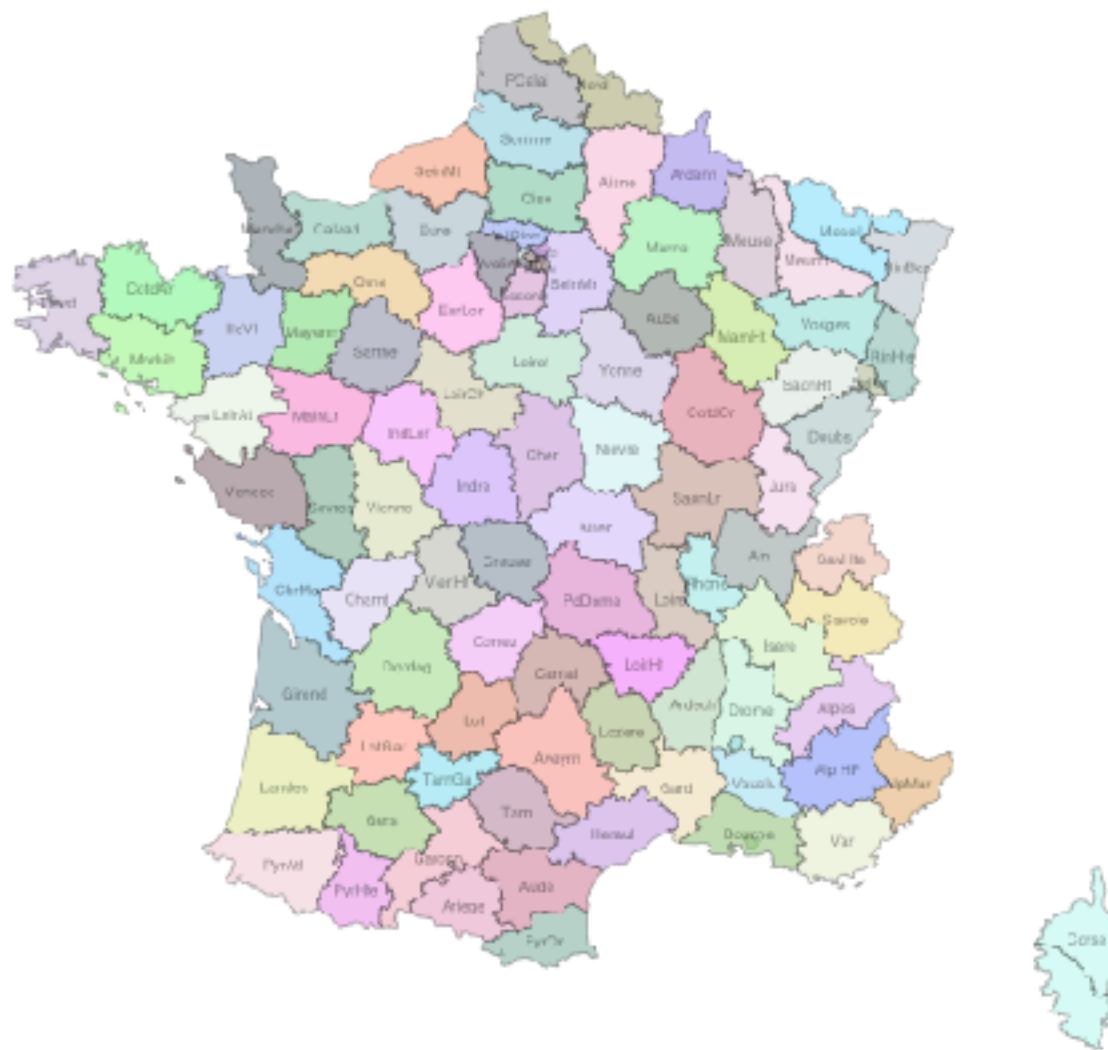
# Spatially ordered treemaps

- quasi-geographic positions

*Spatially ordered treemaps.*

*Wood and Dykes.*

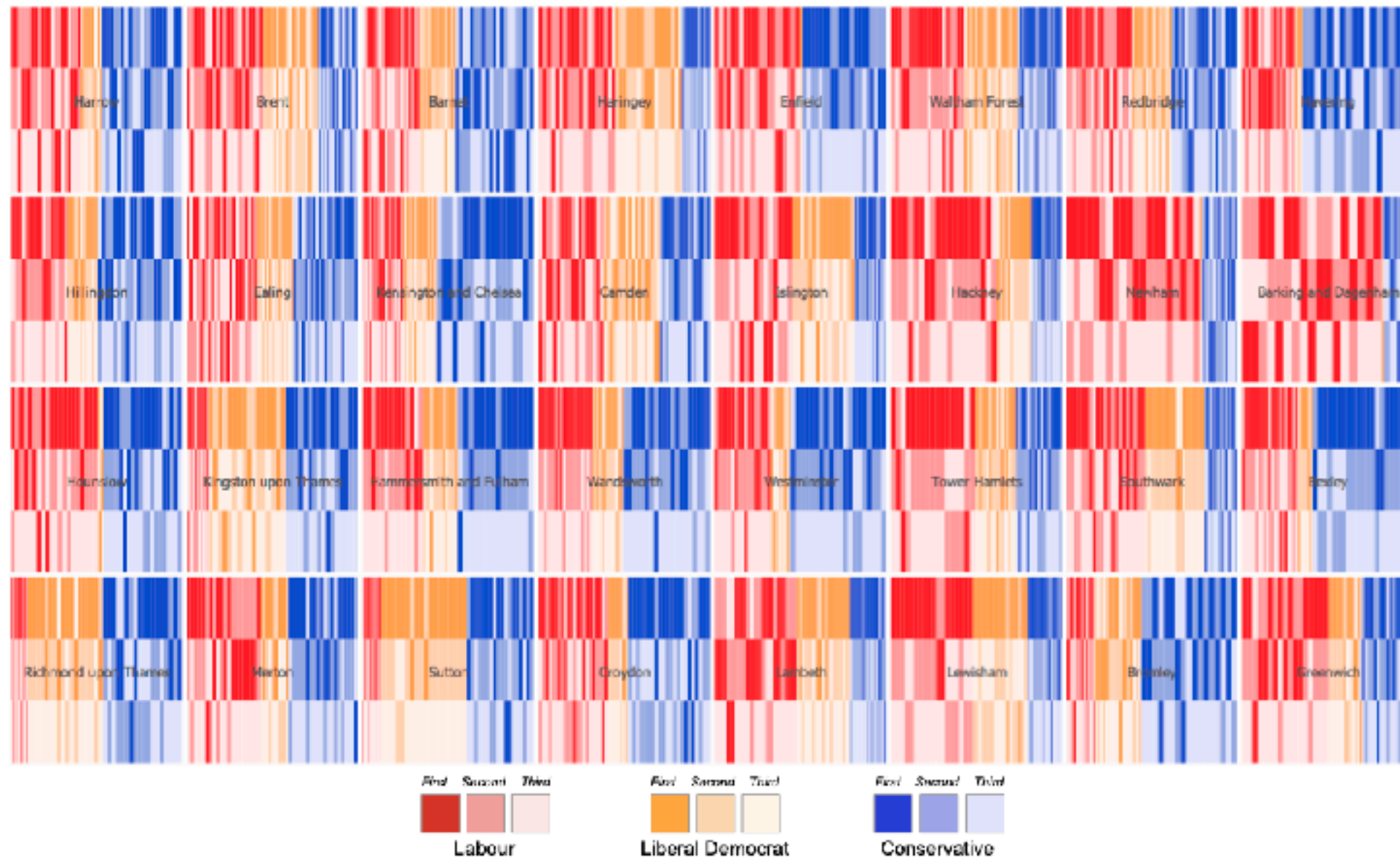
*IEEE TVCG (Proc. InfoVis) 14(6):1348-1355, 2008.*





# AcrossDistrib, Take 2

- *Does it vary in different wards? Does it depend on party affiliation?*



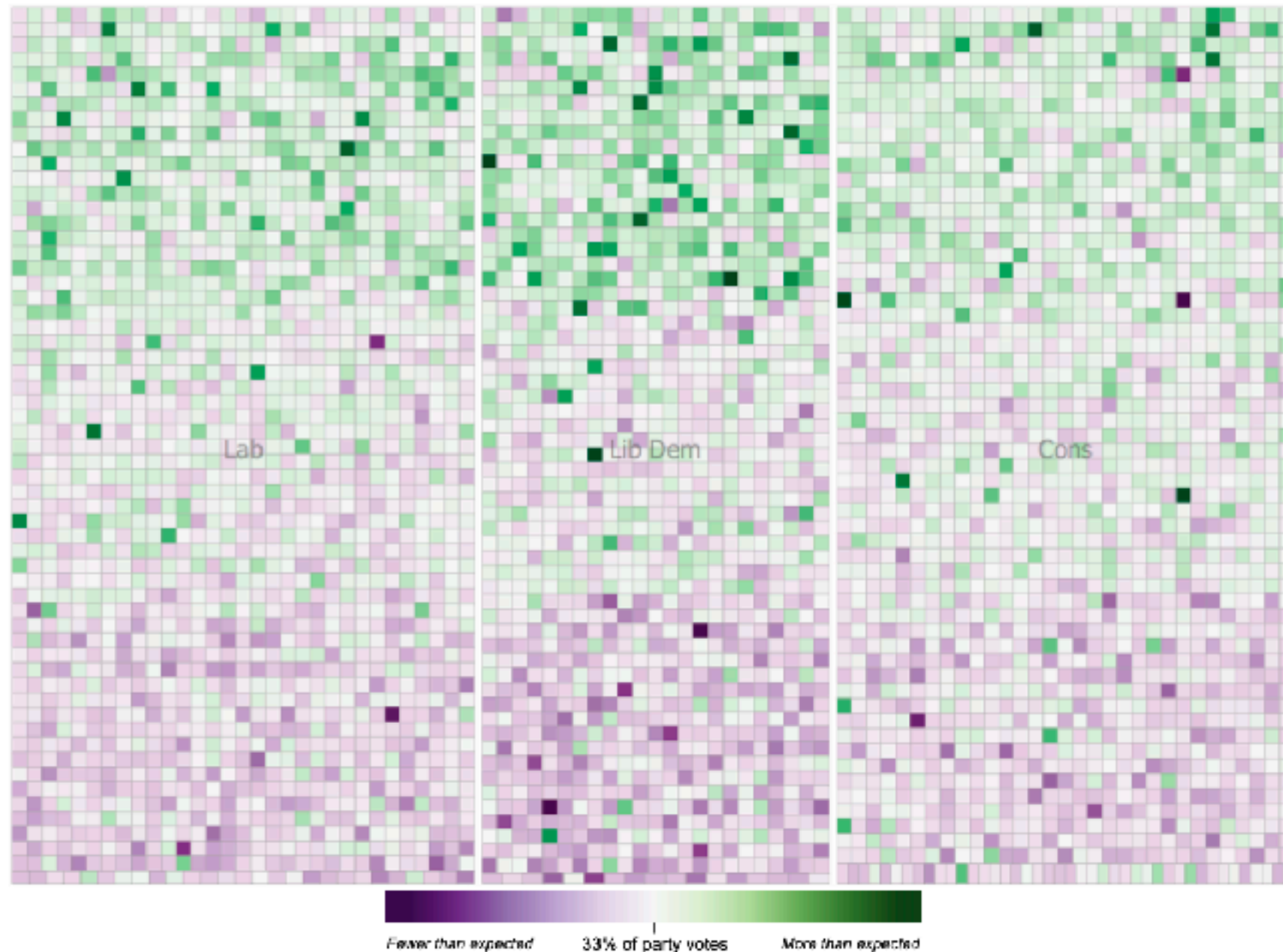
- Position\_Within: 1-3

[Fig 4. *BallotMaps: Detecting name bias in alphabetically ordered ballot papers.* Wood, Badawood, Dykes, and Slingsby. *IEEE TVCG (Proc. InfoVis)* 17(12):2384-2391, 2011.]

Fig. 4. Alpha position within party (vertical position) and voting rank within party for the three main parties in each ward (vertical bars) in each borough (grid squares). If no name order bias existed, dark and light cells would be randomly distributed in the top, middle and bottom thirds of each borough. Actual voting data show that darker cells (indicating a candidate most votes within their party) are more common in the upper third (listed first on the ballot paper within their party) and lighter cells (least votes within party) are more common in the lower third (listed third within their on the ballot paper).

# PosWithin

- *To what extent does the position in the ballot influence the number of votes a candidate gets **within** their party?*



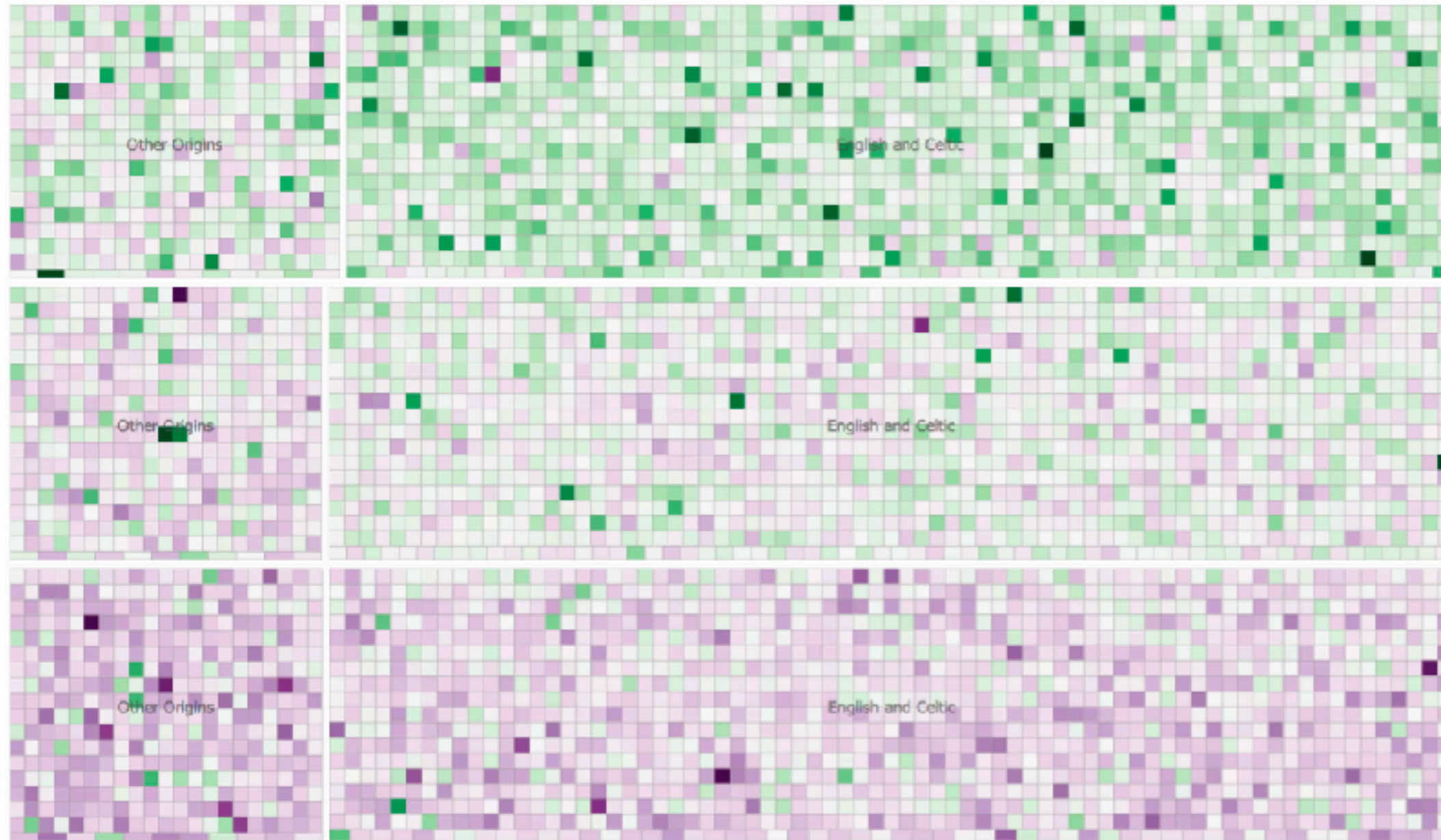
- Signed\_Chi: -1 to 1
- Residual: -1 to 1

[Fig 6. *BallotMaps: Detecting name bias in alphabetically ordered ballot papers.* Wood, Badawood, Dykes, and Slingsby. *IEEE TVCG (Proc. InfoVis)* 17(12):2384-2391, 2011.]

Fig. 6. Signed chi values for each candidate arranged by party (left to right) and ballot position within party (top to bottom, row by row). The top third represents candidates ordered first in their party, then ordered by absolute position on the ballot paper; the middle third represents candidate ordered second within their party etc. If no name order bias existed, purple and green cells would be randomly distributed within the ballotMap.

# Name Ethnicity

- *To what extent does the perceived ethnicity of candidate's name matter?*



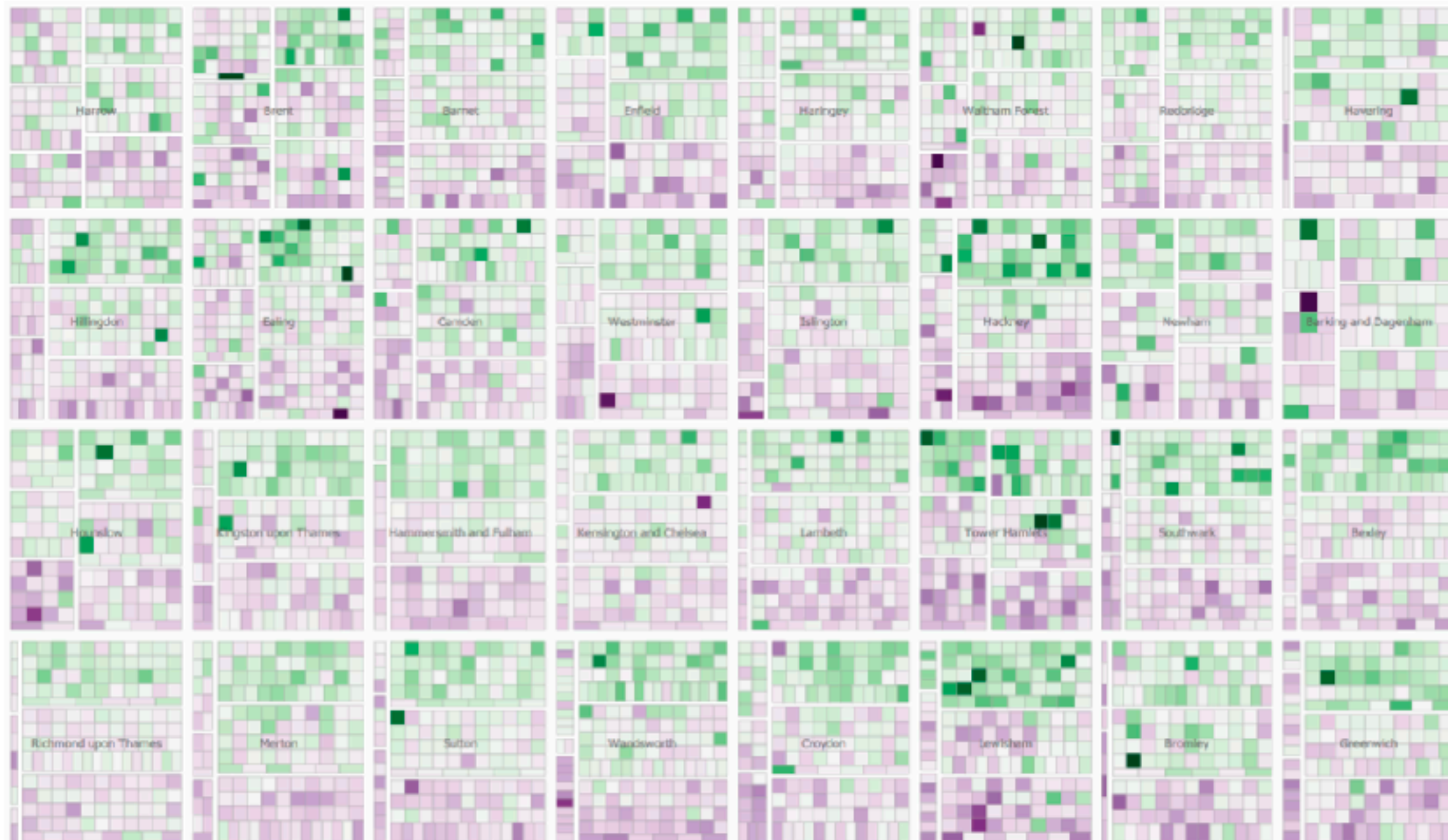
Name: English or Celtic  
vs Other Origins

[Fig 7. *BallotMaps: Detecting name bias in alphabetically ordered ballot papers.* Wood, Badawood, Dykes, and Slingsby. *IEEE TVCG (Proc. InfoVis)* 17(12):2384-2391, 2011.]

Fig. 7. Signed chi values for candidates arranged by binary classification of name origin and ballot position within party (top to bottom, row by row). The top third represents candidates ordered first in their party, then ordered by absolute position on the ballot paper; the middle third represents candidate ordered second within their party etc. Name order bias (tendency for green cells in the upper third and purple in the lower third) is stronger for 'English or Celtic' names than for other names where candidates listed first are not so likely to get more votes than expected.

# Name Ethnicity

- *Does this effect vary with geography?*



[Fig 8. *BallotMaps: Detecting name bias in alphabetically ordered ballot papers. Wood, Badawood, Dykes, and Slingsby. IEEE TVCG (Proc. InfoVis) 17(12):2384-2391, 2011.*]

Fig. 8. Signed chi values for each candidate in each borough arranged by binary classification of name origin ('other name origins' left, 'English or Celtic' right) and ballot position within party (top to bottom). The degree of name order bias is indicated by the strength of separation of green (more votes than expected) and purple (fewer votes than expected) cells. This varies by borough and by ethnic origin of candidate names.

# Q&A / Backup Slides

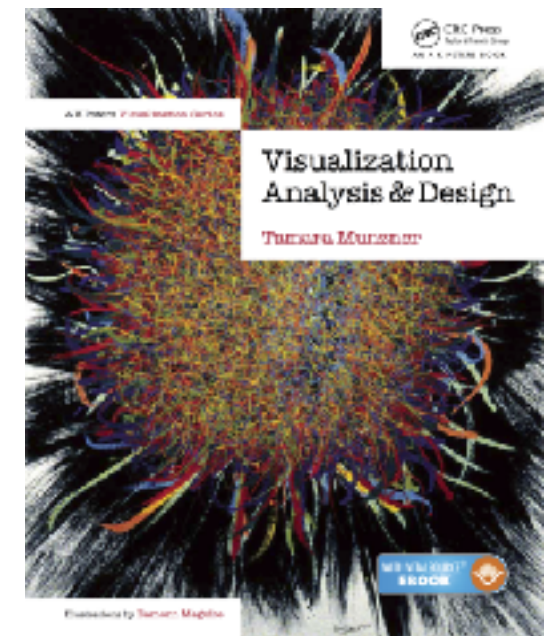
# Visualization Analysis & Design

## *Network Data (Ch 9)*

**Tamara Munzner**

Department of Computer Science  
University of British Columbia

[@tamaramunzner](#)



# Network data

- networks

- model relationships between things

- aka graphs

- two kinds of items, both can have attributes

- nodes

- links

- tree

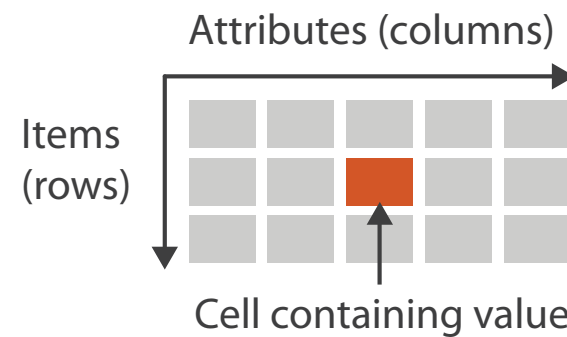
- special case

- no cycles

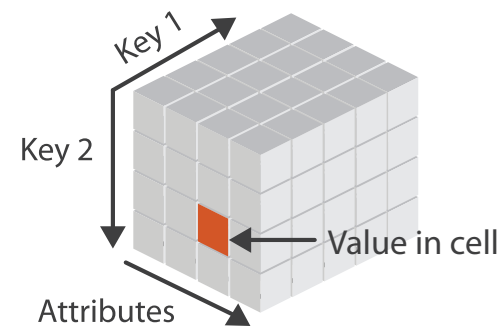
- one parent per node

## ➔ Dataset Types

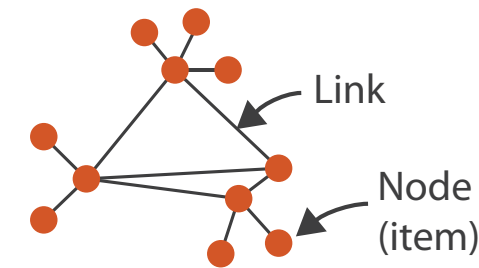
➔ Tables



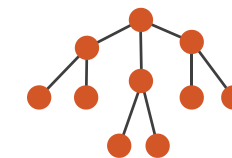
➔ *Multidimensional Table*



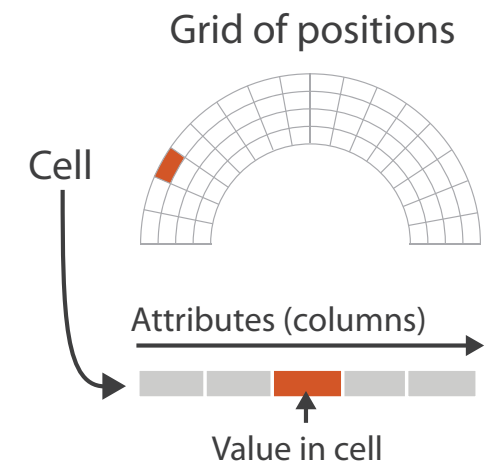
➔ Networks



➔ Trees

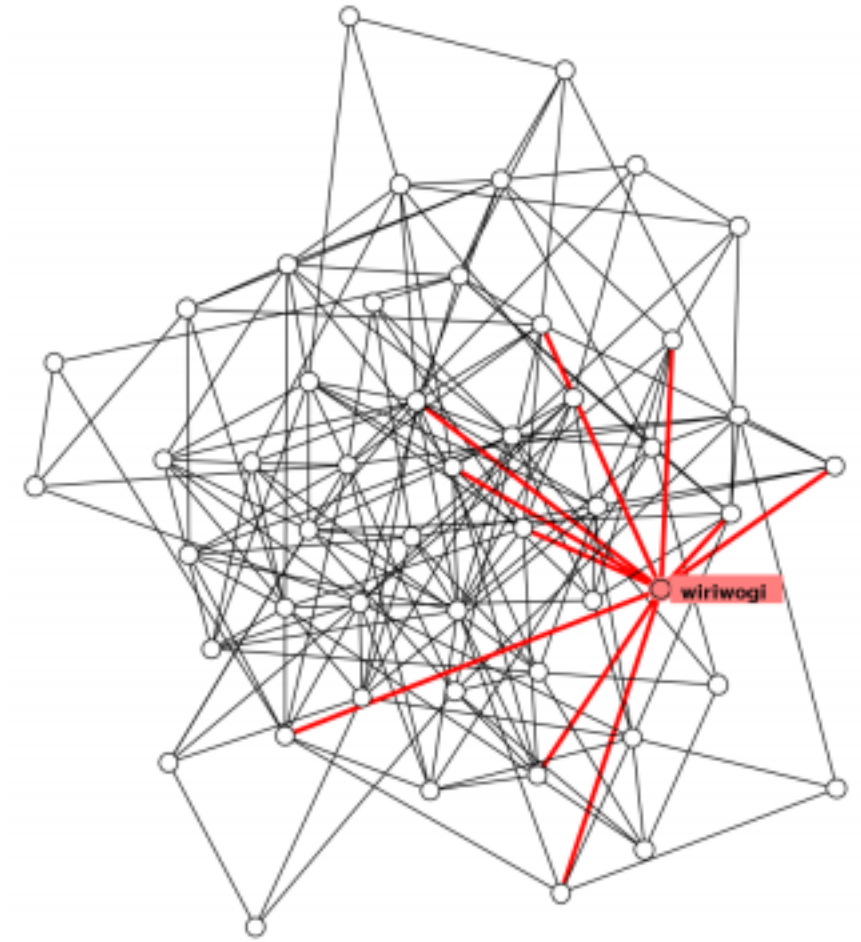


➔ Spatial  
➔ Fields (Continuous)



# Network tasks: topology-based and attribute-based

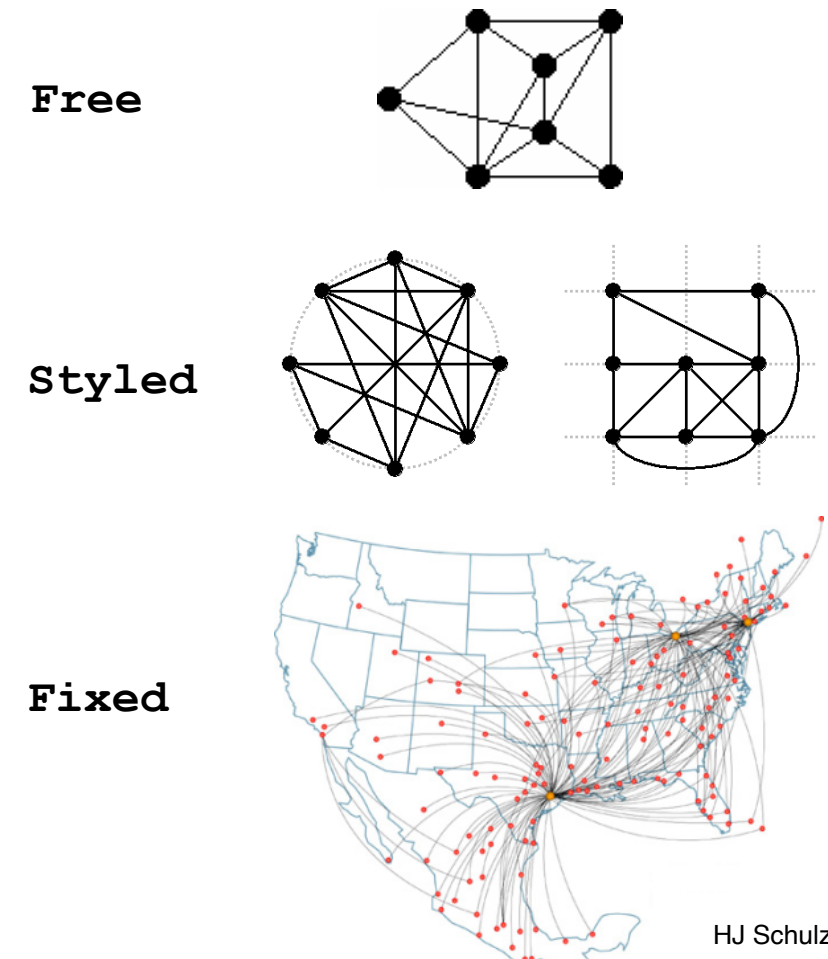
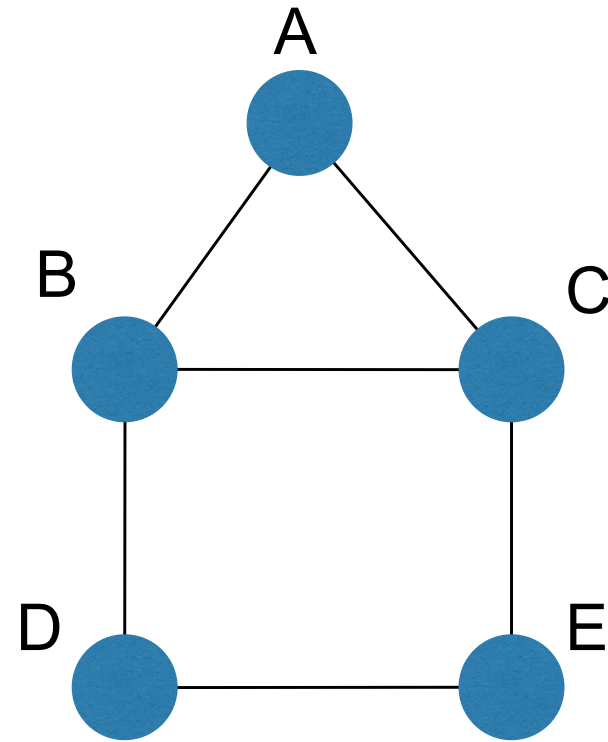
- topology based tasks
  - find paths
  - find (topological) neighbors
  - compare centrality/importance measures
  - identify clusters / communities
- attribute based tasks (similar to table data)
  - find distributions, ...
- combination tasks, incorporating both
  - example: find friends-of-friends who like cats
    - topology: find all adjacent nodes of given node
    - attributes: check if has-pet (node attribute) == cat





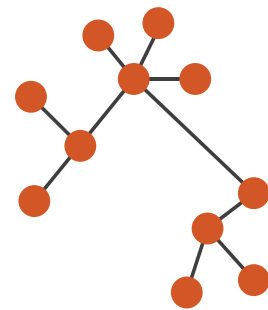
# Node-link diagrams

- nodes: point marks
- links: line marks
  - straight lines or arcs
  - connections between nodes
- intuitive & familiar
  - most common
  - many, many variants



