

# Visualization Analysis & Design

*Reduce: Aggregation & Filtering (Ch 13)*

**Tamara Munzner**

Department of Computer Science  
University of British Columbia

@tamaramunzner

# How to handle complexity: 3 previous strategies

→ *Derive*



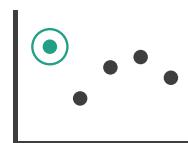
## Manipulate

→ **Change**



- derive new data to show within view
- change view over time
- facet across multiple views

→ **Select**

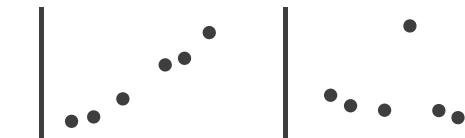


→ **Navigate**

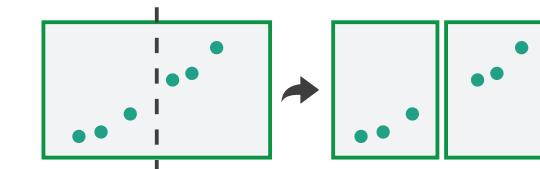


## Facet

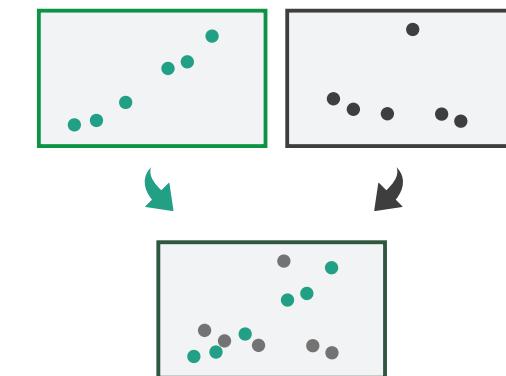
→ **Juxtapose**



→ **Partition**

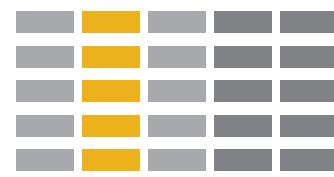


→ **Superimpose**



# How to handle complexity: 3 previous strategies + 1 more

→ *Derive*



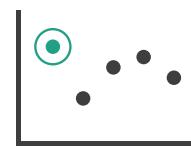
**Manipulate**

→ **Change**



- derive new data to show within view
- change view over time
- facet across multiple views
- reduce items/attributes within single view

→ **Select**



→ **Navigate**

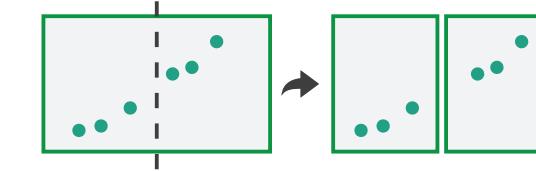


**Facet**

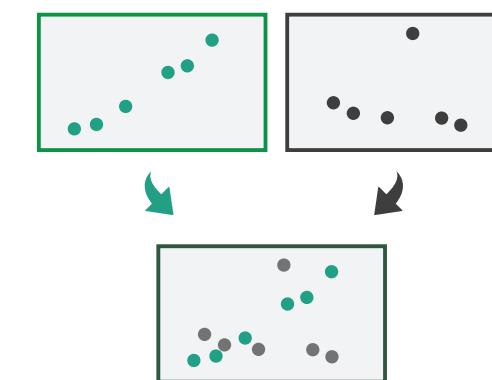
→ **Juxtapose**



→ **Partition**



→ **Superimpose**



**Reduce**

→ **Filter**



→ **Aggregate**



→ **Embed**



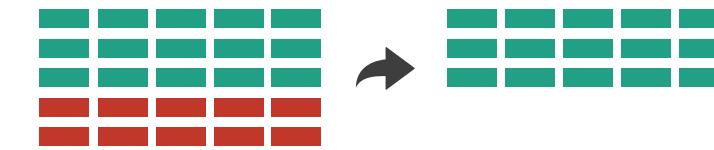
# Reduce items and attributes

- reduce/increase: inverses
- filter
  - pro: straightforward and intuitive
    - to understand and compute
  - con: out of sight, out of mind

## Reducing Items and Attributes

→ Filter

→ Items



→ Attributes



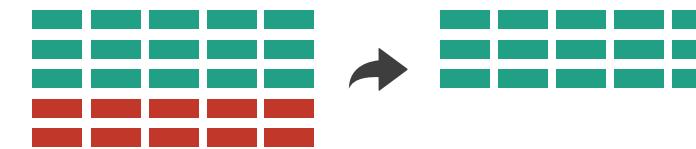
# Reduce items and attributes

- reduce/increase: inverses
- filter
  - pro: straightforward and intuitive
    - to understand and compute
  - con: out of sight, out of mind
- aggregation
  - pro: inform about whole set
  - con: difficult to avoid losing signal
- not mutually exclusive
  - combine filter, aggregate
  - combine reduce, change, facet

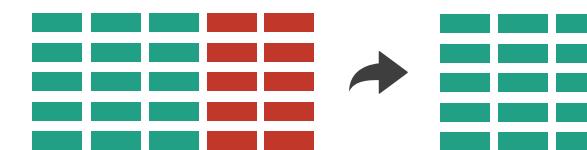
## Reducing Items and Attributes

### → Filter

→ Items

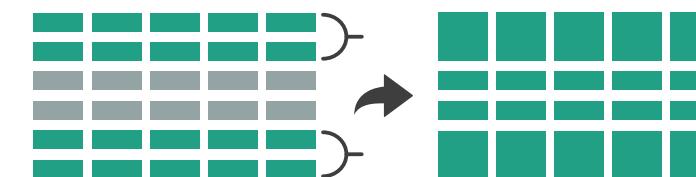


→ Attributes

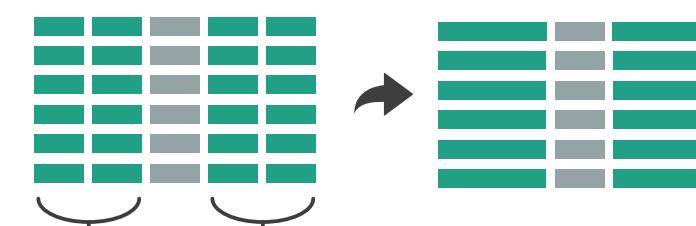


### → Aggregate

→ Items



→ Attributes



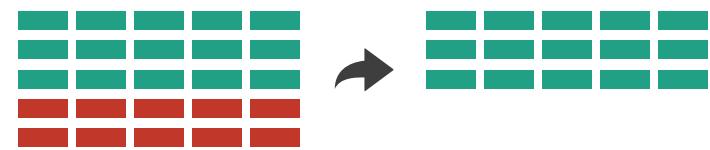
# Filter

- eliminate some elements
  - either items or attributes
- according to what?
  - any possible function that partitions dataset into two sets
    - attribute values bigger/smaller than x
    - noise/signal
- filters vs queries
  - query: start with nothing, add in elements
  - filters: start with everything, remove elements
  - best approach depends on dataset size

