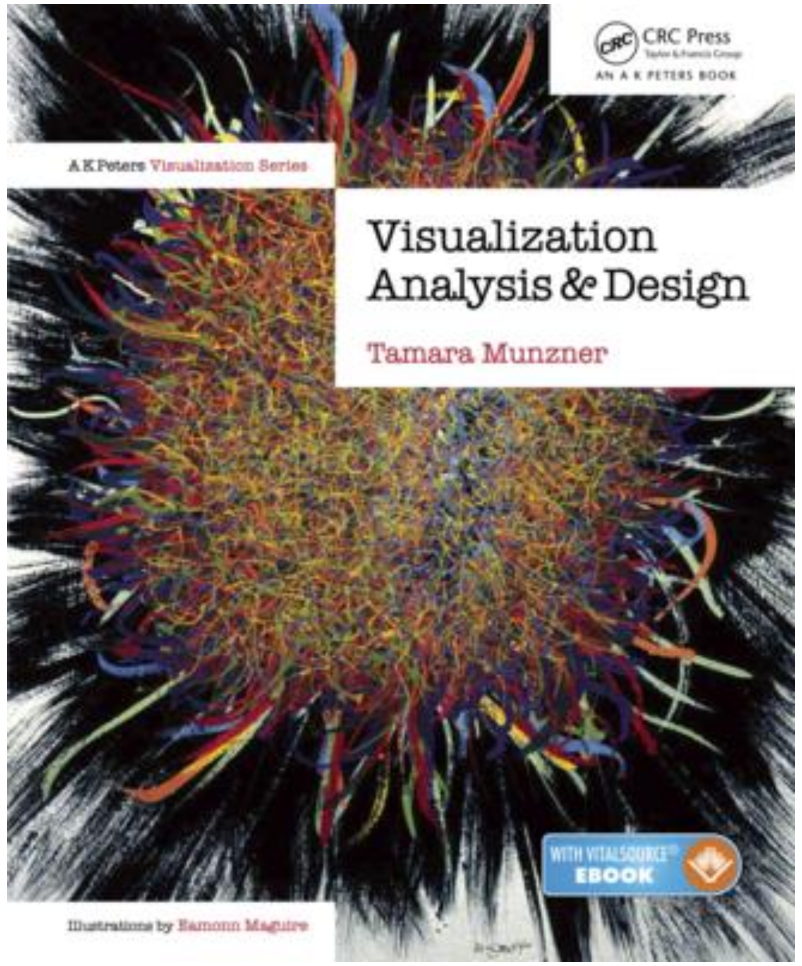
Reduce & Embed

Yu-Shuen Wang, CS, NCTU

- Slides refer to https://www.cs.ubc.ca/~tmm/



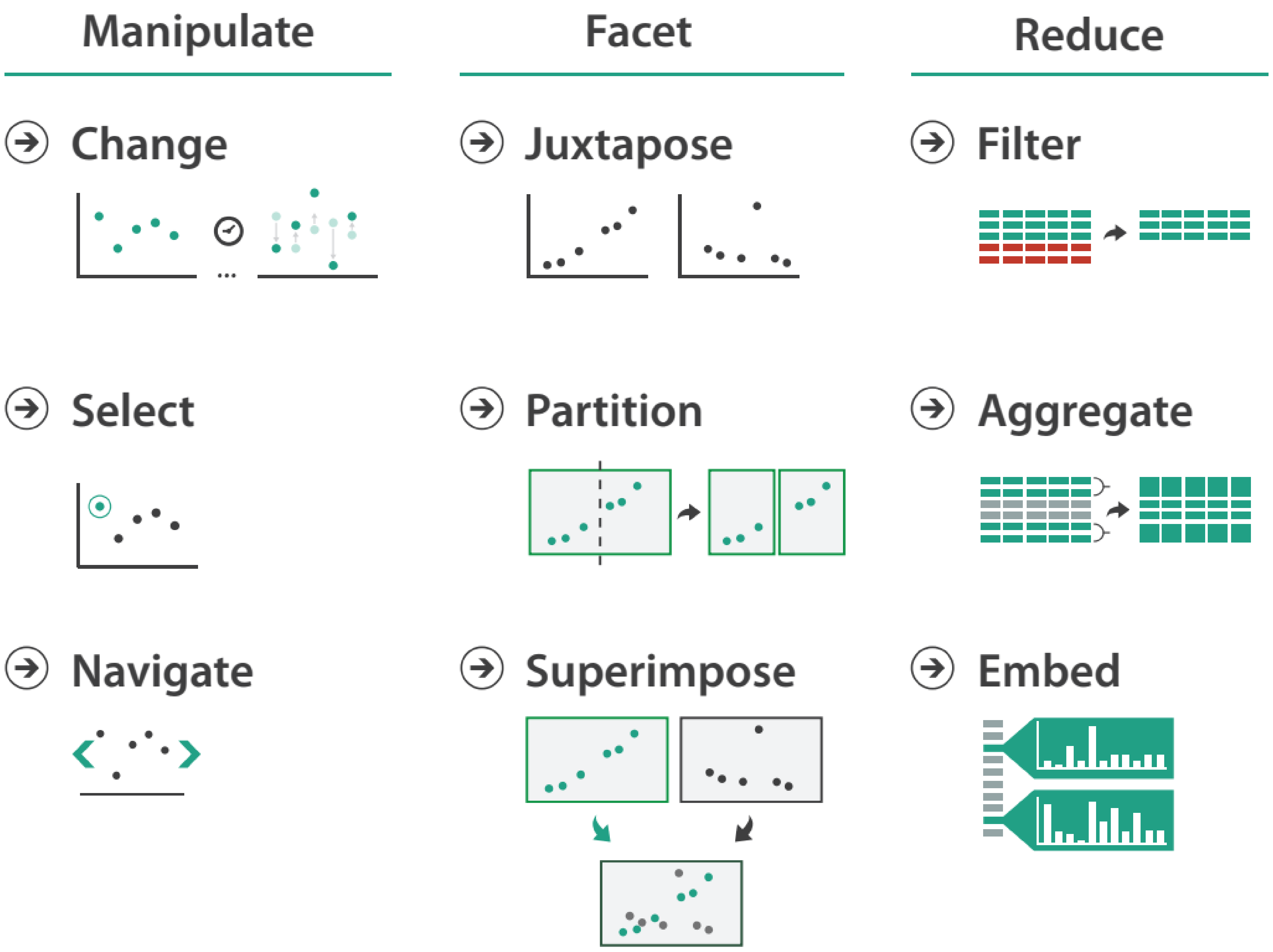
Guidelines

- Reduce: Filter, Aggregate (CH. 13)

- Embed: Focus + Context (CH. 14)

How to handle complexity: 1 previous strategy + 3 more

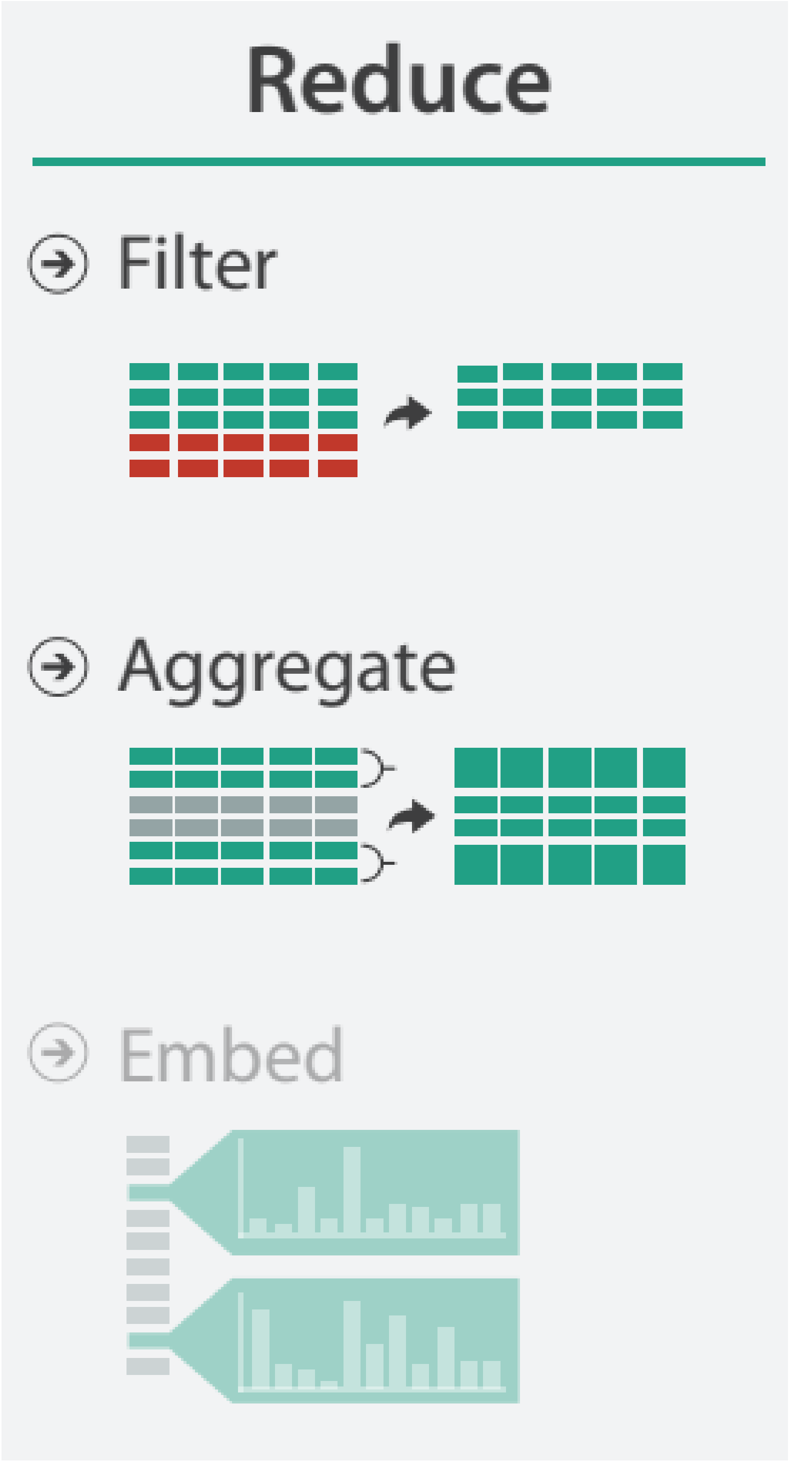
- derive new data to show within view
- change view over time
- facet across multiple views
- reduce items/attributes within single view
- embed focus and context