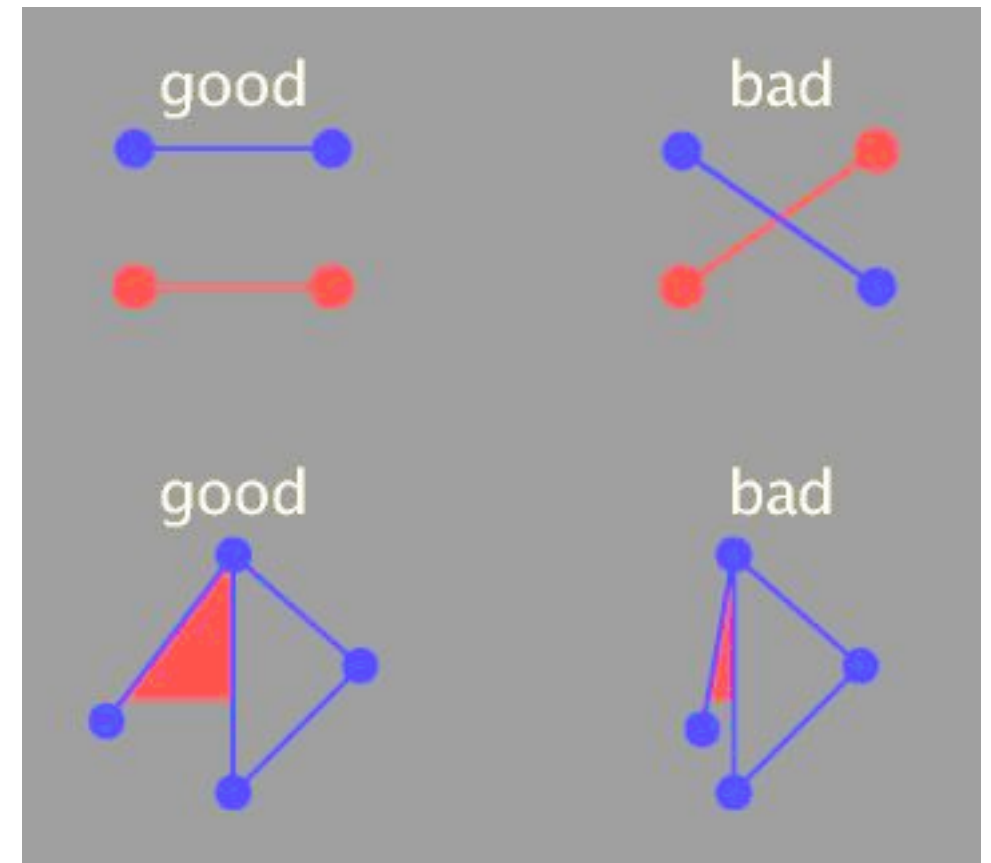
## ➔ Node-Link Diagrams Connection Marks

✓ NETWORKS    ✓ TREES



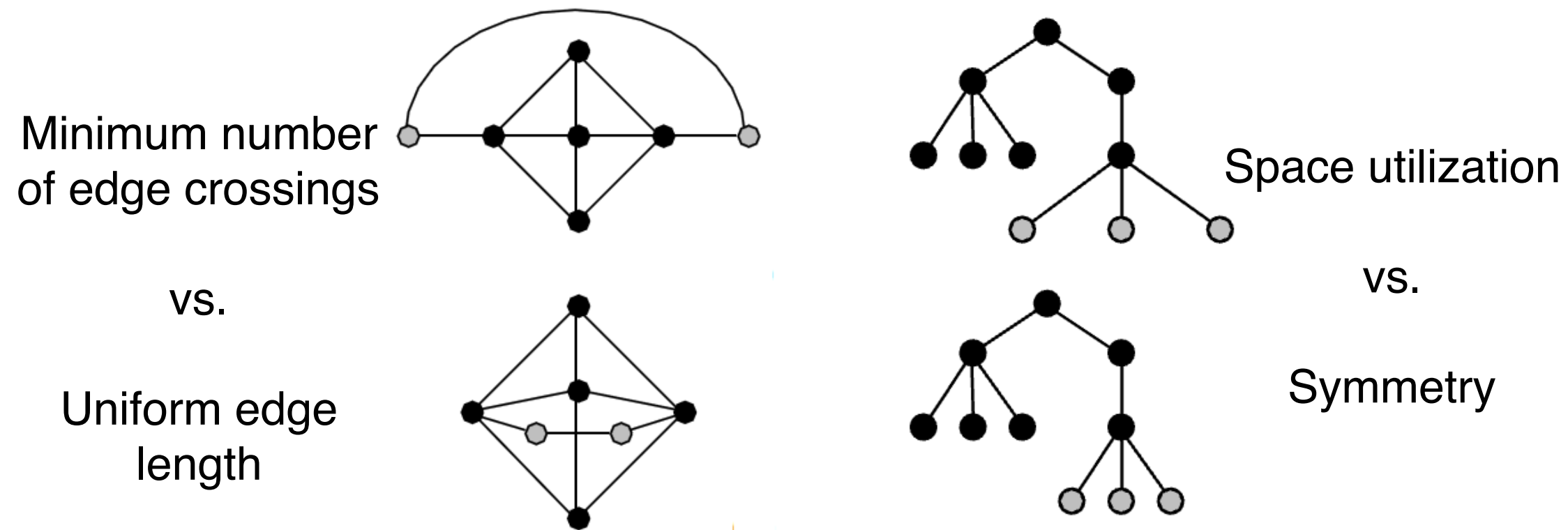
# Criteria for good node-link layouts

- minimize
  - edge crossings, node overlaps
  - distances between topological neighbor nodes
  - total drawing area
  - edge bends
- maximize
  - angular distance between different edges
  - aspect ratio disparities
- emphasize symmetry
  - similar graph structures should look similar in layout



# Criteria conflict

- most criteria NP-hard individually
- many criteria directly conflict with each other



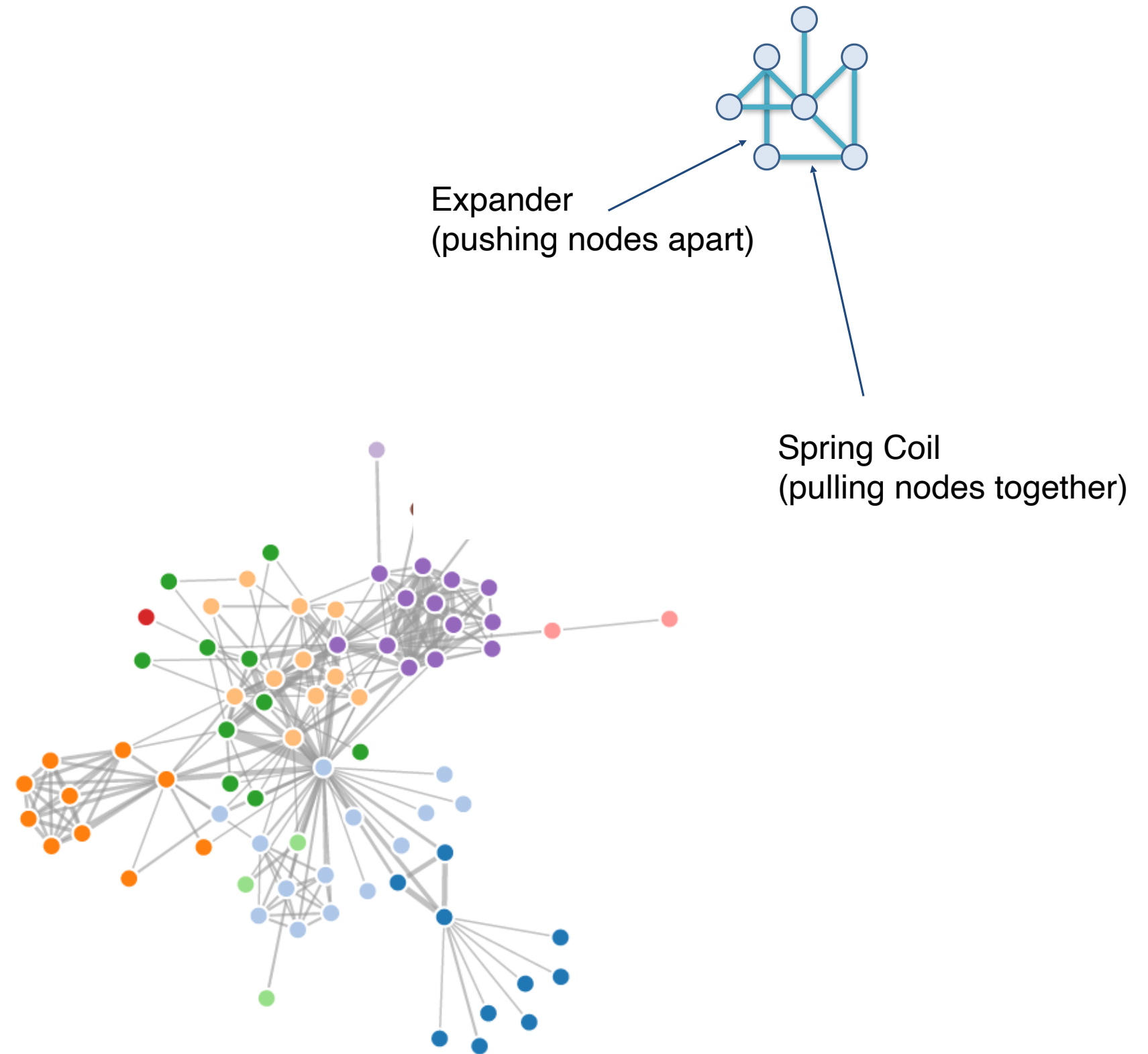
Schulz 2004

# Optimization-based layouts

- formulate layout problem as optimization problem
- convert criteria into weighted cost function
  - $F(\text{layout}) = a * [\text{crossing counts}] + b * [\text{drawing space used}] + \dots$
- use known optimization techniques to find layout at minimal cost
  - energy-based physics models
  - force-directed placement
  - spring embedders

# Force-directed placement

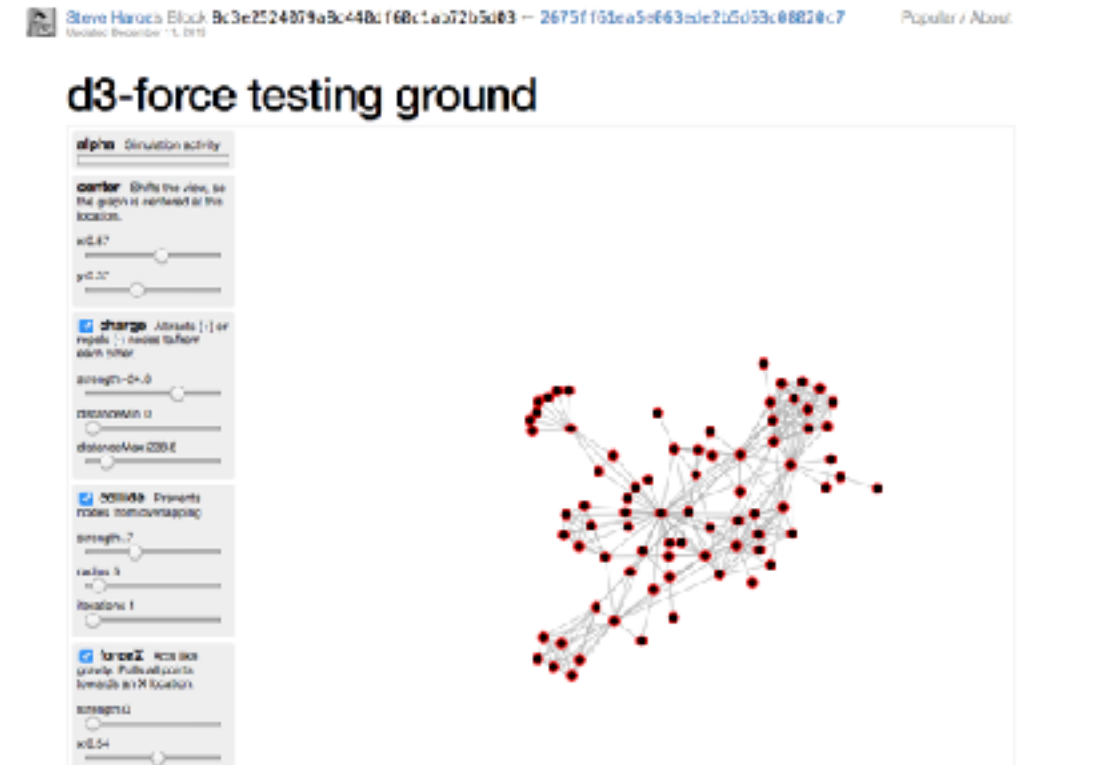
- physics model
  - links = springs pull together
  - nodes = magnets repulse apart
- algorithm
  - place vertices in random locations
  - while not equilibrium
    - calculate force on vertex
      - sum of
        - » pairwise repulsion of all nodes
        - » attraction between connected nodes
    - move vertex by  $c * \text{vertex\_force}$



<http://mbostock.github.com/d3/ex/force.html>

# Force-directed placement properties

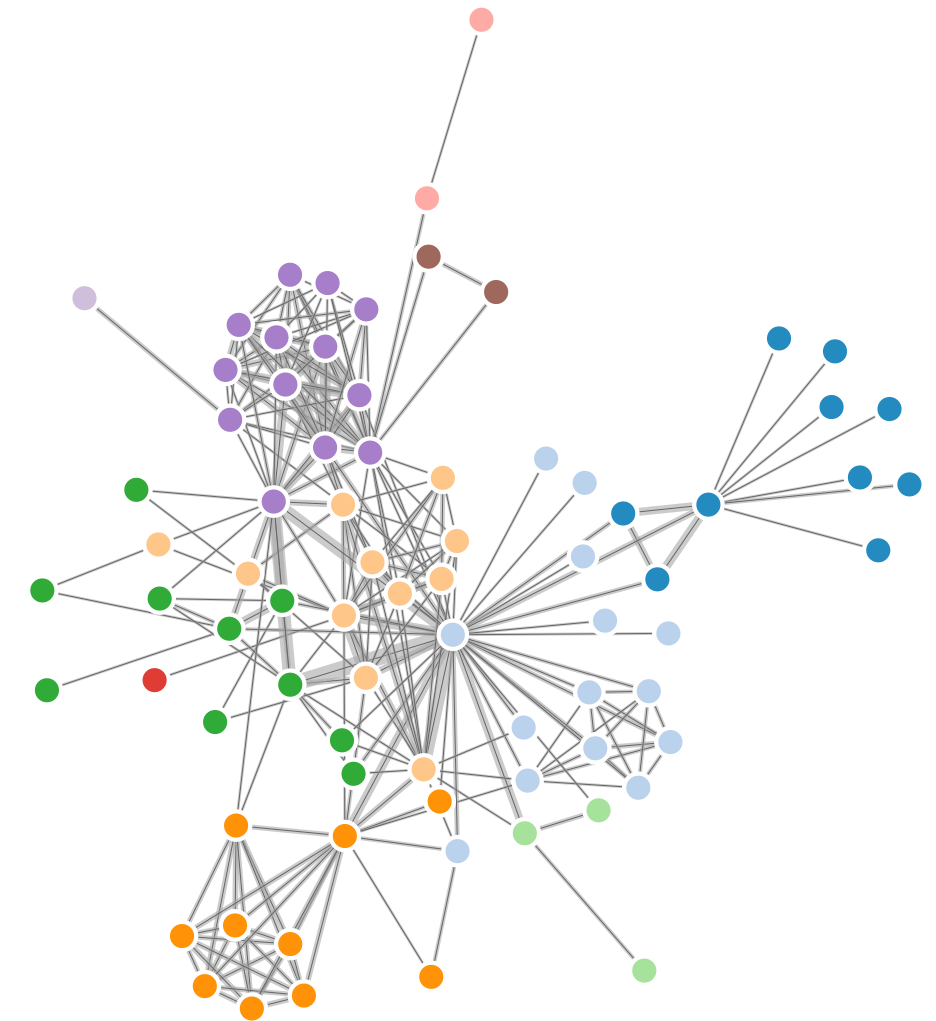
- strengths
  - reasonable layout for small, sparse graphs
  - clusters typically visible
  - edge length uniformity
- weaknesses
  - nondeterministic
  - computationally expensive:  $O(n^3)$  for  $n$  nodes
    - each step is  $n^2$ , takes  $\sim n$  cycles to reach equilibrium
  - naive FD doesn't scale well beyond 1K nodes
  - iterative progress: engaging but distracting



<https://bl.ocks.org/steveharoz/8c3e2524079a8c440df60c1ab72b5d03>

# Idiom: **force-directed placement**

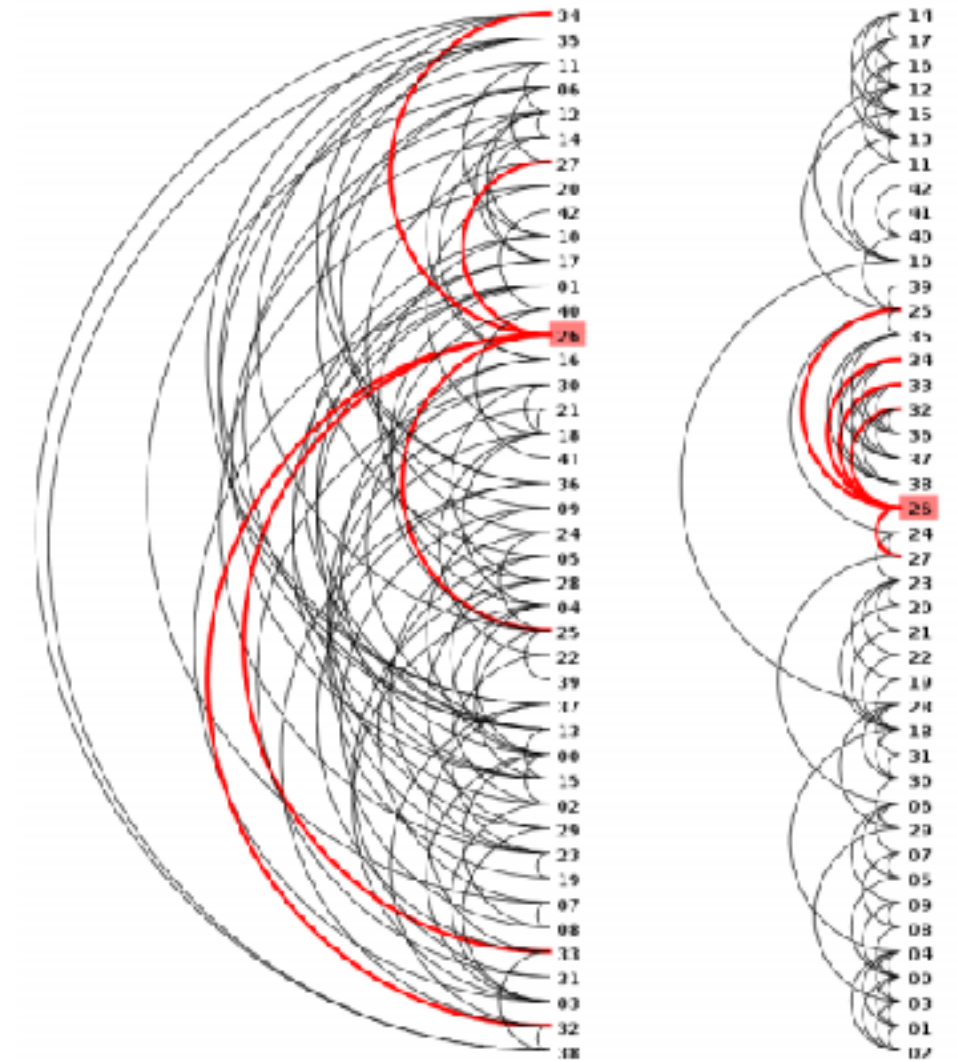
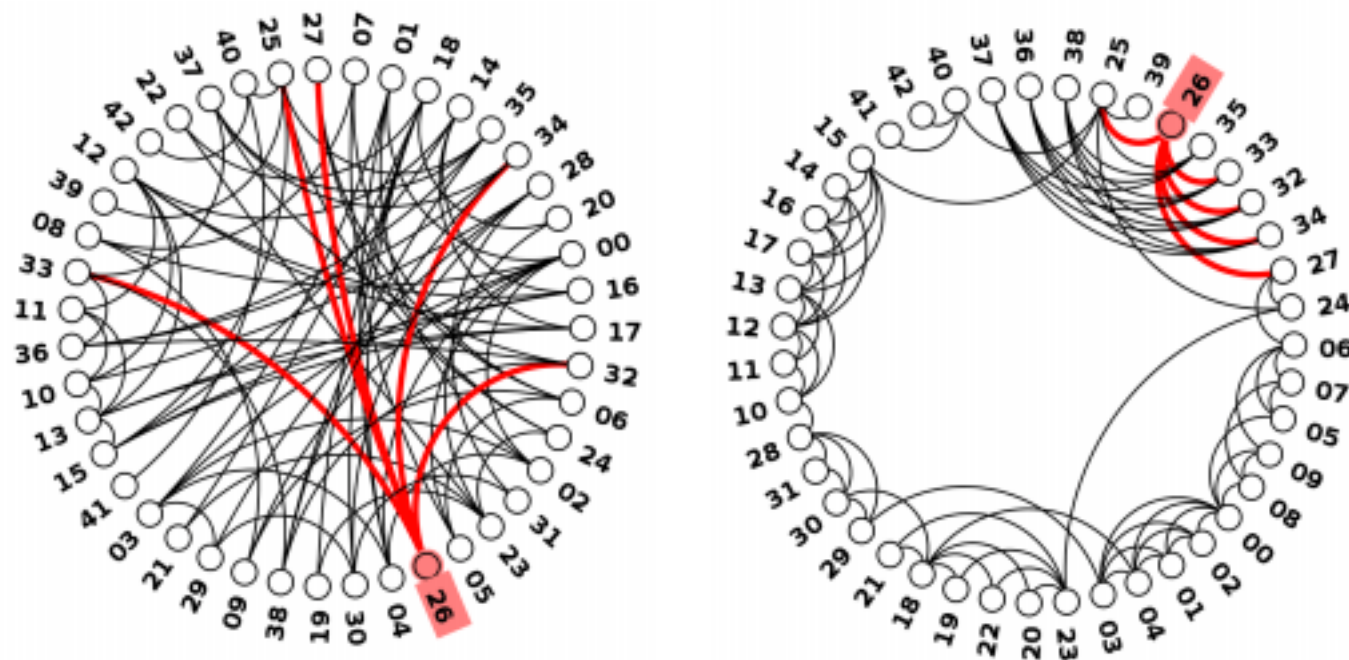
- visual encoding
  - link connection marks, node point marks
- considerations
  - spatial position: no meaning directly encoded
    - left free to minimize crossings
  - proximity semantics?
    - sometimes meaningful
    - sometimes arbitrary, artifact of layout algorithm
    - tension with length
      - long edges more visually salient than short
- tasks
  - explore topology; locate paths, clusters
- scalability
  - node/edge density  $E < 4N$



<http://mbostock.github.com/d3/ex/force.html>

# Idiom: circular layouts / arc diagrams (node-link)

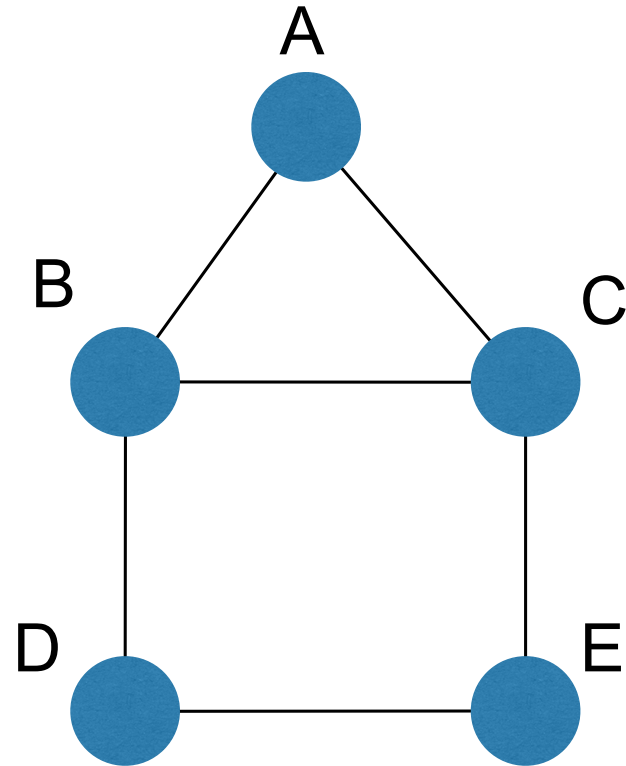
- restricted node-link layouts: lay out nodes around circle or along line
- data
  - original: network
  - derived: node ordering attribute (global computation)
- considerations: node ordering crucial to avoid excessive clutter from edge crossings
  - examples: before & after barycentric ordering





# Adjacency matrix representations

- derive adjacency matrix from network



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

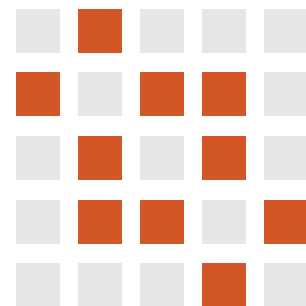


## Adjacency Matrix

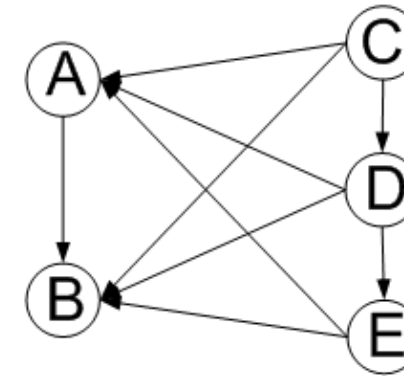
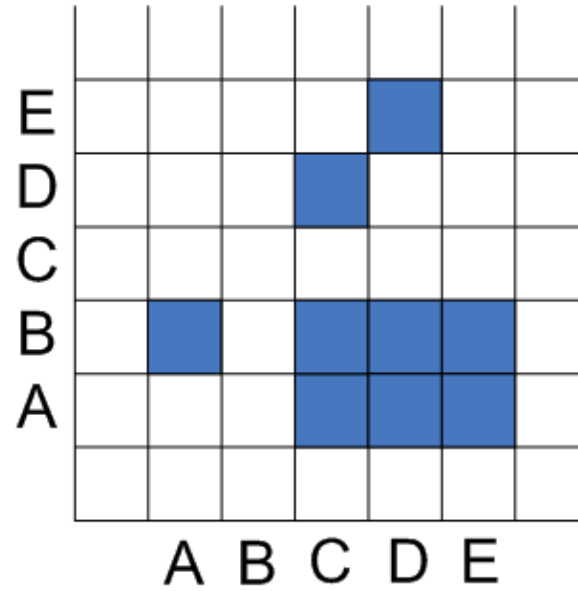
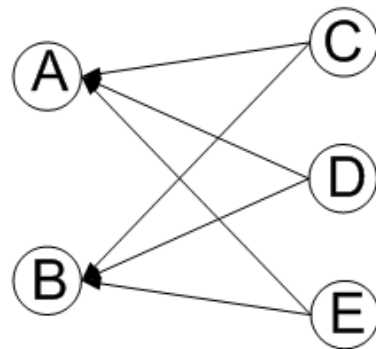
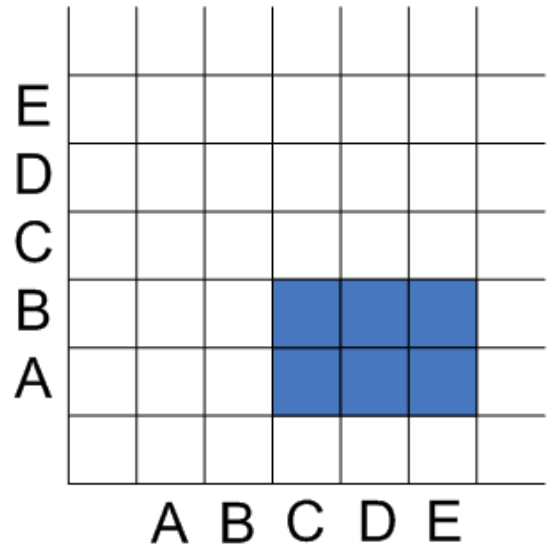
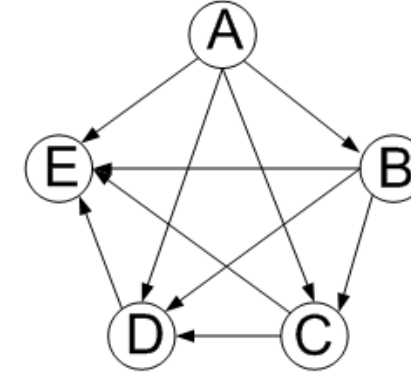
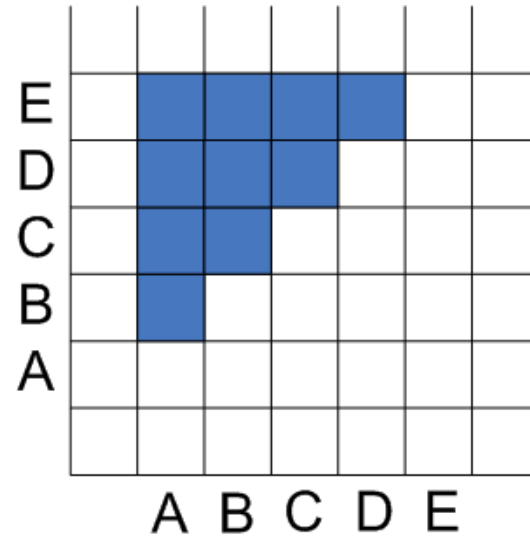
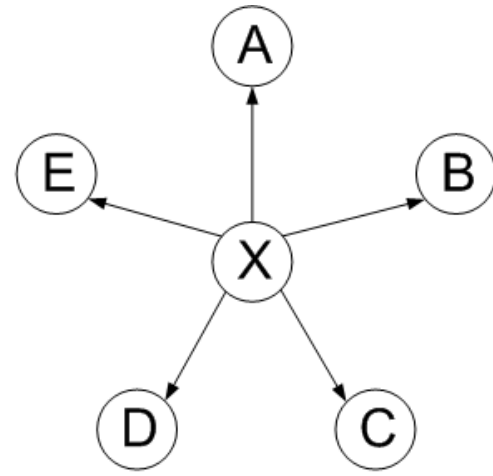
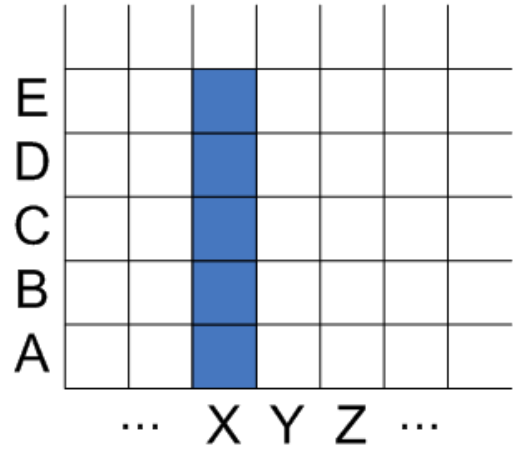
Derived Table

✓ NETWORKS

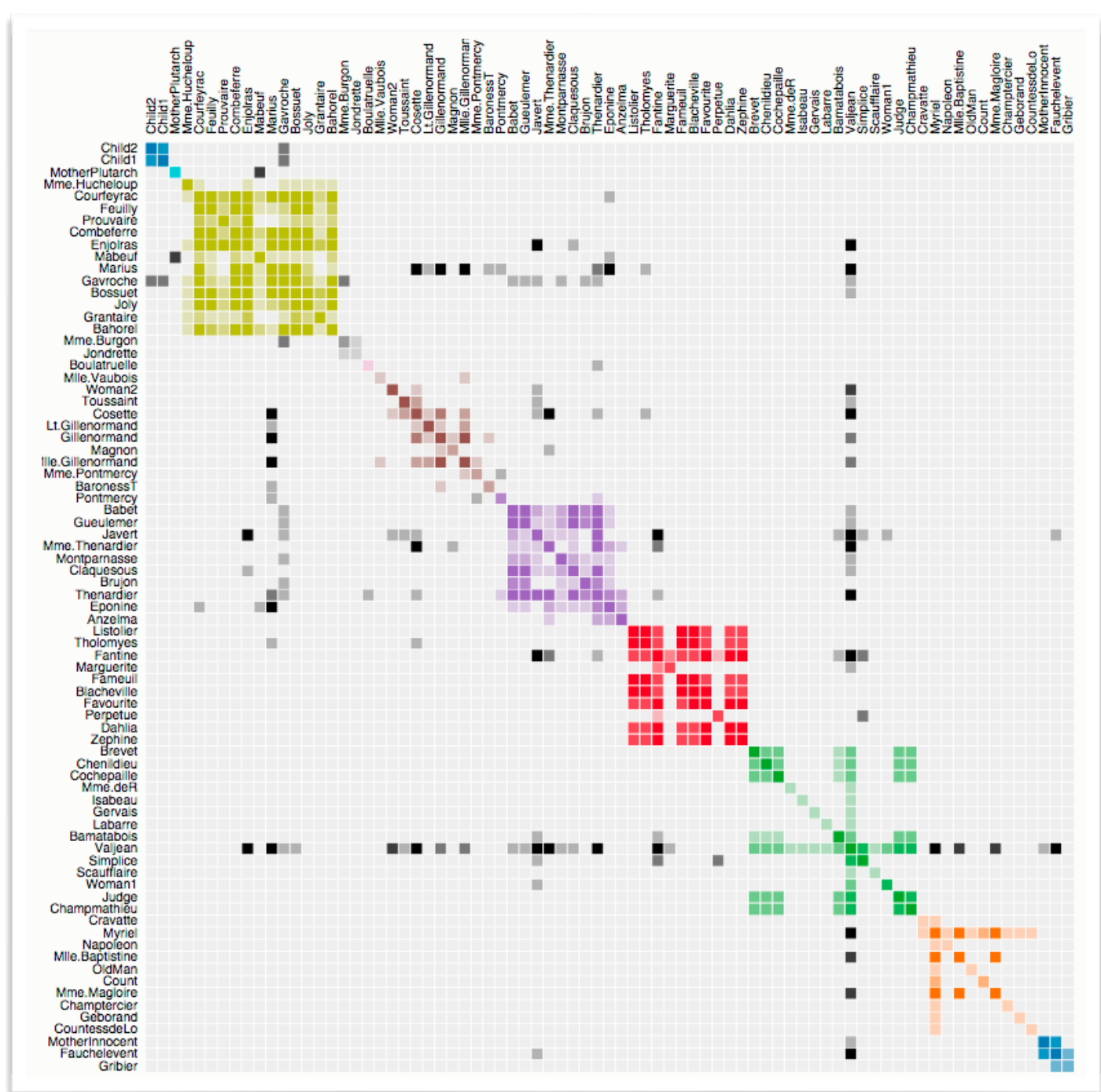
✓ TREES



# Adjacency matrix examples



# Node order is crucial: Reordering

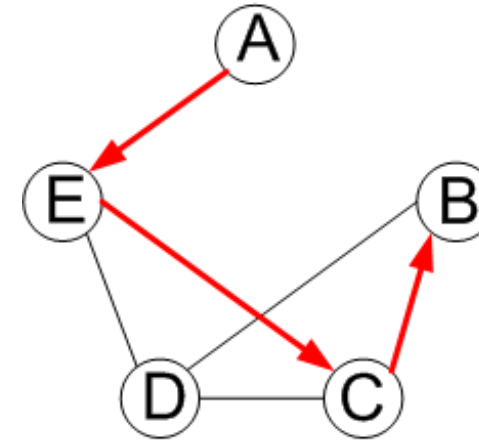


<https://bost.ocks.org/mike/miserables/>

# Adjacency matrix

		TO							
		A	B	C	D	E	F	G	H
FROM	A		■	■					
	B	■		■	■				
	C		■				■		
	D								■
	E				■		■	■	
	F					■		■	
	G						■		■
	H					■		■	