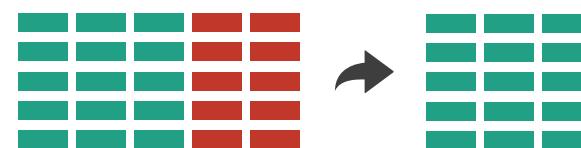
## Reducing Items and Attributes

### ④ Filter

→ Items

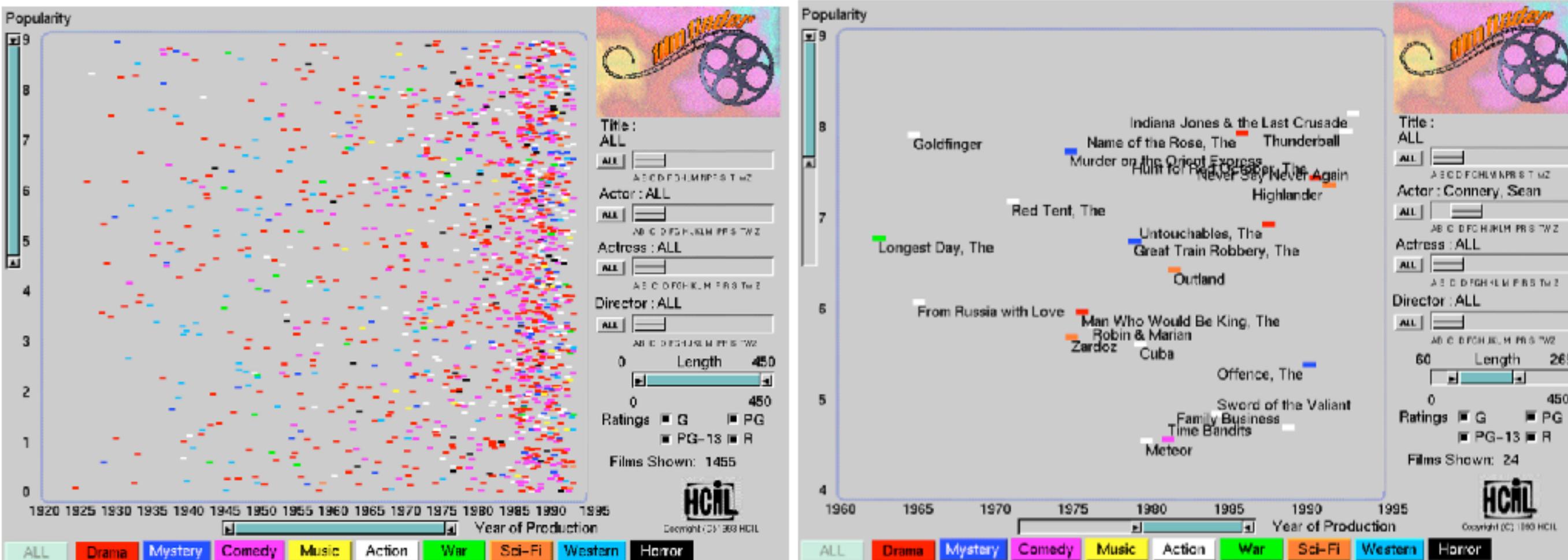


→ Attributes



# Idiom: FilmFinder

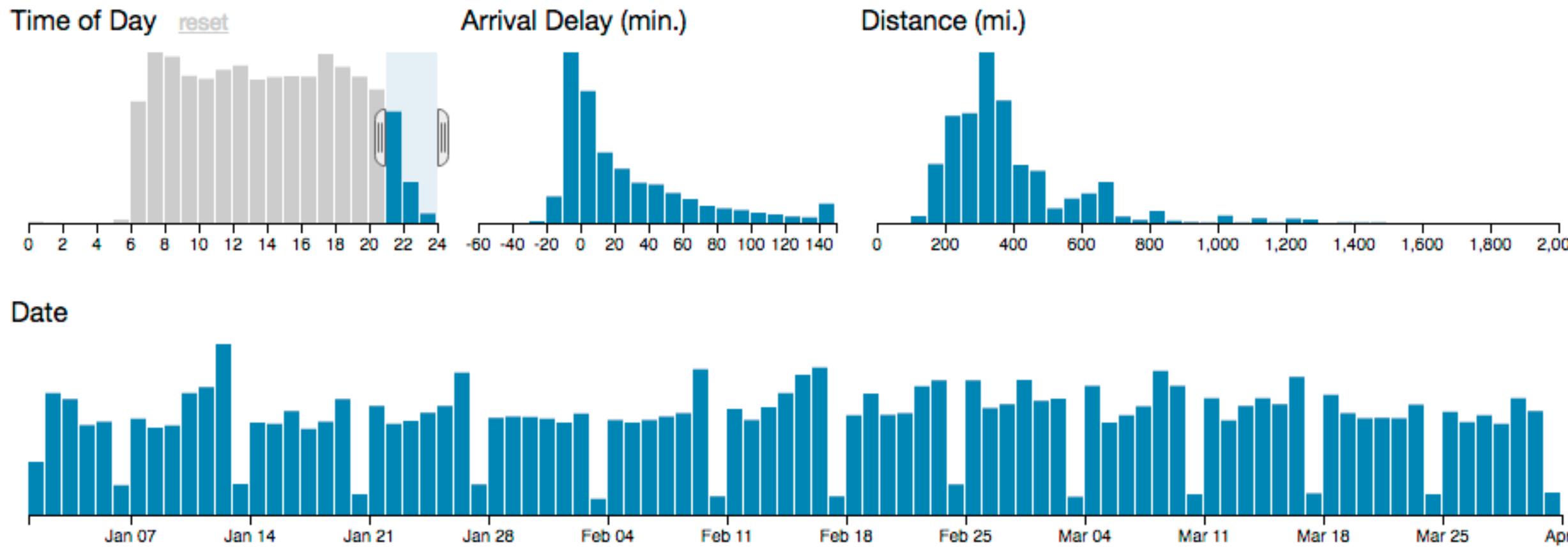
- dynamic queries/filters for items
  - tightly coupled interaction and visual encoding idioms, so user can immediately see results of action



# Idiom: cross filtering

# System: Crossfilter

- item filtering
- coordinated views/controls combined
  - all selected histogram sliders update when any ranges change



<http://square.github.io/crossfilter/>

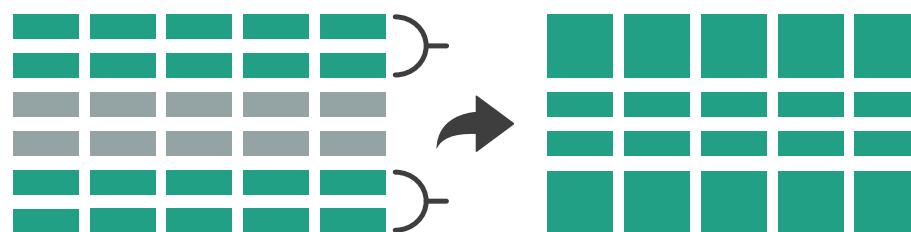
<https://observablehq.com/@uwdata/interaction>

# Aggregate

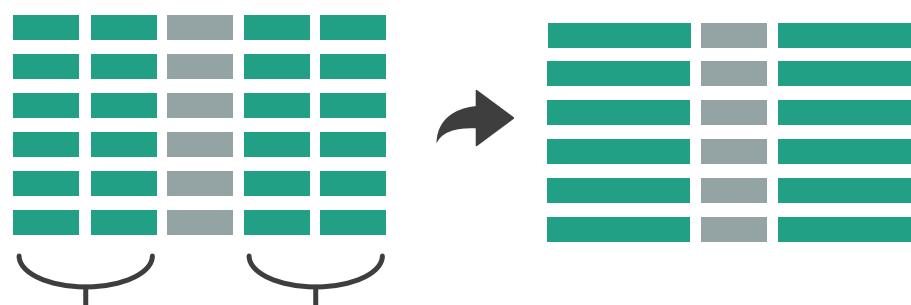
- a group of elements is represented by a smaller number of derived elements

➔ Aggregate

→ Items

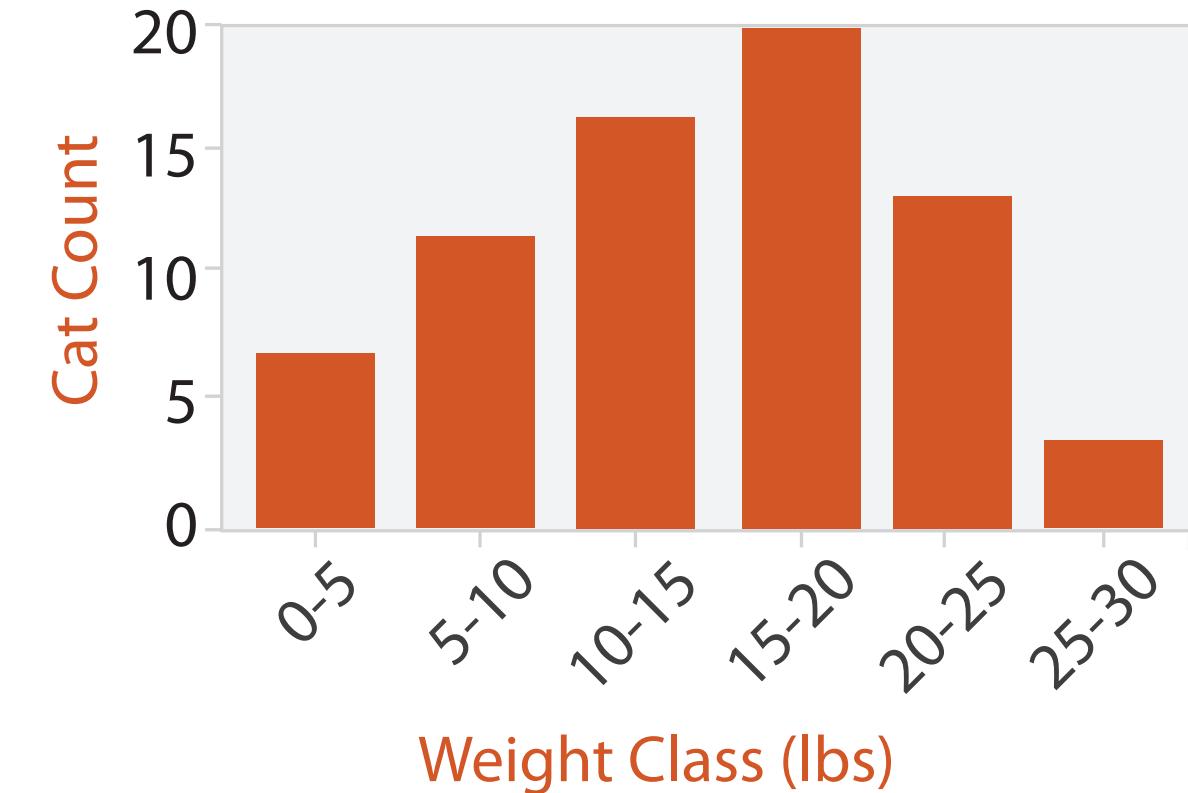


→ Attributes



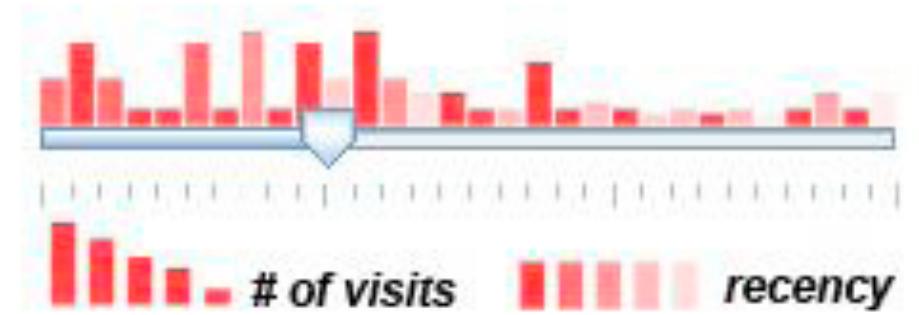
# Idiom: histogram

- static item aggregation
- task: find distribution
- data: table
- derived data
  - new table: keys are bins, values are counts
- bin size crucial
  - pattern can change dramatically depending on discretization
  - opportunity for interaction: control bin size on the fly



# Idiom: scented widgets

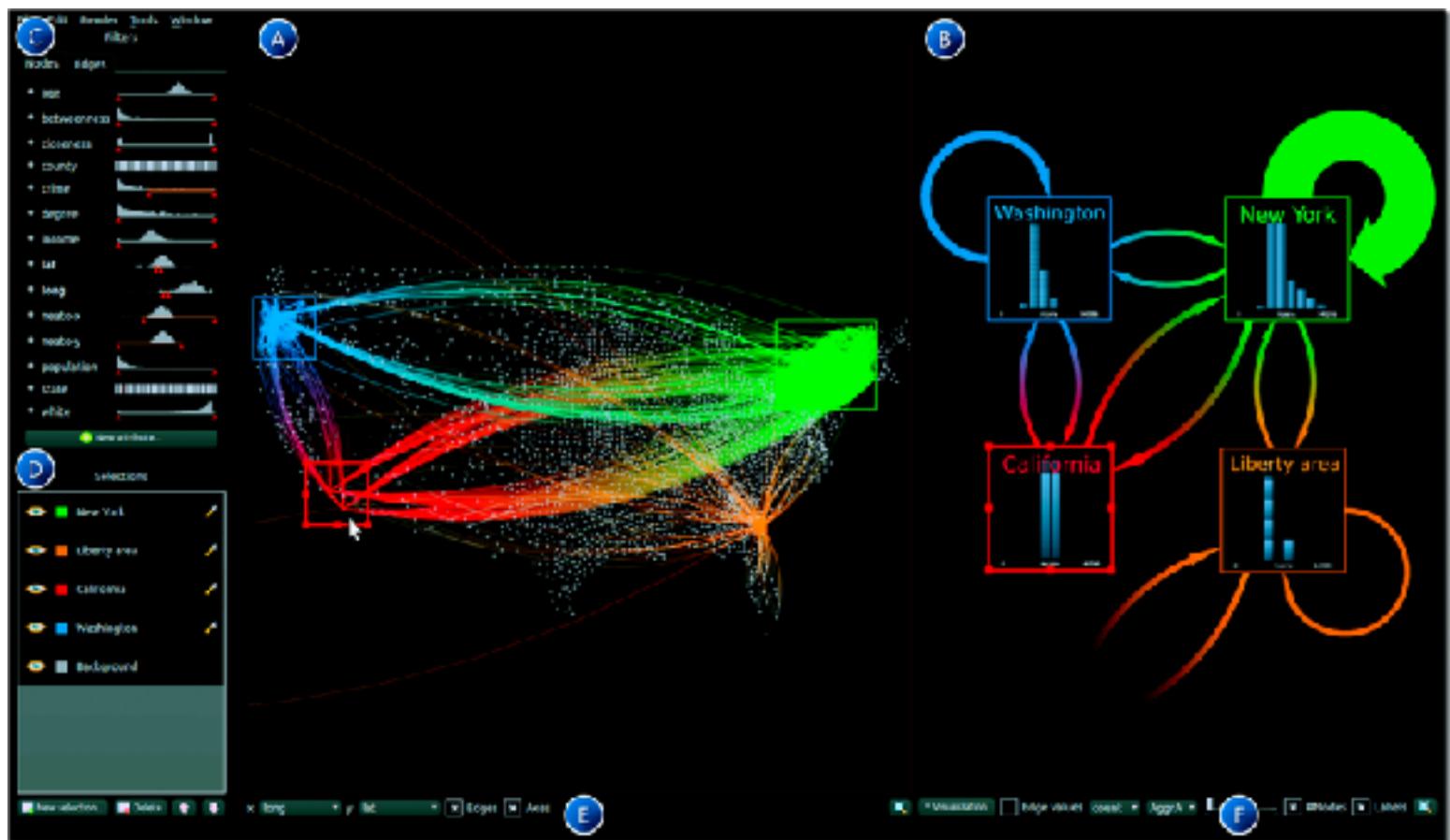
- augmented widgets show *information scent*
  - better cues for *information foraging*: show whether value in drilling down further vs looking elsewhere
- concise use of space: histogram on slider