Reduce:

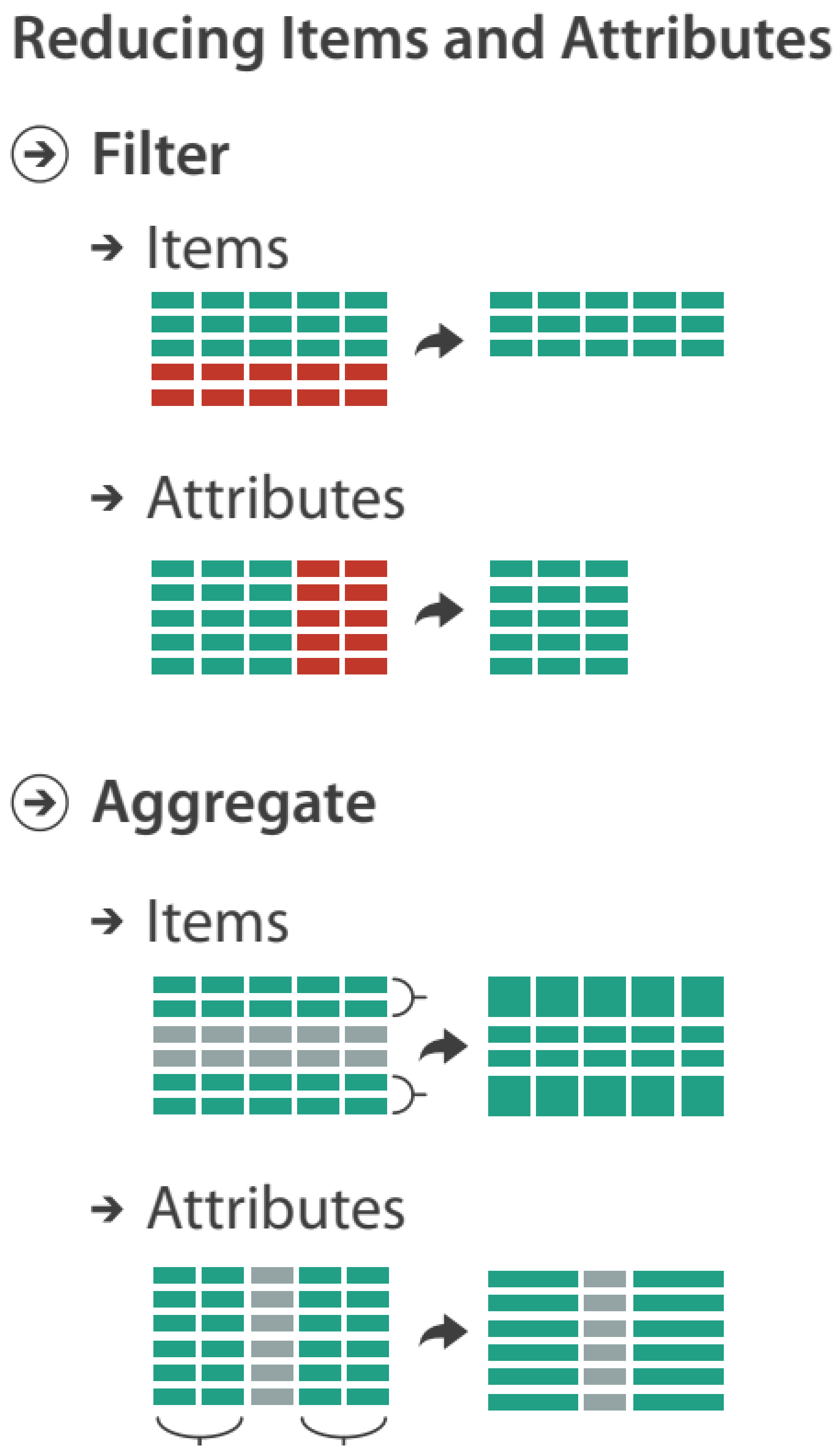
Filter,

Aggregate



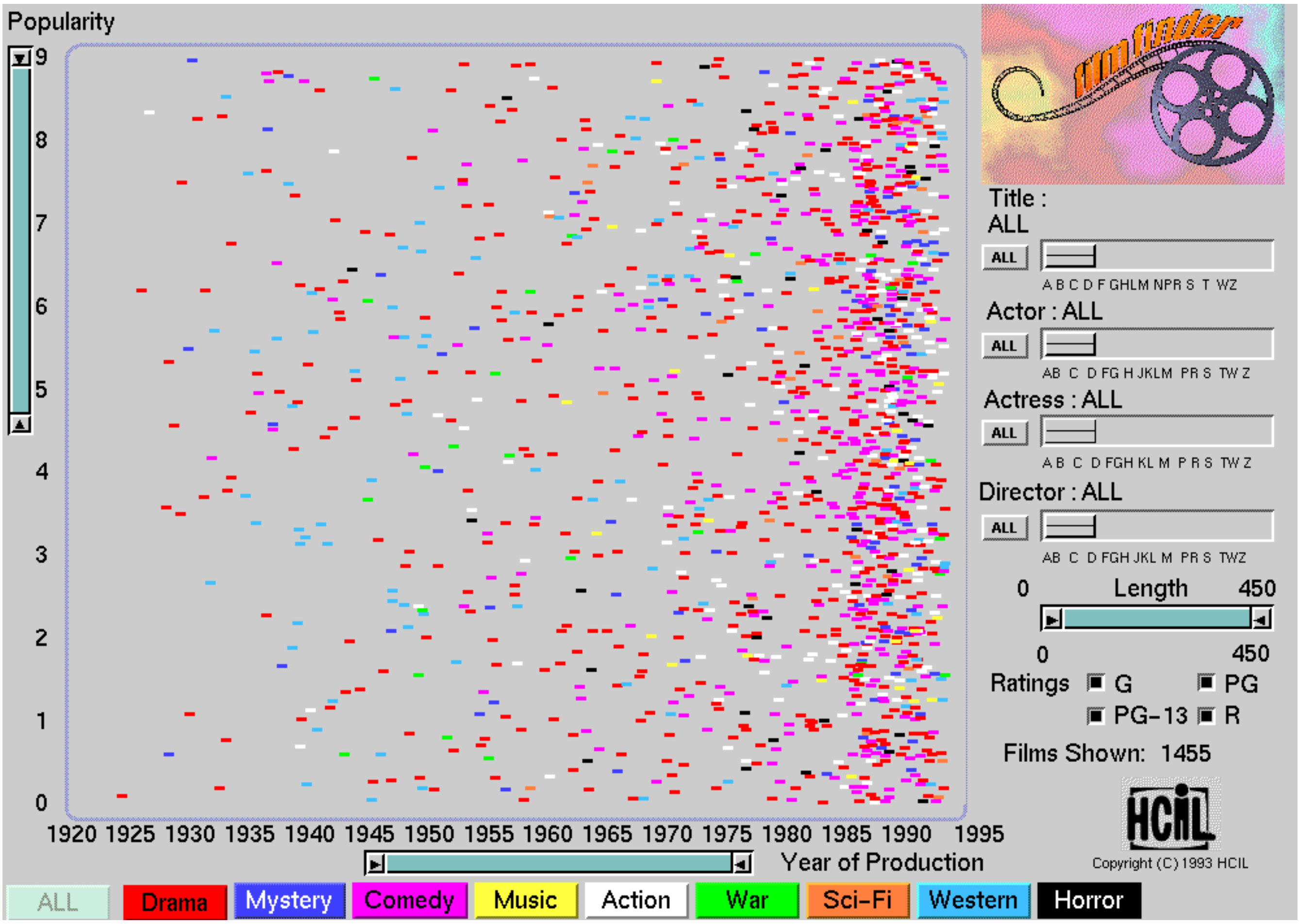
Reduce items and attributes

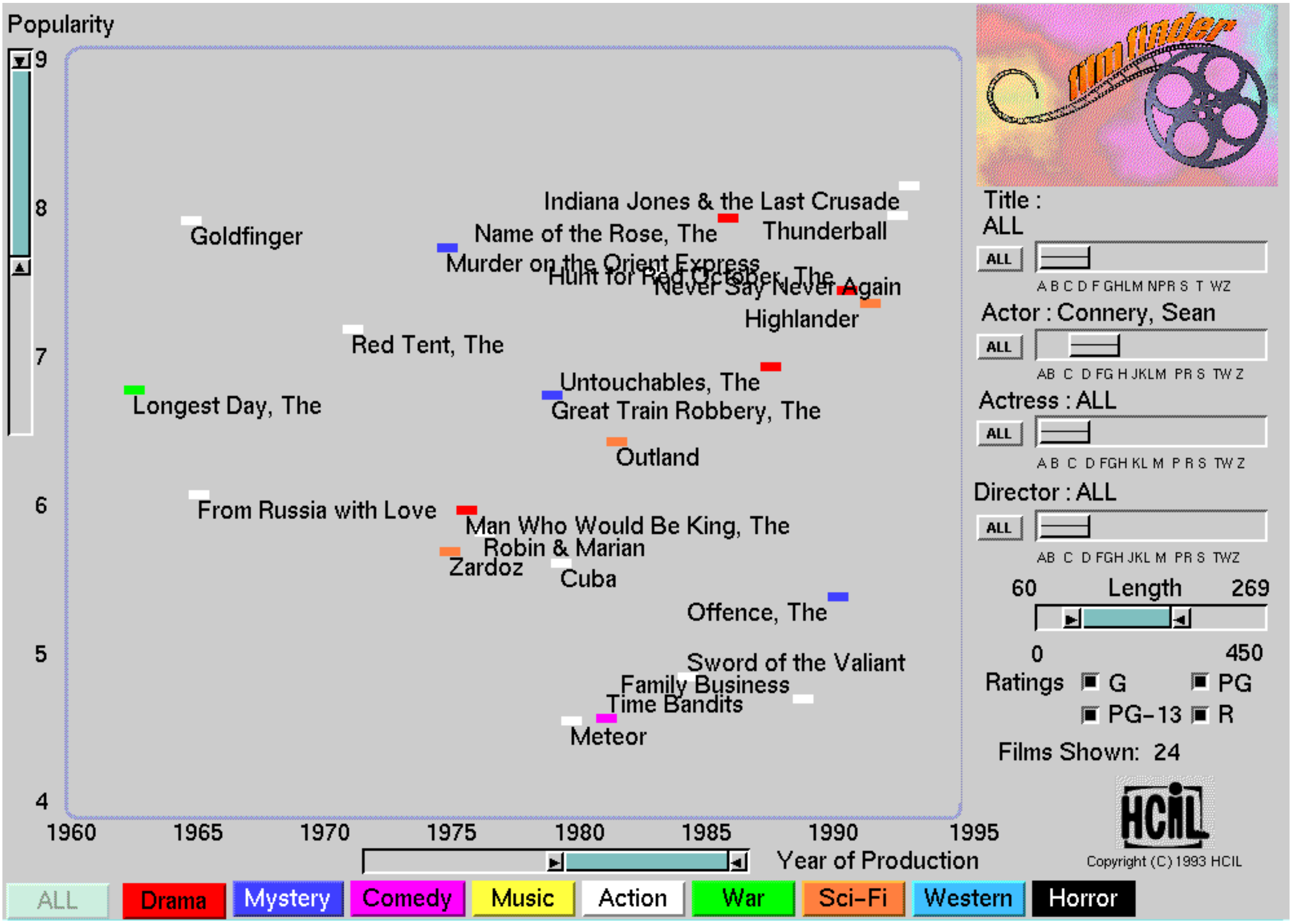
- reduce/increase: inverses
- filter
  - –pro: straightforward and intuitive
    - to understand and compute
  - –con: out of sight, out of mind
- aggregation
  - –pro: inform about whole set
  - –con: difficult to avoid losing signal
- not mutually exclusive
  - –combine filter, aggregate
  - –combine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



Idiom: **dynamic filtering** System: **FilmFinder**

- item filtering
- browse through tightly coupled interaction