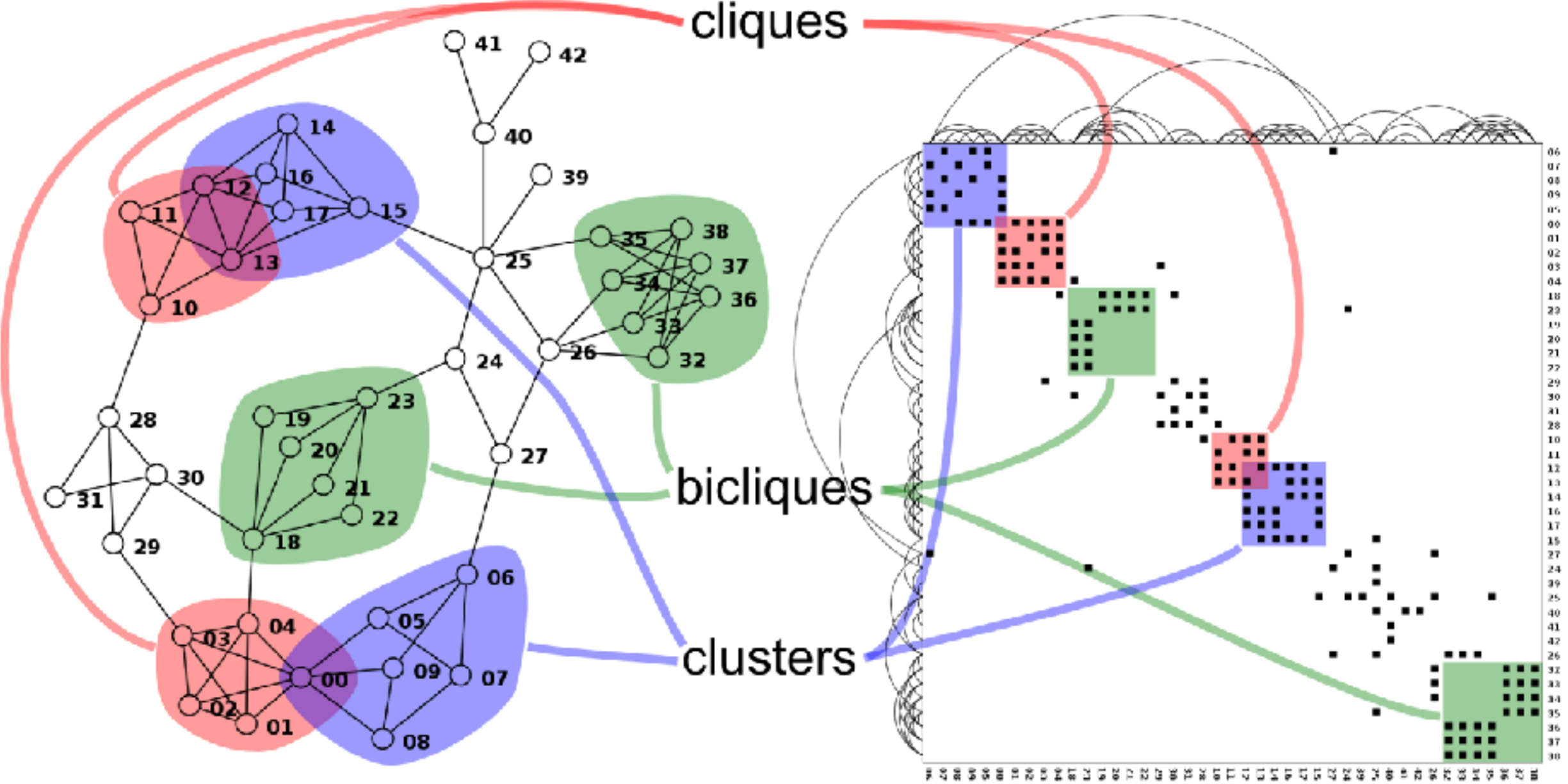
good for topology tasks  
related to neighborhoods  
(node 1-hop neighbors)



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

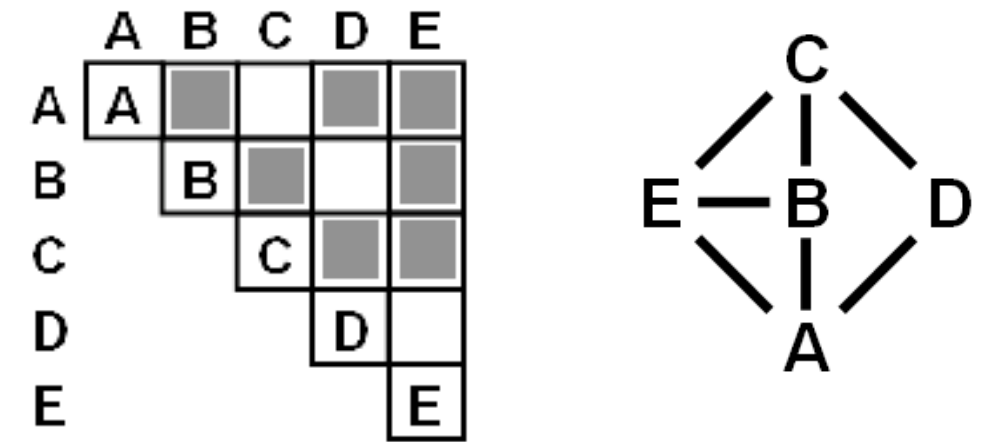
bad for topology tasks  
related to paths

# Structures visible in both

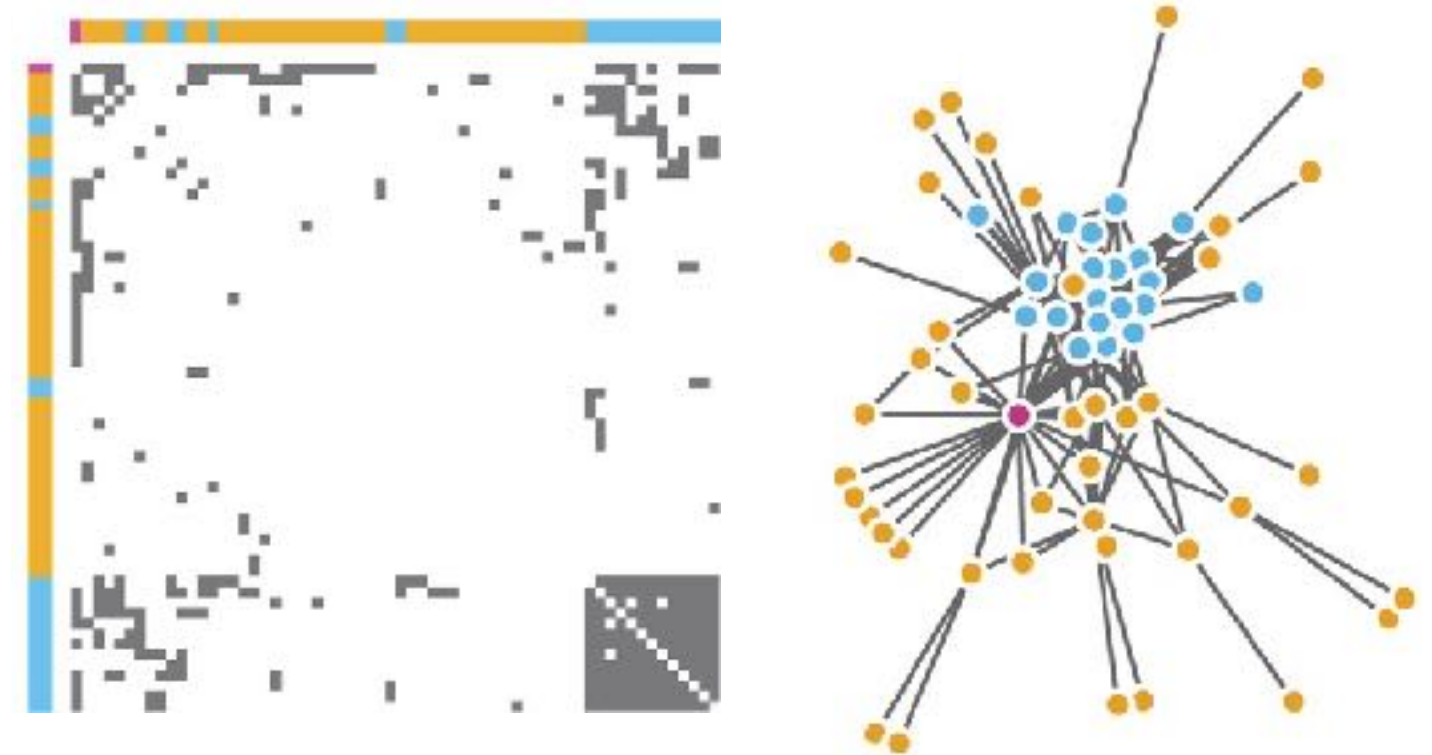


# Idiom: adjacency matrix view

- data: network
  - transform into same data/encoding as heatmap
- derived data: table from network
  - 1 quant attrib
    - weighted edge between nodes
  - 2 categ attribs: node list x 2
- visual encoding
  - cell shows presence/absence of edge
- scalability
  - 1K nodes, 1M edges



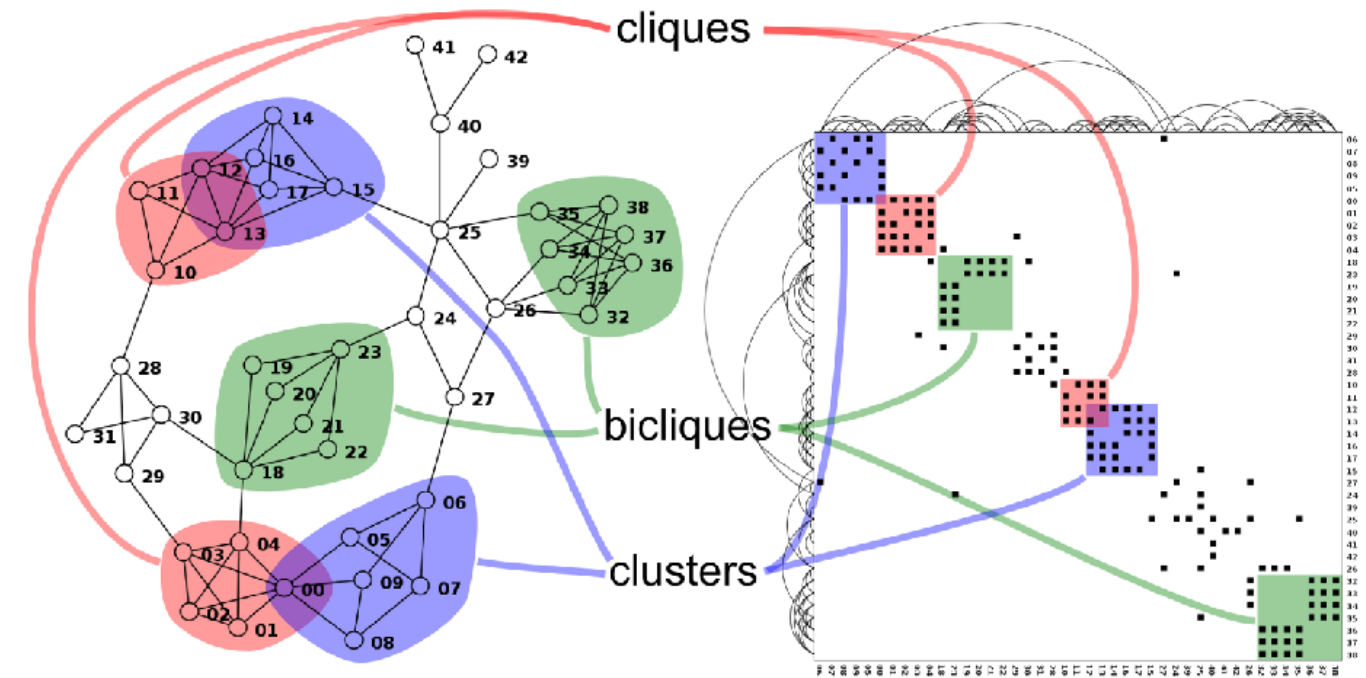
[NodeTrix: a Hybrid Visualization of Social Networks. Henry, Fekete, and McGuffin. IEEE TVCG (Proc. InfoVis) 13(6):1302-1309, 2007.]



[Points of view: Networks. Gehlenborg and Wong. Nature Methods 9:115.]

# Node-link vs. matrix comparison

- node-link diagram strengths
  - topology understanding, path tracing
  - intuitive, flexible, no training needed
- adjacency matrix strengths
  - focus on edges rather than nodes
  - layout straightforward (reordering needed)
  - predictability, scalability
  - some topology tasks trainable
- empirical study
  - node-link best for small networks
  - matrix best for large networks
    - if tasks don't involve path tracing!

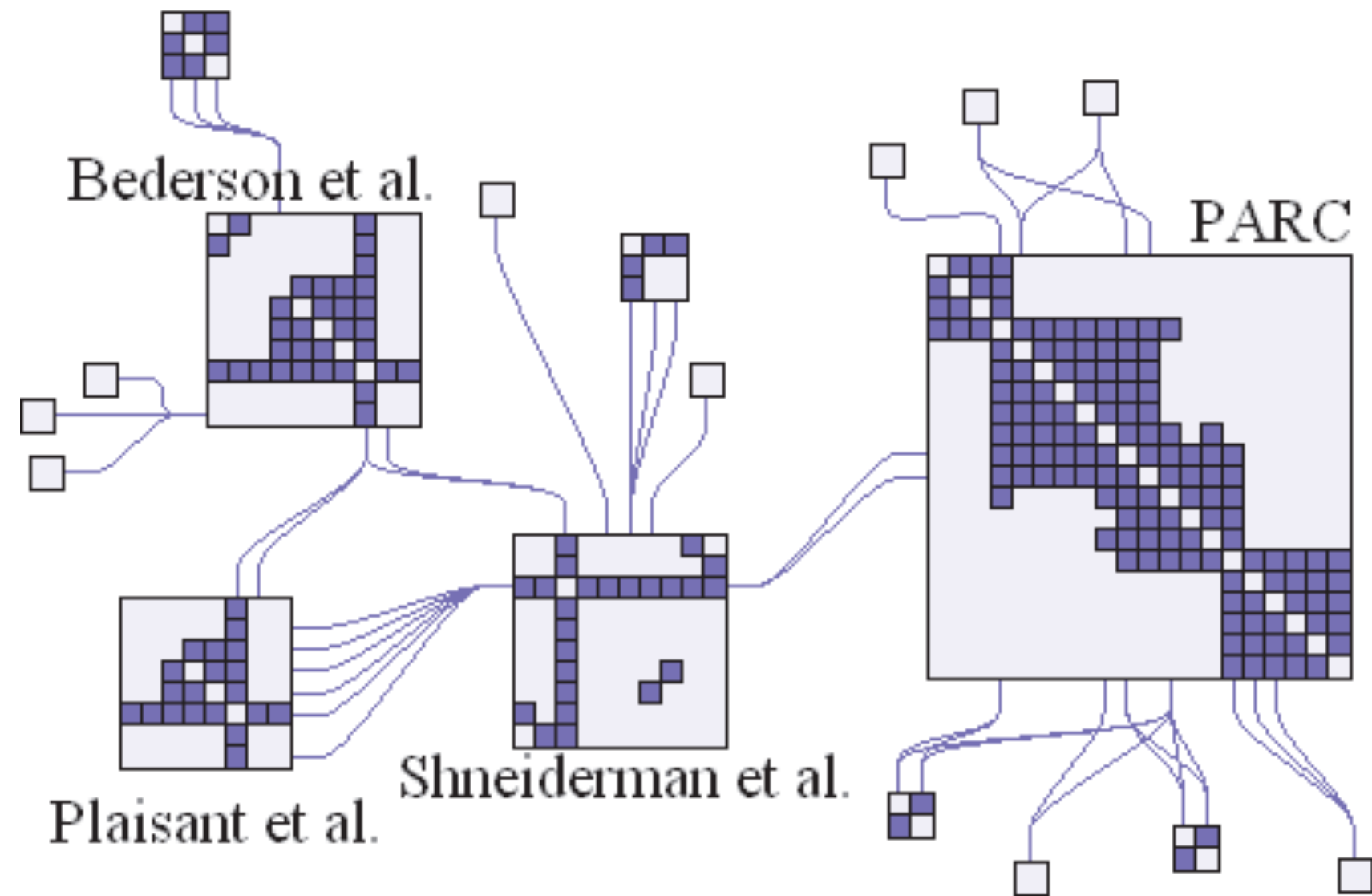


<http://www.michaelmcguffin.com/courses/vis/patternsInAdjacencyMatrix.png>

*[On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis. Ghoniem, Fekete, and Castagliola. Information Visualization 4:2 (2005), 114–135.]*

# Idiom: NodeTrix

- hybrid nodelink/matrix
- capture strengths of both



[NodeTrix: a Hybrid Visualization of Social Networks.  
Henry, Fekete, and McGuffin. IEEE TVCG (Proc. InfoVis)  
13(6):1302-1309, 2007.]



# Trees

# Node-link trees

- Reingold-Tilford

- tidy drawings of trees

- exploit parent/child structure

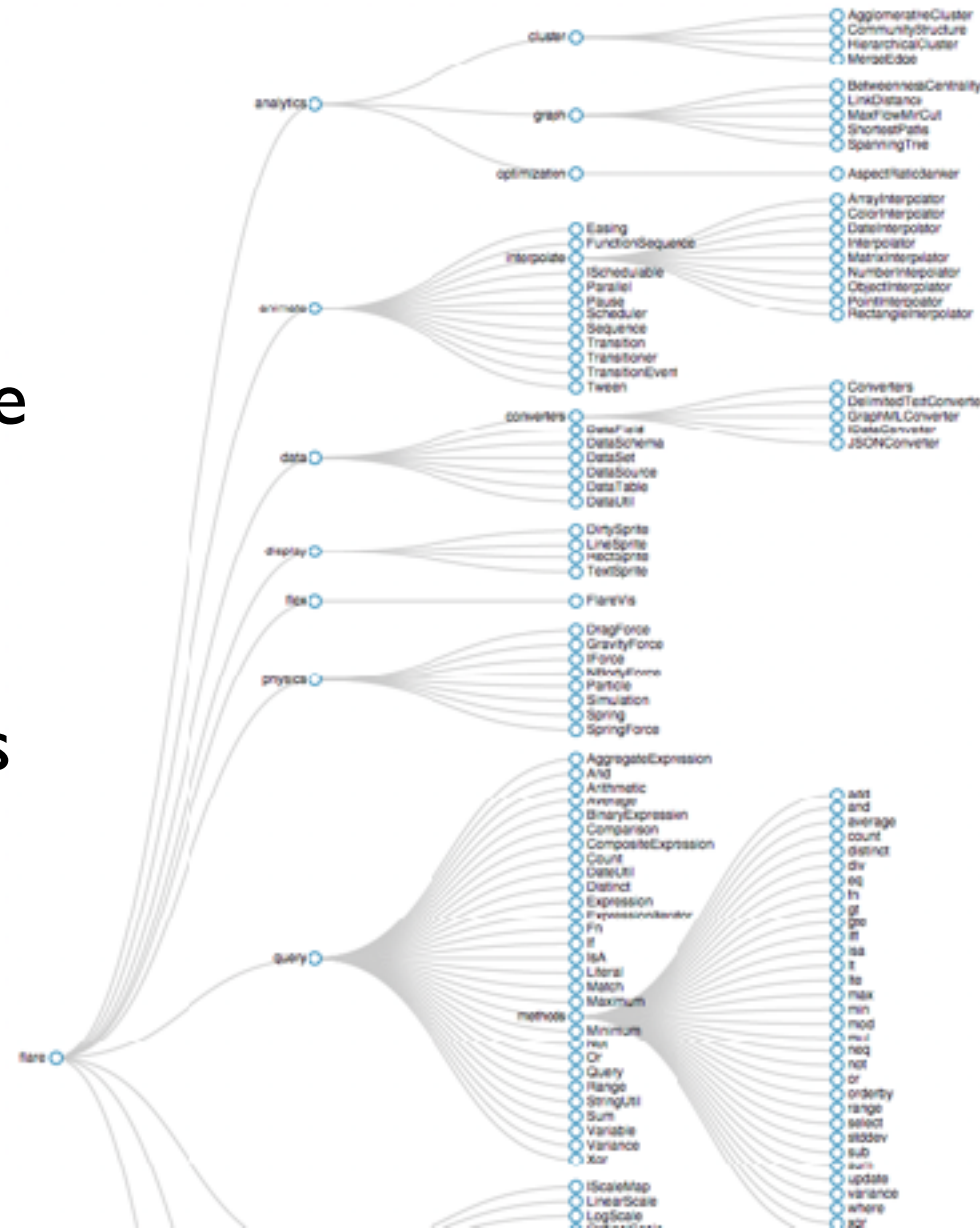
- allocate space: compact but without overlap

- rectilinear and radial variants

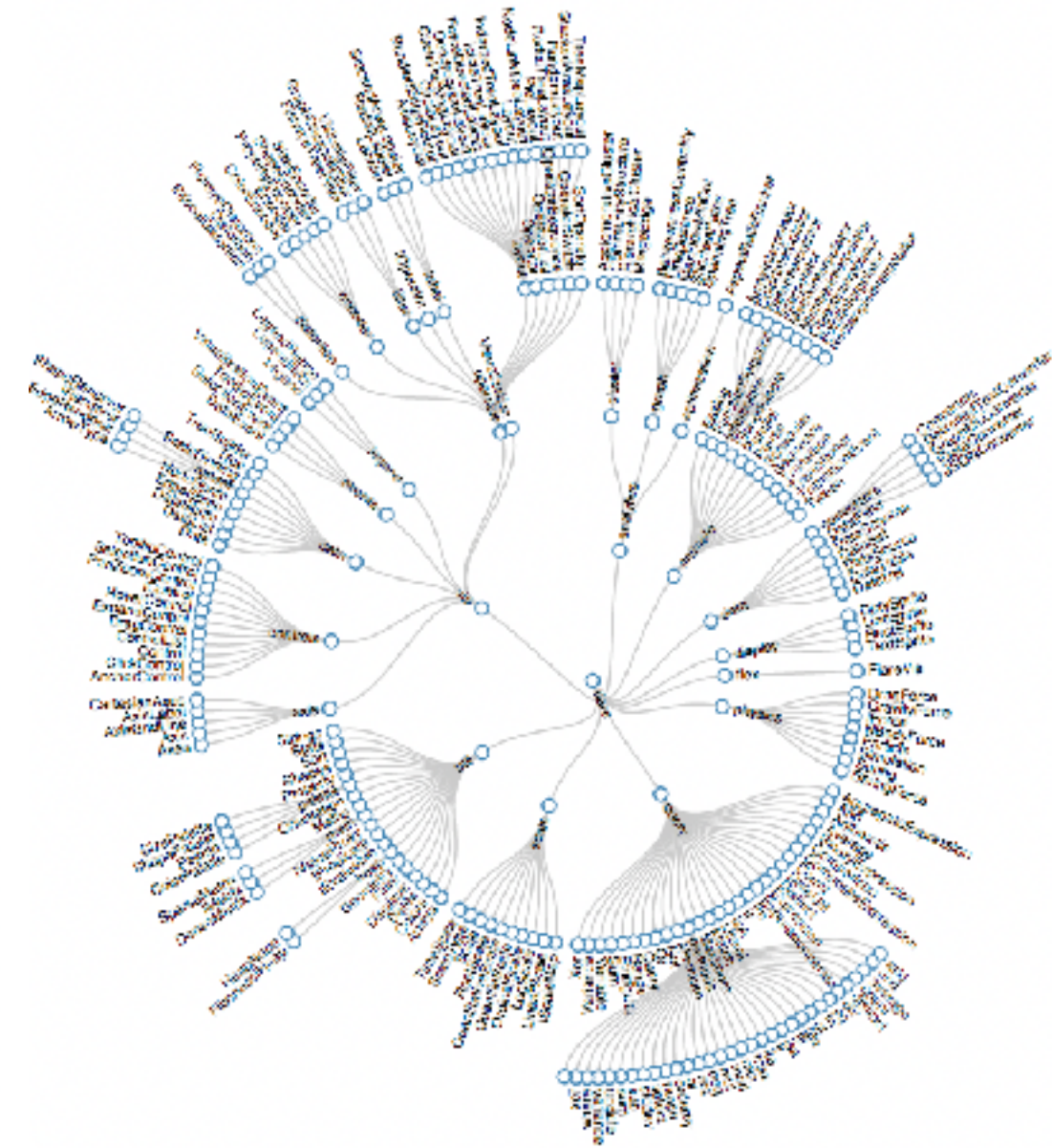
[Tidier drawing of trees. Reingold and Tilford. IEEE Trans. Software Eng., SE-7(2):223–228, 1981.]

- nice algorithm writeup

<http://billmill.org/pymag-trees/>



<http://bl.ocks.org/mbostock/4339184>

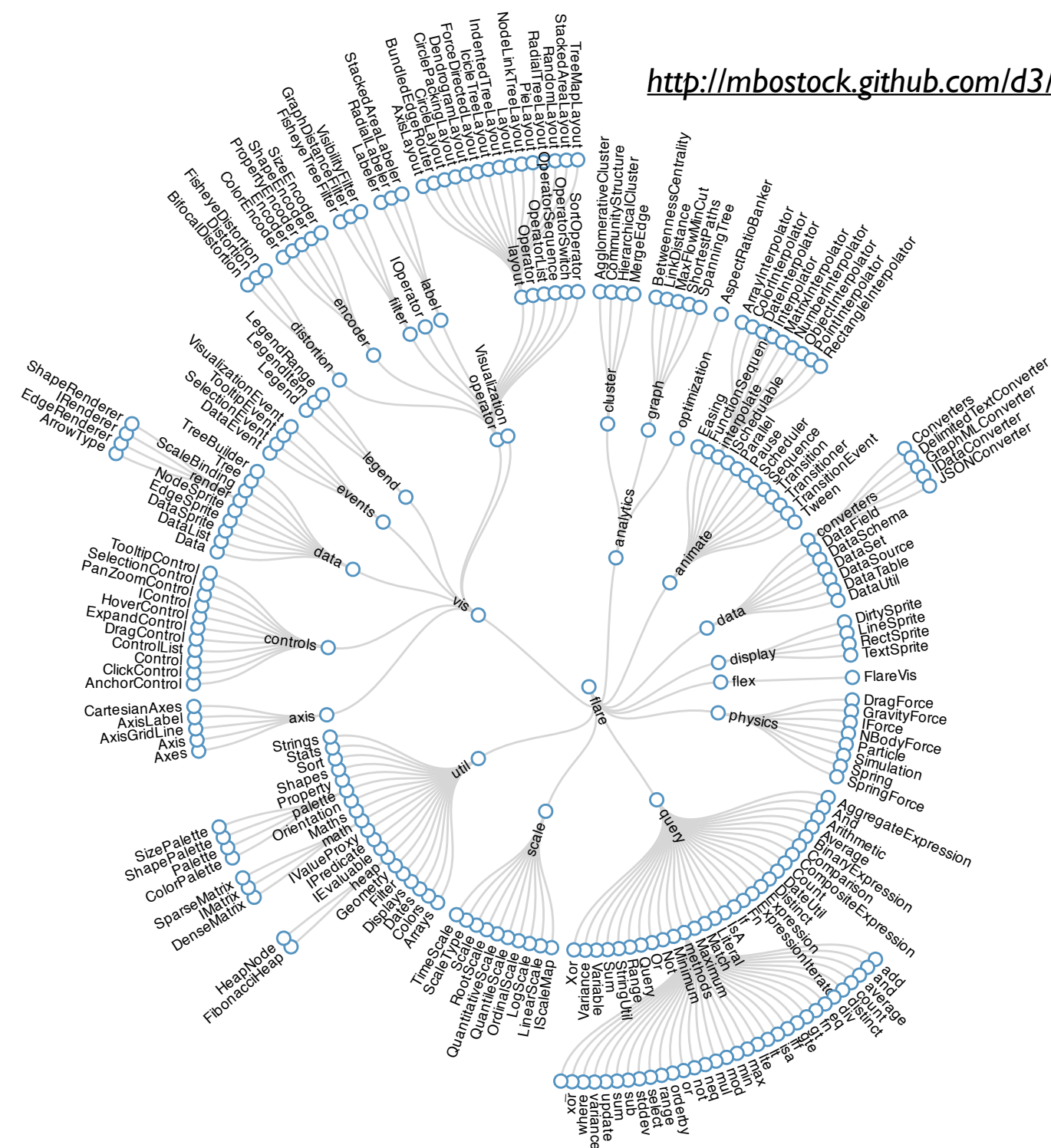


<http://bl.ocks.org/mbostock/4063550>

# Idiom: radial node-link tree

<http://mbostock.github.com/d3/ex/tree.html>

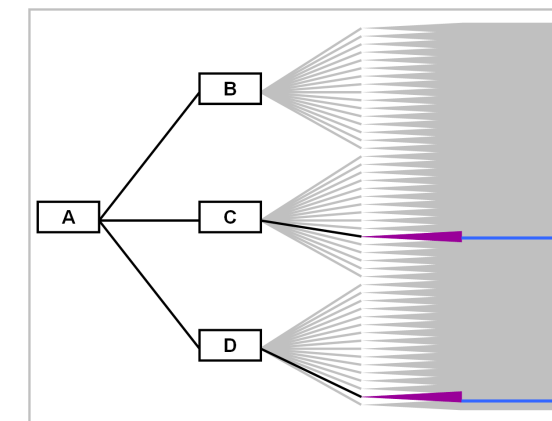
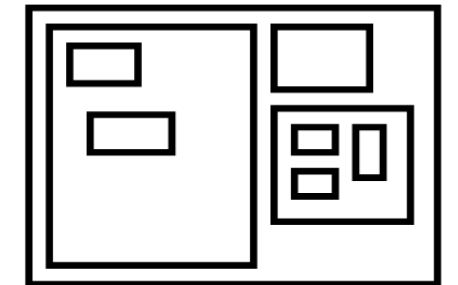
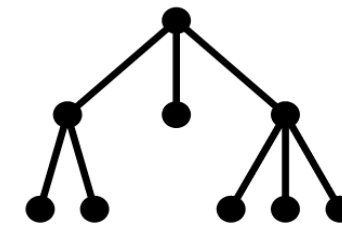
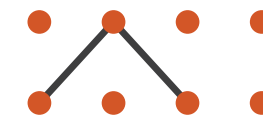
- data
  - tree
- encoding
  - link connection marks
  - point node marks
  - radial axis orientation
    - angular proximity: siblings
    - distance from center: depth in tree
- tasks
  - understanding topology, following paths
- scalability
  - 1K - 10K nodes (with/without labels)



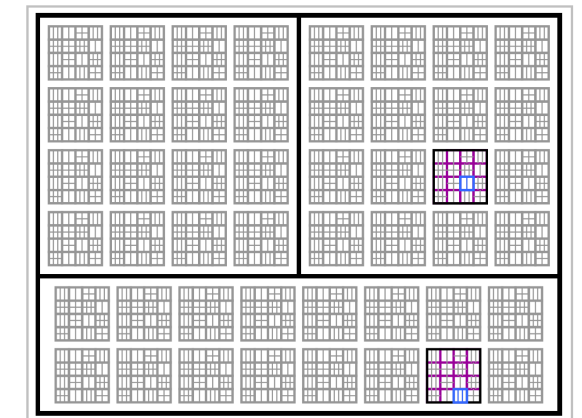
# Link marks: Connection and containment

- marks as links (vs. nodes)
  - common case in network drawing
  - 1D case: connection
    - ex: all node-link diagrams
    - emphasizes topology, path tracing
    - networks and trees
  - 2D case: containment
    - ex: all treemap variants
    - emphasizes attribute values at leaves (size coding)
    - only trees

→ Connection → Containment



**Node-Link Diagram**



**Treemap**

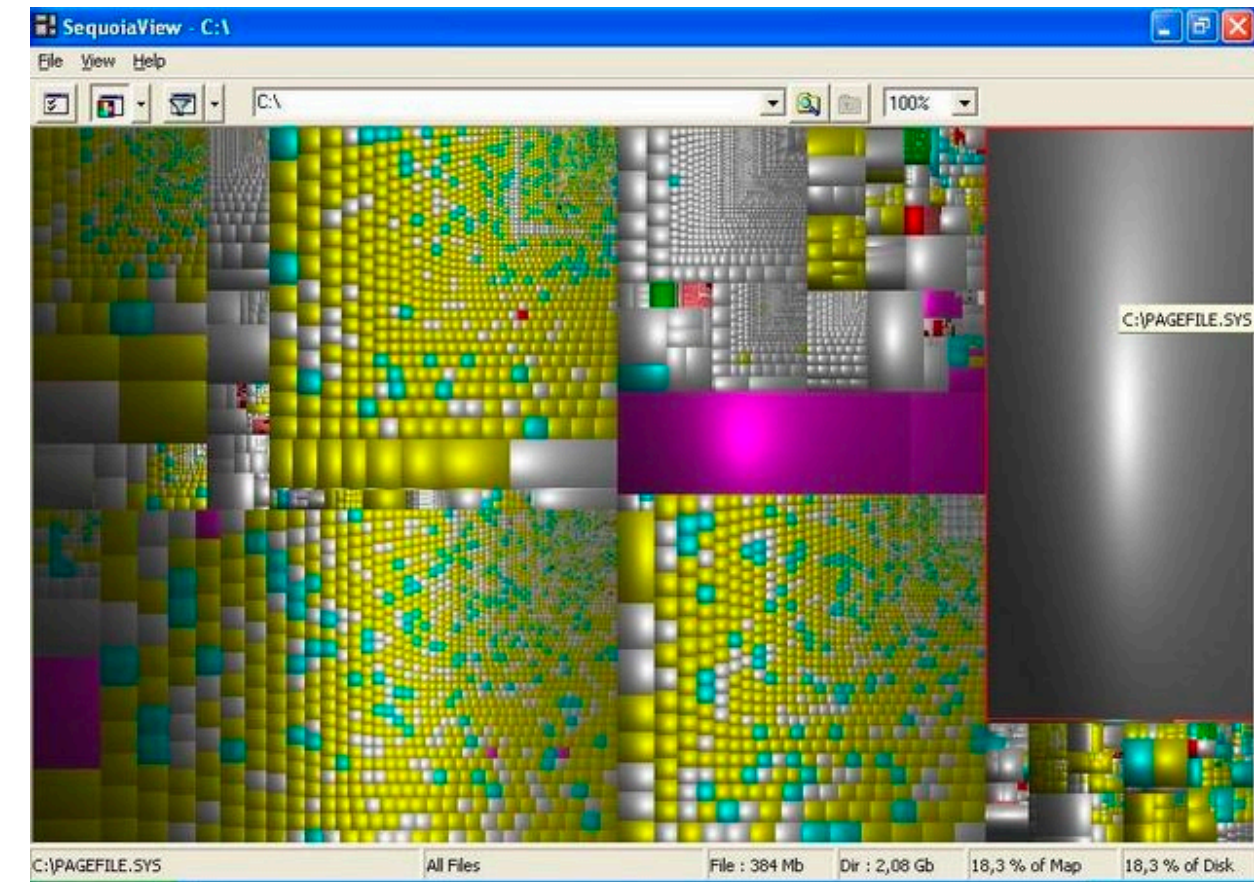
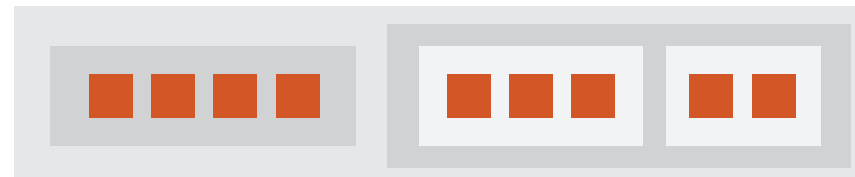
[Elastic Hierarchies: Combining Treemaps and Node-Link Diagrams.  
Dong, McGuffin, and Chignell. Proc. InfoVis 2005, p. 57-64.]