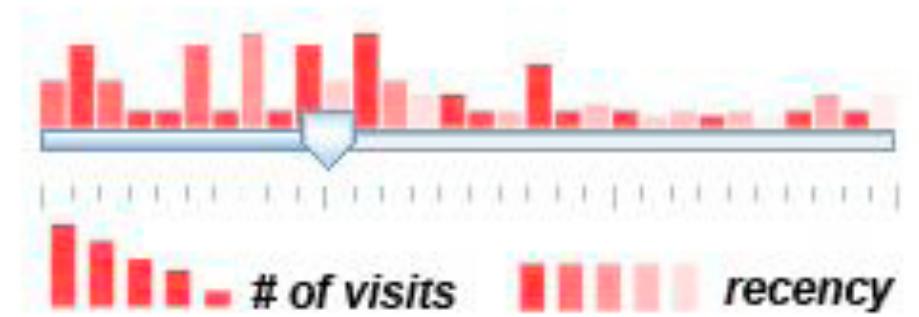
[Scented Widgets: Improving Navigation Cues with Embedded Visualizations. Willett, Heer, and Agrawala. IEEE TVCG (Proc. InfoVis 2007) 13:6 (2007), 1129–1136.]

# Idiom: scented widgets

- augmented widgets show *information scent*
  - better cues for *information foraging*: show whether value in drilling down further vs looking elsewhere
- concise use of space: histogram on slider



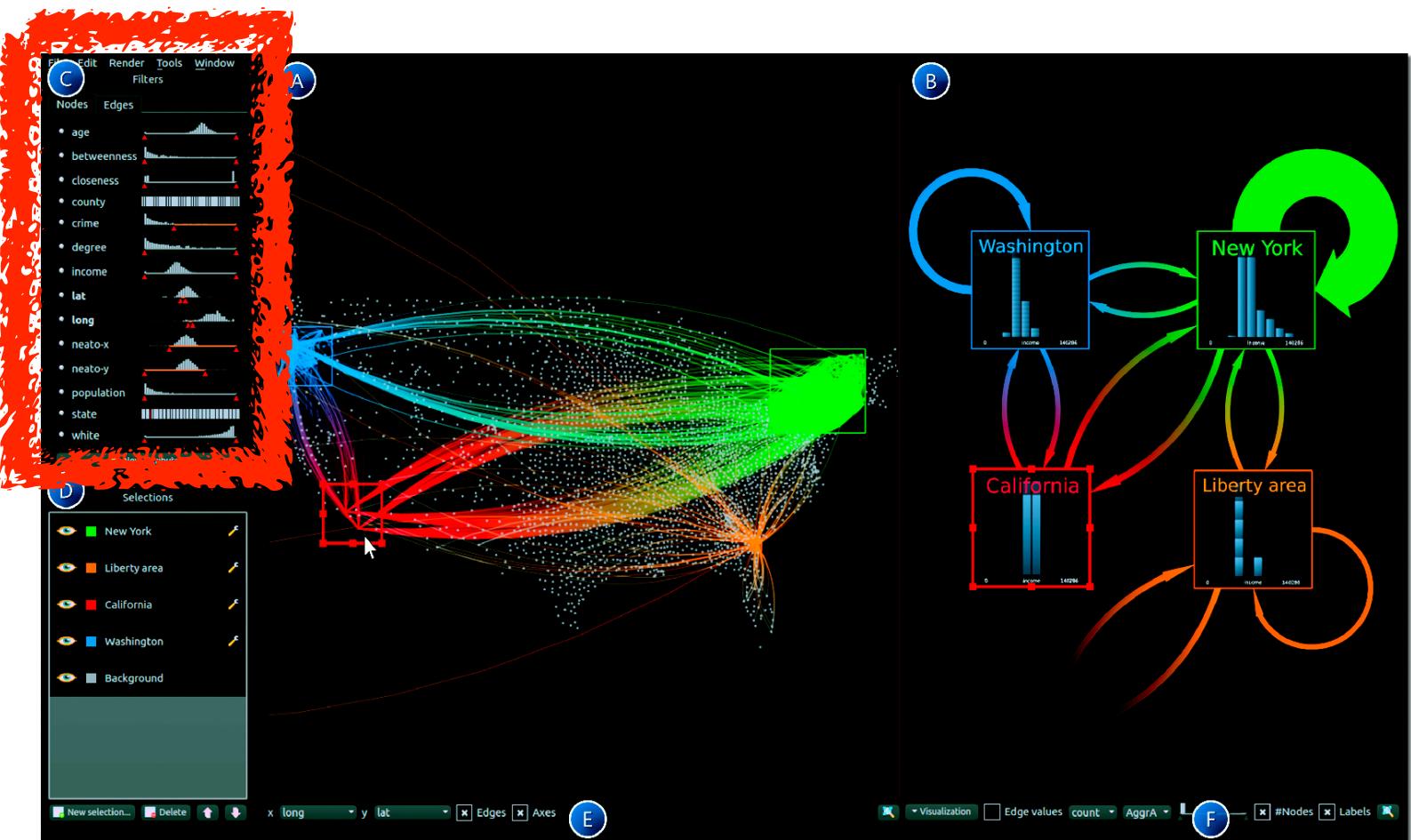
[Multivariate Network Exploration and Presentation: From Detail to Overview via Selections and Aggregations. van den Elzen, van Wijk, IEEE TVCG 20(12): 2014 (Proc. InfoVis 2014).]



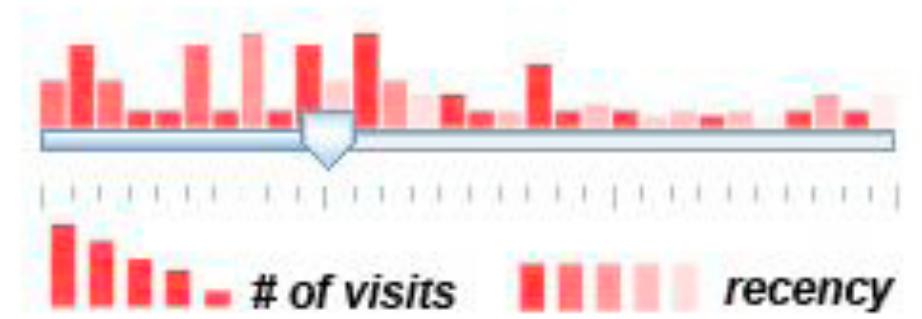
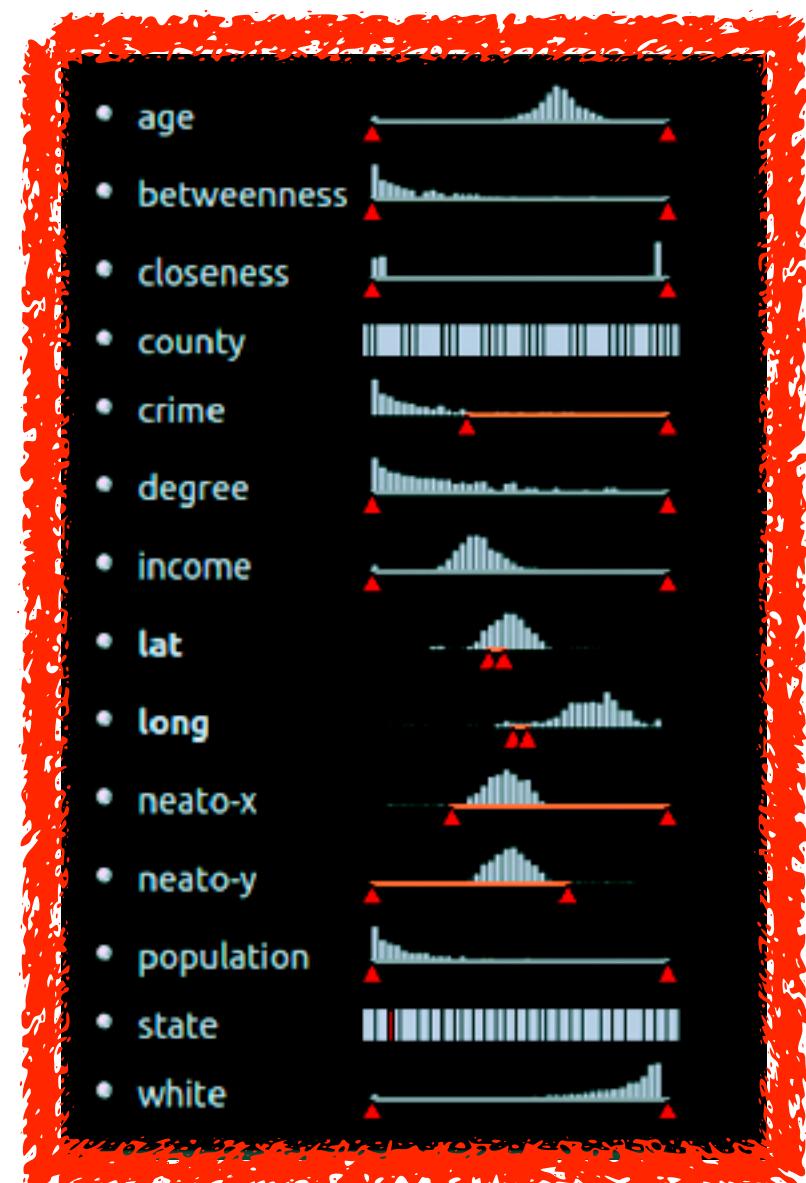
[Scented Widgets: Improving Navigation Cues with Embedded Visualizations. Willett, Heer, and Agrawala. IEEE TVCG (Proc. InfoVis 2007) 13:6 (2007), 1129–1136.]