  - –alternative to queries that might return far too many or too few

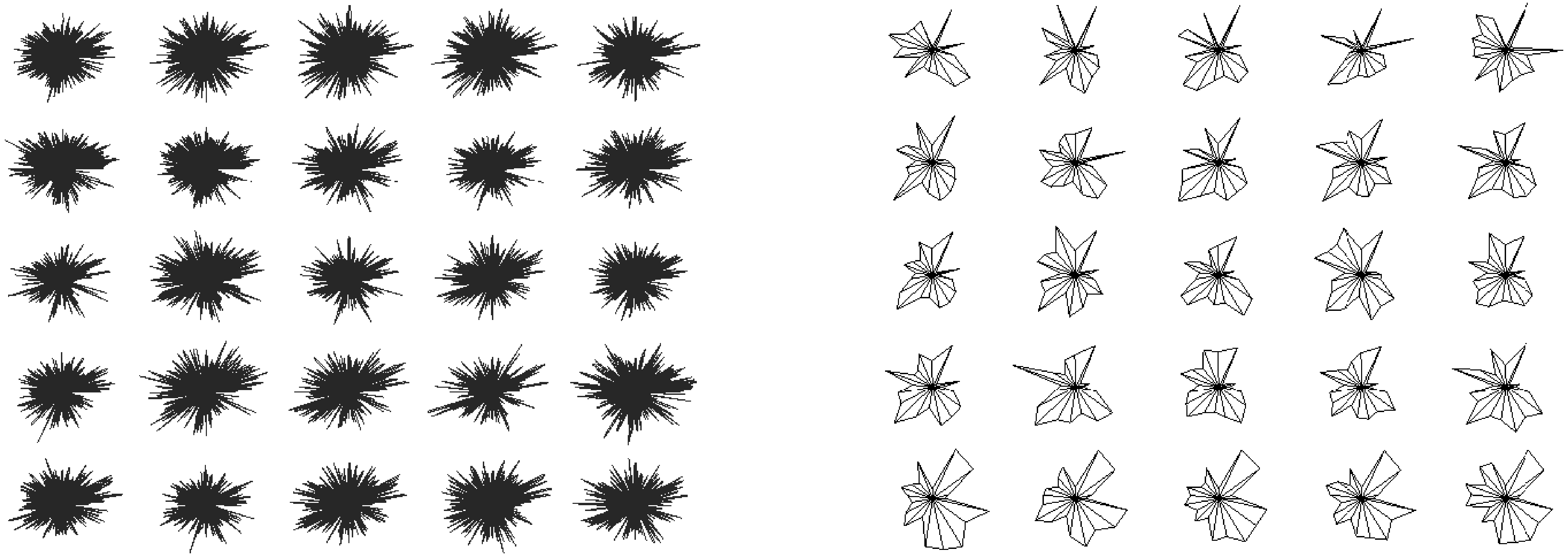




[Visual information seeking: Tight coupling of dynamic query filters with starfield displays. Ahlberg and Shneiderman. Proc. ACM Conf. on Human Factors in Computing Systems (CHI), pp. 313–317, 1994.]

Idiom: **DOSFA**

- attribute filtering
- encoding: star glyphs

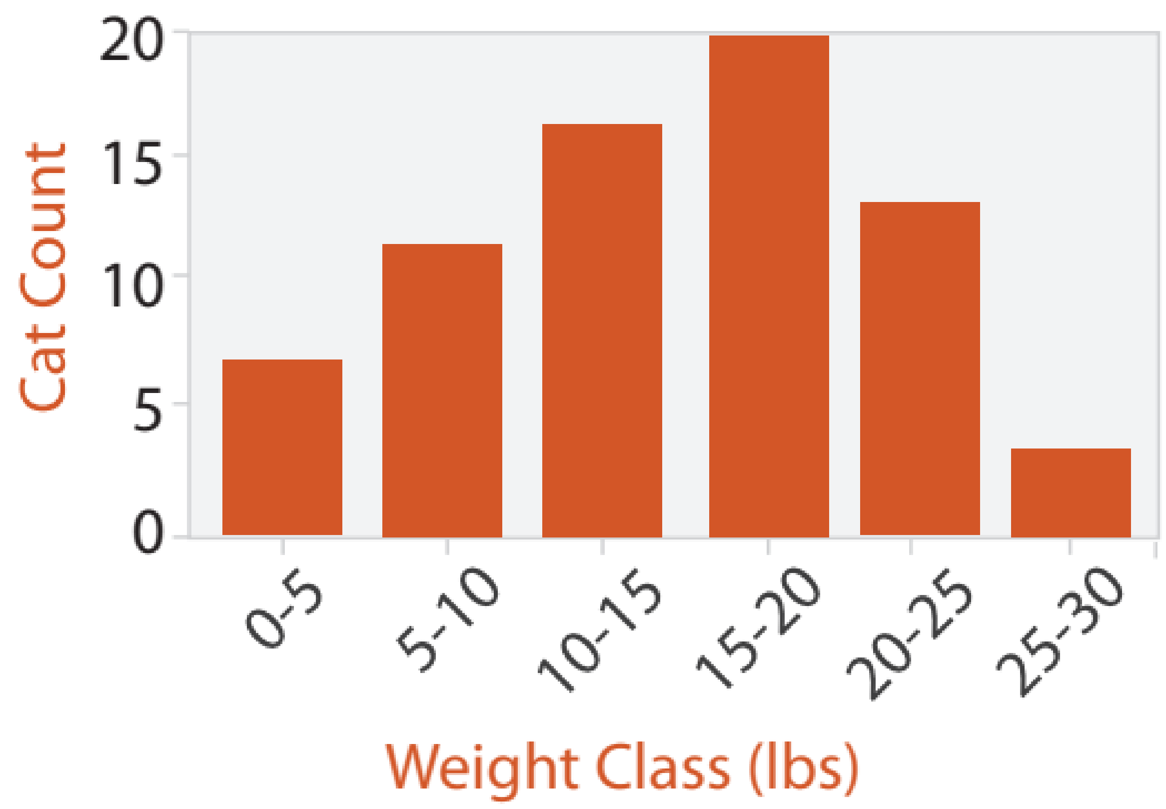


full dataset(298dim) after

[Interactive Hierarchical Dimension Ordering, Spacing and Filtering for Exploration Of High Dimensional Datasets. Yang, Peng,Ward, and. Rundensteiner. Proc. IEEE Symp. Information Visualization (InfoVis), pp. 105–112, 2003.]