# Idiom: treemap

- data
  - tree
  - 1 quant attrib at leaf nodes
- encoding
  - area containment marks for hierarchical structure
  - rectilinear orientation
  - size encodes quant attrib
- tasks
  - query attribute at leaf nodes
  - ex: disk space usage within filesystem
- scalability
  - IM leaf nodes

➔ **Enclosure**  
Containment Marks

NETWORKS  TREES



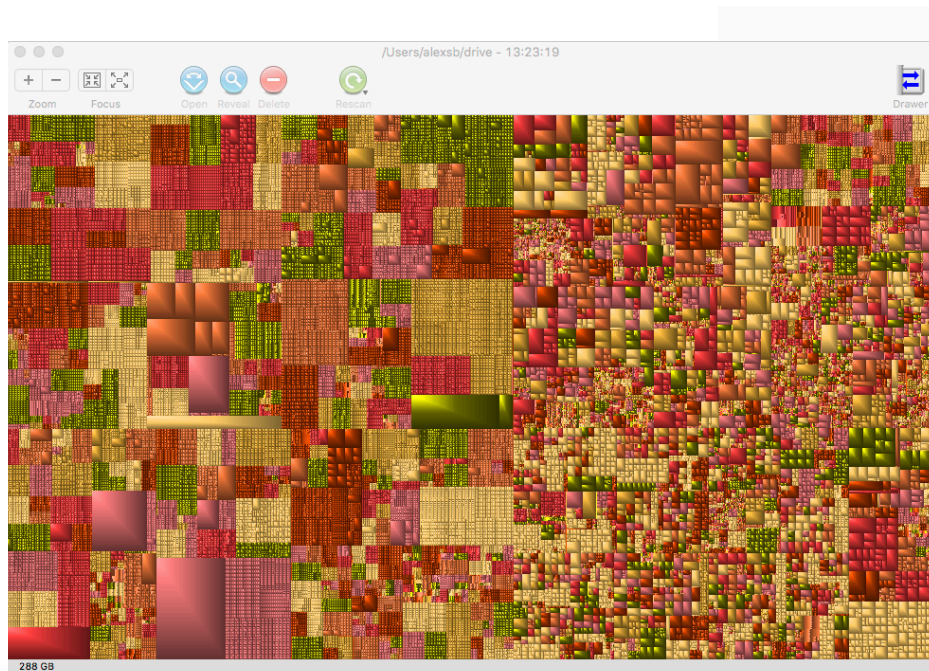
<https://www.win.tue.nl/sequoiaview/>

[Cushion Treemaps. van Wijk and van de Wetering.  
Proc. Symp. InfoVis 1999, 73-78.]

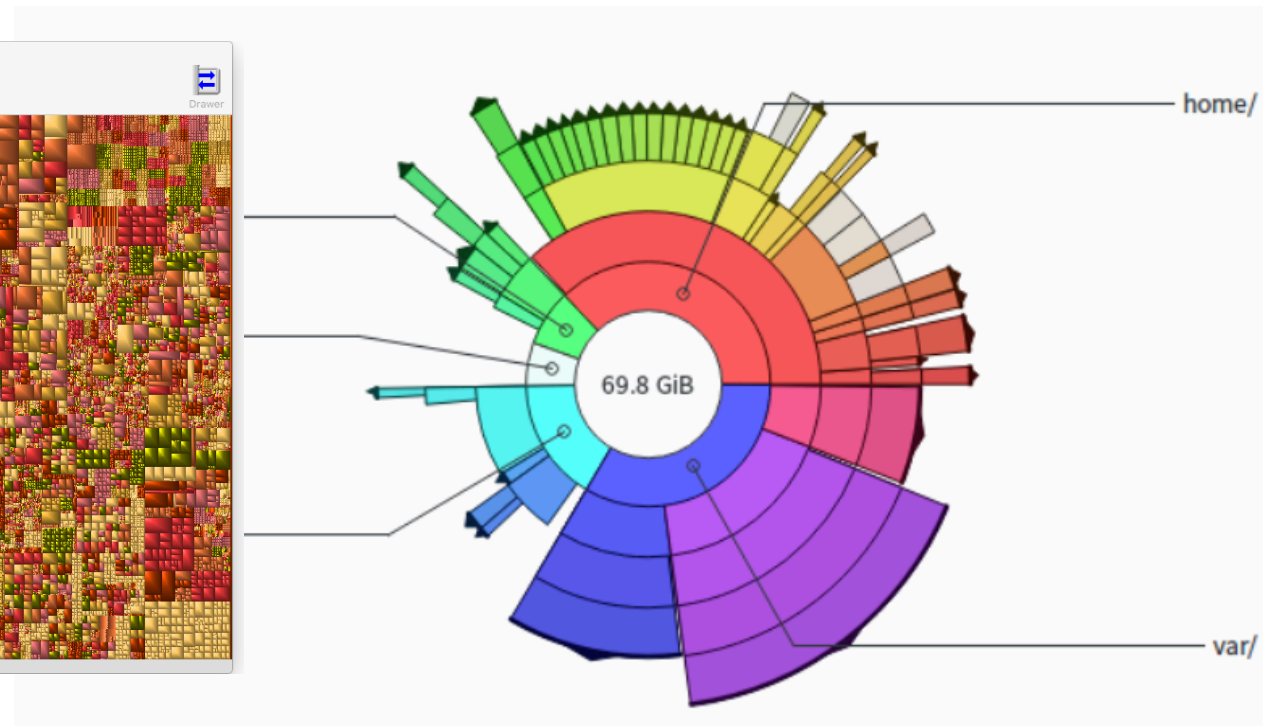
# Idiom: implicit tree layouts (sunburst, icicle plot)

- alternative to connection and containment: position
  - show parent-child relationships only through relative positions

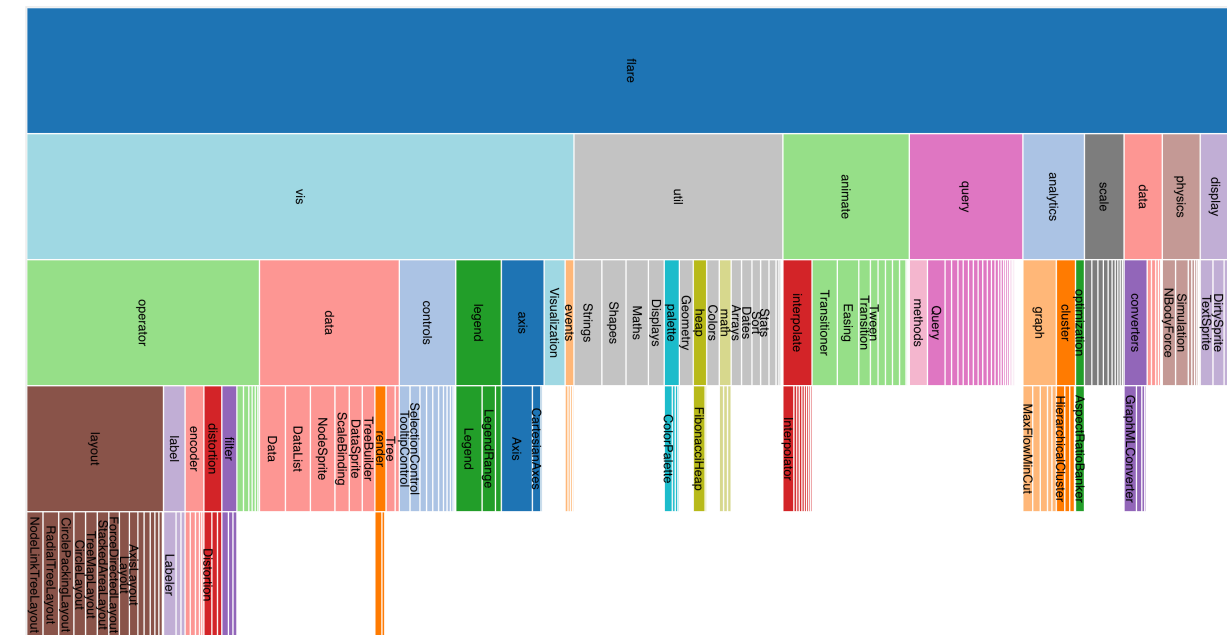
Treemap  
containment



Sunburst  
position (radial)



Icicle Plot  
position (rectilinear)





# Idiom: implicit tree layouts (sunburst, icicle plot)

- alternative to connection and containment: position
  - show parent-child relationships only through relative positions

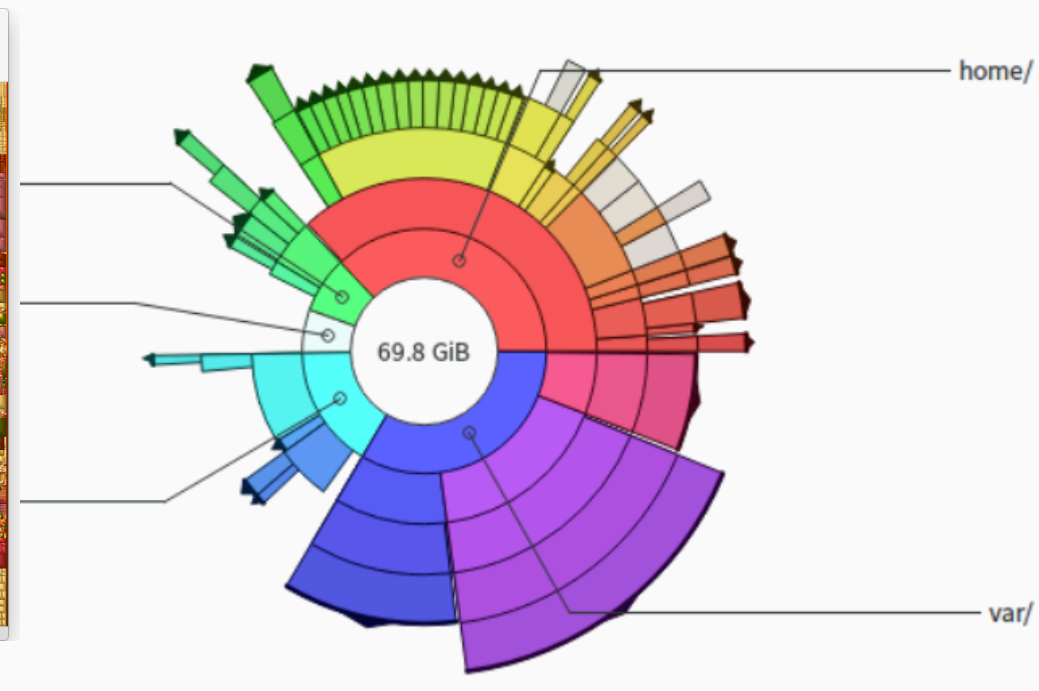
## Treemap

containment  
only leaves visible



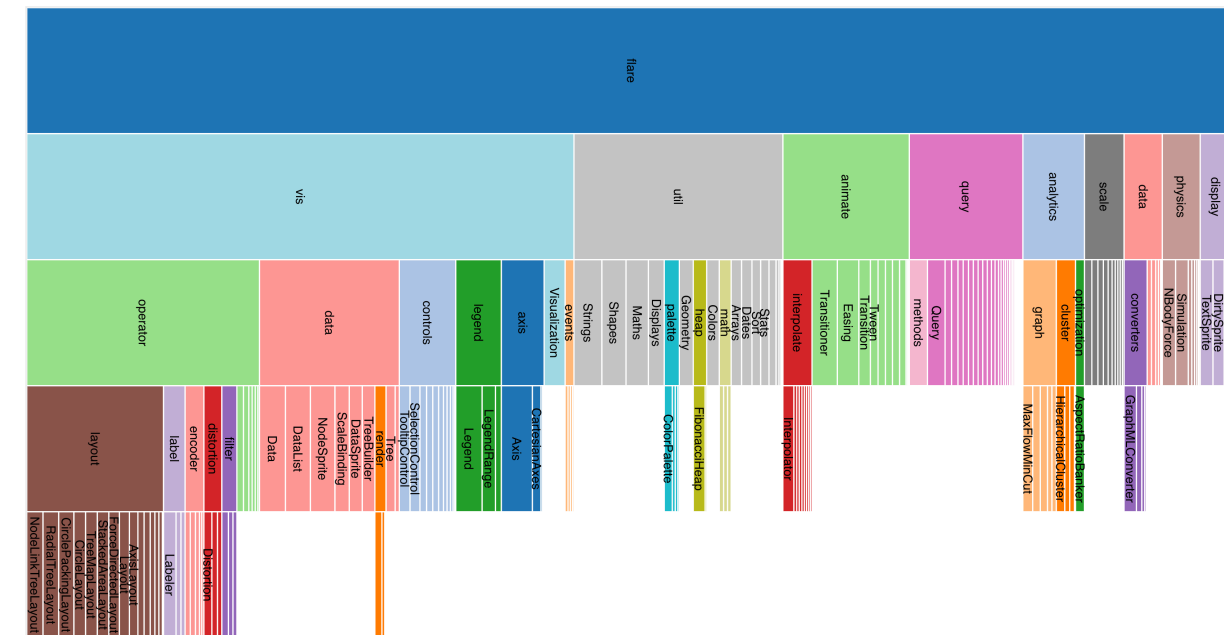
## Sunburst

position (radial)  
inner nodes & leaves visible



## Icicle Plot

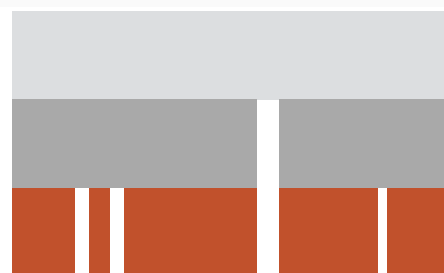
position (rectilinear)  
inner nodes & leaves visible



**Implicit**  
Spatial Position

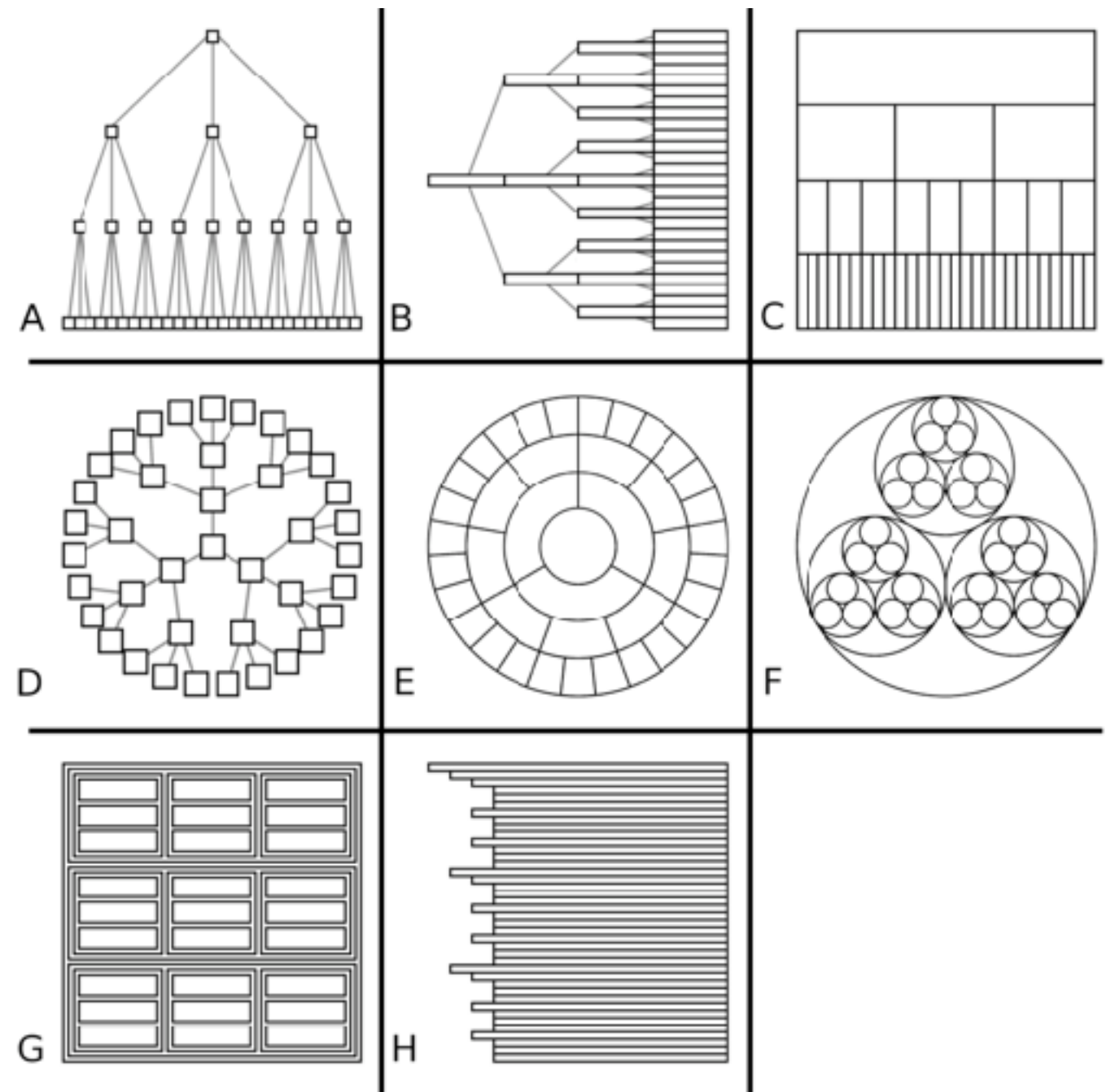
✗ NETWORKS

✓ TREES



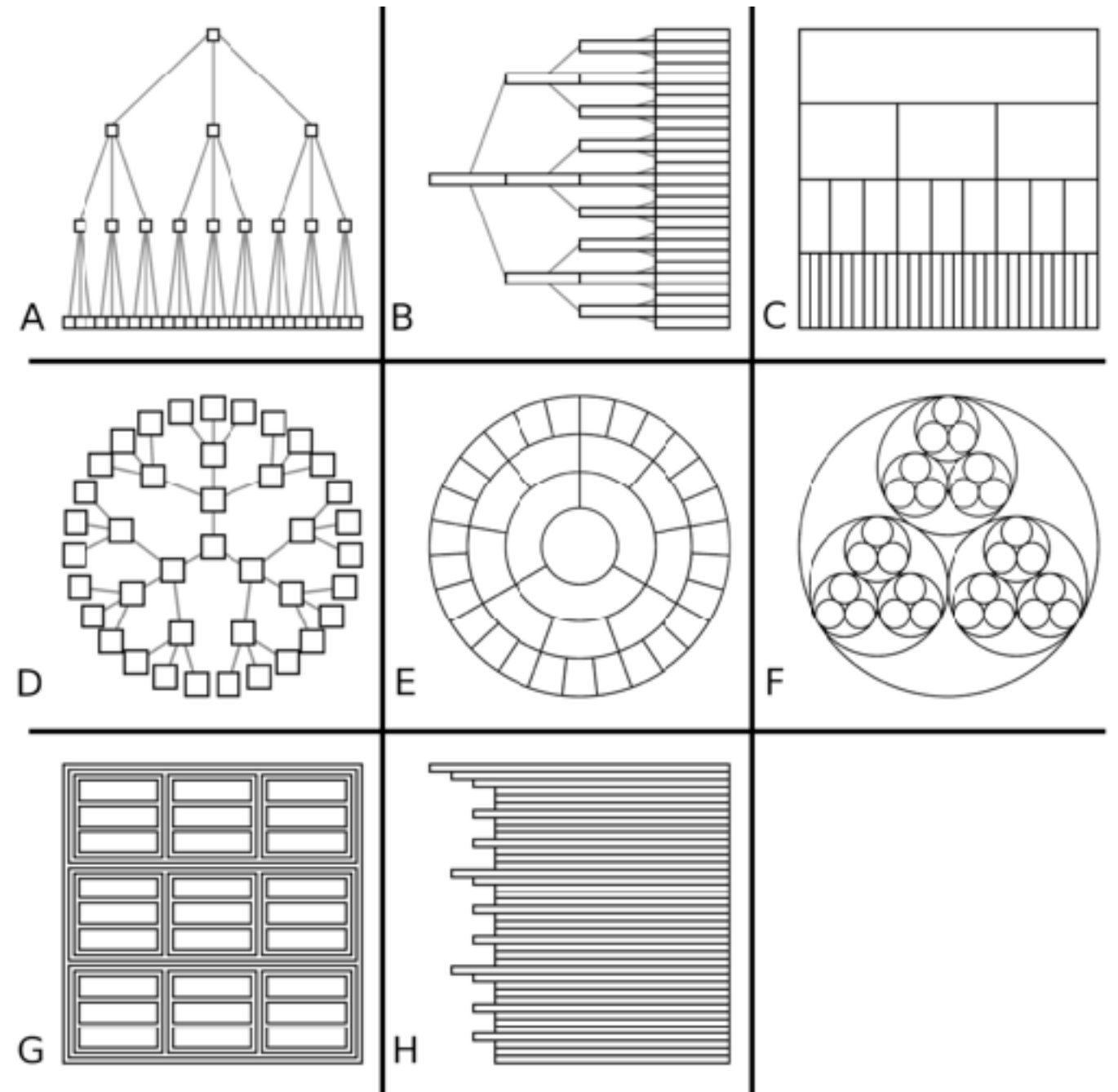


# Tree drawing idioms comparison



# Comparison: tree drawing idioms

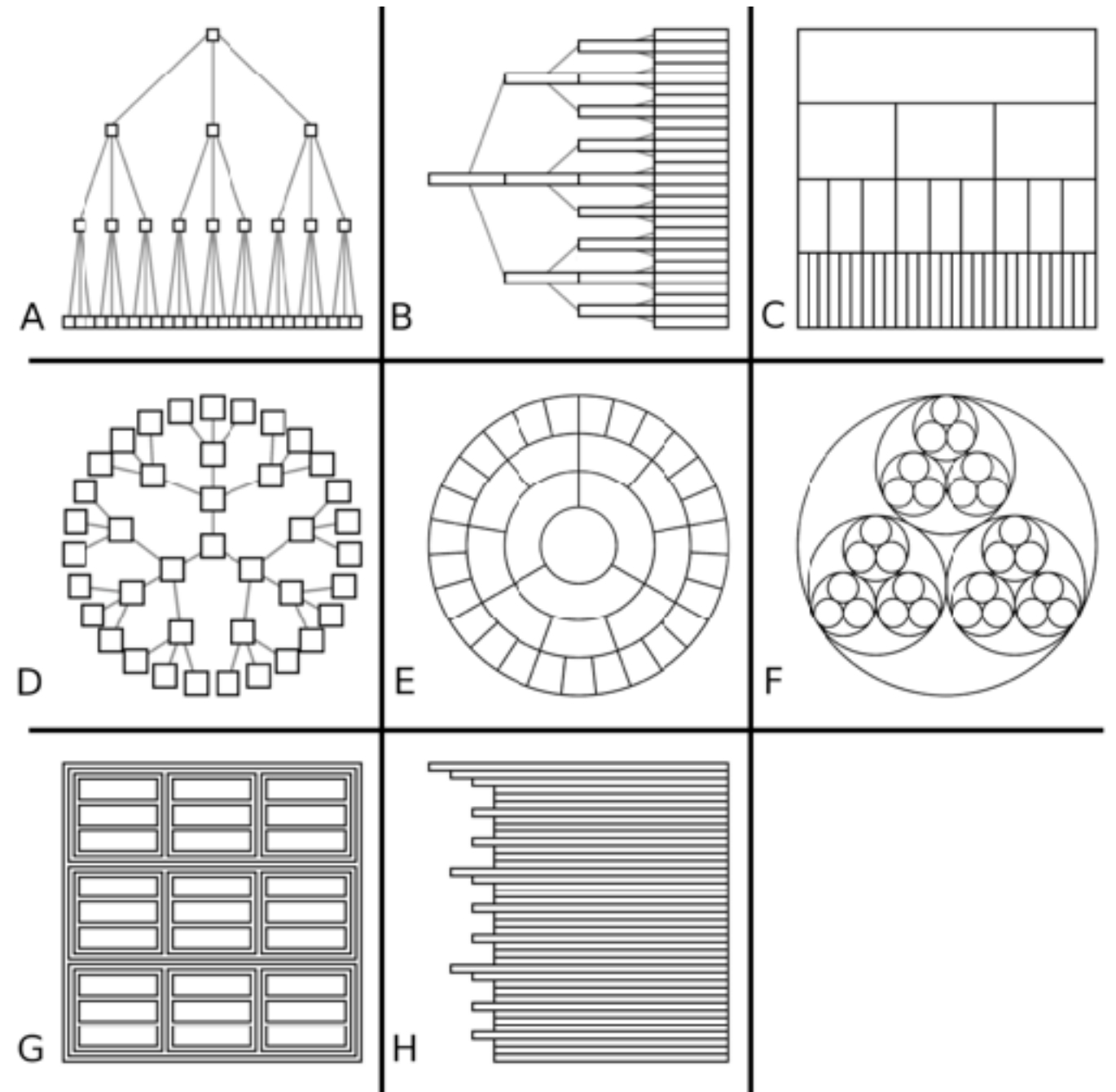
- data shown
  - link relationships
  - tree depth
  - sibling order



[Quantifying the Space-Efficiency of 2D Graphical Representations of Trees.  
McGuffin and Robert. *Information Visualization* 9:2 (2010), 115–140.]

# Comparison: tree drawing idioms

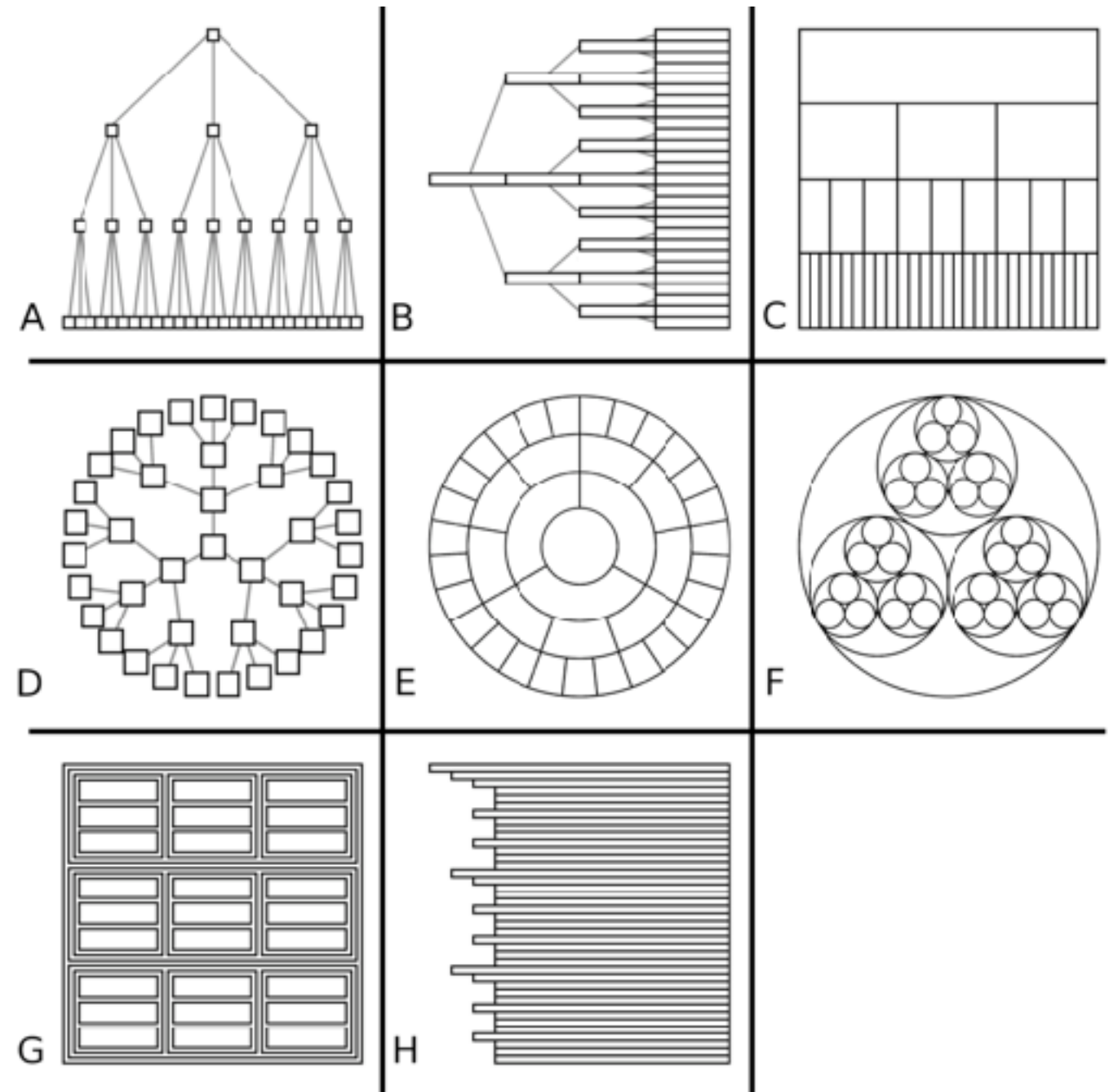
- data shown
  - link relationships
  - tree depth
  - sibling order
- design choices
  - connection vs containment link marks
  - rectilinear vs radial layout
  - spatial position channels



[Quantifying the Space-Efficiency of 2D Graphical Representations of Trees.  
McGuffin and Robert. *Information Visualization* 9:2 (2010), 115–140.]

# Comparison: tree drawing idioms

- data shown
  - link relationships
  - tree depth
  - sibling order
- design choices
  - connection vs containment link marks
  - rectilinear vs radial layout
  - spatial position channels
- considerations
  - redundant? arbitrary?
  - information density?
    - avoid wasting space
    - consider where to fit labels!