# Idiom: scented widgets

- augmented widgets show *information scent*
  - better cues for *information foraging*: show whether value in drilling down further vs looking elsewhere
- concise use of space: histogram on slider

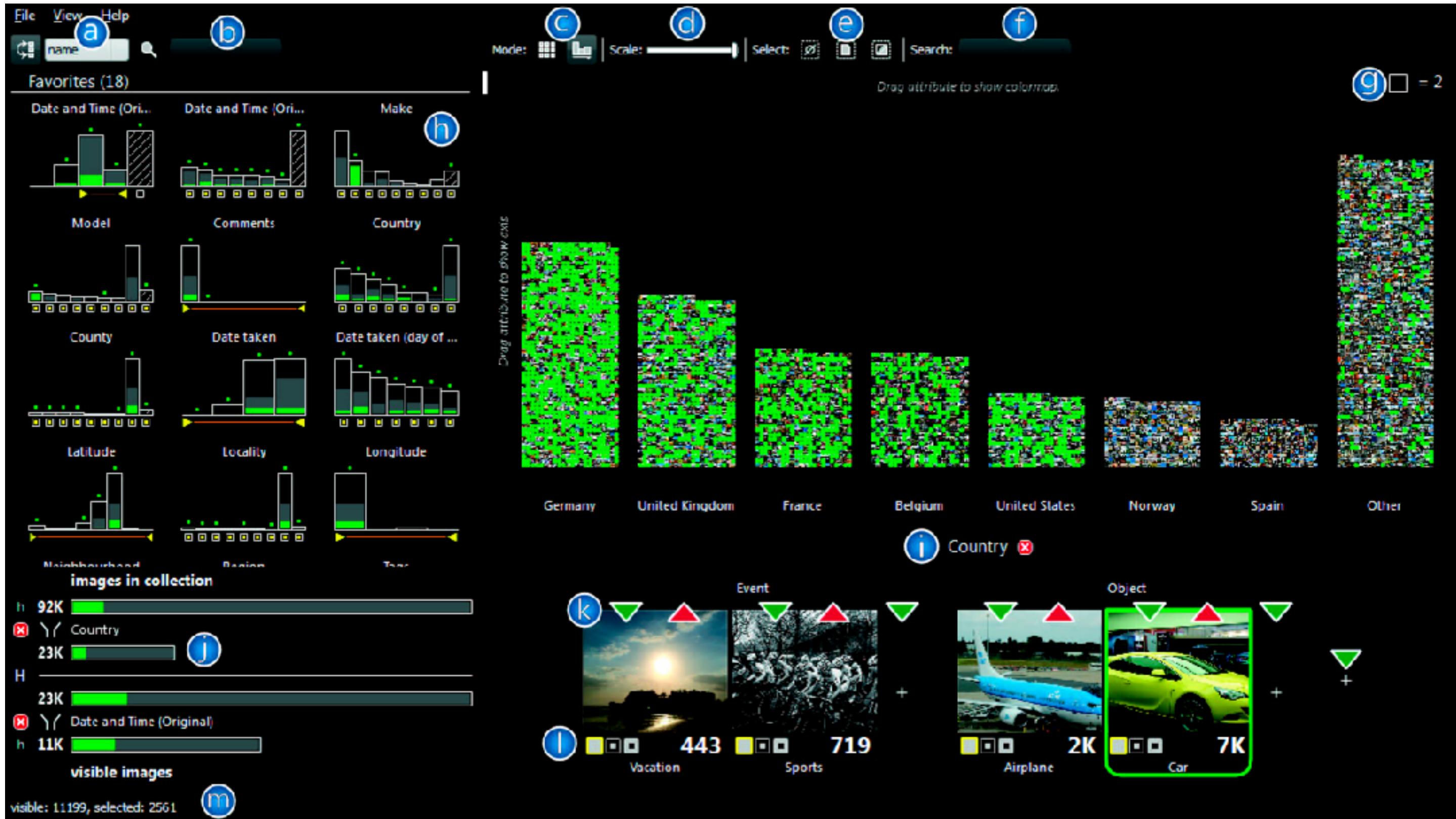


[Multivariate Network Exploration and Presentation: From Detail to Overview via Selections and Aggregations. van den Elzen, van Wijk, IEEE TVCG 20(12): 2014 (Proc. InfoVis 2014).]



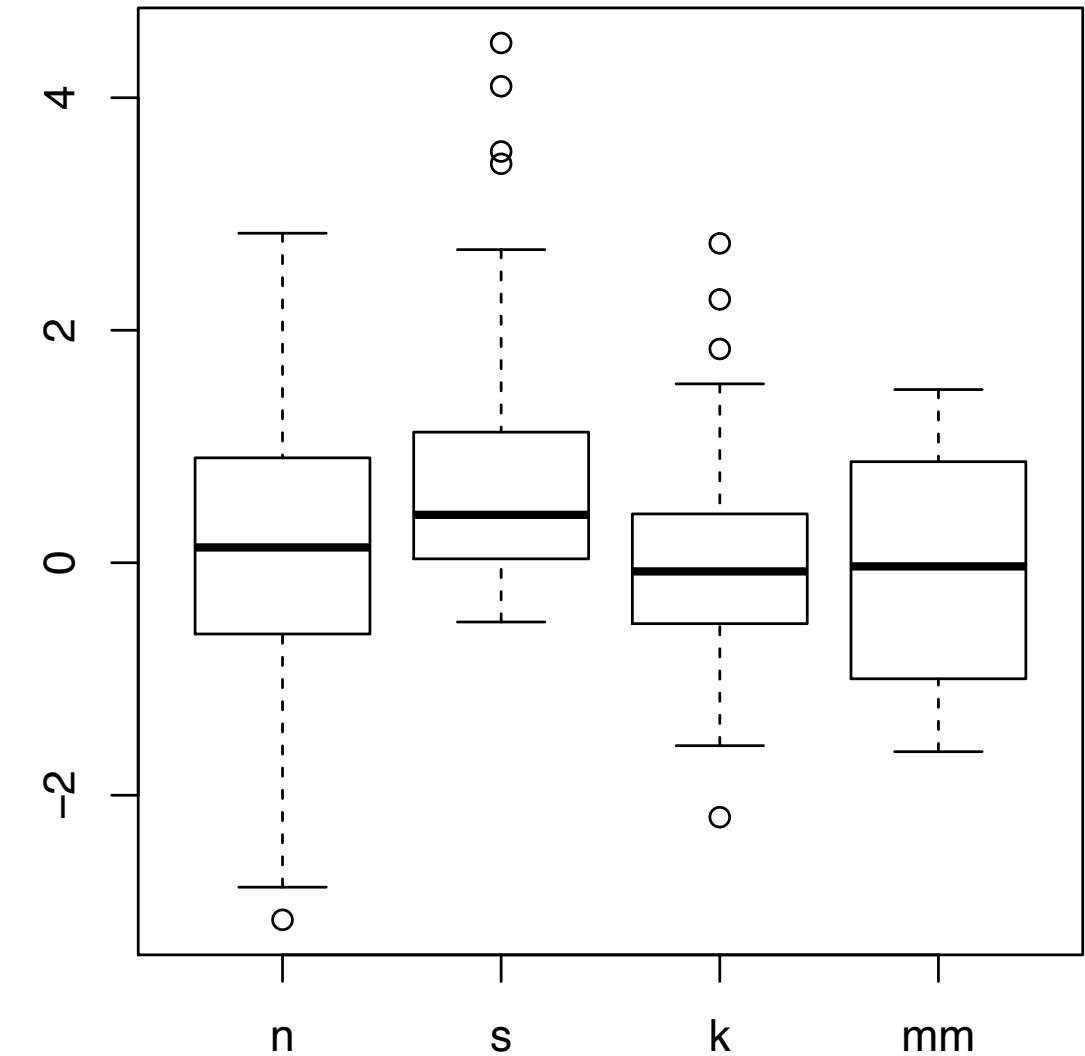
[Scented Widgets: Improving Navigation Cues with Embedded Visualizations. Willett, Heer, and Agrawala. IEEE TVCG (Proc. InfoVis 2007) 13:6 (2007), 1129–1136.]

# Scented histogram bisliders: detailed



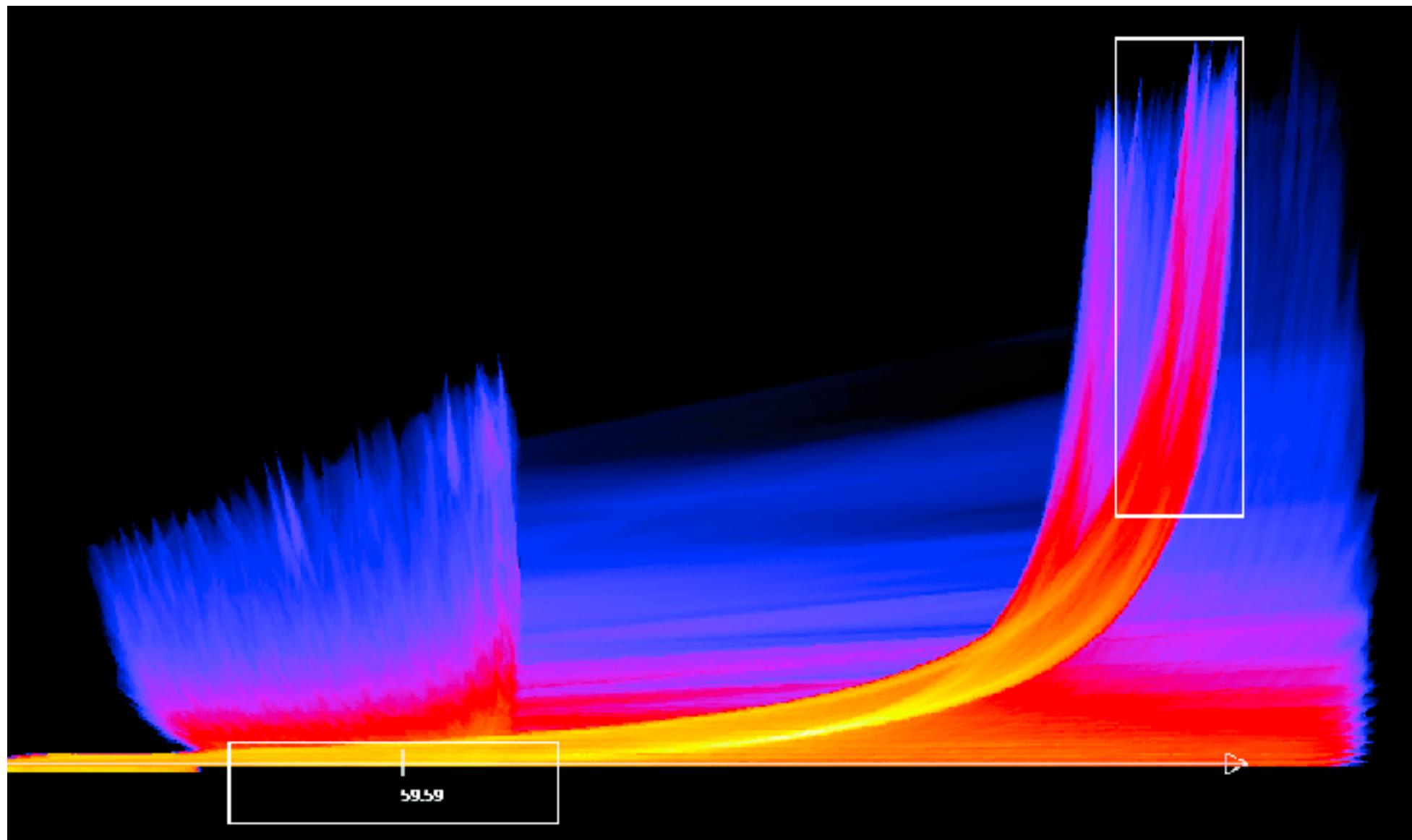
# Idiom: **boxplot**

- static item aggregation
- task: find distribution
- data: table
- derived data
  - 5 quant attrs
    - median: central line
    - lower and upper quartile: boxes
    - lower upper fences: whiskers
      - values beyond which items are outliers
  - outliers beyond fence cutoffs explicitly shown
- scalability
  - unlimited number of items!