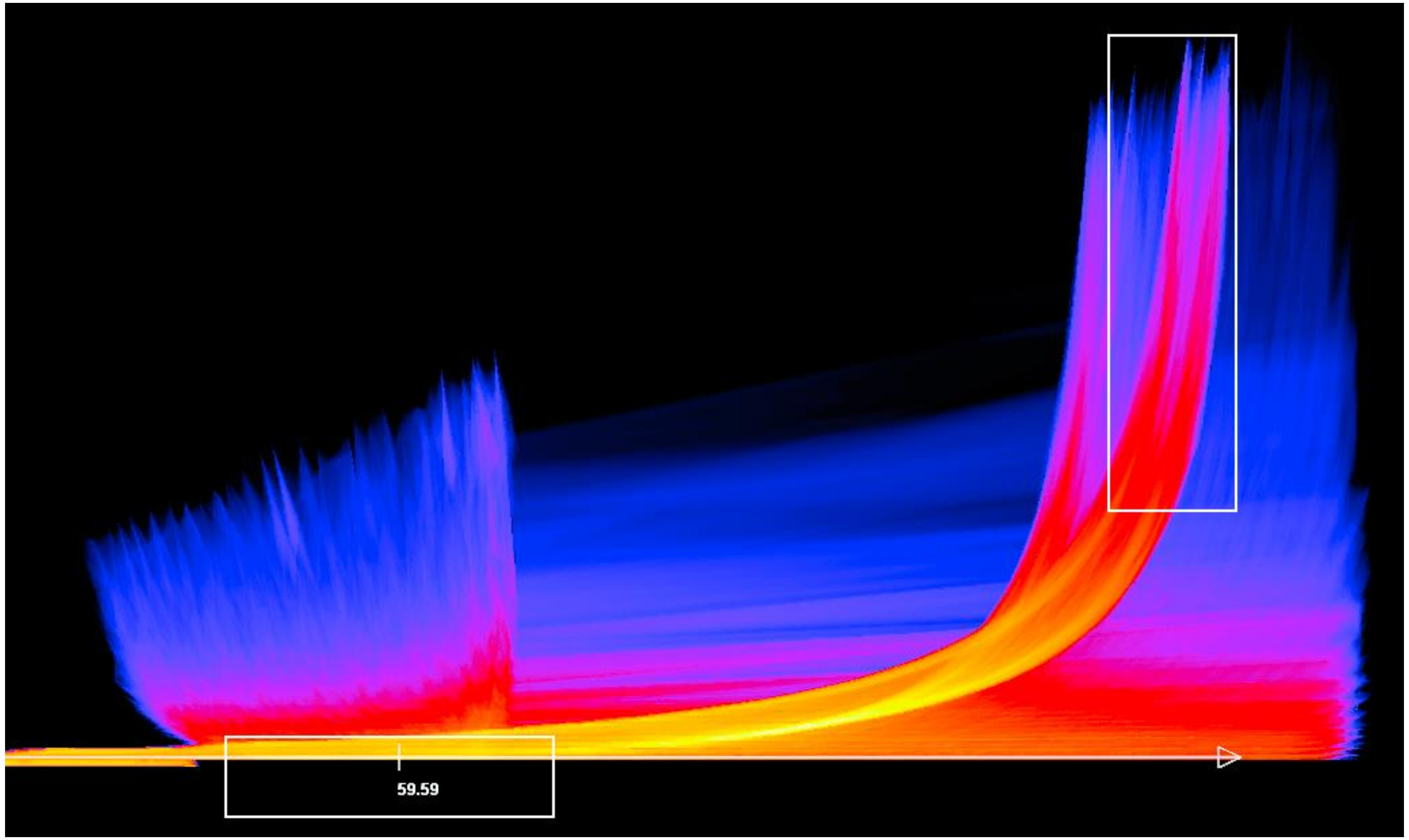
Idiom: **histogram**

- 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



Continuous scatterplot

- static item aggregation
- data: table
- derived data: table
  - key attribs x,y for pixels
  - quant attrib: overplot density
- dense space-filling 2D matrix
- color: sequential categorical hue + ordered luminance colormap



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

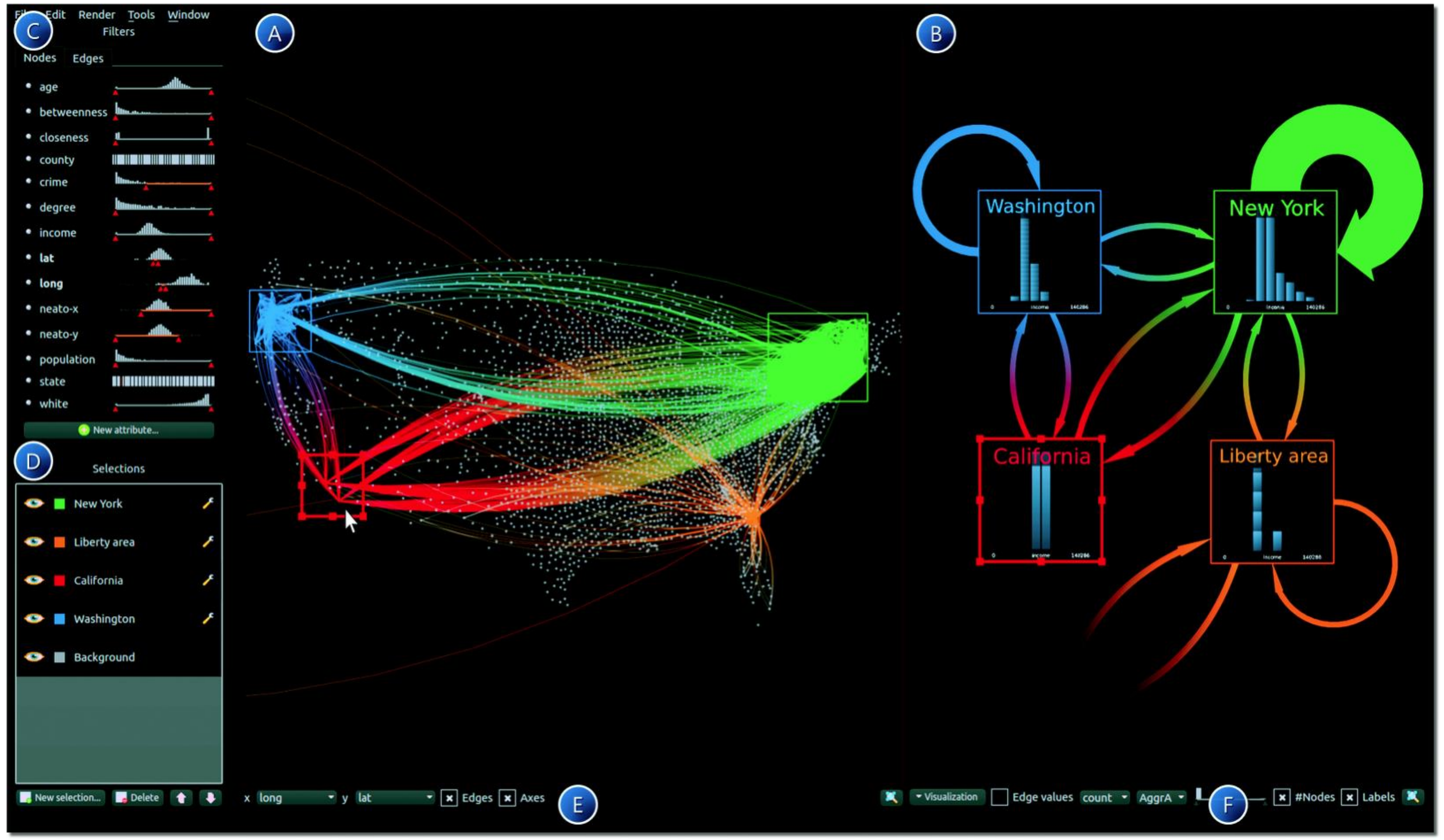
Idiom: **scented widgets**

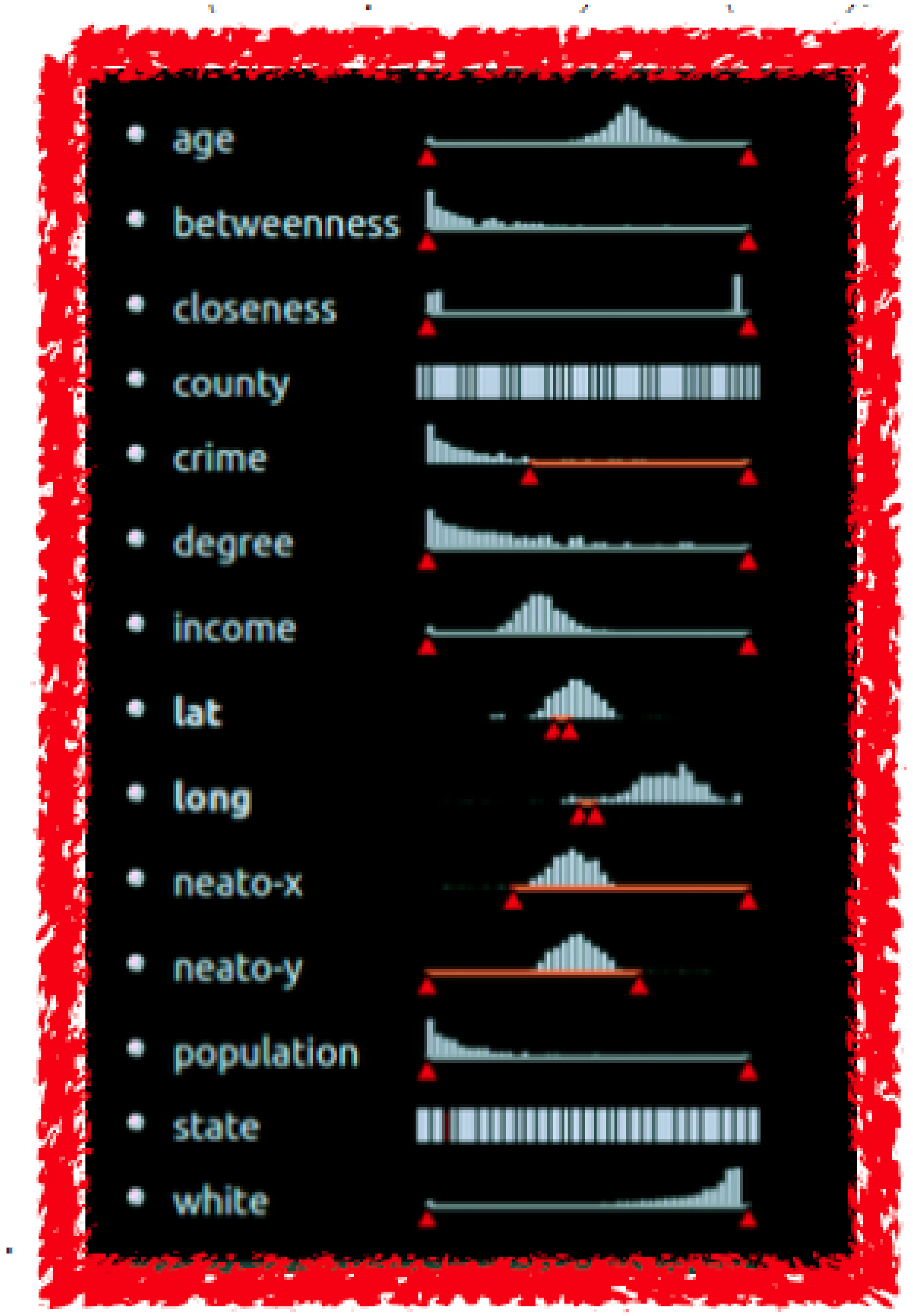
- augment widgets for filtering to show ***information scent***
  - –cues to show whether value in drilling down further vs looking elsewhere
- concise, in part of screen normally considered control panel