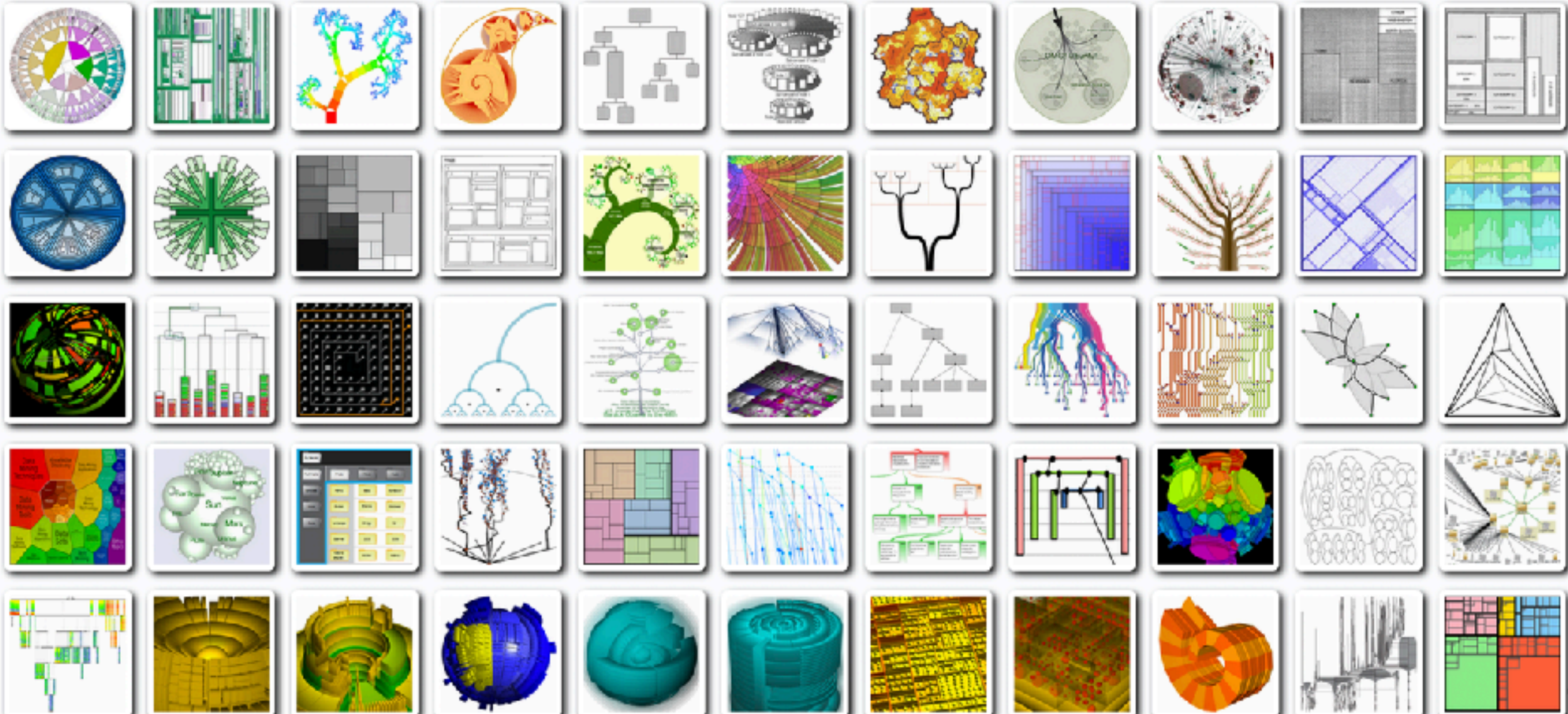


[Quantifying the Space-Efficiency of 2D Graphical Representations of Trees.  
McGuffin and Robert. *Information Visualization* 9:2 (2010), 115–140.]




# [treevis.net](http://treevis.net): Many, many options!




How to cite this site? v.21-OCT-2014

treevis.net - A Visual Bibliography of Tree Visualization 2.0 by Hans-Jörg Schulz

Check out other surveys!

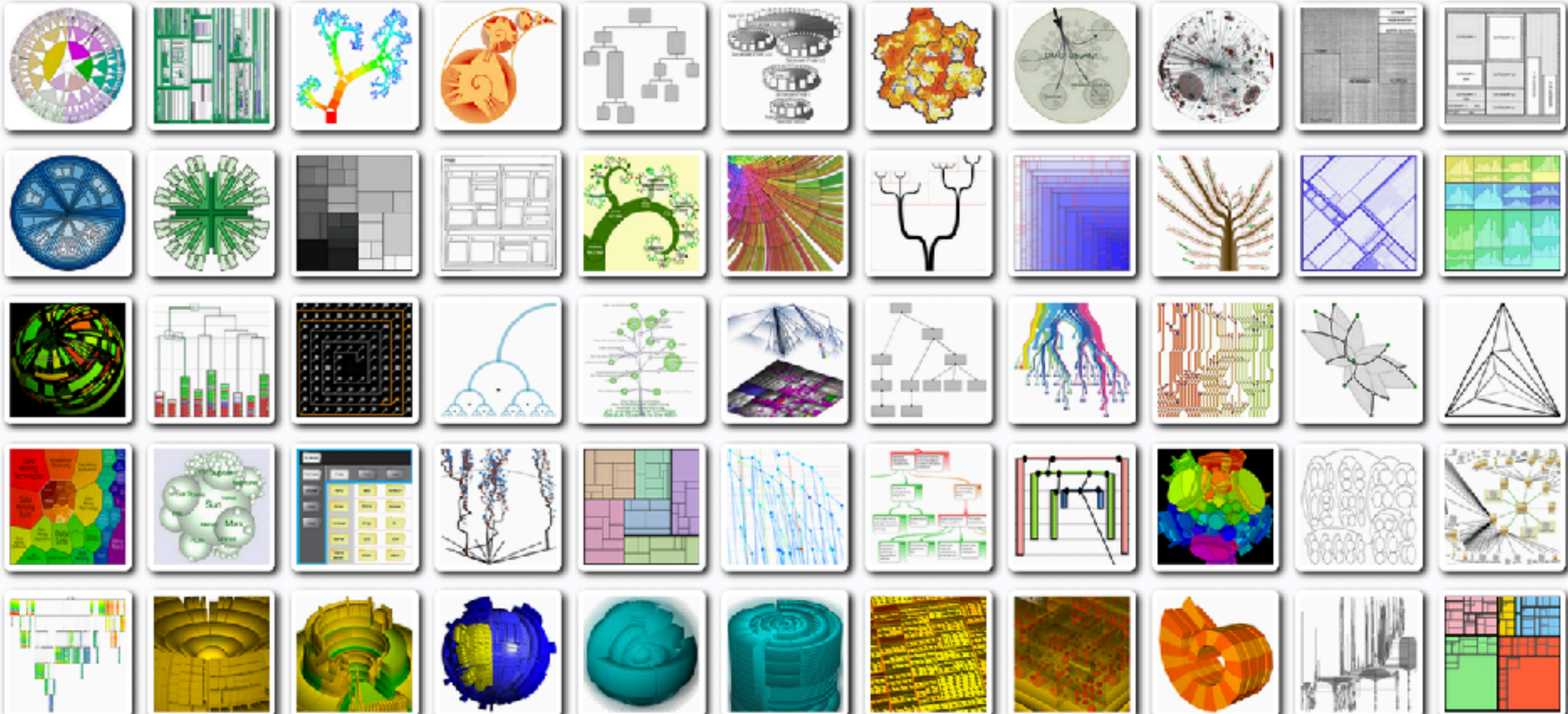
Dimensionality: All   

Representation: All   

Alignment: All   

Fulltext Search:  ×

Techniques Shown: 277

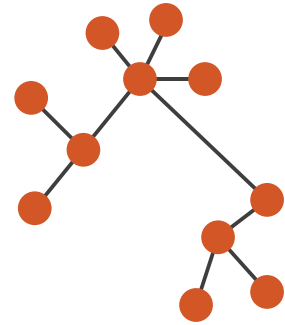


<https://treevis.net/>

# Arrange networks and trees

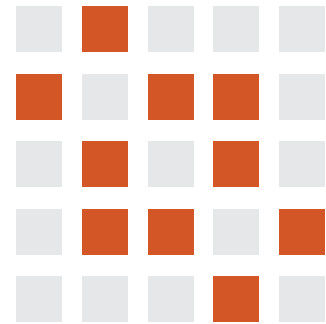
## → Node–Link Diagrams Connection Marks

✓ NETWORKS ✓ TREES



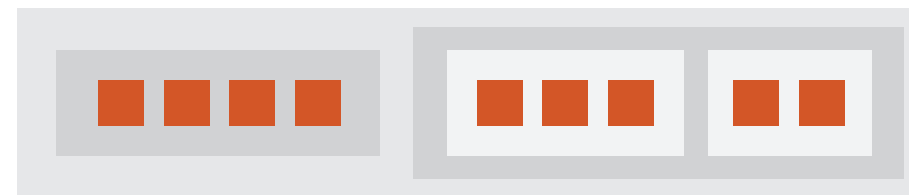
## → Adjacency Matrix Derived Table

✓ NETWORKS ✓ TREES



## → Enclosure Containment Marks

✗ NETWORKS ✓ TREES



## → Implicit Spatial Position

✗ NETWORKS ✓ TREES



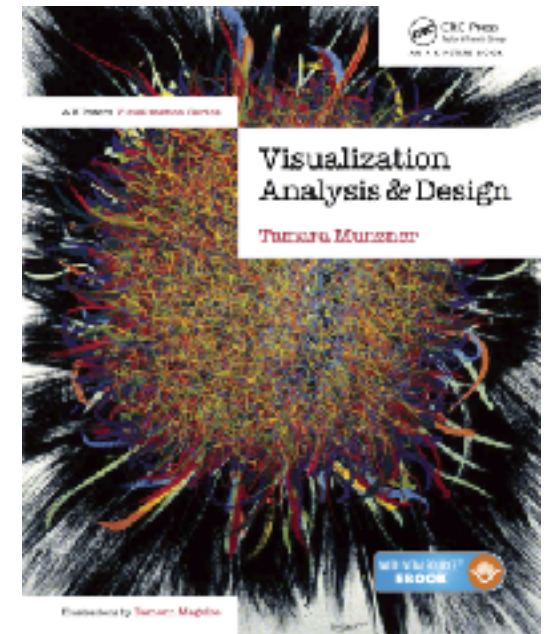
# Visualization Analysis & Design

## *Network Data (Ch 9) II*

**Tamara Munzner**

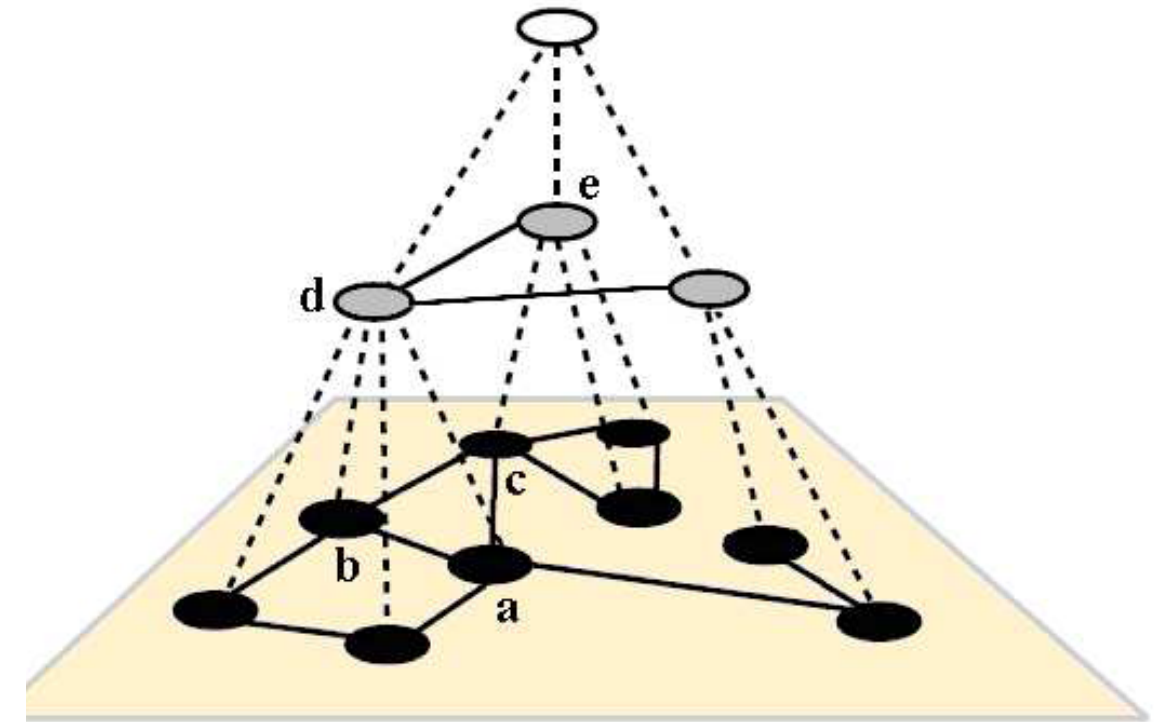
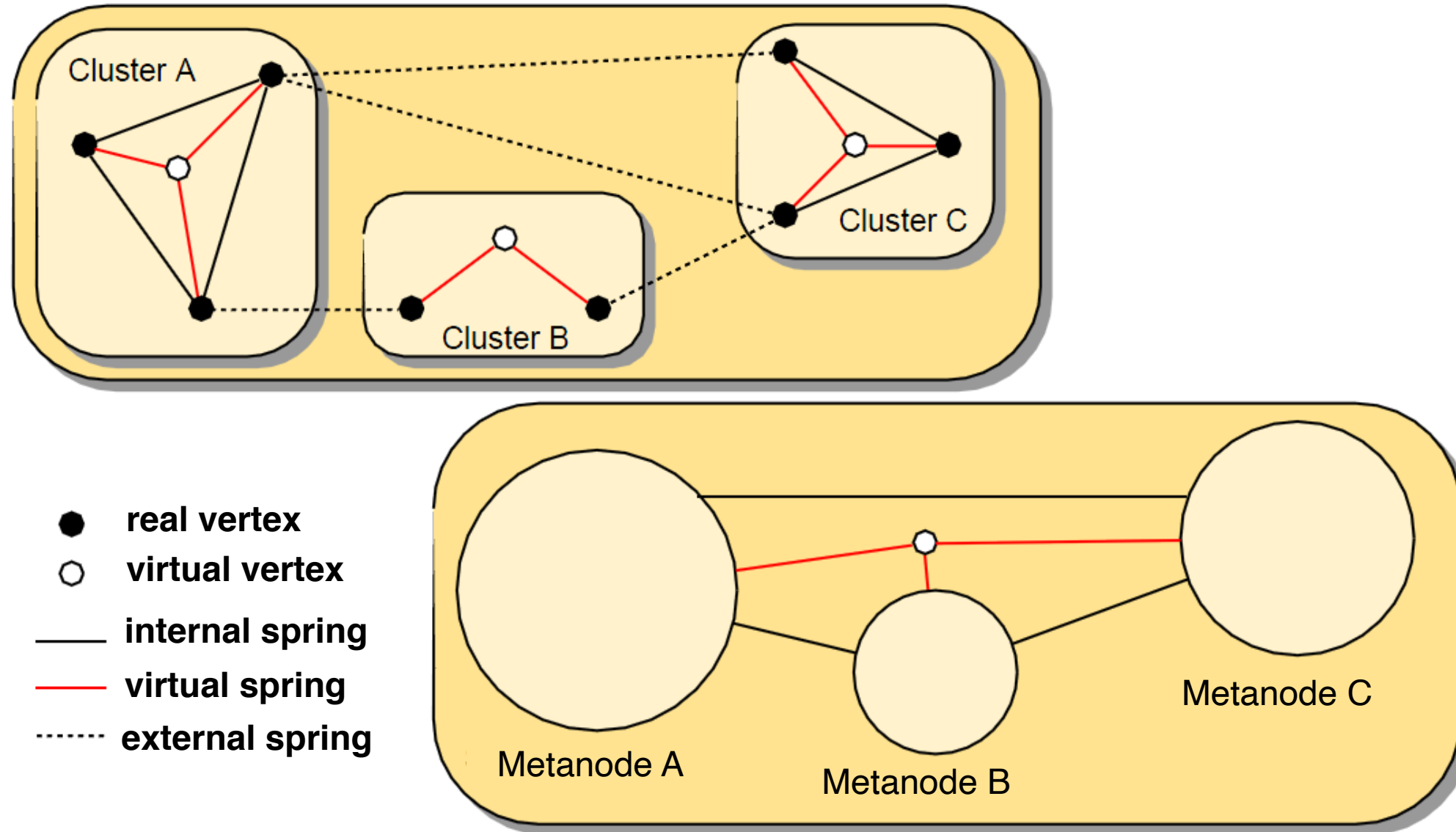
Department of Computer Science  
University of British Columbia

[@tamaramunzner](#)



# Multilevel networks

- derive cluster hierarchy of metanodes on top of original graph nodes

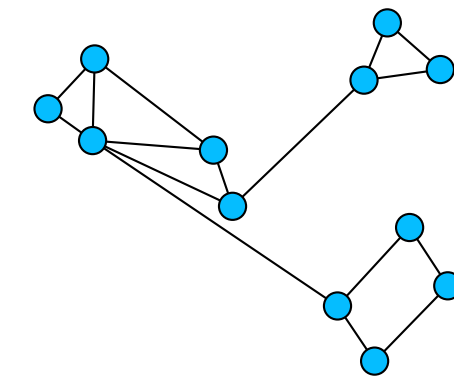


[Schulz 2004]

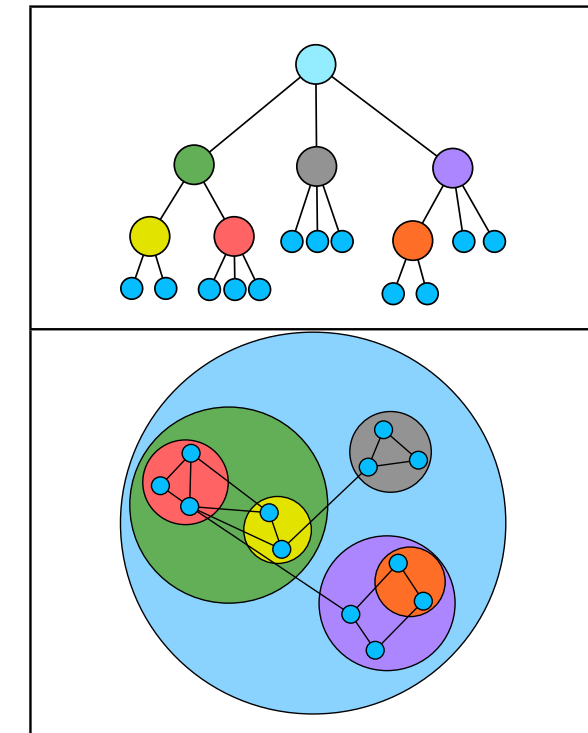


# Idiom: GrouseFlocks

- data: compound network
  - network
  - cluster hierarchy atop it
    - derived or interactively chosen
- visual encoding
  - connection marks for network links
  - containment marks for hierarchy
  - point marks for nodes
- dynamic interaction
  - select individual metanodes in hierarchy to expand/contract



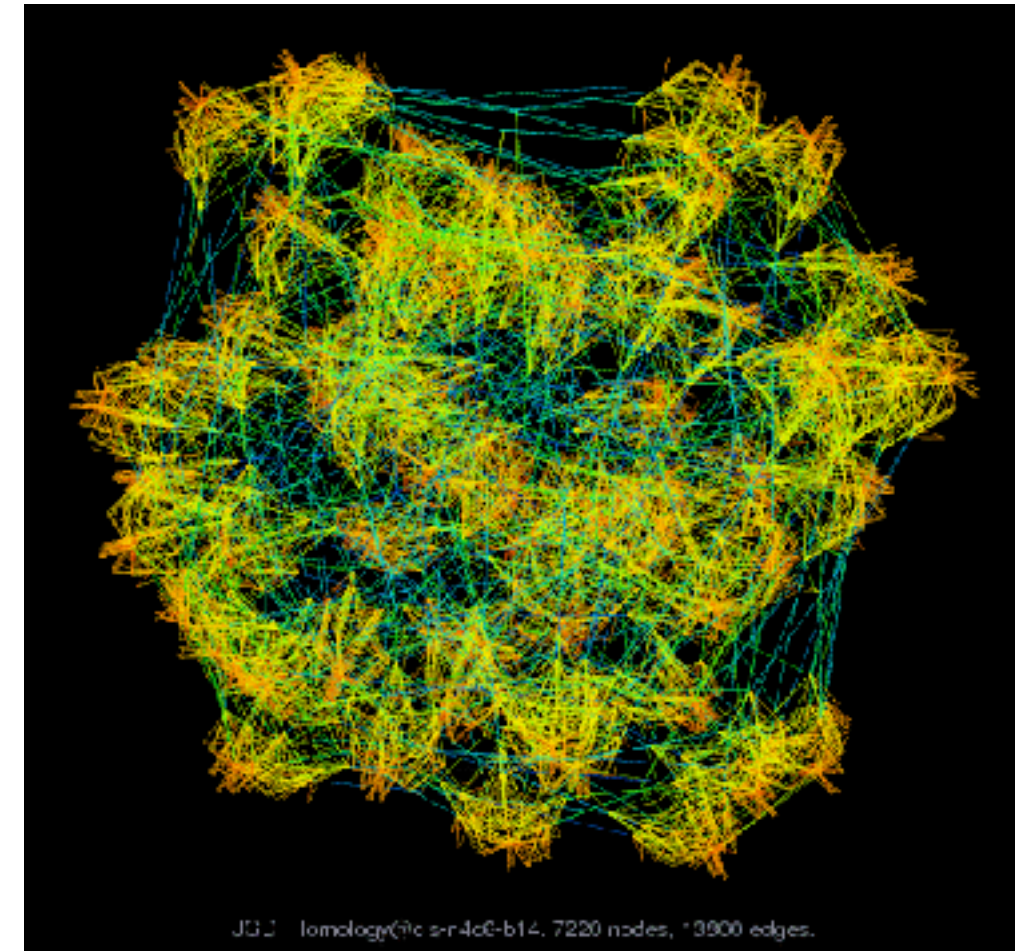
Graph Hierarchy 1



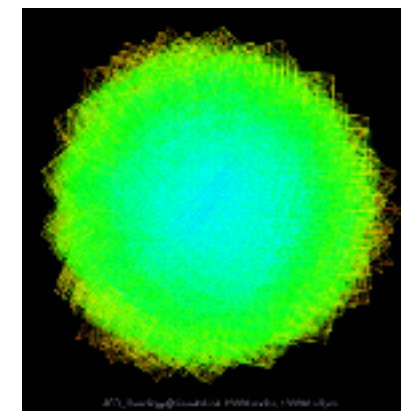
[GrouseFlocks: Steerable Exploration of Graph Hierarchy Space. Archambault, Munzner, and Auber. *IEEE TVCG* 14(4):900-913, 2008.]

# Idiom: **sfdp** (multi-level force-directed placement)

- data: compound graph
  - original: network
  - derived: cluster hierarchy atop it
- considerations
  - better algorithm for same encoding technique
    - same: fundamental use of space
    - hierarchy used for algorithm speed/quality but not shown explicitly
- scalability
  - nodes, edges: 1K-10K
  - hairball problem eventually hits

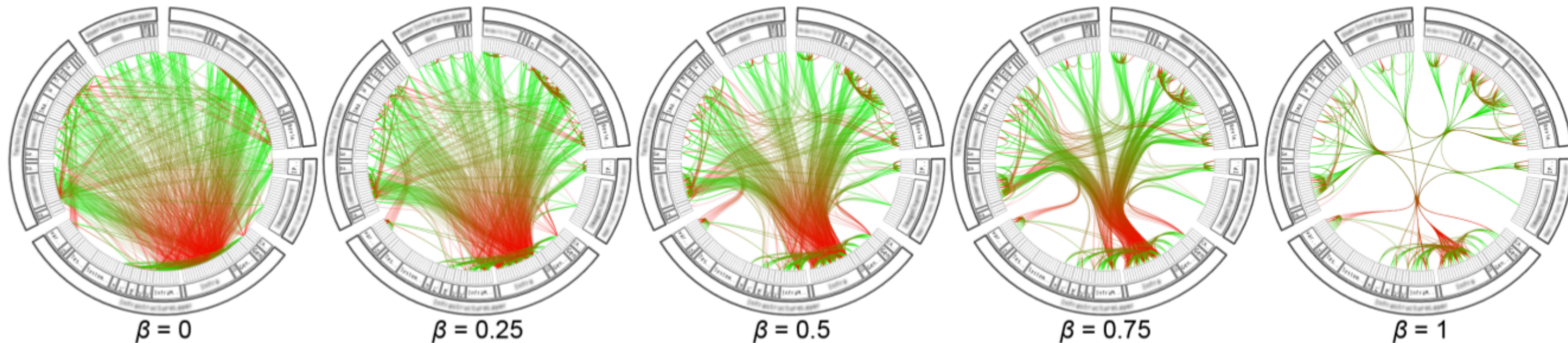


*[Efficient and high quality force-directed graph drawing. Hu. The Mathematica Journal 10:37–71, 2005.]*



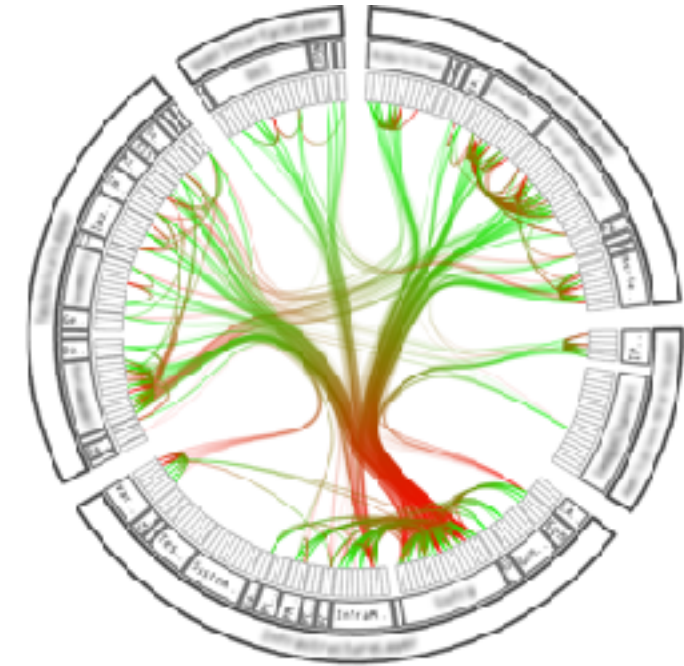
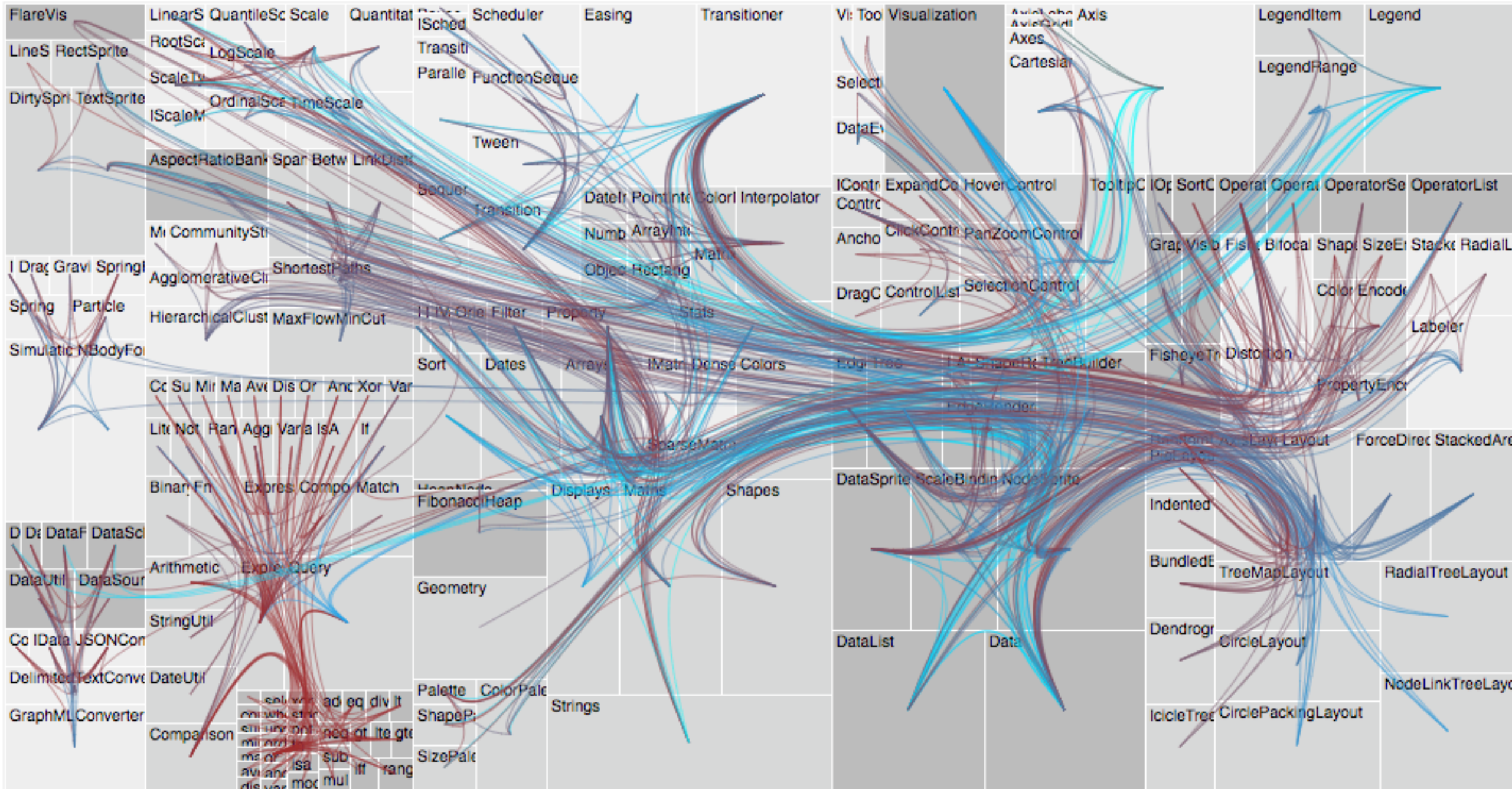
# Idiom: hierarchical edge bundling

- data
  - any layout of compound network
    - network: software classes (nodes), import/export between classes (links)
    - cluster hierarchy: class package structure
  - derived: bundles of edges with same source/destination (multi-level)
- idiom: curve edge routes according to bundles
- task: edge clutter reduction



# Hierarchical edge bundling

- works for any layout: treemap vs radial



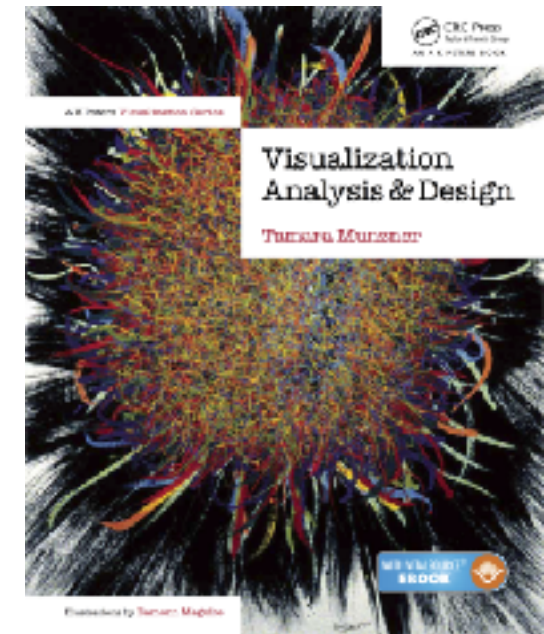
# Visualization Analysis & Design

## *Spatial Data (Ch 9)*

**Tamara Munzner**

Department of Computer Science  
University of British Columbia

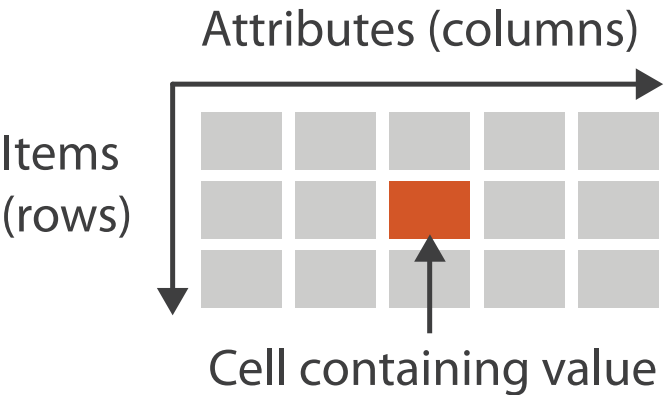
[@tamaramunzner](#)



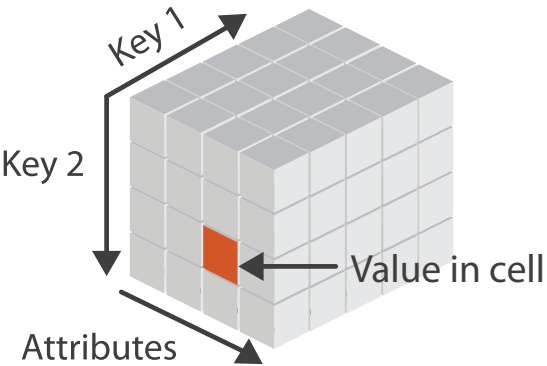
# Focus on Spatial

## → Dataset Types

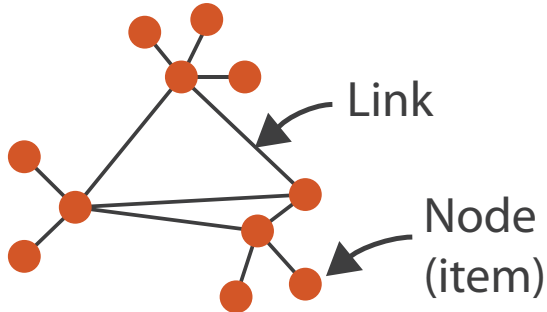
### → Tables



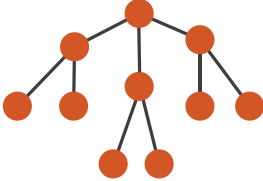
### → Multidimensional Table



### → Networks

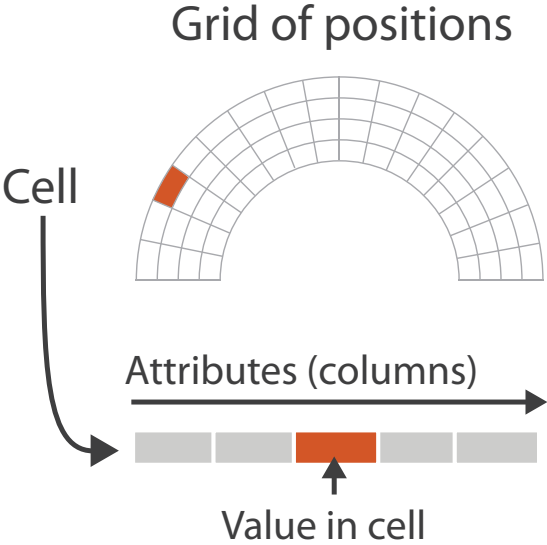


### → Trees

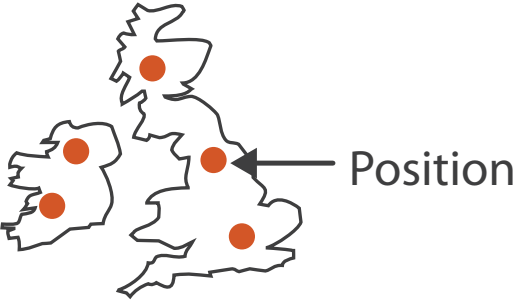


### → Spatial

#### → Fields (Continuous)



#### → Geometry (Spatial)



# How?

## Encode

### → Arrange

→ Express



→ Separate



→ Order



→ Align



→ Use



What?