# Idiom: Continuous scatterplot

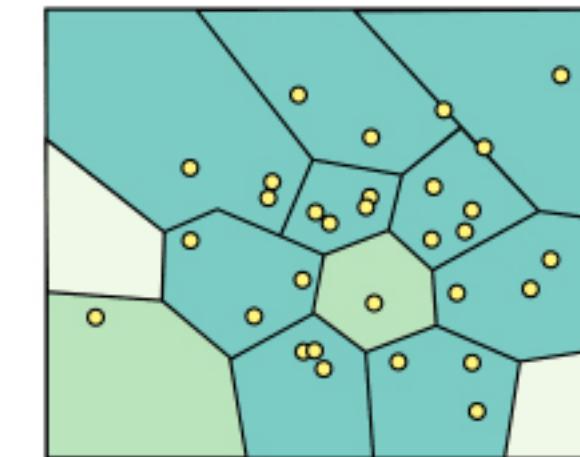
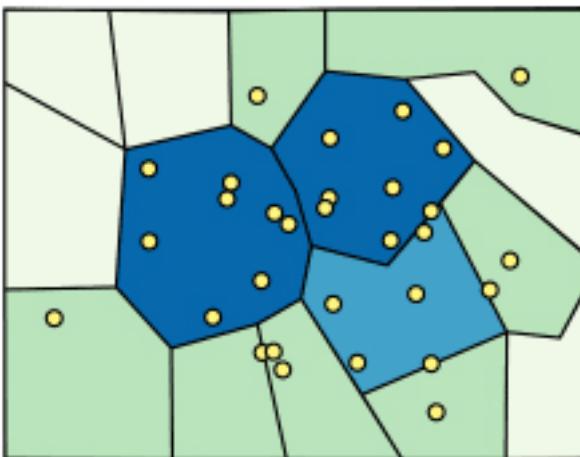
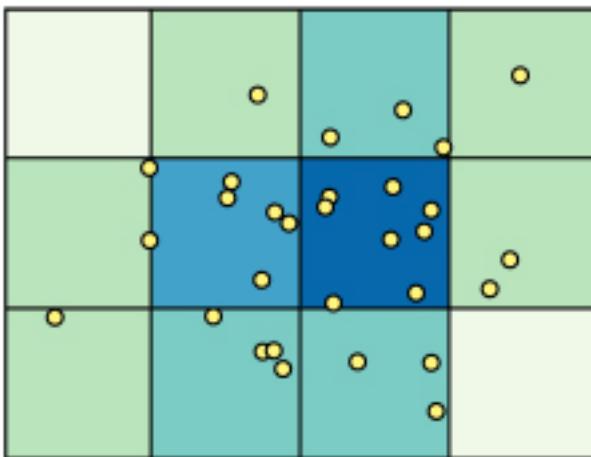
- static item aggregation
- data: table
- derived data: table
  - key attrs x,y for pixels
  - quant attrib: overplot density
- dense space-filling 2D matrix
- color:  
sequential categorical hue +  
ordered luminance colormap
- scalability
  - no limits on overplotting:  
millions of items



[Continuous Scatterplots. Bachthaler and Weiskopf.  
IEEE TVCG (Proc.Vis 08) 14:6 (2008), 1428–1435. 2008.]

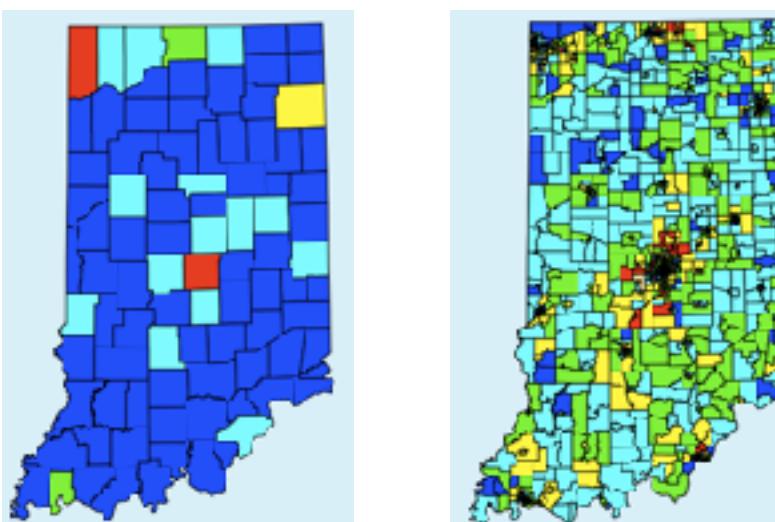
# Spatial aggregation

- MAUP: Modifiable Areal Unit Problem
  - changing boundaries of cartographic regions can yield dramatically different results
  - zone effects



[[http://www.e-education.psu.edu/geog486/l4\\_p7.html](http://www.e-education.psu.edu/geog486/l4_p7.html), Fig 4.cg.6]

- scale effects

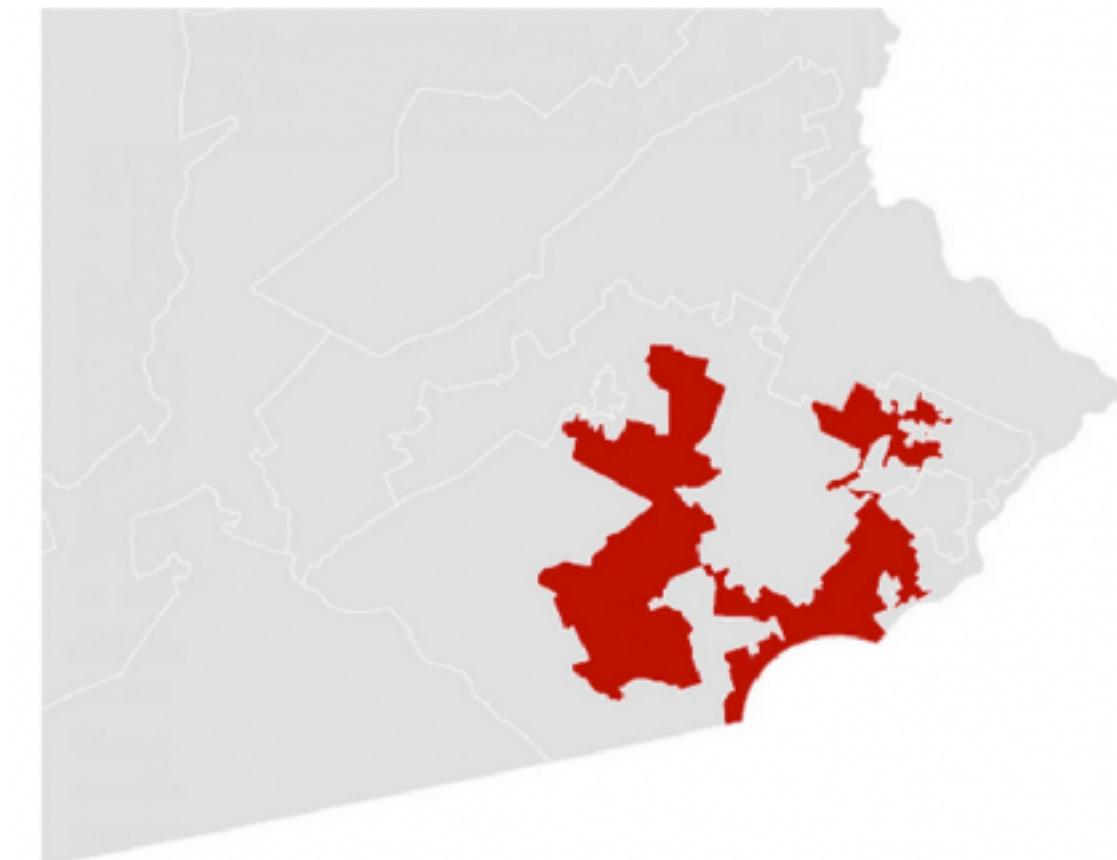
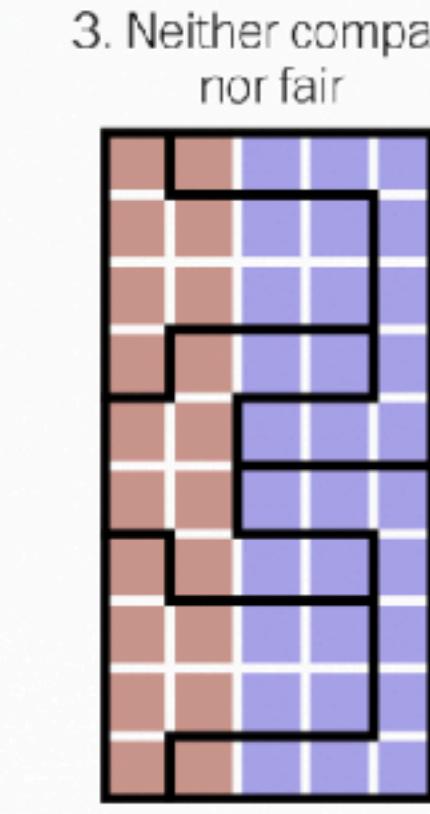
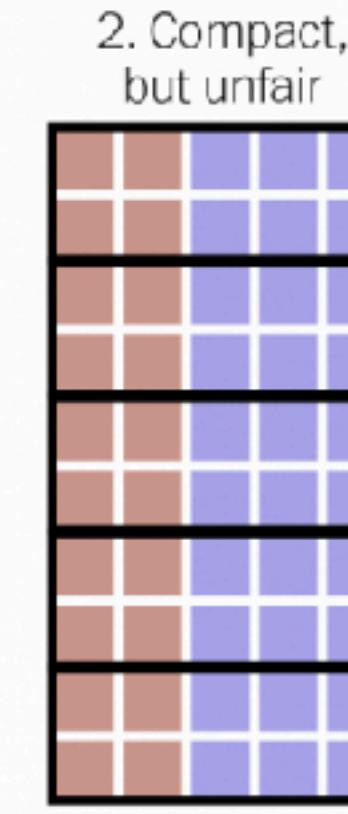
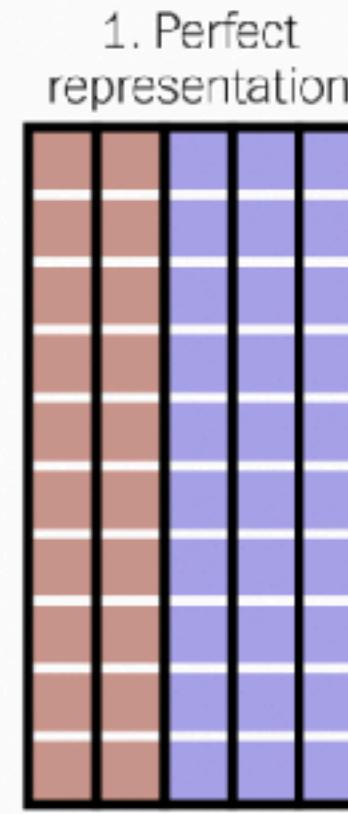
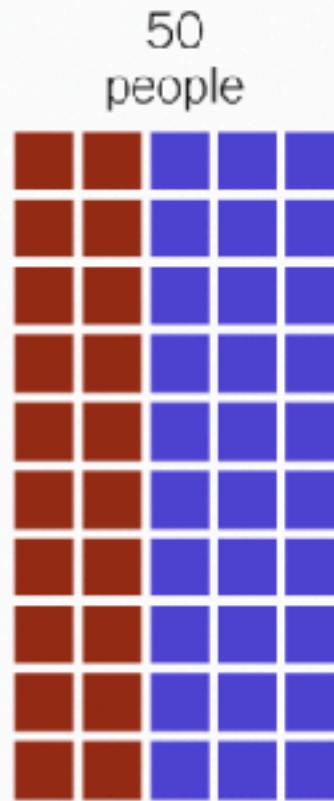