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

Idiom: **scented widgets**

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

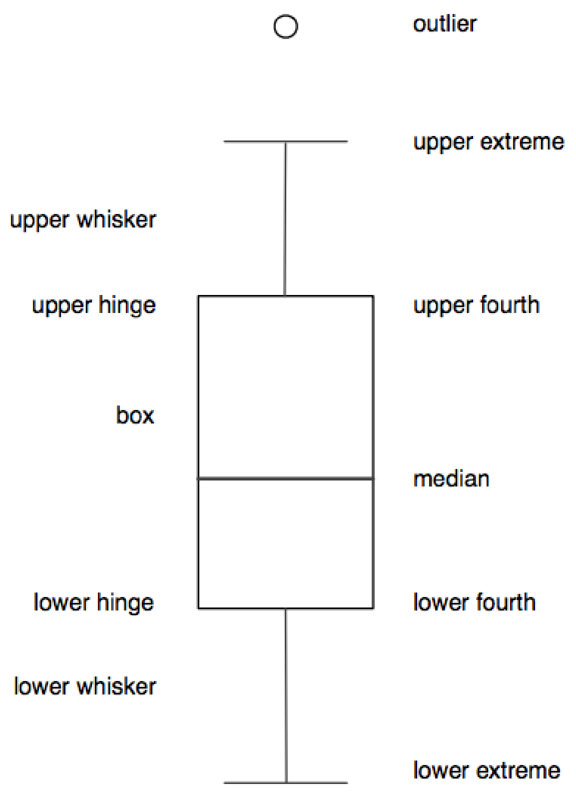


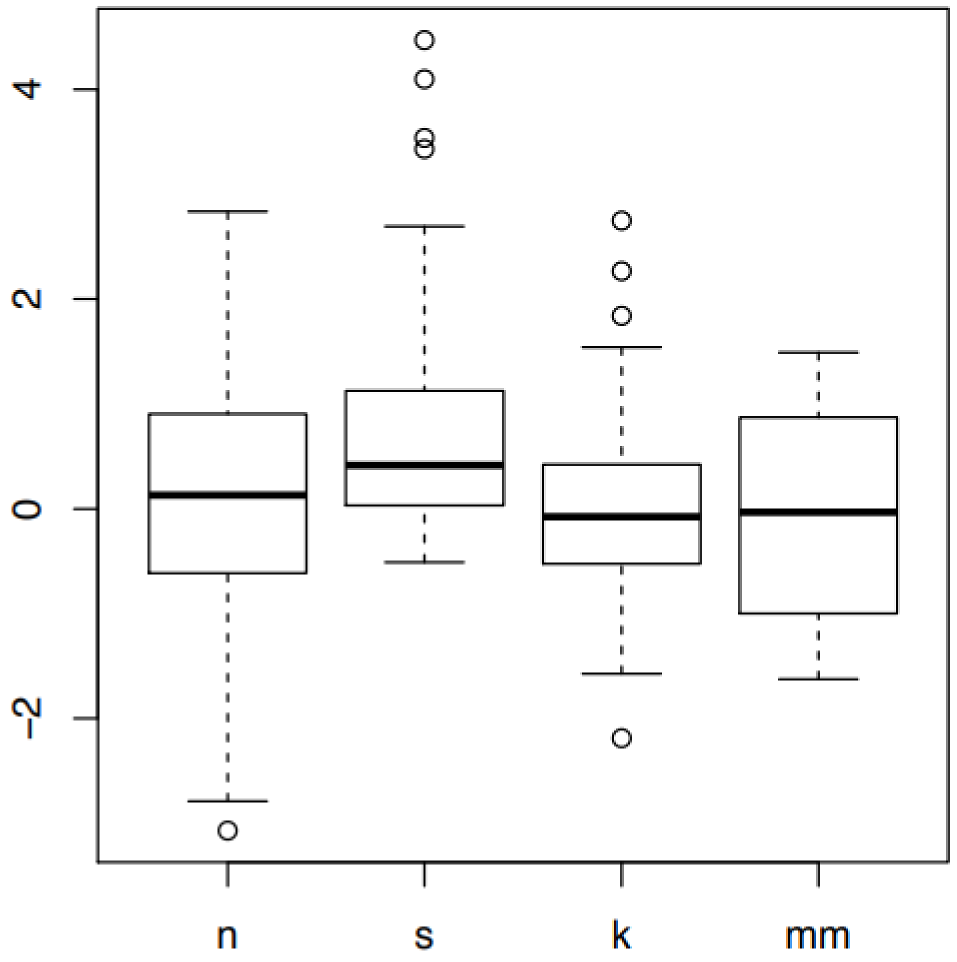


[Multivariate Network Exploration and Presentation: From Detail to Overview via Selections and Aggregations. van den Elzen and van Wijk, TVCG 20(12) 2014.] [video]

Idiom: **boxplot**

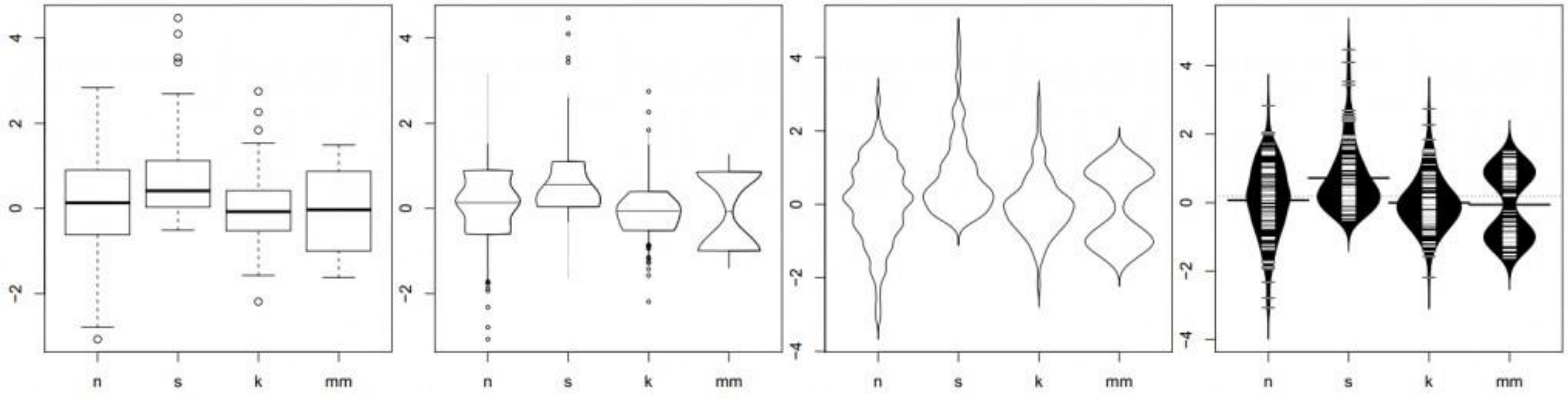
- static item aggregation
- task: find distribution
- data: table
- derived data
  - 5 quant attribs
    - median: central line
    - lower and upper quartile: boxes
    - lower upper fences: whiskers
      - values beyond which items are outliers
  - outliers beyond fence cutoffs explicitly shown





[40 years of boxplots. Wickham and Stryjewski. 2012. had.co.nz]