Why?

How?

### → Map

from **categorical** and **ordered** attributes

→ Color

→ Hue



→ Saturation



→ Luminance



→ Size, Angle, Curvature, ...



→ Shape



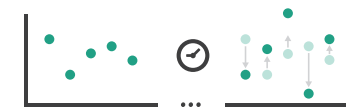
→ Motion

*Direction, Rate, Frequency, ...*

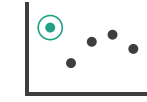


## Manipulate

### → Change



### → Select

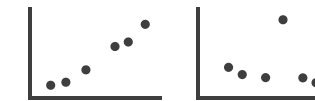


### → Navigate

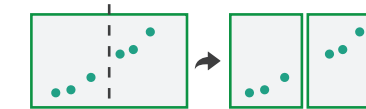


## Facet

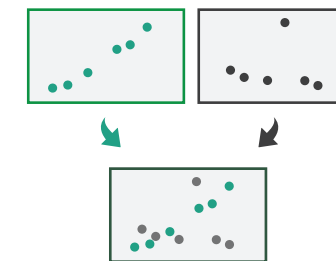
### → Juxtapose



### → Partition



### → Superimpose



## Reduce

### → Filter



### → Aggregate



### → Embed



# How?

## Encode

### → Arrange

→ Express



→ Order



→ Use



→ Separate



→ Align



### → Map

from **categorical** and **ordered** attributes

→ Color

→ Hue



→ Saturation



→ Luminance



→ Size, Angle, Curvature, ...



→ Shape



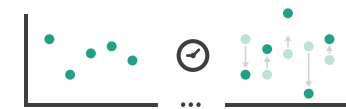
→ Motion

*Direction, Rate, Frequency, ...*

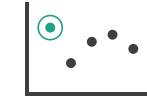


## Manipulate

### → Change



### → Select



### → Navigate

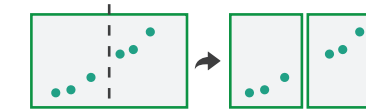


## Facet

### → Juxtapose



### → Partition



### → Superimpose



## Reduce

### → Filter



### → Aggregate



### → Embed



What?

Why?

How?



# Spatial data

- use given spatial position
- when?
  - dataset contains spatial attributes and they have primary importance
  - central tasks revolve around understanding spatial relationships
- examples
  - geographical/cartographic data
  - sensor/simulation data

# Geographic Maps

# Geographic Map



## Interlocking marks

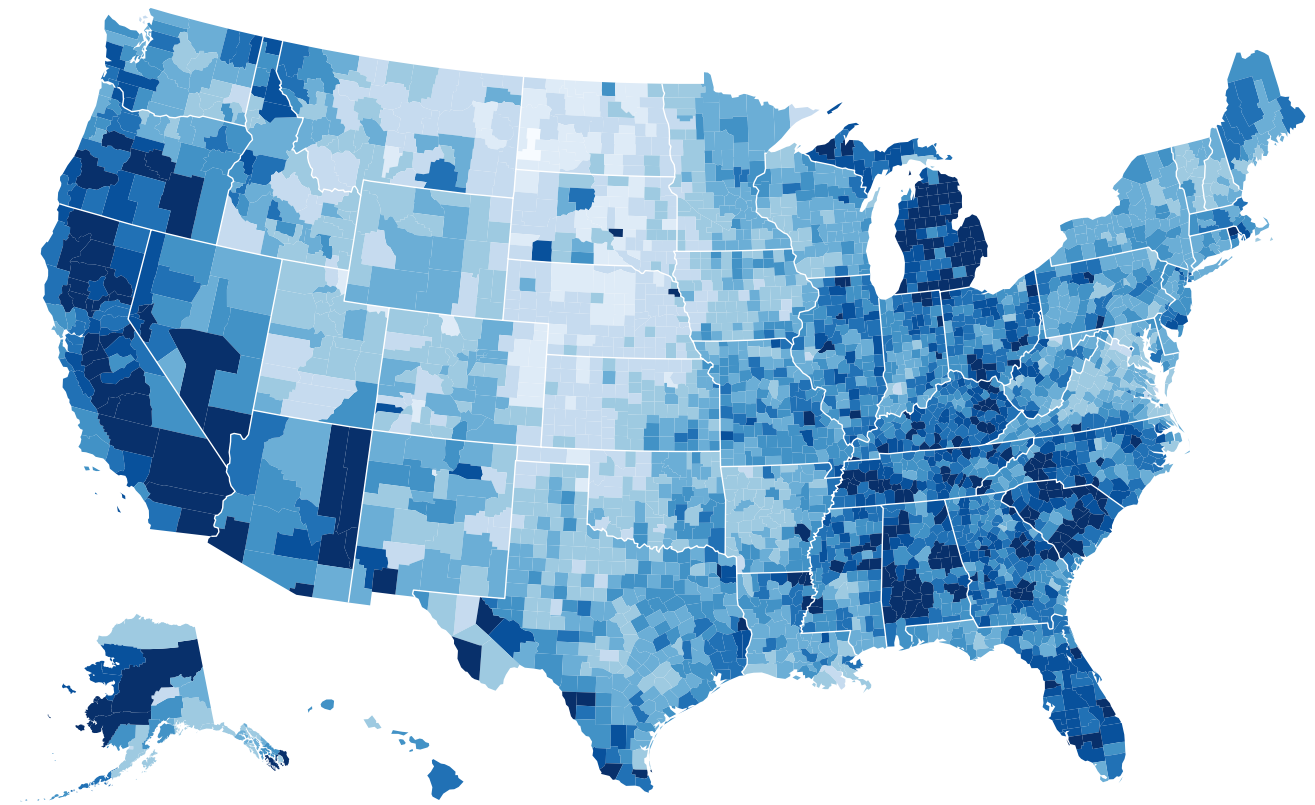
- **shape coded**
  - **area coded**
  - **position coded**
- 
- cannot encode another attribute with these channels, they're "taken"

# Thematic maps

- show spatial variability of attribute ("theme")
  - combine geographic / reference map with (simple, flat) tabular data
  - join together
    - region: interlocking area marks (provinces, countries with outline shapes)
      - also could have point marks (cities, locations with 2D lat/lon coords)
    - region: categorical key attribute in table
      - use to look up value attributes
- major idioms
  - choropleth
  - symbol maps
  - cartograms
  - dot density maps

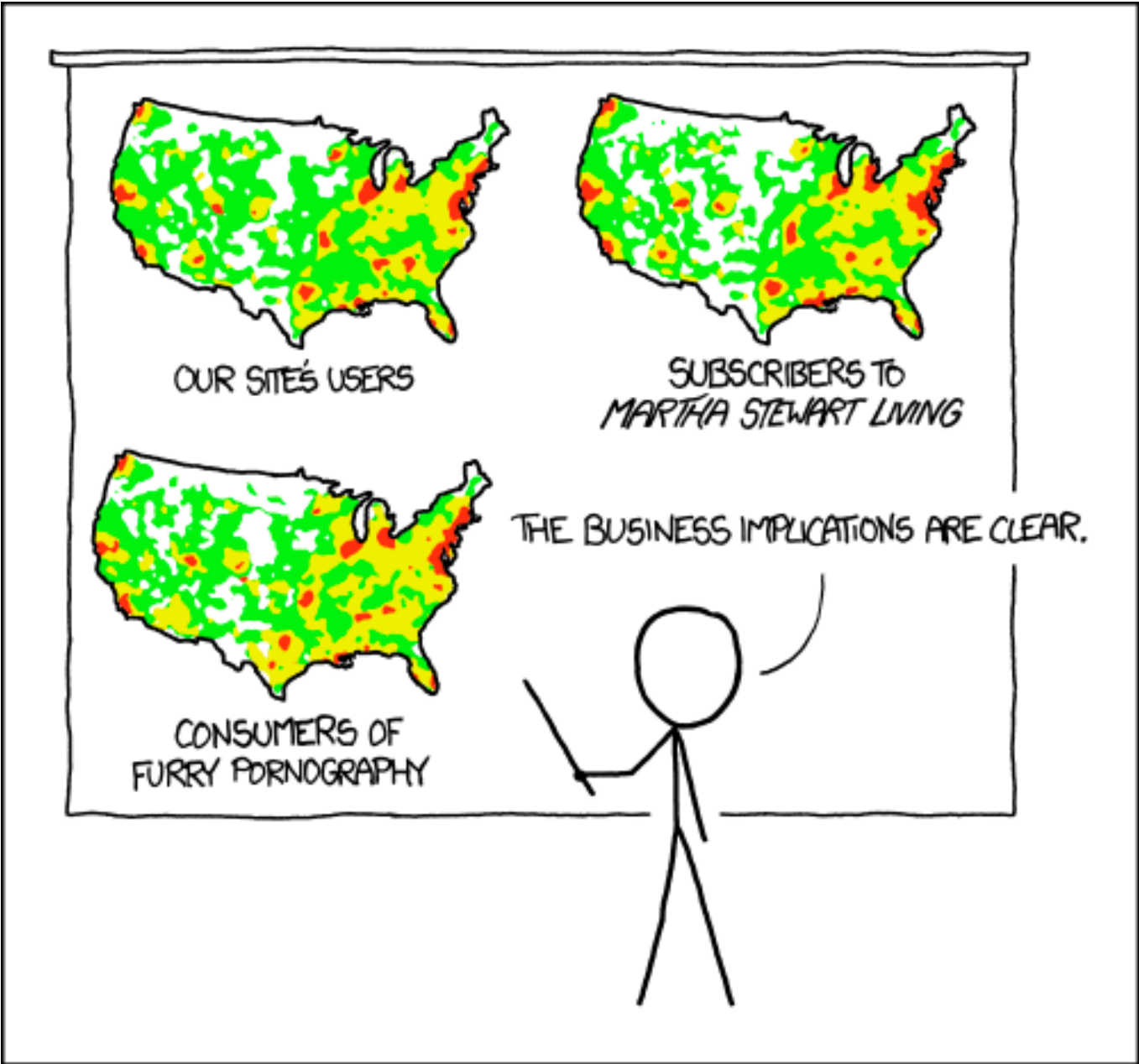
# Idiom: **choropleth map**

- use given spatial data
  - when central task is understanding spatial relationships
- data
  - geographic geometry
  - table with 1 quant attribute per region
- encoding
  - position:
    - use given geometry for area mark boundaries
  - color:
    - sequential segmented colormap



<http://bl.ocks.org/mbostock/4060606>

# Beware: Population maps trickiness!

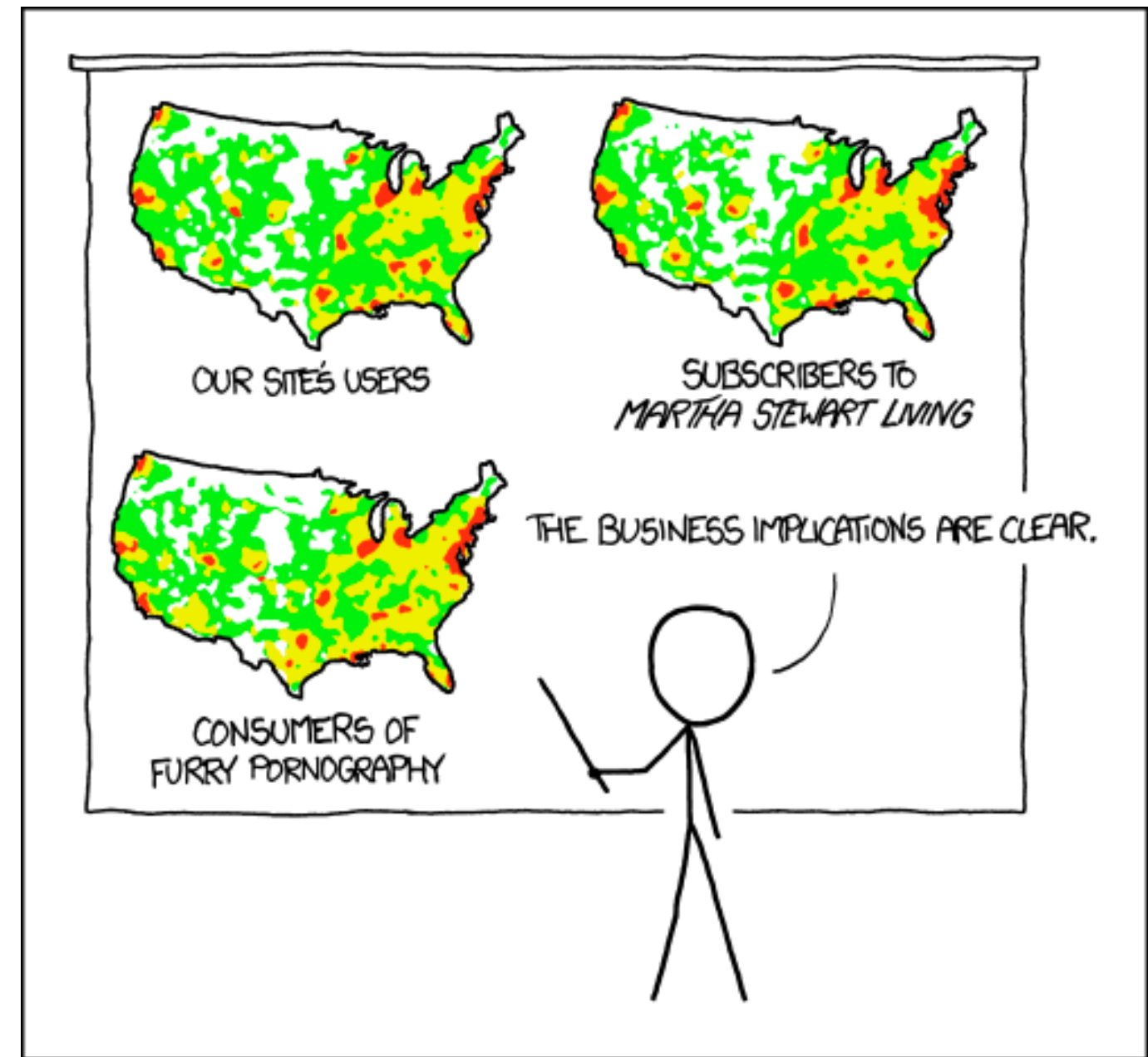


PET PEEVE #208:  
GEOGRAPHIC PROFILE MAPS WHICH ARE  
BASICALLY JUST POPULATION MAPS

[ <https://xkcd.com/1138> ]

# Beware: Population maps trickiness!

- spurious correlations: most attributes just show where people live

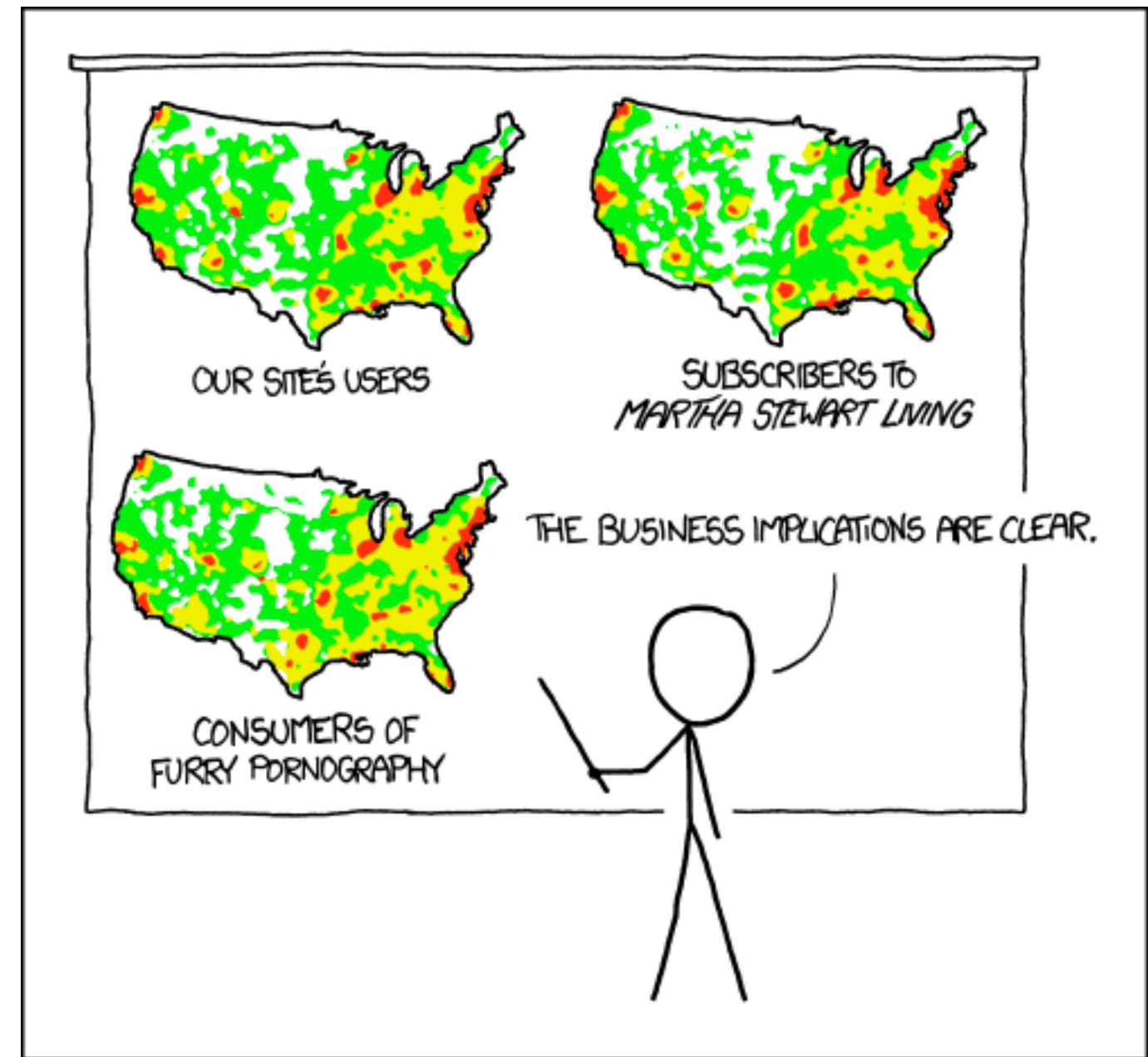


PET PEEVE #208:  
GEOGRAPHIC PROFILE MAPS WHICH ARE  
BASICALLY JUST POPULATION MAPS

[ <https://xkcd.com/1138> ]

# Beware: Population maps trickiness!

- spurious correlations: most attributes just show where people live
- consider when to normalize by population density
  - encode raw data values
    - tied to underlying population
  - but should use normalized values
    - unemployed people per 100 citizens, mean family income



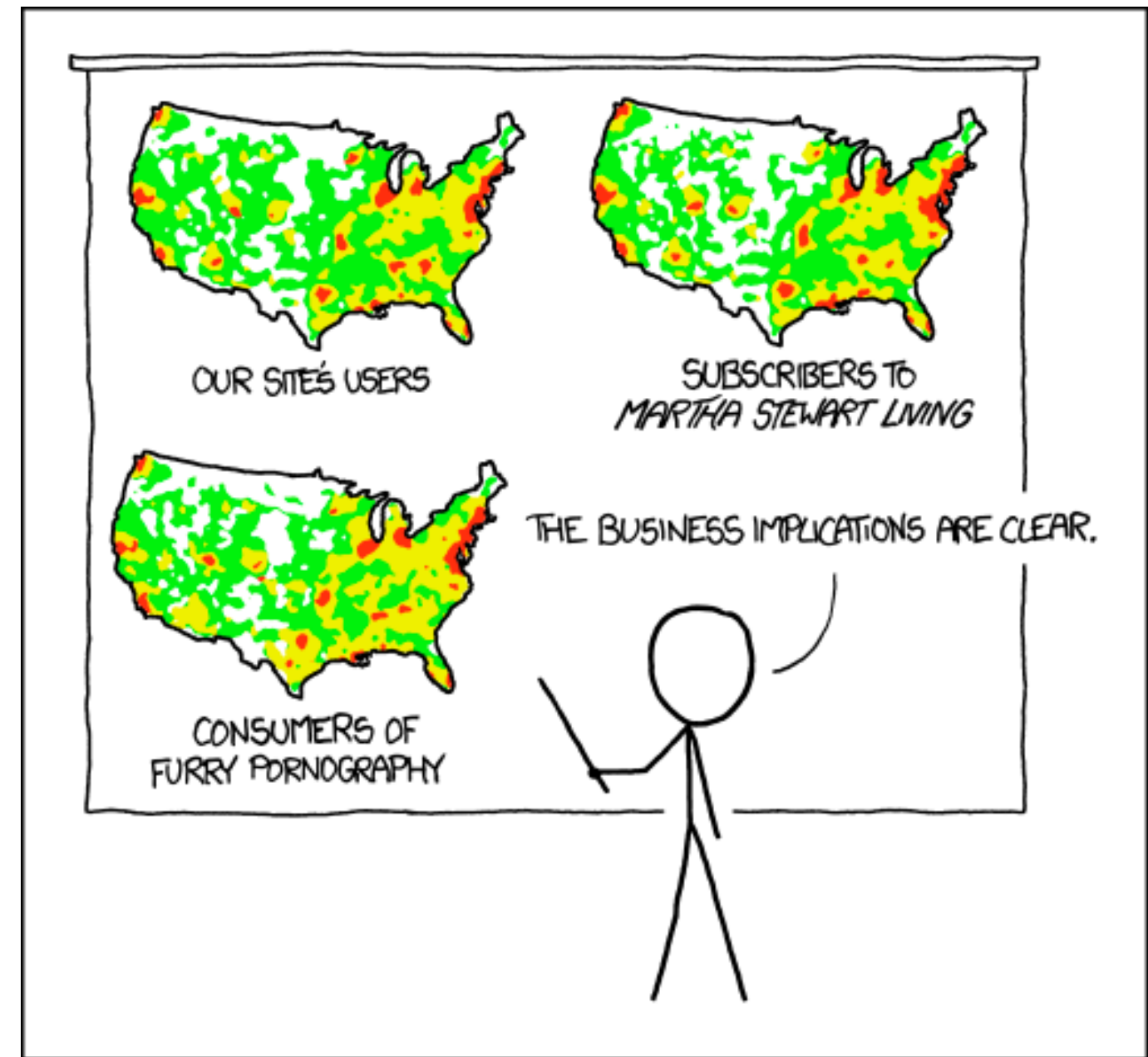
PET PEEVE #208:  
GEOGRAPHIC PROFILE MAPS WHICH ARE  
BASICALLY JUST POPULATION MAPS

[ <https://xkcd.com/1138> ]



# Beware: Population maps trickiness!

- spurious correlations: most attributes just show where people live
- consider when to normalize by population density
  - encode raw data values
    - tied to underlying population
  - but should use normalized values
    - unemployed people per 100 citizens, mean family income
- general issue
  - absolute counts vs relative/normalized data
  - failure to normalize is common error



[ <https://xkcd.com/1138> ]

# Choropleth maps: Recommendations

- only use when central task is understanding spatial relationships
- show only one variable at a time
- normalize when appropriate
- be careful when choosing colors & bins
- best case: regions are roughly equal sized

# Choropleth map: Pros & cons

- pros

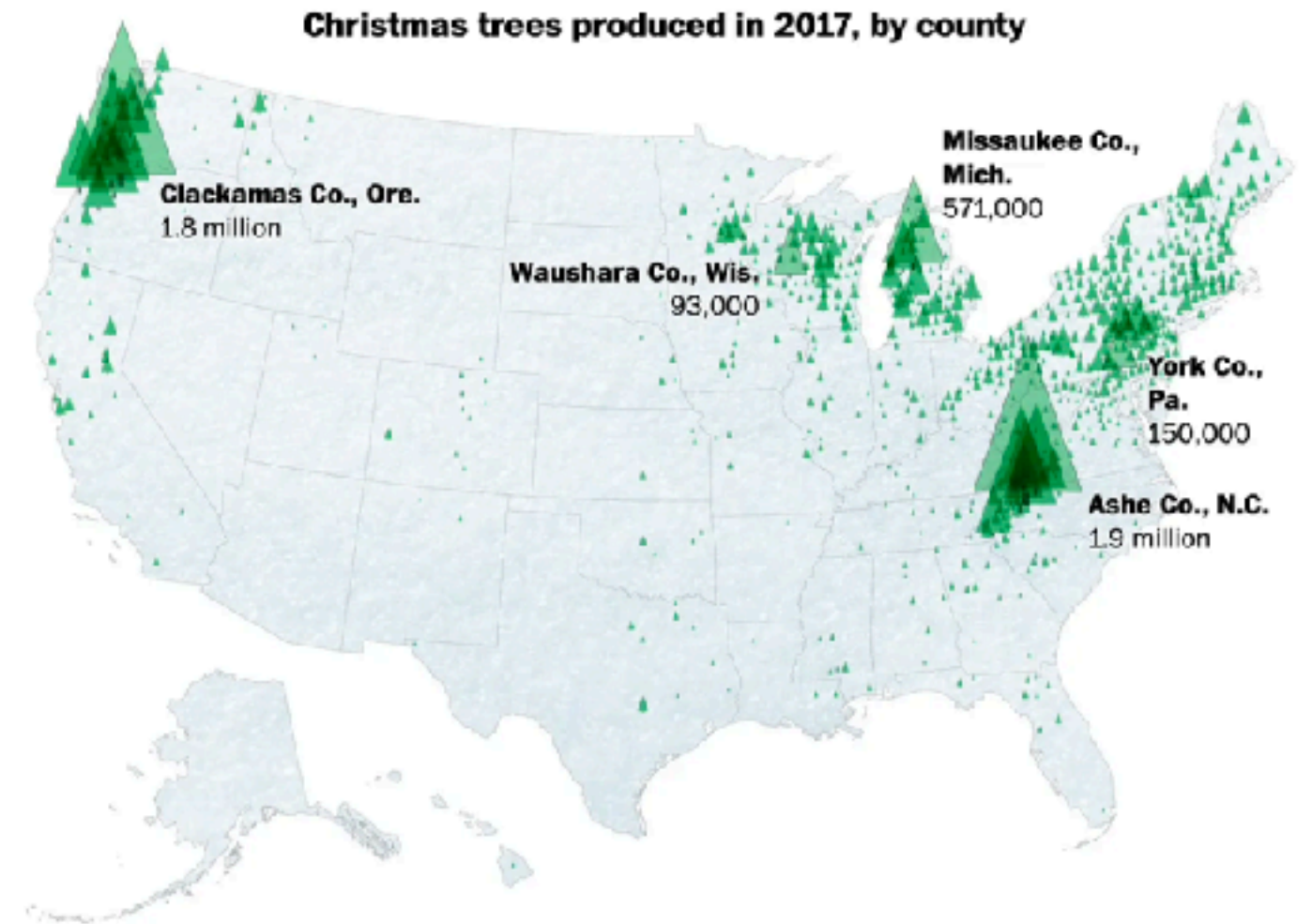
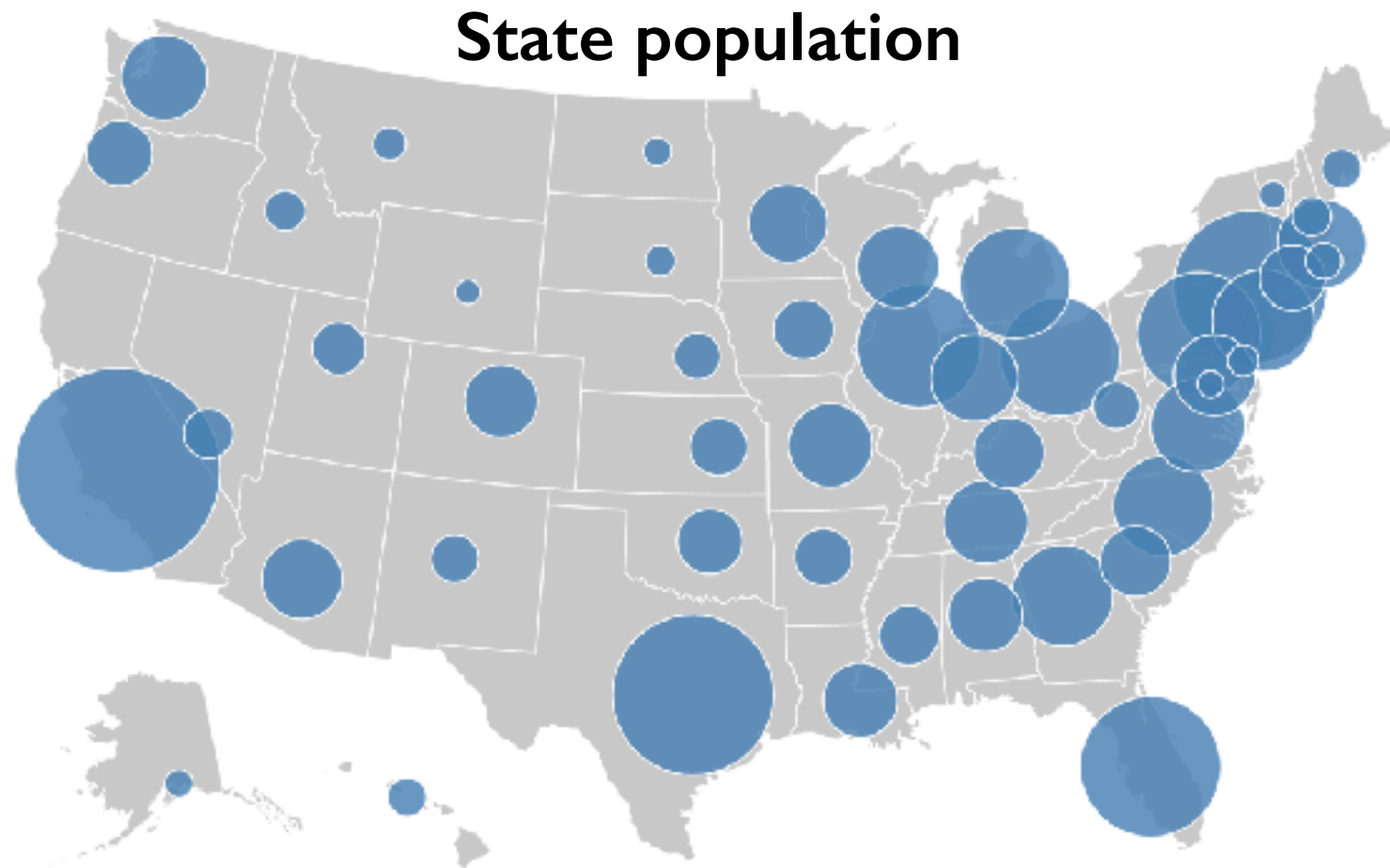
- easy to read and understand
- well established visualization (no learning curve)
- data is often collected and aggregated by geographical regions