<https://blog.cartographica.com/blog/2011/5/19/the-modifiable-areal-unit-problem-in-gis.html>

# Gerrymandering: MAUP for political gain

## Gerrymandering, explained

Three different ways to divide 50 people into five districts



A real district in Pennsylvania:  
Democrats won 51% of the vote but only 5 out of  
18 house seats

WASHINGTONPOST.COM/WONKBLOG

Adapted from Stephen Nass

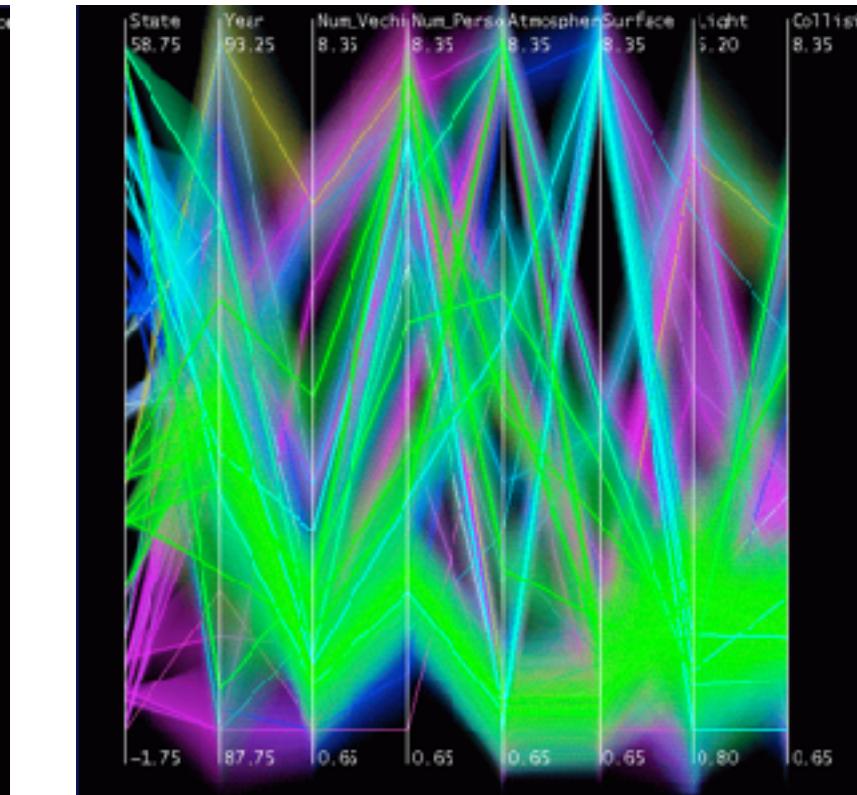
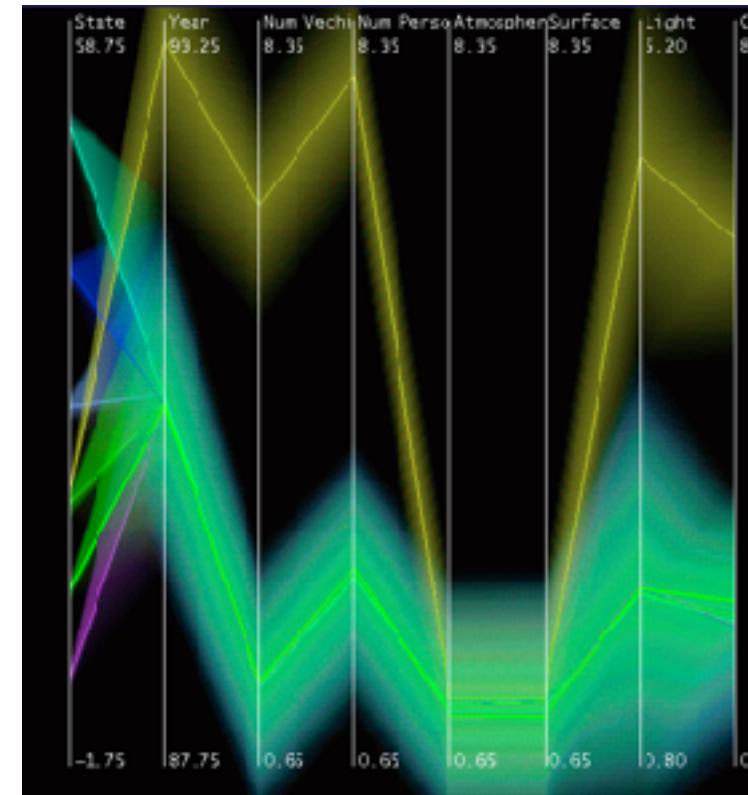
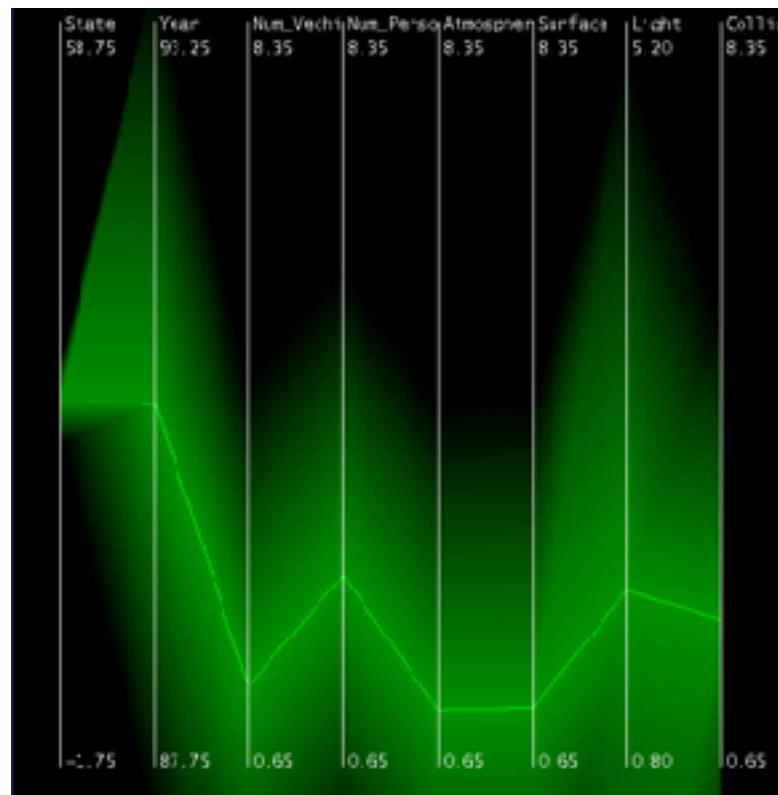
<https://www.washingtonpost.com/news/wonk/wp/2015/03/01/this-is-the-best-explanation-of-gerrymandering-you-will-ever-see/>

# Dynamic aggregation: Clustering

- clustering: classification of items into similar bins
  - based on similarity measure
  - hierarchical algorithms produce "similarity tree": cluster hierarchy
    - agglomerative clustering: start w/ each node as own cluster, then iteratively merge
- cluster hierarchy: derived data used w/ many dynamic aggregation idioms
  - cluster more homogeneous than whole dataset
    - statistical measures & distribution more meaningful

# Idiom: Hierarchical parallel coordinates

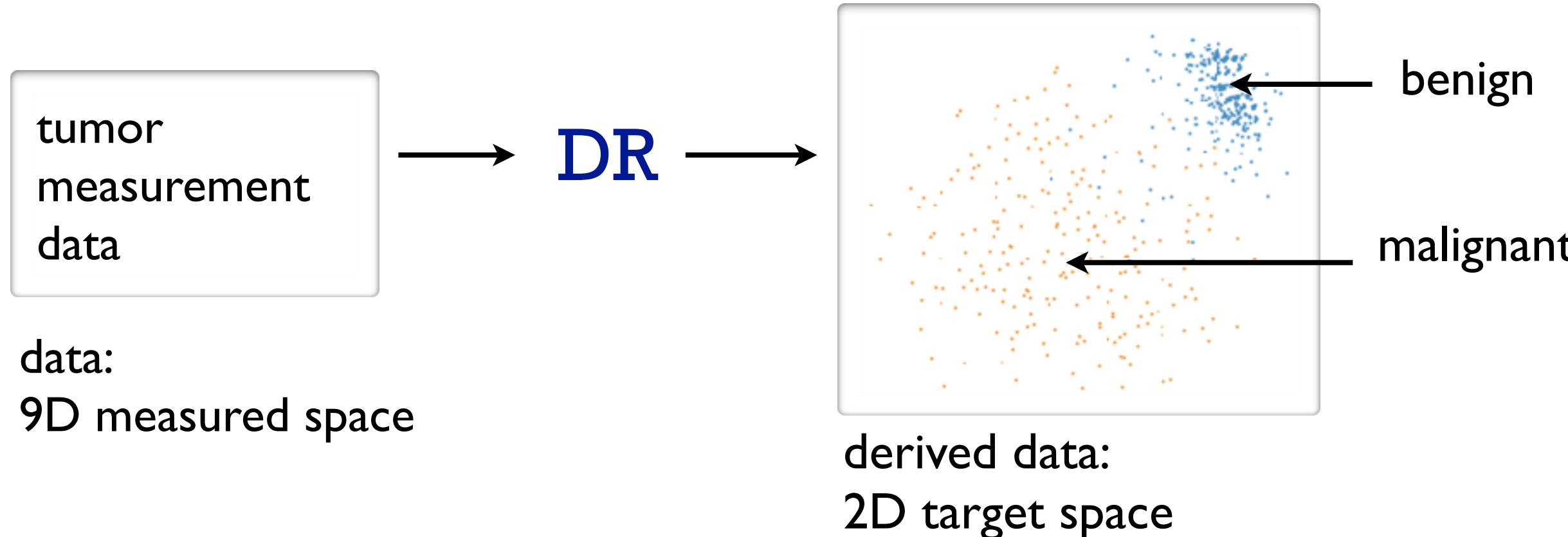
- dynamic item aggregation
- derived data: **cluster hierarchy**
- encoding:
  - cluster band with variable transparency, line at mean, width by min/max values
  - color by proximity in hierarchy



[Hierarchical Parallel Coordinates for Exploration of Large Datasets. Fua, Ward, and Rundensteiner.  
Proc. IEEE Visualization Conference (Vis '99), pp. 43– 50, 1999.]