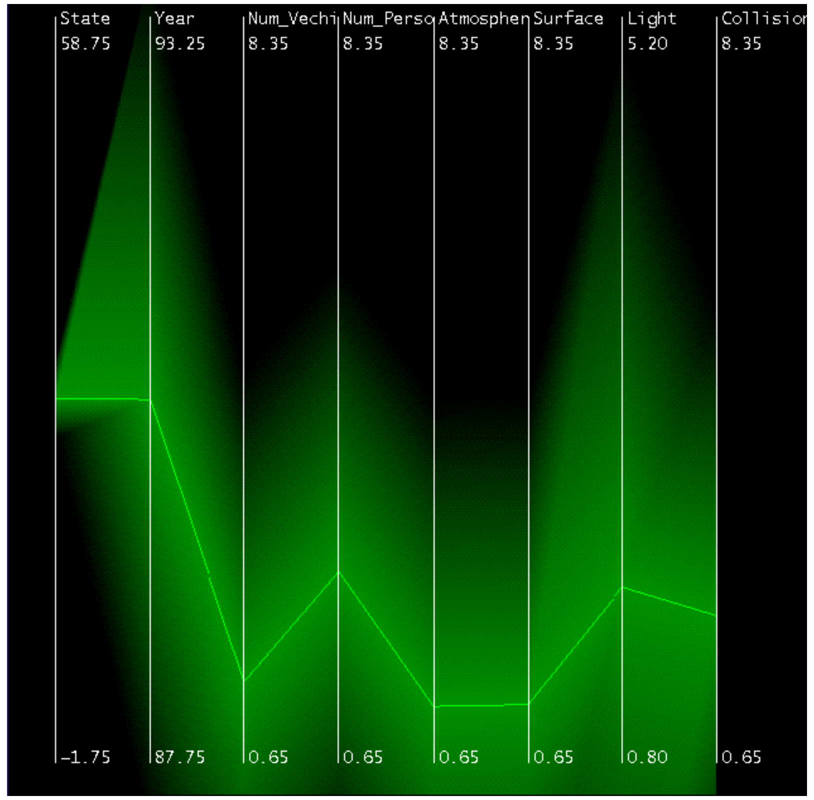
Idiom: **variants of boxplot**

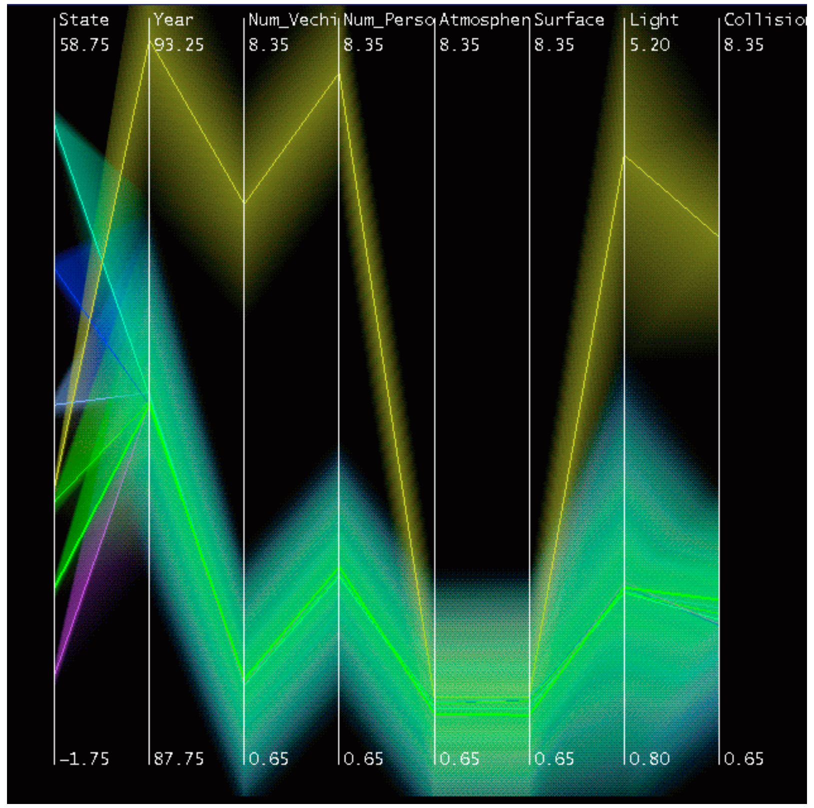


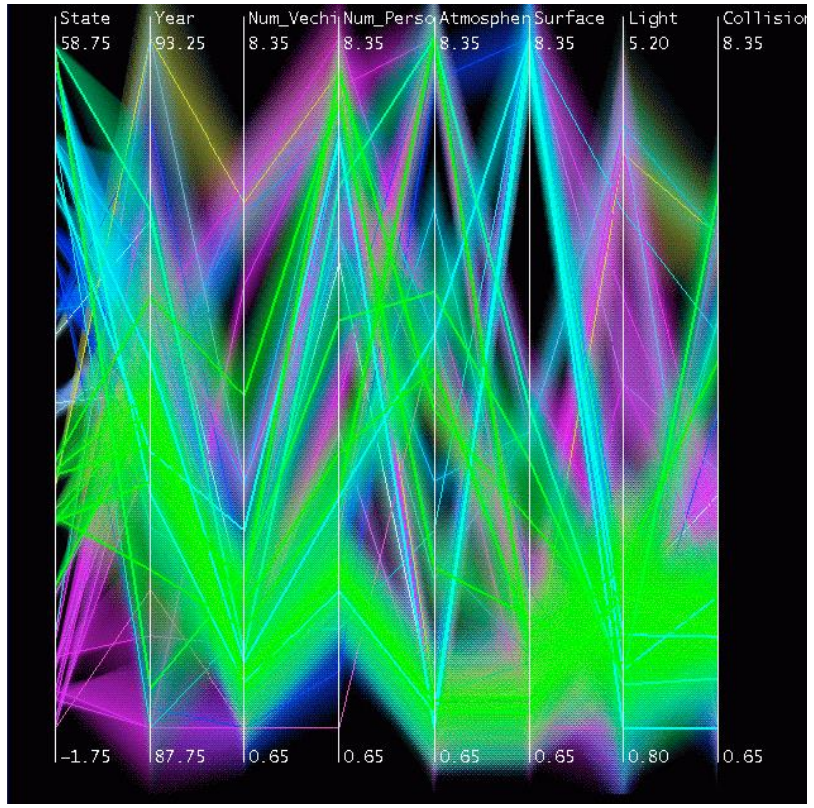
box plot, vase plot, violin plot, bean plot

Idiom: **Hierarchical parallel coordinates**

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



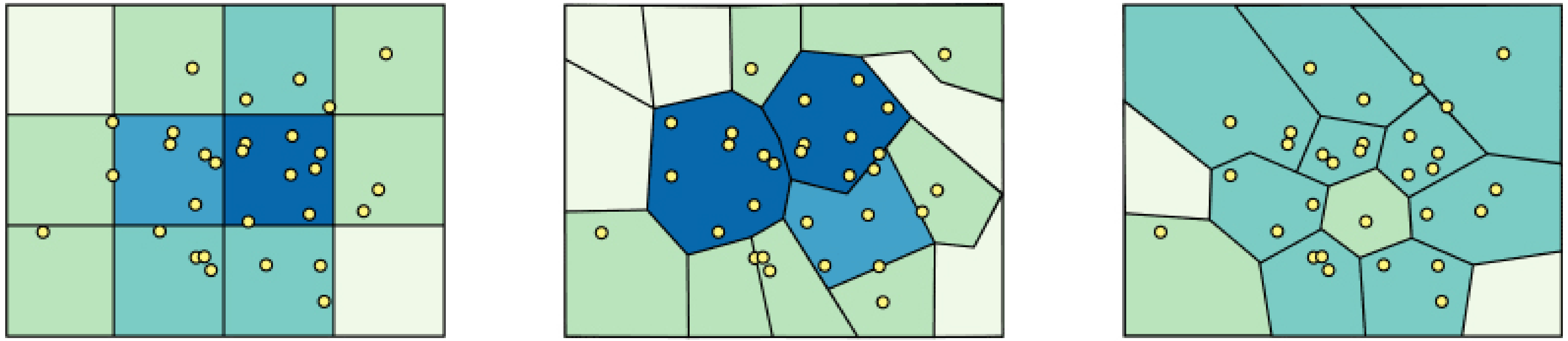




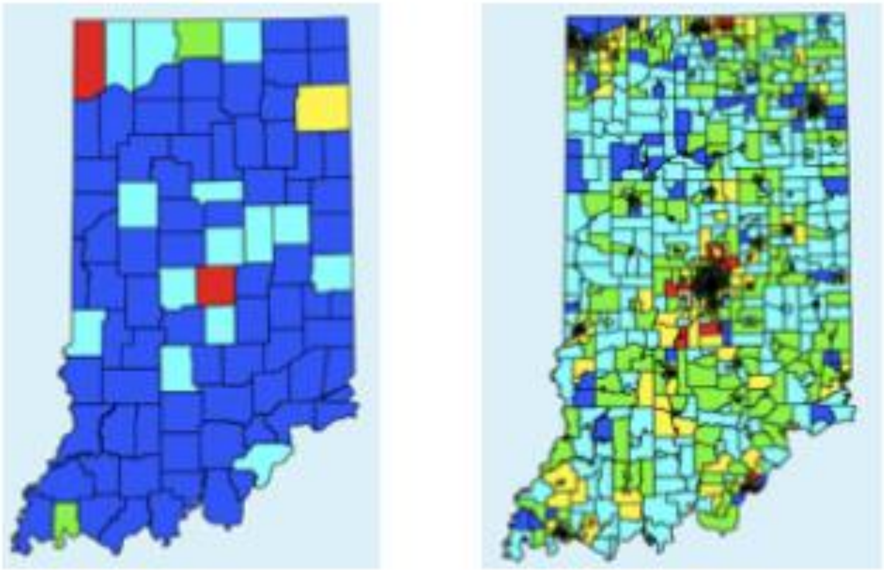
[Hierarchical Parallel Coordinates for Exploration of Large Datasets. Fua, Ward, and Rundensteiner. Proc. IEEE Visualization Conference (Vis '99), pp. 43– 50, 1999.]

Spatial aggregation

- **MAUP**: Modifiable Areal Unit Problem
  - gerrymandering (manipulating voting district boundaries) is one example!
  - Zone effects