- cons

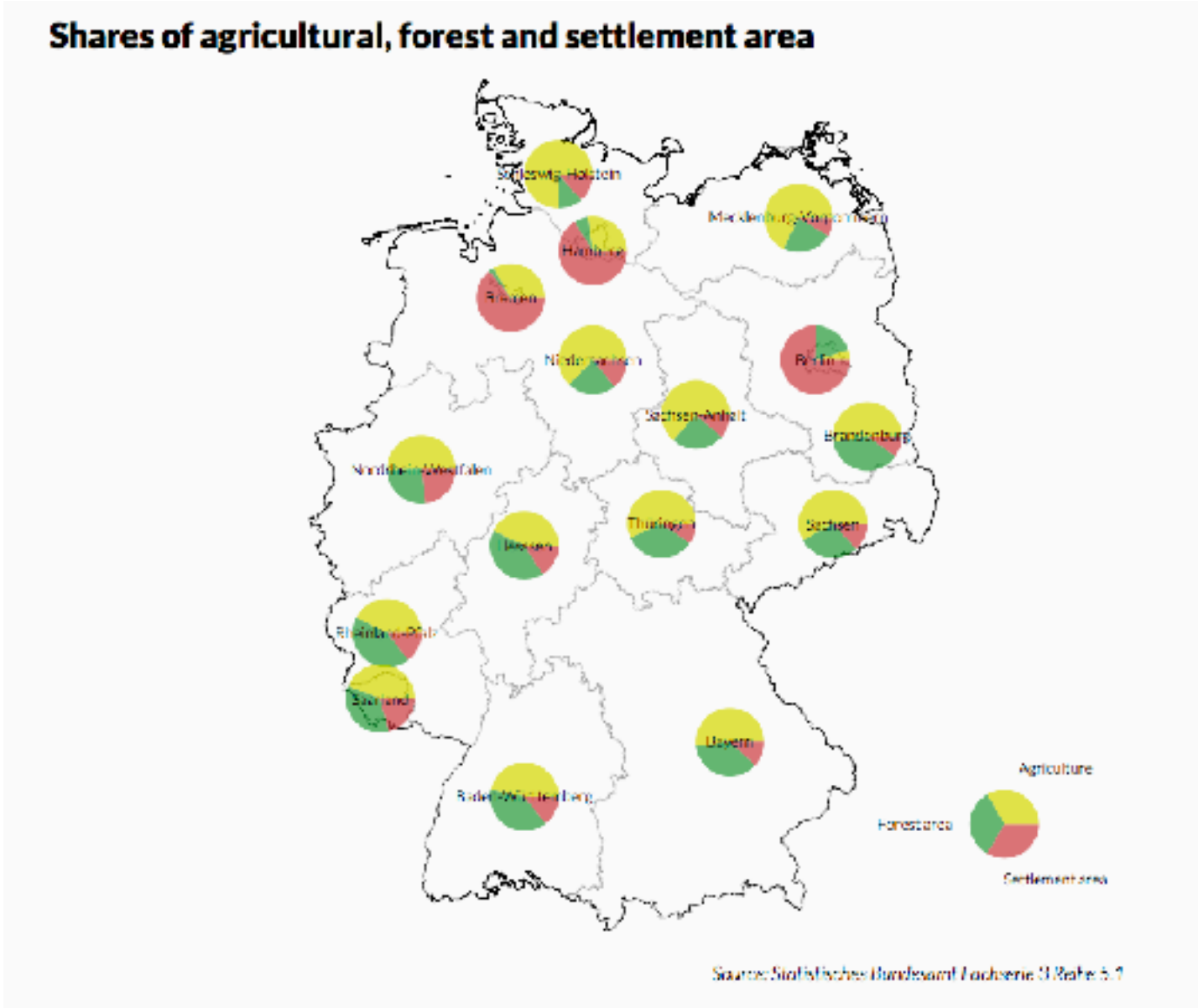
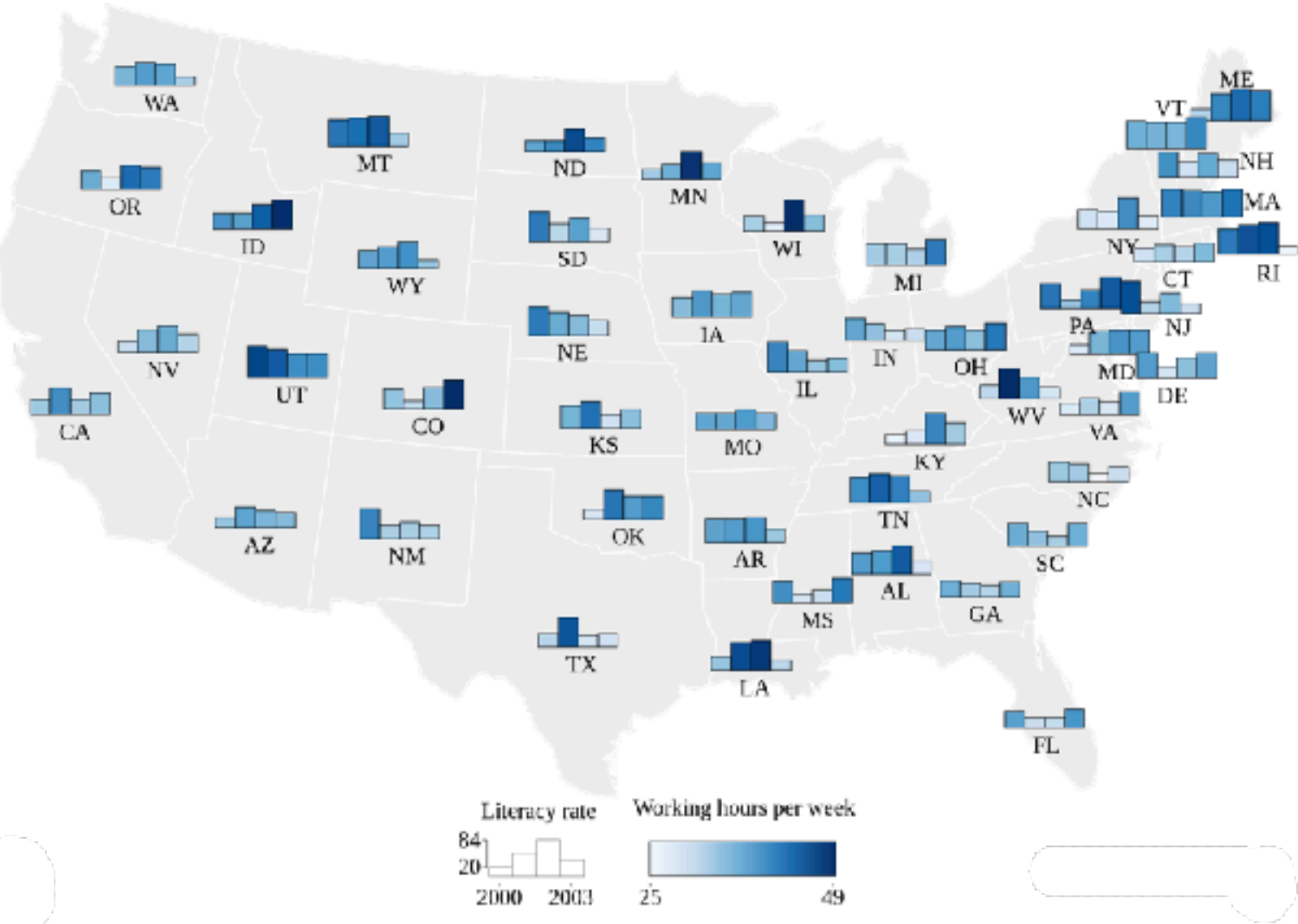
- most effective visual variable used for geographic location
- visual salience depends on region size, not true importance wrt attribute value
  - large regions appear more important than small ones
- color palette choice has a huge influence on the result

# Idiom: **Symbol maps**

- symbol is used to represent aggregated data (mark or glyph)
  - allows use of size and shape and color channels
    - aka proportional symbol maps, graduated symbol maps
- keep original spatial geometry in the background
- often a good alternative to choropleth maps



# Symbol maps with glyphs

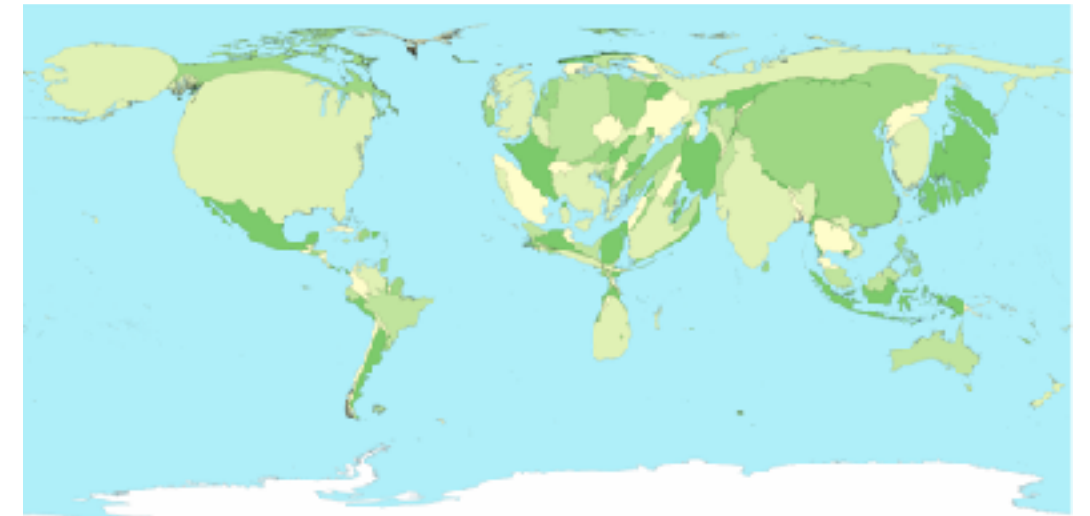


# Symbol map: Pros & cons

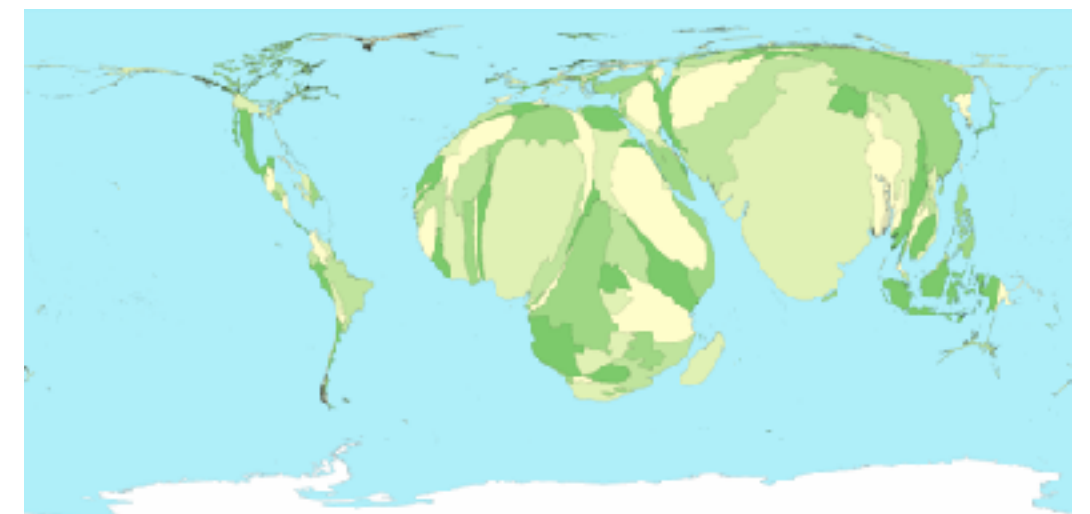
- **pros**
  - somewhat intuitive to read and understand
  - mitigate problems with region size vs data salience
    - marks: symbol size follows attribute value
    - glyphs: symbol size can be uniform
- **cons**
  - possible occlusion / overlap
    - symbols could overlap each other
    - symbols could occlude region boundaries
  - complex glyphs may require explanation / training

# Idiom: **Contiguous cartogram**

- interlocking marks:  
shape, area, and position coded
- derive new interlocking marks
  - based on combination of original interlocking marks and new quantitative attribute
- algorithm to create new marks
  - input: target size
  - goal: shape as close to the original as possible
  - requirement: maintain constraints
    - relative position
    - contiguous boundaries with their neighbours

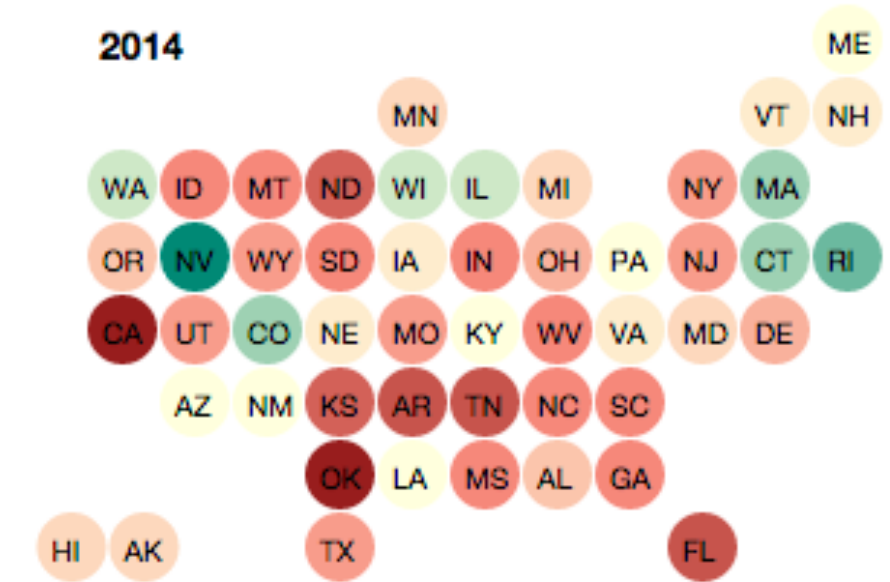
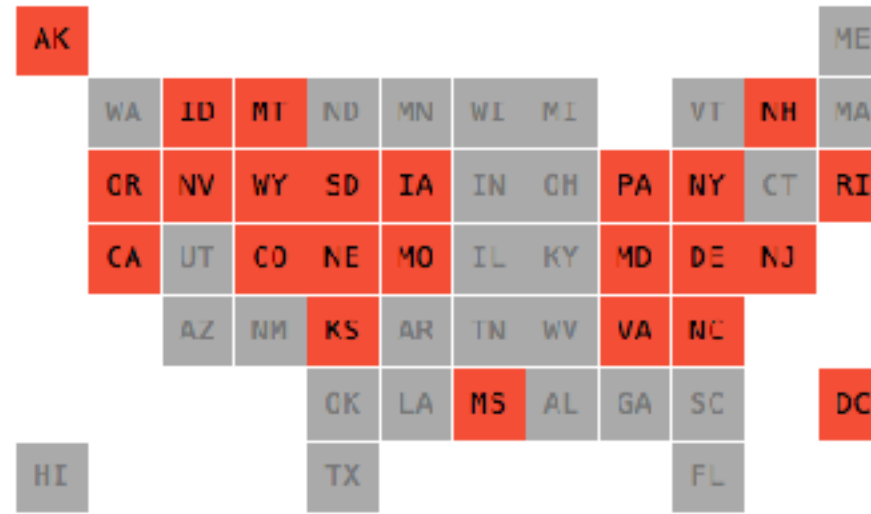
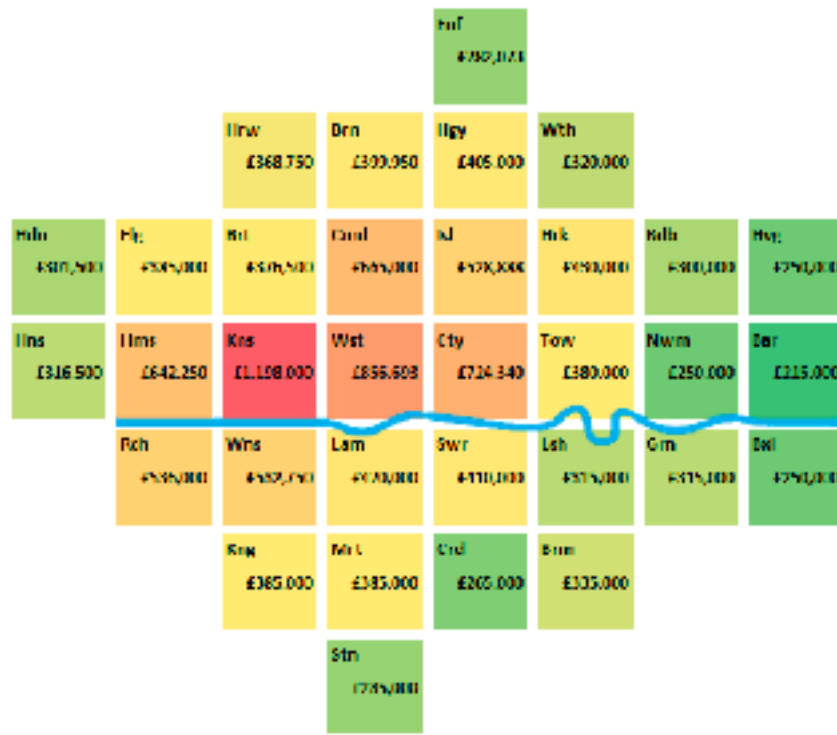


Greenhouse Emissions



Child Mortality

# Idiom: Grid Cartogram



- uniform-sized shapes arranged in rectilinear grid
- maintain approximate spatial position and arrangement



# Cartogram: Pros & cons

- pros

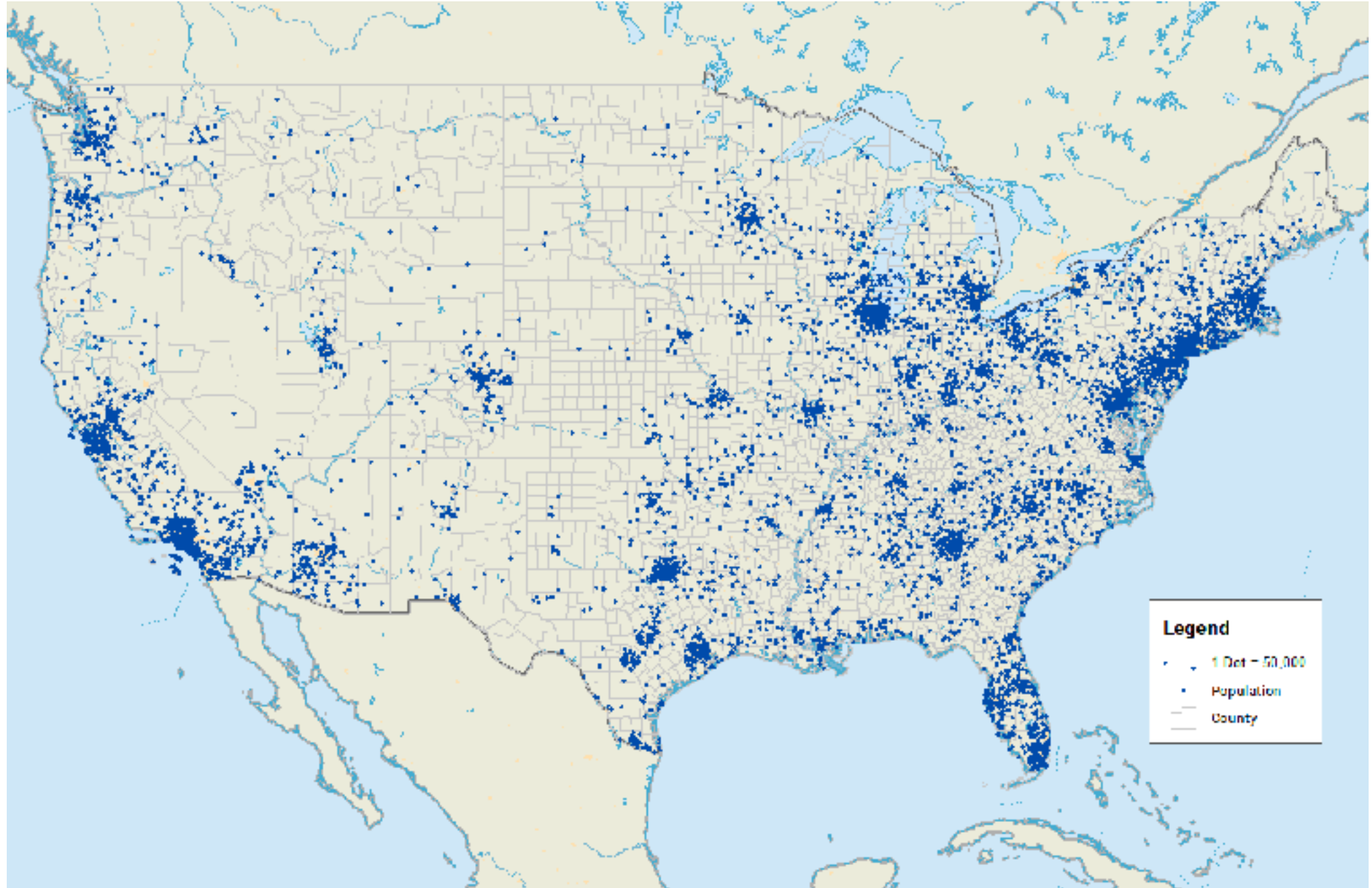
- can be intriguing and engaging
- best case: strong and surprising size disparities
- non-contiguous cartograms often easier to understand

- cons

- require substantial familiarity with original dataset & use of memory
  - compare distorted marks to memory of original marks
  - mitigation strategies: transitions or side by side views
- major distortion is problematic
  - may be aesthetically displeasing
  - may result in unrecognizable marks
- difficult to extract exact quantities

# Idiom: **Dot density maps**

- visualize distribution of a phenomenon by placing dots
- one symbol represents a constant number of items
  - dots have uniform size & shape
  - allows use of color channel
- task:  
show spatial patterns, clusters



# Dot density maps: Pros and cons

- pros

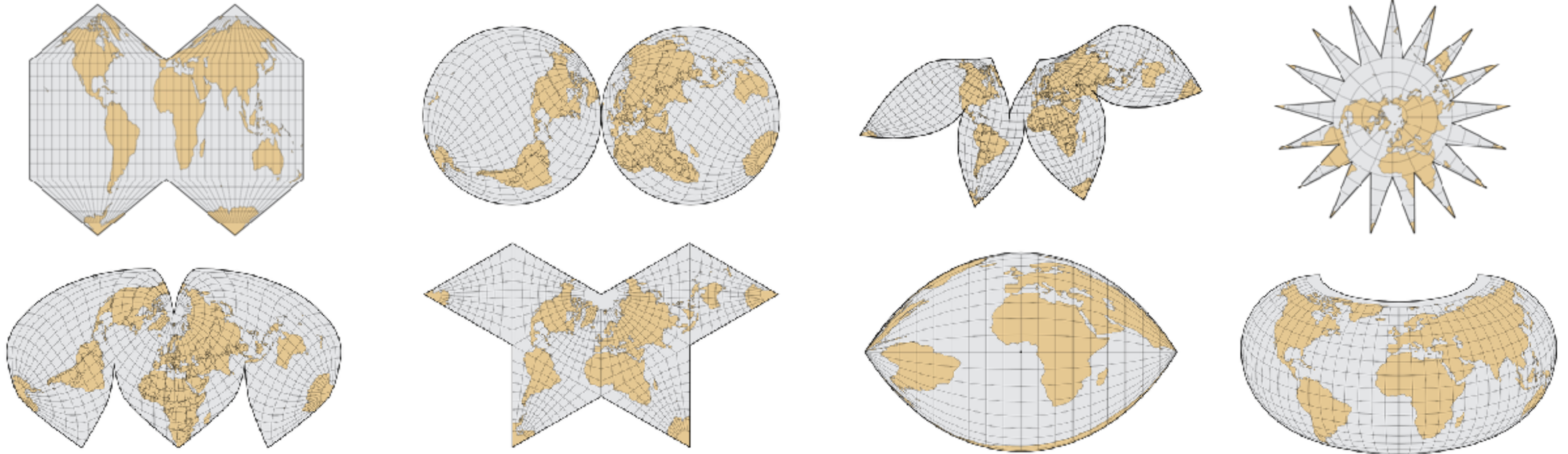
- straightforward to understand
- avoids choropleth non-uniform region size problems

- cons

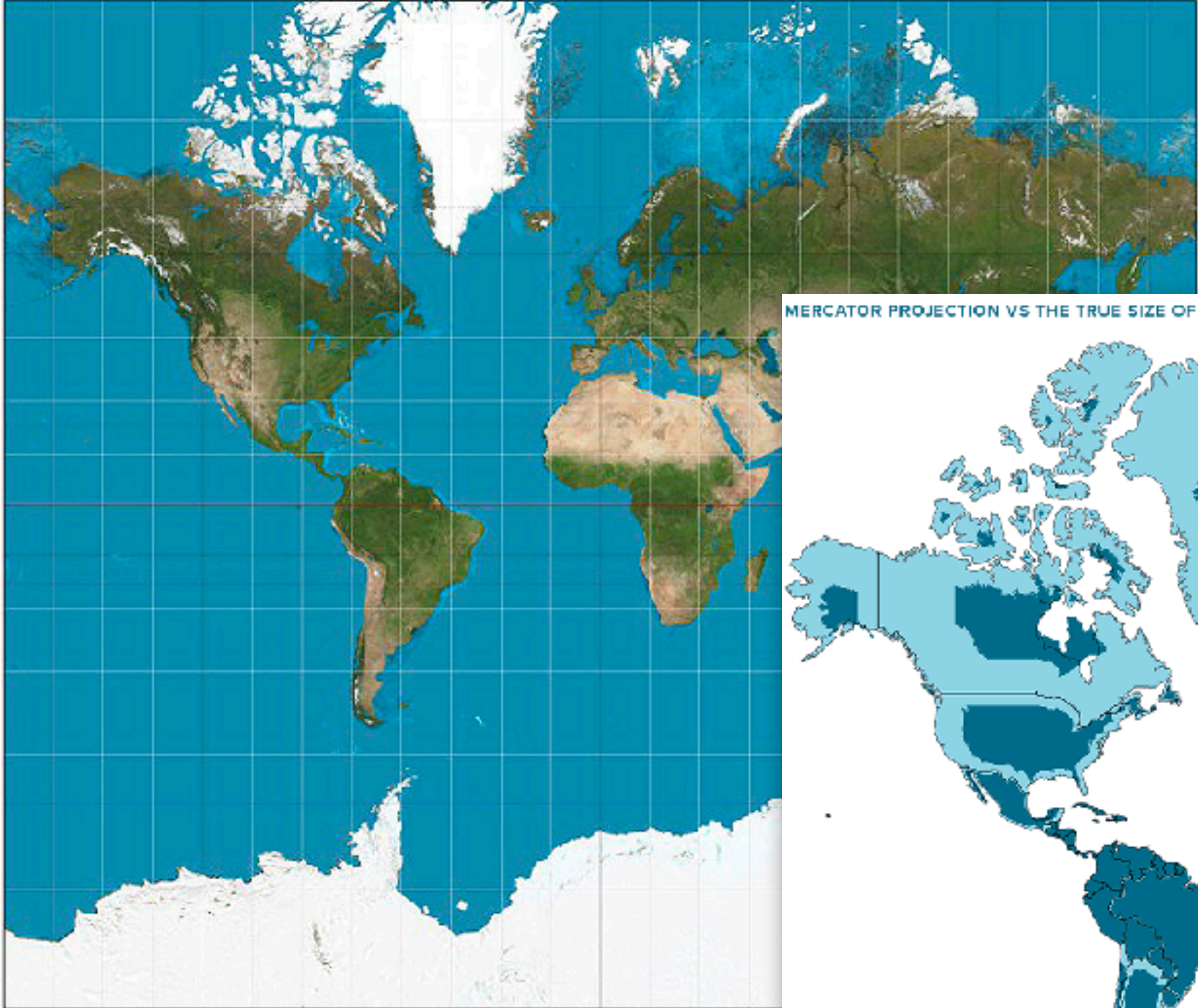
- challenge: normalization, just like choropleths
  - show population density (correlated with attribute), not effect of interest
- perceptual disadvantage:  
difficult to extract quantities
- performance disadvantage:  
rendering many dots can be slow

# Map Projections

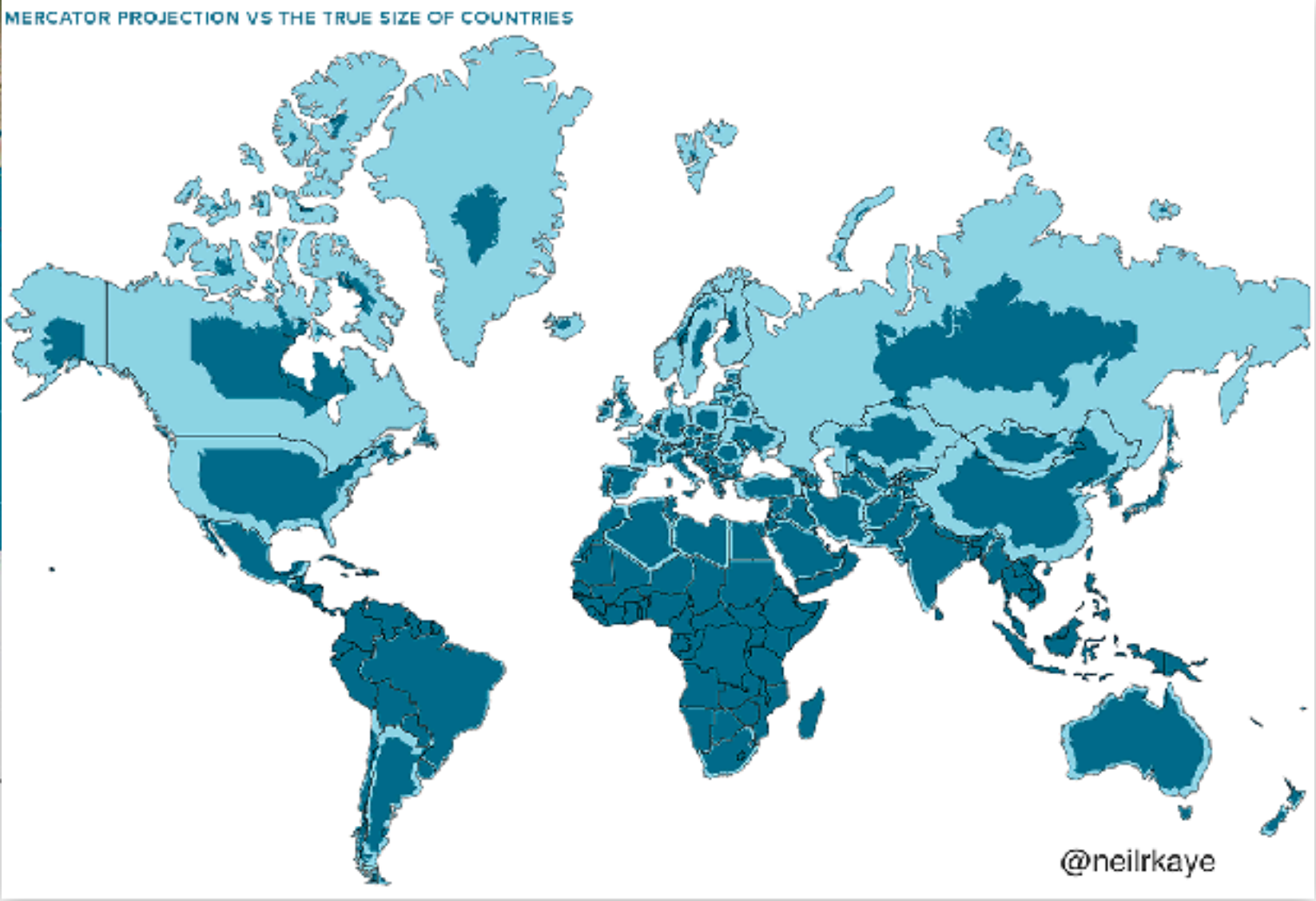
- mathematical functions that map 3D surface geometry of the Earth to 2D maps
- all projections of sphere on plane necessarily distort surface in some way
- **interactive:** [philogb.github.io/page/myriahedral/](http://philogb.github.io/page/myriahedral/) and [jasondavies.com/maps/](http://jasondavies.com/maps/)



# Mercator Projection



» Heavily distorts country sizes; particularly close to the poles.