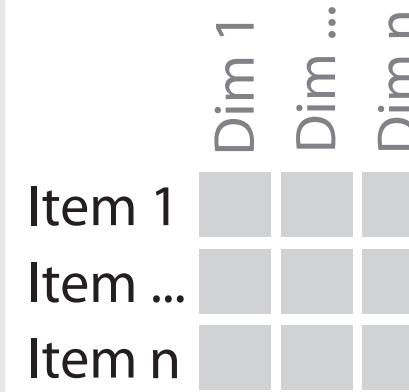
# Attribute aggregation: Dimensionality reduction

- attribute aggregation
  - derive low-dimensional target space from high-dimensional measured space
    - capture most of variance with minimal error
  - use when you can't directly measure what you care about
    - true dimensionality of dataset conjectured to be smaller than dimensionality of measurements
    - latent factors, hidden variables

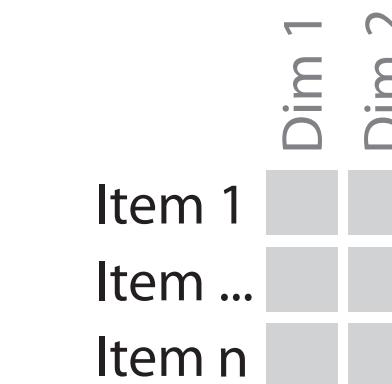


# Idiom: Dimensionality reduction for documents

Task 1

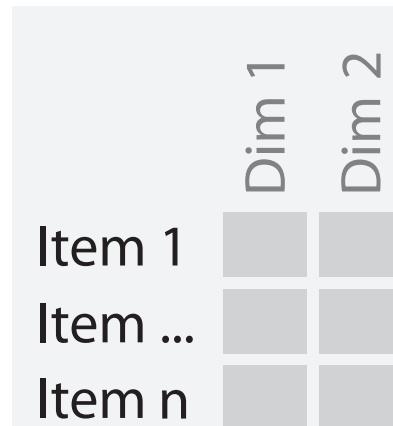


In  
HD data

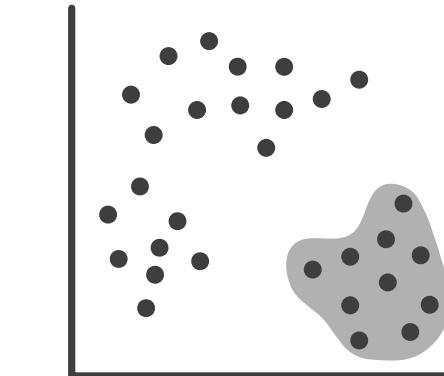


Out  
2D data

Task 2

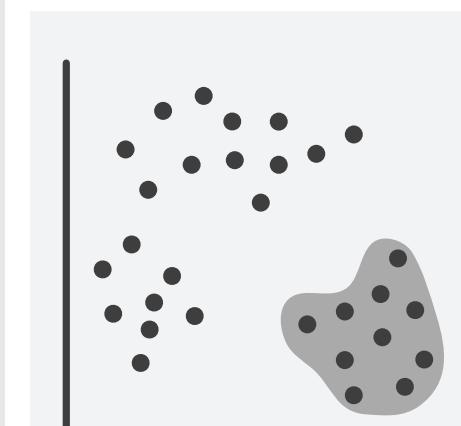


In  
2D data

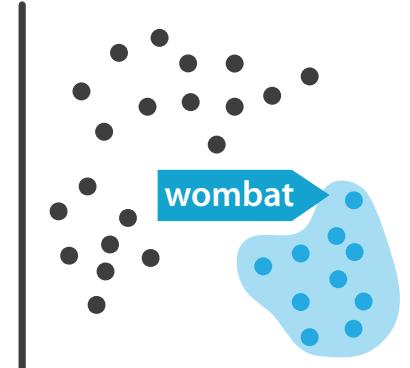


Out  
Scatterplot  
Clusters & points

Task 3



In  
Scatterplot  
Clusters & points



Out  
Labels for  
clusters

What?

- In High-dimensional data
- Out 2D data

Why?

- Produce
- Derive

What?

- In 2D data
- Out Scatterplot
- Out Clusters & points

Why?

- Discover
- Explore
- Identify

How?

- Encode
- Navigate
- Select

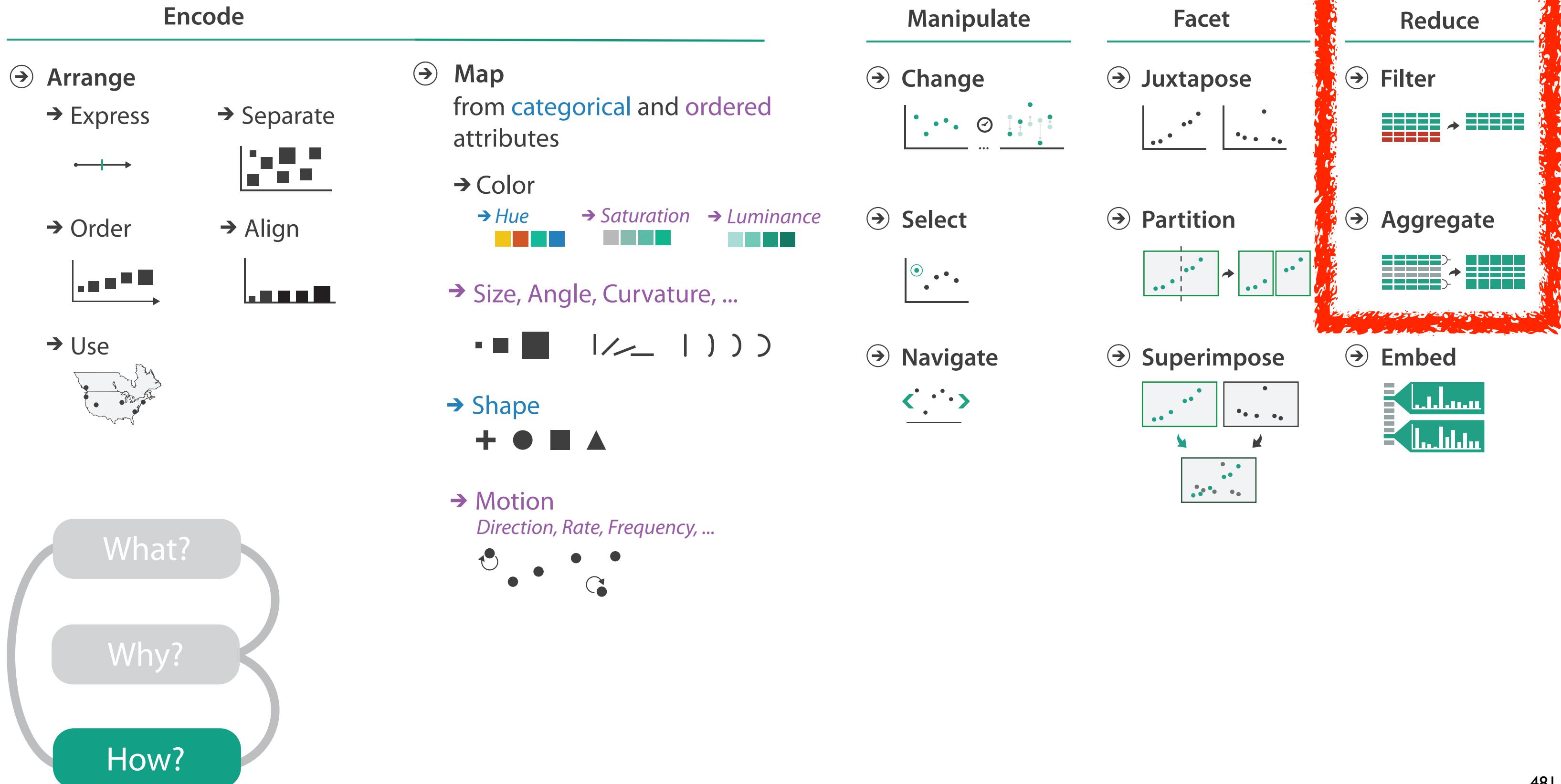
What?

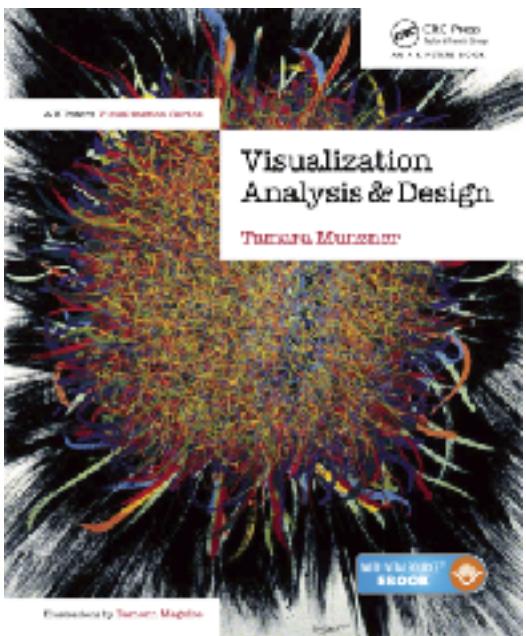
- In Scatterplot
- In Clusters & points
- Out Labels for clusters

Why?