  - Scale effects



Spatial aggregation

– Zone effects

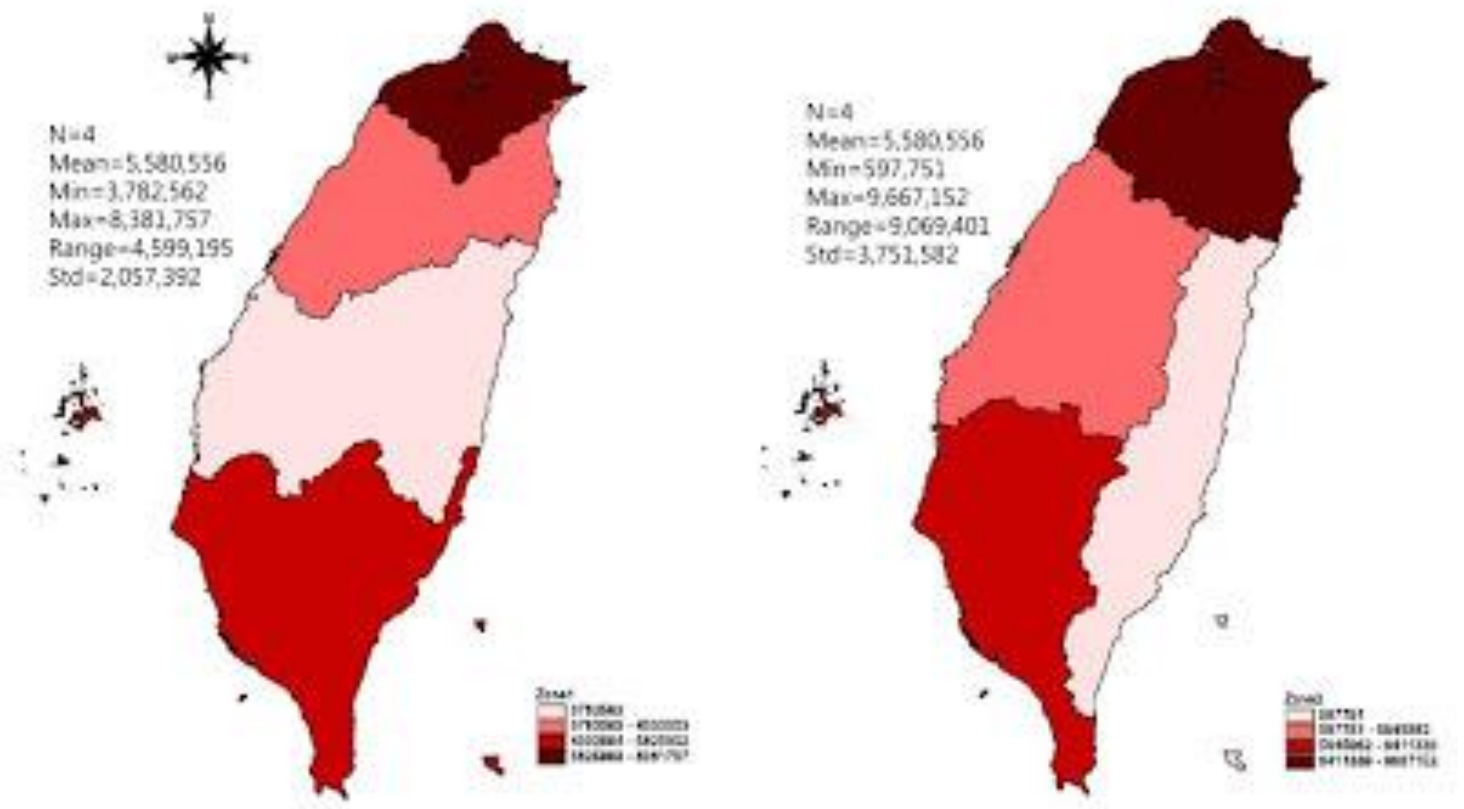


圖4、區域1之人口分布圖 圖5、區域2之人口分布圖

– Scale effects

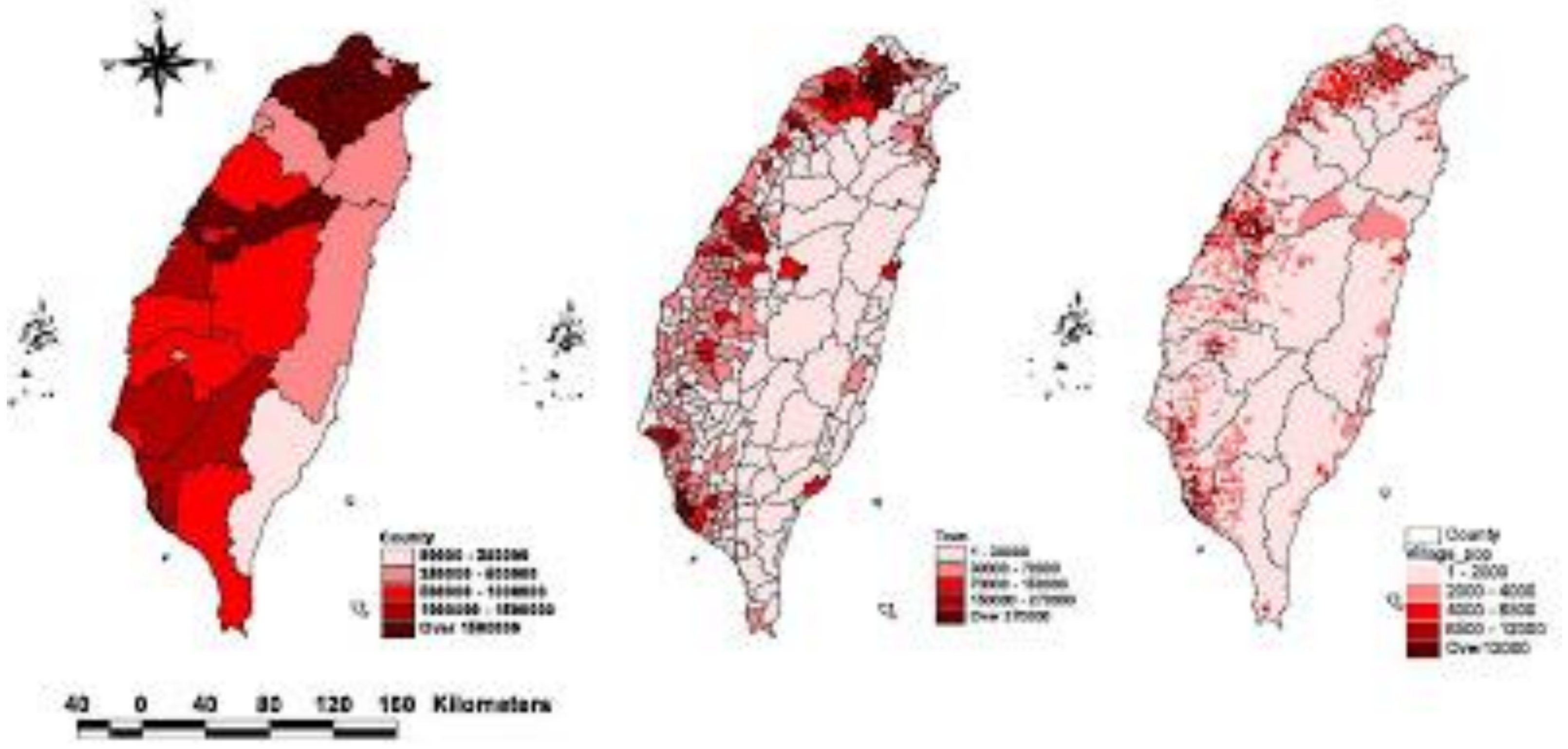
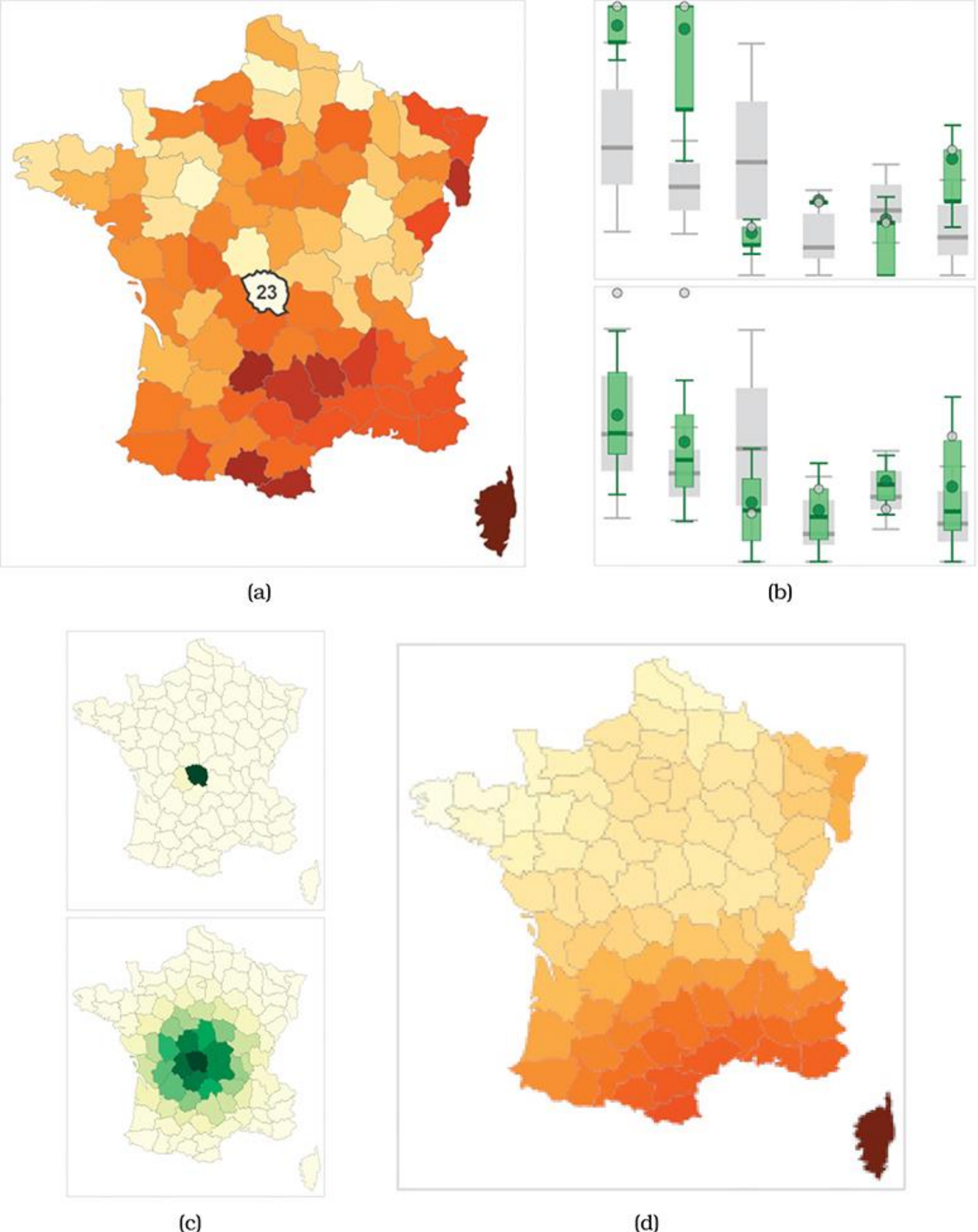


圖1、縣市人口分布圖 圖2、鄉鎮人口分布圖 圖3、村里人口分布圖

可調整地區單元問題(MAUP)

http://utopia1234.blogspot.tw/2009/11/maup.html



A multivariate geographic dataset used to explore social issues in 19th century France.

The six quantitative attributes are:

- population per crime against persons
- population per crime against property
- percentage who can read and write
- donations to the poor
- population per illegitimate birth
- population per suicide.

Geowigs are geographically weighted interactive graphics. (a) A choropleth map showing attribute x1. (b) The set of gw-boxplots for all six attributes at two scales. (c) Weighting maps showing the scales: local and larger. (d) A gw-mean map at the larger scale.