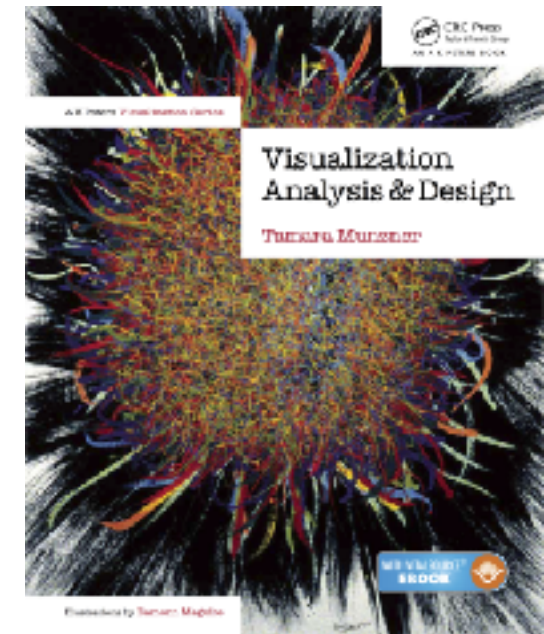
# Visualization Analysis & Design

## *Spatial Data (Ch 9) II*

**Tamara Munzner**

Department of Computer Science  
University of British Columbia

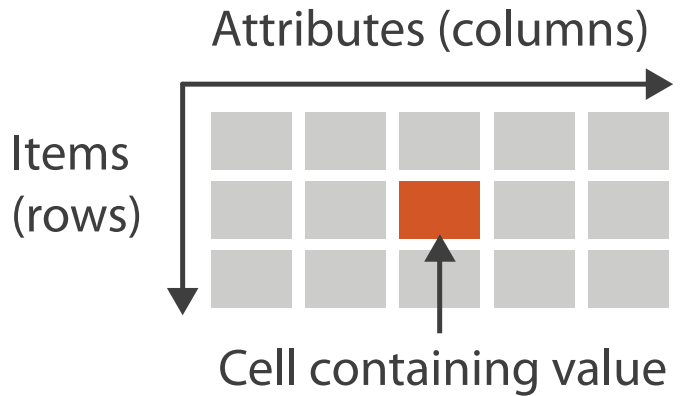
[@tamaramunzner](#)



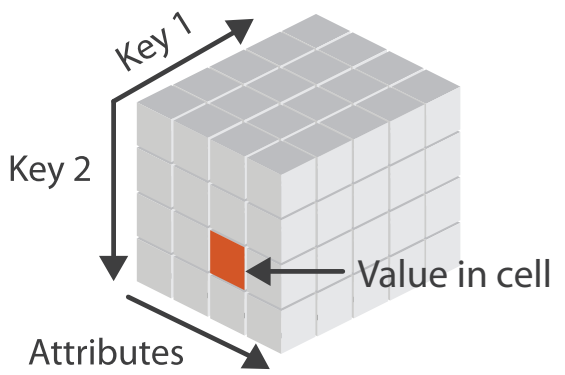
# Focus on Spatial

## → Dataset Types

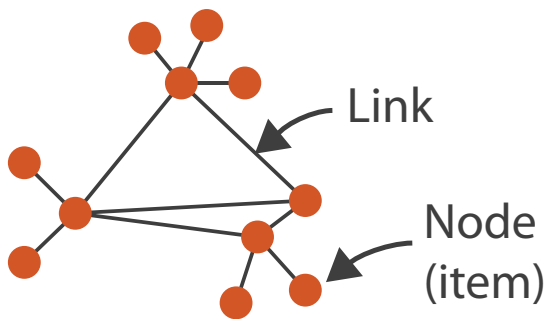
→ Tables



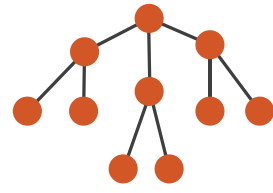
→ *Multidimensional Table*



→ Networks

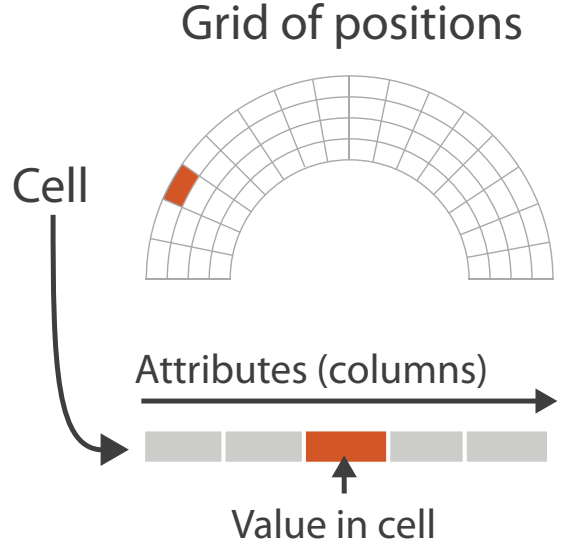


→ Trees




→ Spatial

→ Fields (Continuous)



→ Geometry (Spatial)

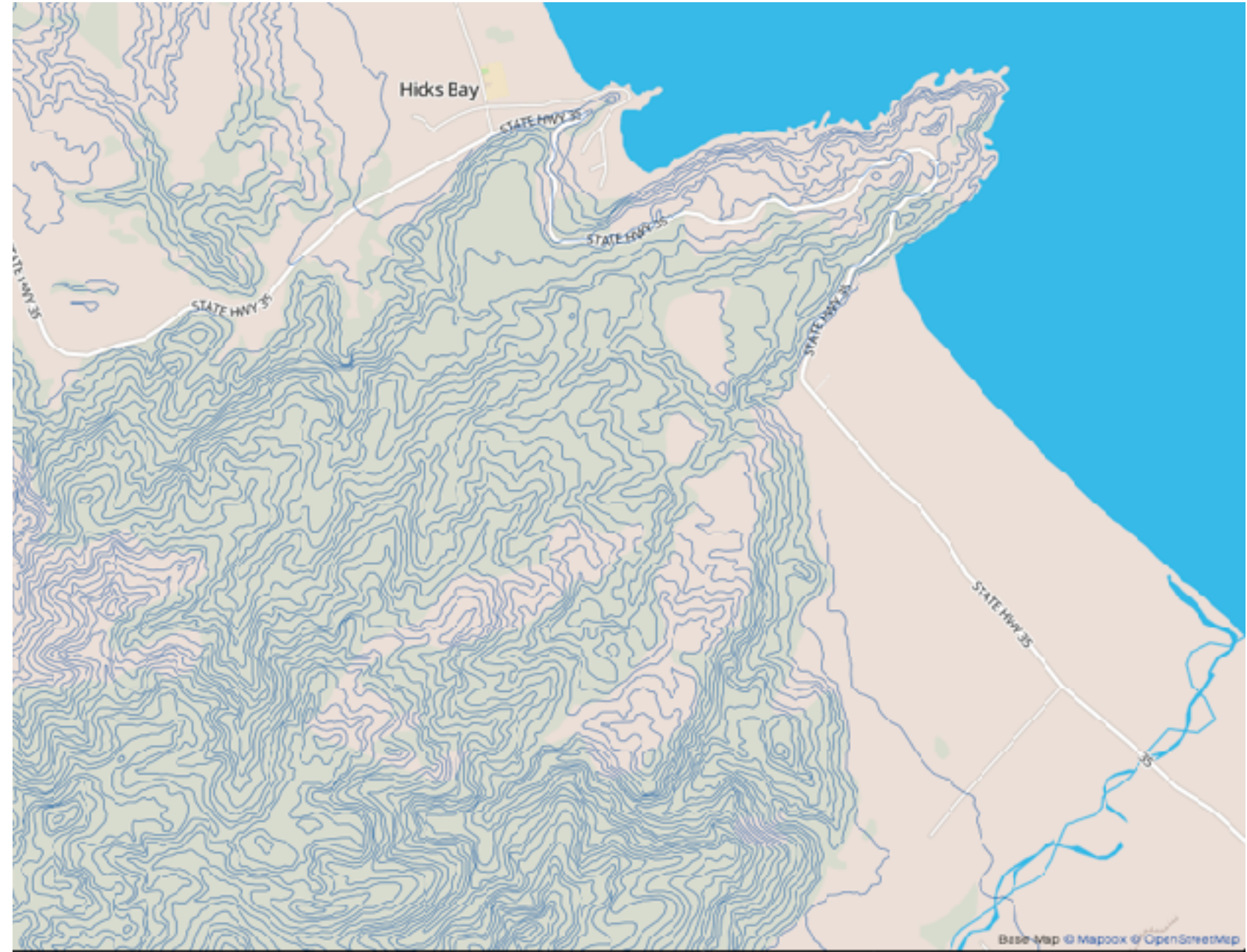


# Spatial Fields



# Idiom: **topographic map**

- data
  - geographic geometry
  - scalar spatial field
    - 1 quant attribute per grid cell
- derived data
  - isoline geometry
    - isocontours computed for specific levels of scalar values
- task
  - understanding terrain shape
    - densely lined regions = steep
- pros
  - use only 2D position, avoid 3D challenges
  - color channel available for other attributes
- cons
  - significant clutter from additional lines



Land Information New Zealand Data Service

# Idioms: **isosurfaces**, **direct volume rendering**

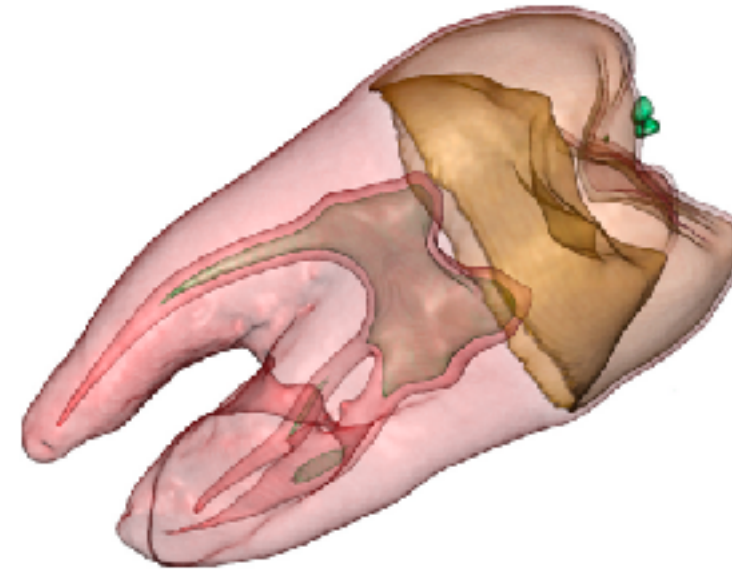
- **data**
  - scalar spatial field (3D volume)
    - 1 quant attribute per grid cell
- **task**
  - shape understanding, spatial relationships

*[Interactive Volume Rendering Techniques. Kniss. Master's thesis, University of Utah Computer Science, 2002.]*

*[Multidimensional Transfer Functions for Volume Rendering. Kniss, Kindlmann, and Hansen. In The Visualization Handbook, edited by Charles Hansen and Christopher Johnson, pp. 189–210. Elsevier, 2005.]*

# Idioms: **isosurfaces**, **direct volume rendering**

- **data**
  - scalar spatial field (3D volume)
    - 1 quant attribute per grid cell
- **task**
  - shape understanding, spatial relationships
- **isosurface**
  - derived data: isocontours computed for specific levels of scalar values

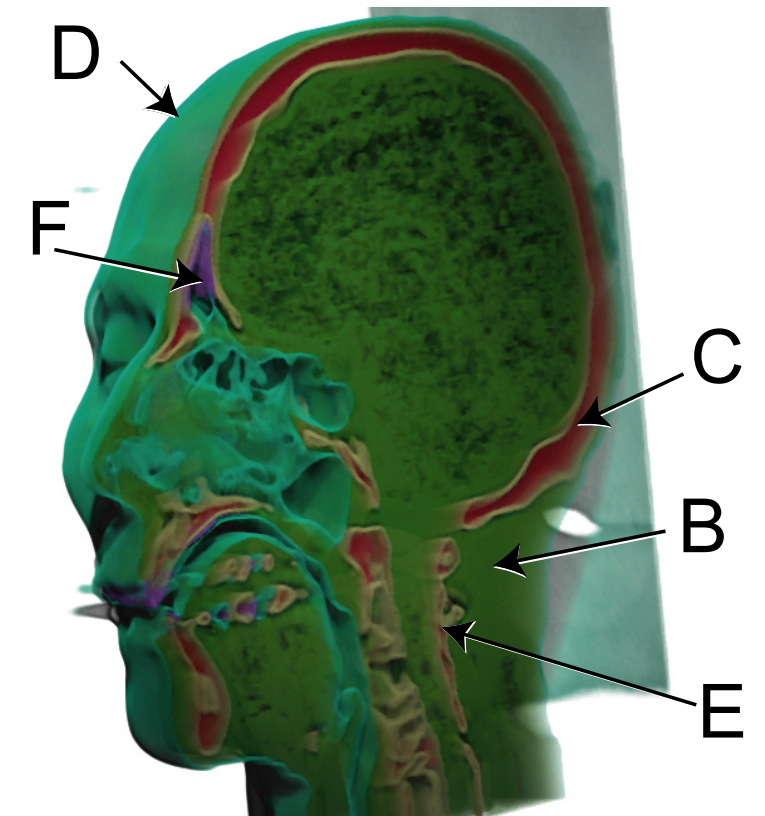


*[Interactive Volume Rendering Techniques. Kniss. Master's thesis, University of Utah Computer Science, 2002.]*

*[Multidimensional Transfer Functions for Volume Rendering. Kniss, Kindlmann, and Hansen. In The Visualization Handbook, edited by Charles Hansen and Christopher Johnson, pp. 189–210. Elsevier, 2005.]*

# Idioms: **isosurfaces**, **direct volume rendering**

- **data**
  - scalar spatial field (3D volume)
    - 1 quant attribute per grid cell
- **task**
  - shape understanding, spatial relationships
- **isosurface**
  - derived data: isocontours computed for specific levels of scalar values
- **direct volume rendering**
  - transfer function maps scalar values to color, opacity
    - no derived geometry

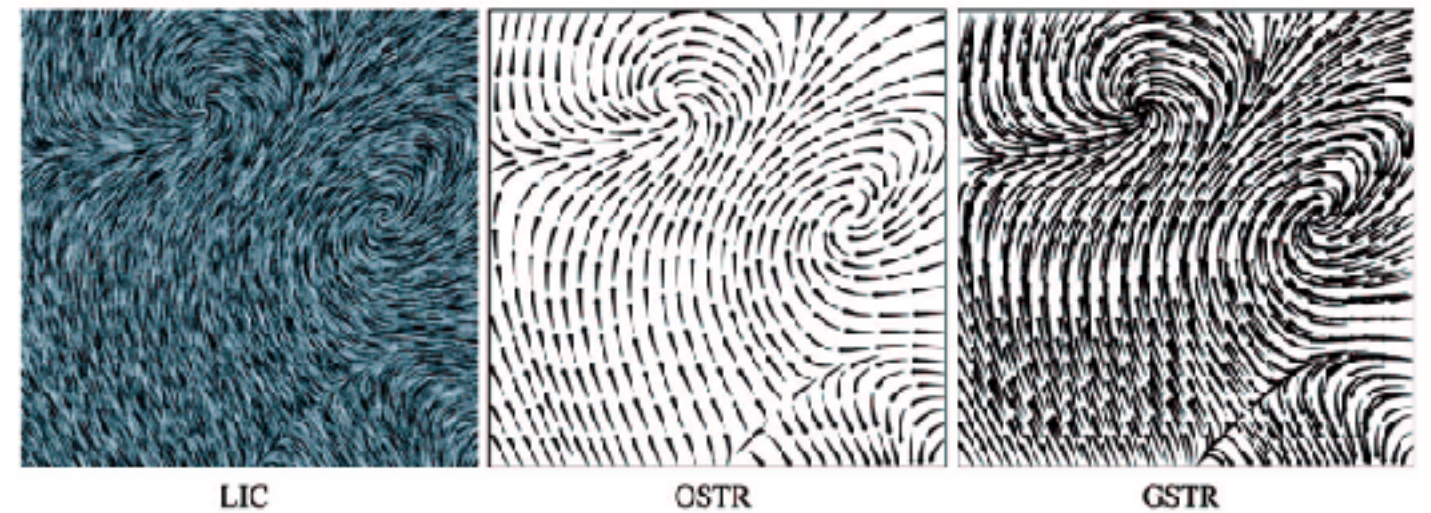
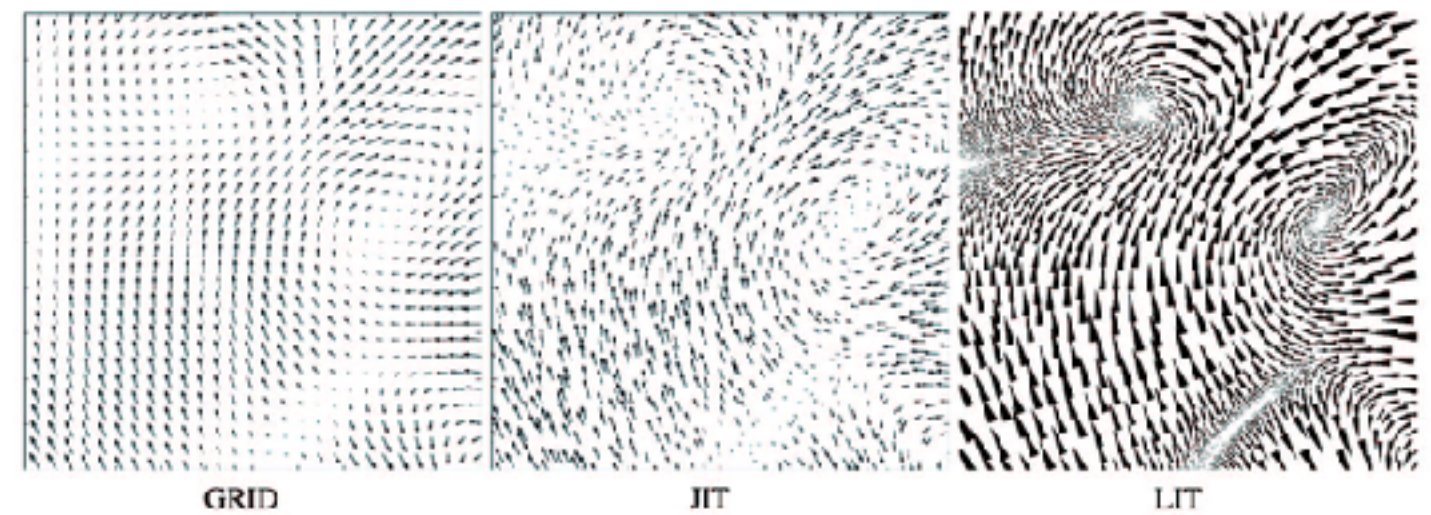


[Interactive Volume Rendering Techniques. Kniss. Master's thesis, University of Utah Computer Science, 2002.]

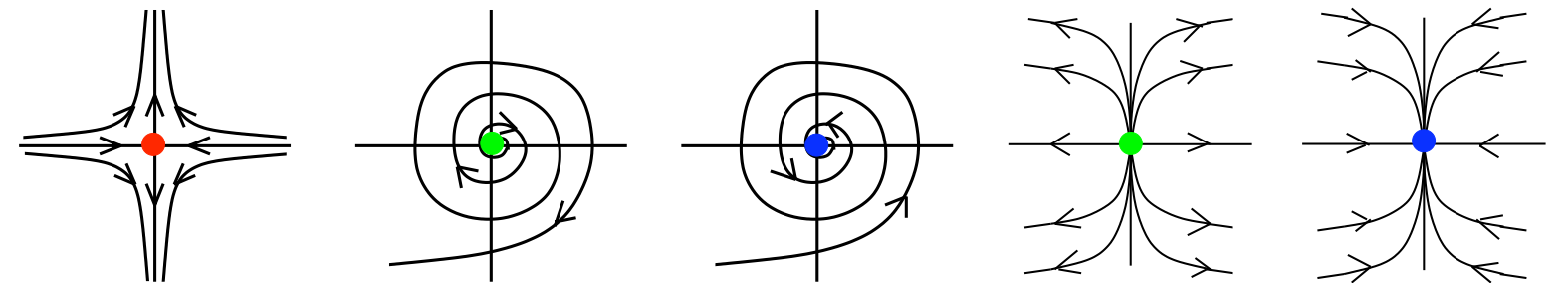
[Multidimensional Transfer Functions for Volume Rendering. Kniss, Kindlmann, and Hansen. In The Visualization Handbook, edited by Charles Hansen and Christopher Johnson, pp. 189–210. Elsevier, 2005.]

# Vector and tensor fields

- data
  - multiple attribs per cell (vector: 2)
- idiom families
  - flow *glyphs*
    - purely local
  - *geometric flow*
    - derived data from tracing particle trajectories
    - sparse set of seed points
  - *texture flow*
    - derived data, dense seeds
  - *feature flow*
    - global computation to detect features



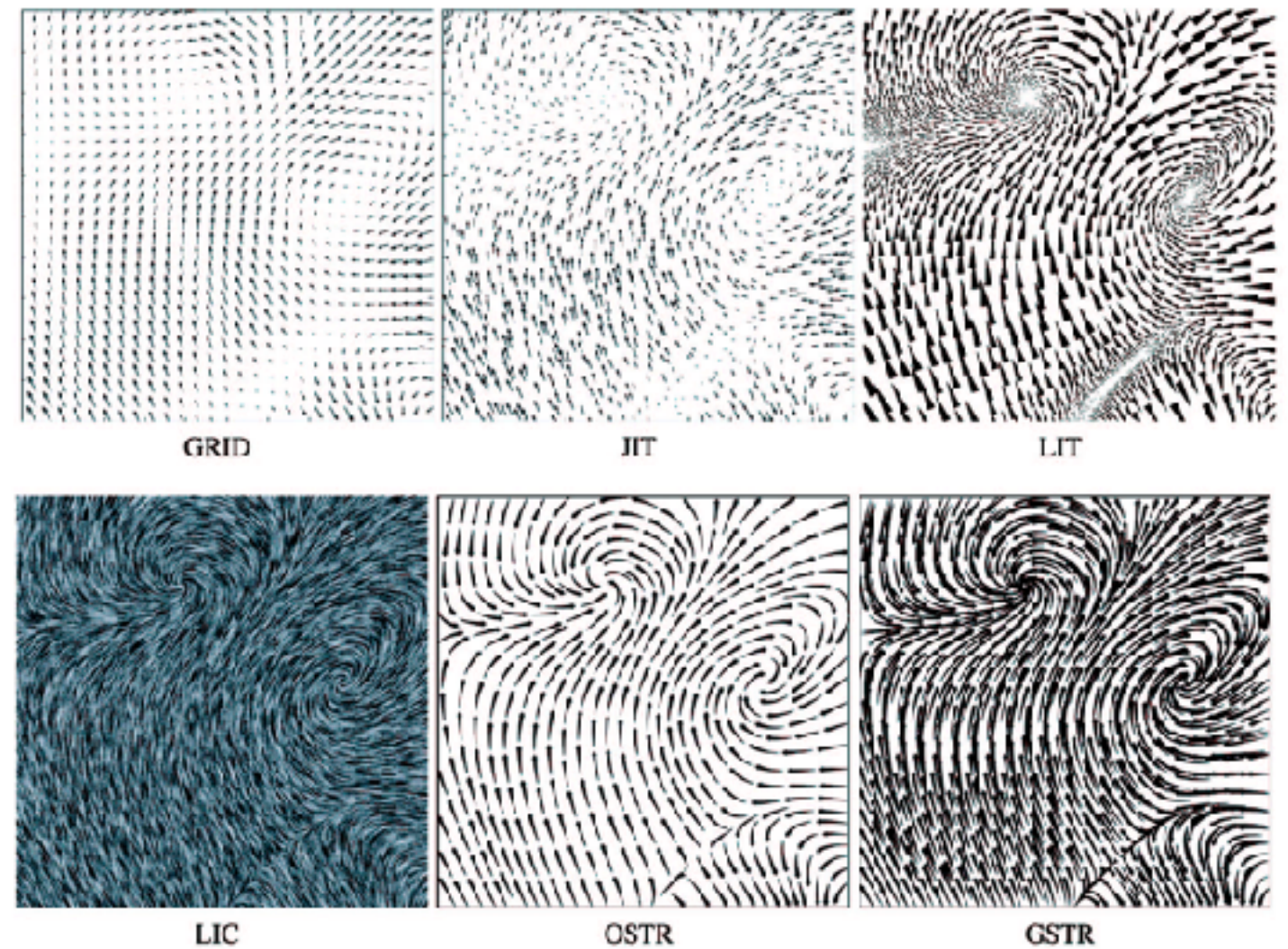
[Comparing 2D vector field visualization methods: A user study. Laidlaw et al. IEEE Trans. Visualization and Computer Graphics (TVCG) 11:1 (2005), 59–70.]



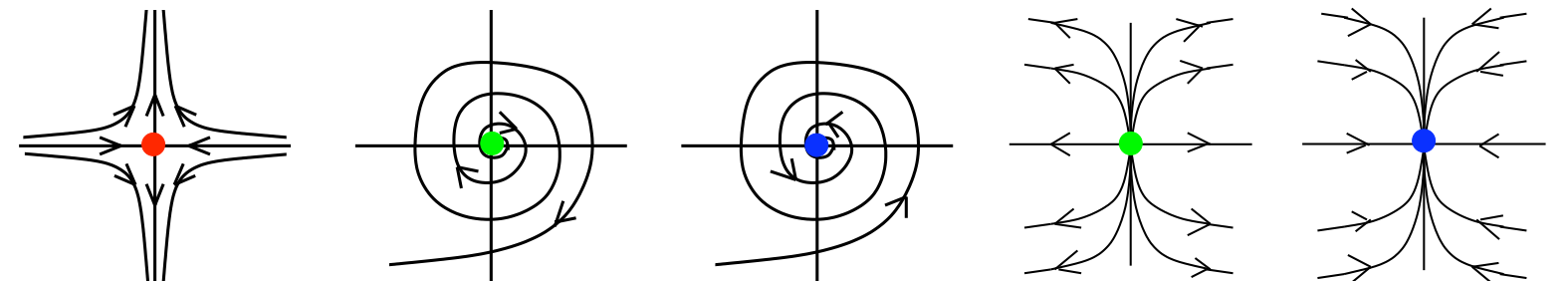
[Topology tracking for the visualization of time-dependent two-dimensional flows. Tricoche, Wischgoll, Scheuermann, and Hagen. Computers & Graphics 26:2 (2002), 249–257.]

# Vector fields

- empirical study tasks
  - finding critical points, identifying their types
  - identifying what type of critical point is at a specific location
  - predicting where a particle starting at a specified point will end up (advection)



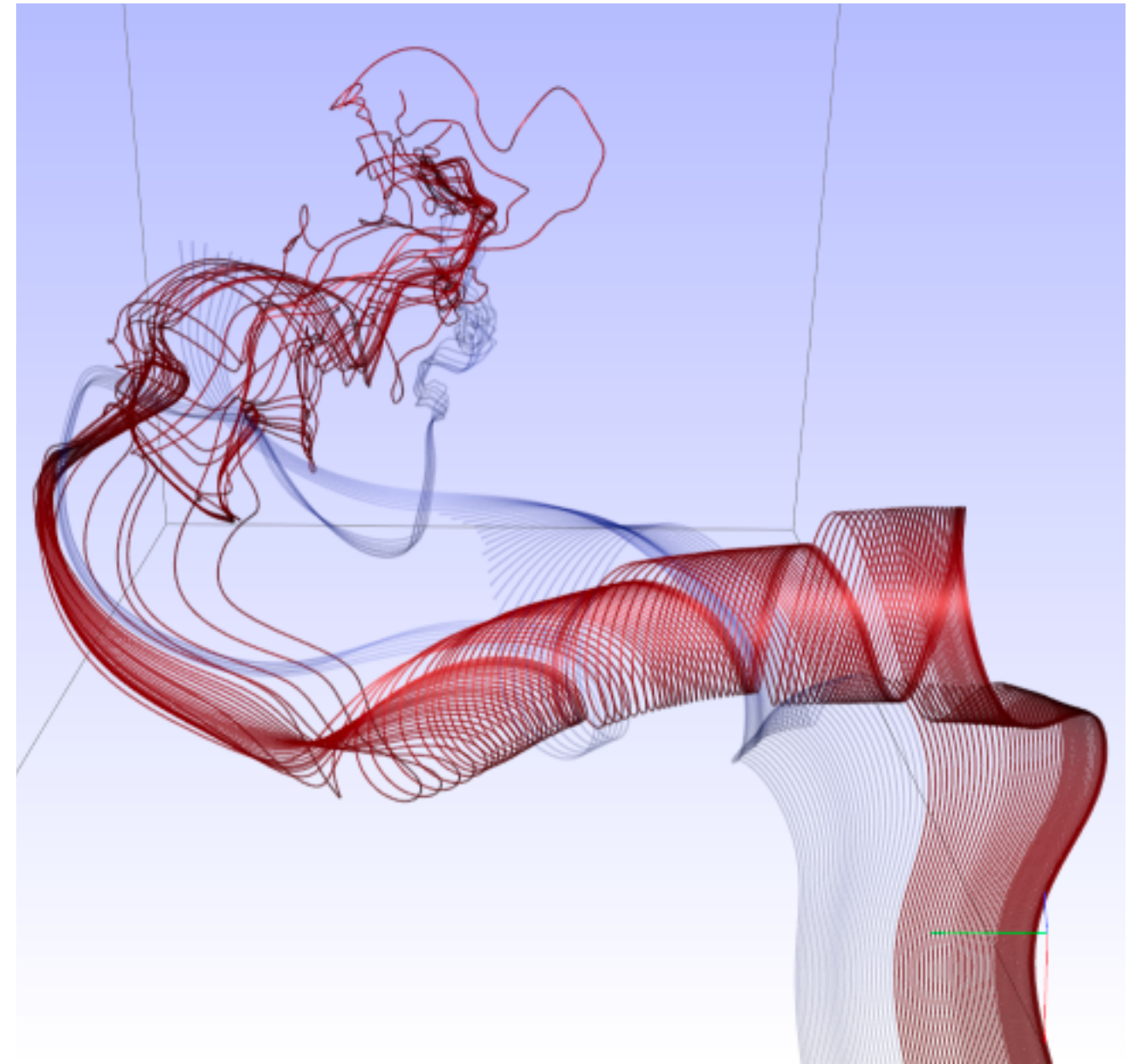
[Comparing 2D vector field visualization methods: A user study. Laidlaw et al. *IEEE Trans. Visualization and Computer Graphics (TVCG)* 11:1 (2005), 59–70.]



[Topology tracking for the visualization of time-dependent two-dimensional flows. Tricoche, Wischgoll, Scheuermann, and Hagen. *Computers & Graphics* 26:2 (2002), 249–257.]

# Idiom: **similarity-clustered streamlines**

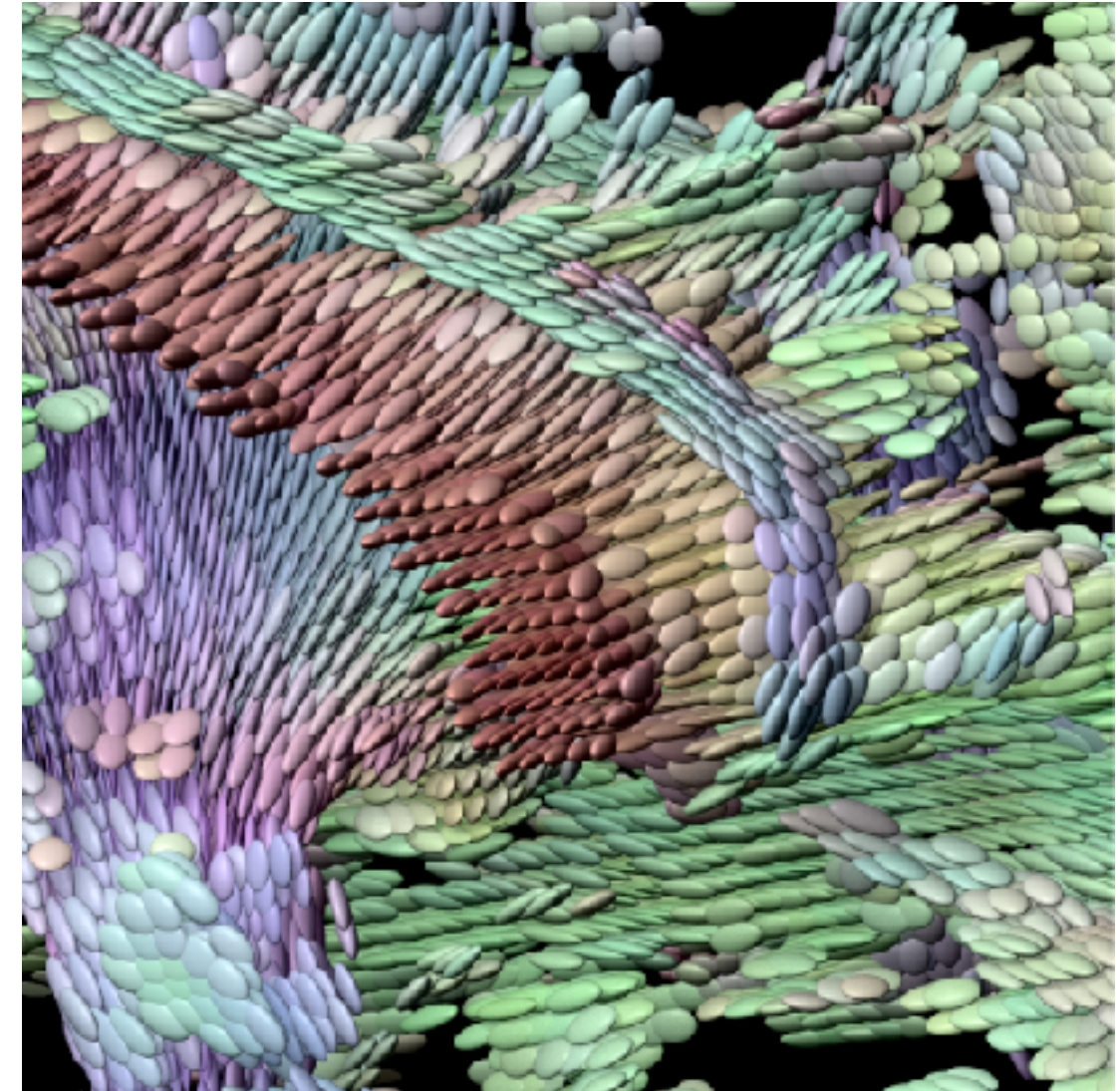
- data
  - 3D vector field
- derived data (from field)
  - streamlines: trajectory particle will follow
- derived data (per streamline)
  - curvature, torsion, tortuosity
  - signature: complex weighted combination
  - compute cluster hierarchy across all signatures
  - encode: color and opacity by cluster
- tasks
  - find features, query shape
- scalability
  - millions of samples, hundreds of streamlines



*[Similarity Measures for Enhancing Interactive Streamline Seeding. McLoughlin, Jones, Laramee, Malki, Masters, and Hansen. IEEE Trans. Visualization and Computer Graphics 19:8 (2013), 1342–1353.]*

# Idiom: Ellipsoid Tensor Glyphs

- data
  - tensor field: multiple attributes at each cell (entire matrix)
    - stress, conductivity, curvature, diffusivity...
  - derived data:
    - shape (eigenvalues)
    - orientation (eigenvectors)
- visual encoding
  - glyph: 3D ellipsoid





# Arrange spatial data

## ① Use Given

### → Geometry

→ *Geographic*

### → Spatial Fields

→ *Scalar Fields (one value per cell)*

→ *Isocontours*

→ *Direct Volume Rendering*

→ *Vector and Tensor Fields (many values per cell)*

→ *Flow Glyphs (local)*

→ *Geometric (sparse seeds)*

→ *Textures (dense seeds)*

→ *Features (globally derived)*