- Produce
- Annotate

# How?





# Visualization Analysis & Design

*Embed: Focus+Context (Ch 14)*

**Tamara Munzner**

Department of Computer Science  
University of British Columbia

@tamaramunzner

# How to handle complexity: 4 strategies

→ *Derive*



## Manipulate

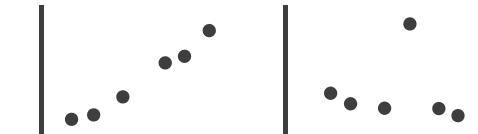
→ Change



- derive new data to show within view
- change view over time
- facet across multiple views
- reduce items/attributes within single view

## Facet

→ Juxtapose

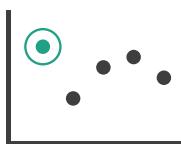


## Reduce

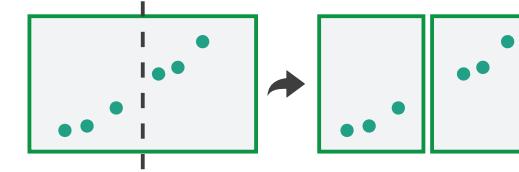
→ Filter



→ Select



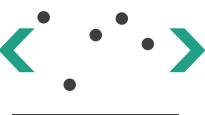
→ Partition



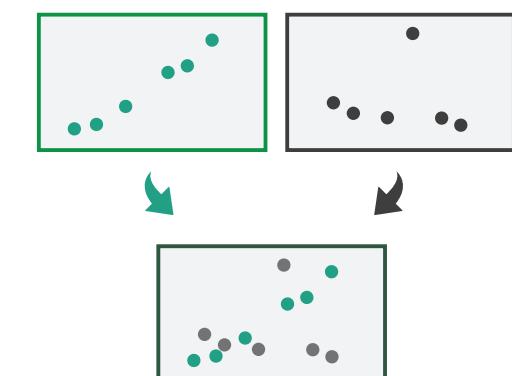
→ Aggregate



→ Navigate



→ Superimpose



→ Embed



# Embed: Focus+Context

- combine focus + context info within single view
  - vs standard navigation within view
  - vs multiple views

# Embed: Focus+Context

- combine focus + context info within single view
  - vs standard navigation within view
  - vs multiple views
- elide data
  - selectively filter and aggregate

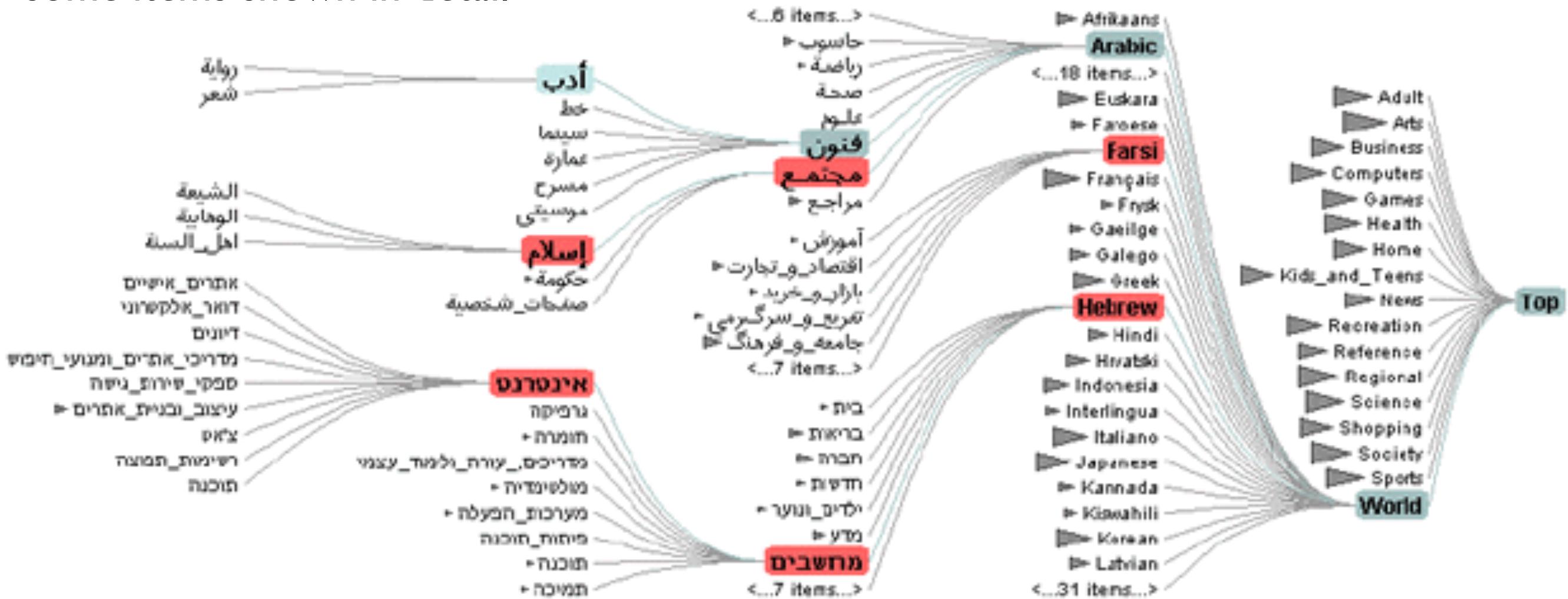
→ Embed

→ Elide Data



# Idiom: DOI Trees Revisited

- focus+context choice: elide
  - some items dynamically filtered out
  - some items dynamically aggregated together
  - some items shown in detail



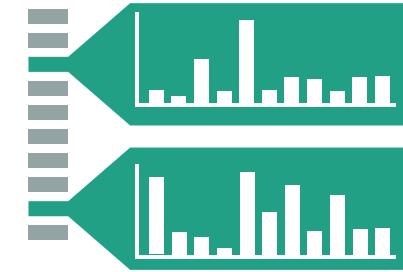
[DOI Trees Revisited: Scalable, Space-Constrained Visualization of Hierarchical Data. Heer and Card.  
Proc. Advanced Visual Interfaces (AVI), pp. 421–424, 2004.]

# Embed: Focus+Context

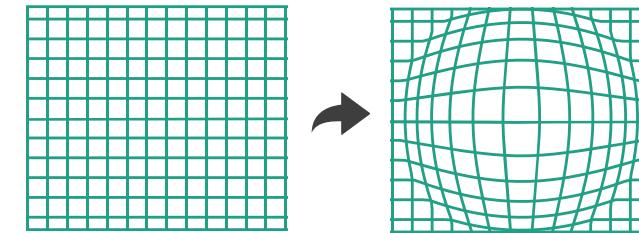
- combine focus + context info within single view
  - vs standard navigation within view
  - vs multiple views
- elide data
  - selectively filter and aggregate
- distort geometry
  - carefully chosen to integrate F+C

→ Embed

→ Elide Data

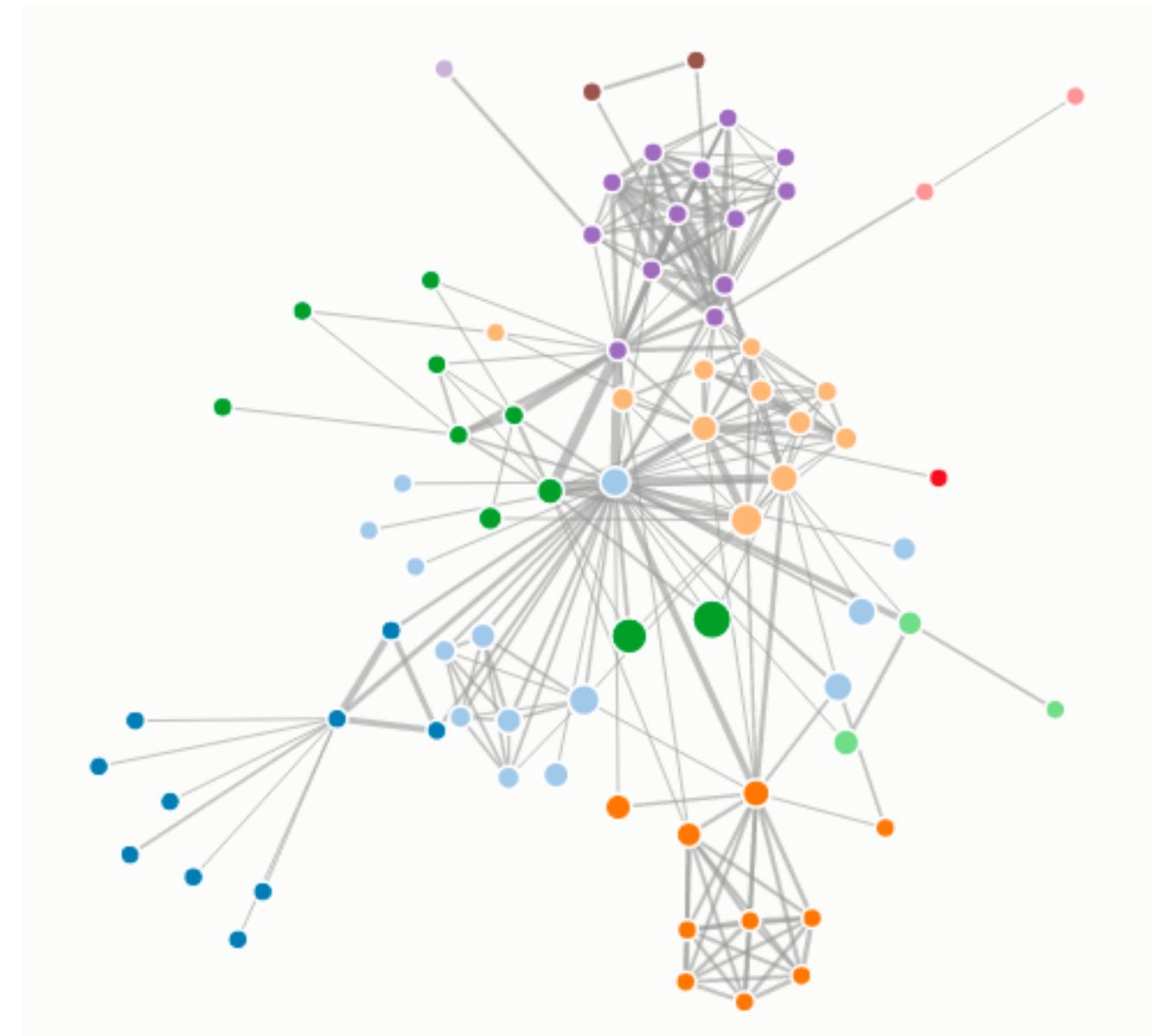


→ Distort Geometry



# Idiom: Fisheye Lens

- F+C choice: distort geometry
  - shape: radial
  - focus: single extent
  - extent: local
  - metaphor: draggable lens



[D3 Fisheye Lens] <https://bostocks.org/mike/fisheye/>

# Embed: Focus+Context

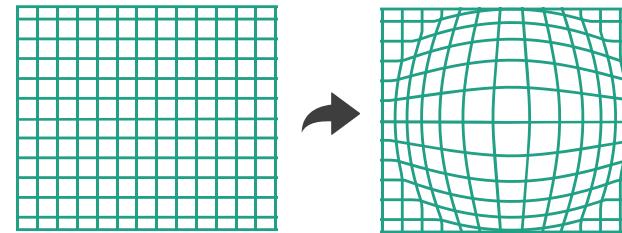
- combine focus + context info within single view
  - vs standard navigation within view
  - vs multiple views
- elide data
  - selectively filter and aggregate
- distort geometry:  
design choices
  - region shape: radial, rectilinear, complex
  - how many regions: one, many
  - region extent: local, global
  - interaction metaphor

→ Embed

→ Elide Data



→ Distort Geometry



# Distortion costs and benefits

- **benefits**
  - combine focus and context information in single view
- **costs**
  - length comparisons impaired
    - topology comparisons unaffected: connection, containment
  - effects of distortion unclear if original structure unfamiliar
  - object constancy/tracking may be impaired