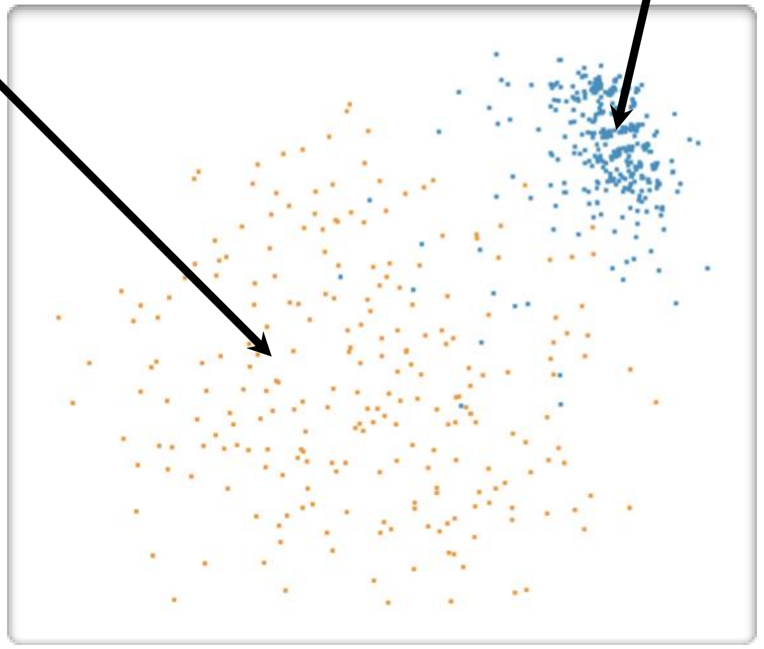
Dimensionality reduction

- attribute aggregation
  - –derive low-dimensional target space from high-dimensional measured space
  - –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

Malignant Benign

Tumor
Measurement Data

⟶ DR ⟶



data: 9D measured space

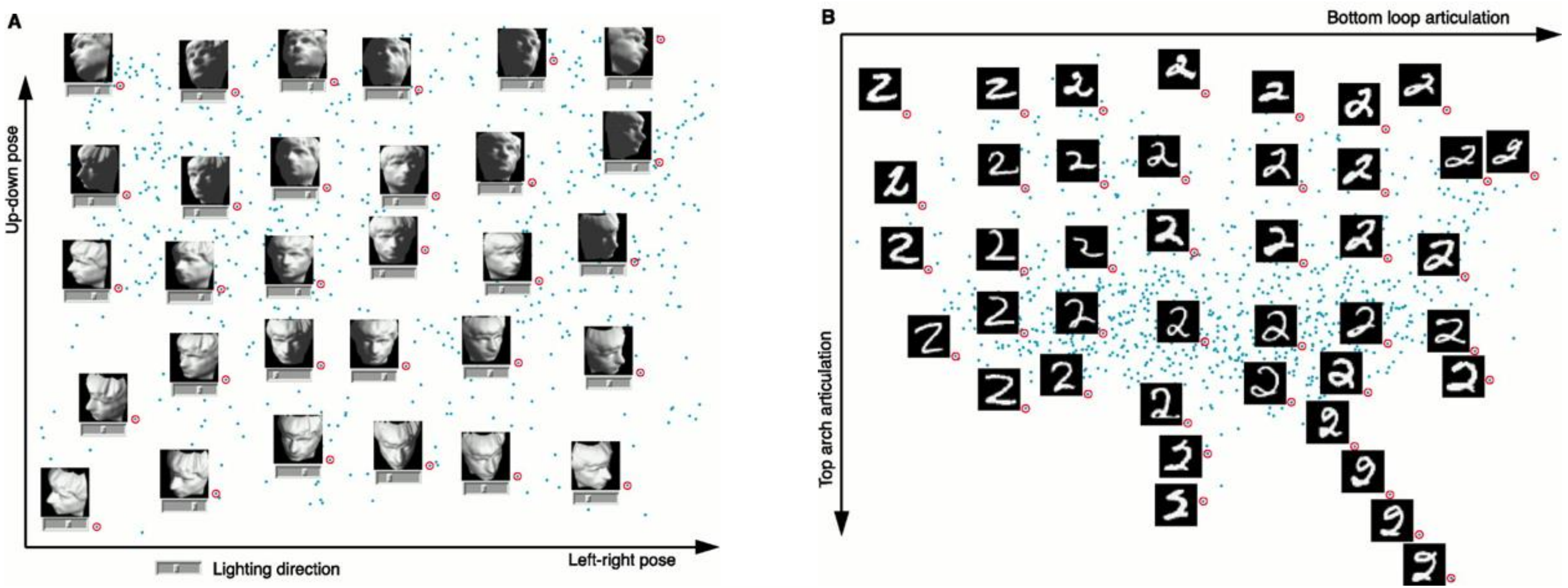
derived data: 2D target space

Dimensionality reduction & visualization

- why do people do DR?
  - improve performance of downstream algorithm
- avoid curse of dimensionality
  - data analysis
- if look at the output: visual data analysis
- abstract tasks when visualizing DR data
  - dimension-oriented tasks
- naming synthesized dims, mapping synthesized dims to original dims
  - cluster-oriented tasks
- verifying clusters, naming clusters, matching clusters and classes

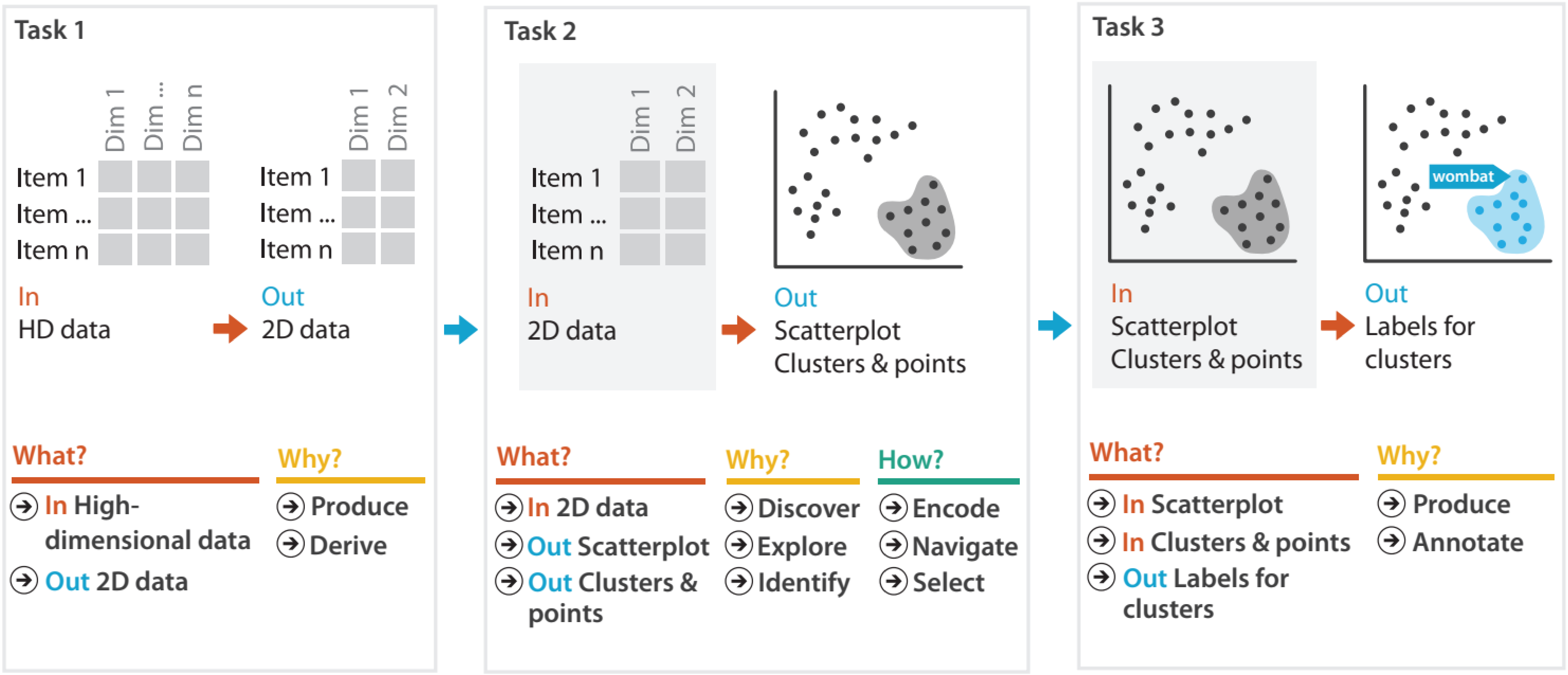
Dimension-oriented tasks

- naming synthesized dims: inspect data represented by lowD points



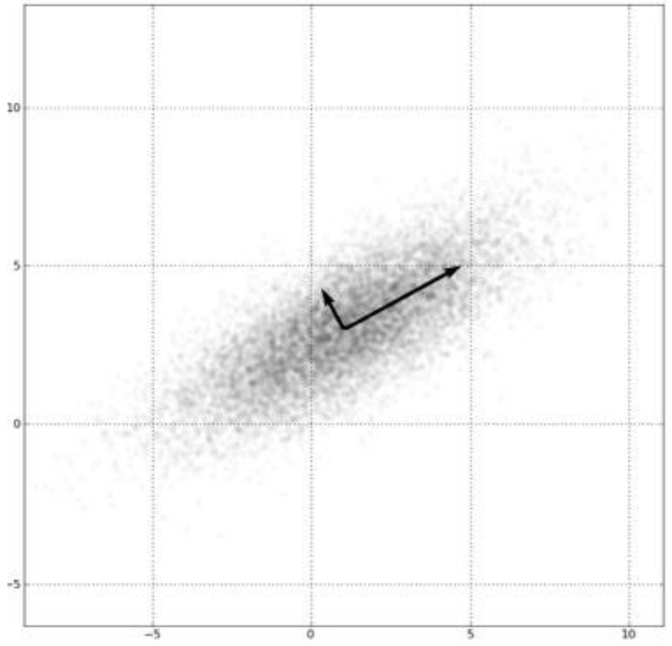
[A global geometric framework for nonlinear dimensionality reduction. Tenenbaum, de Silva, and Langford. Science, 290(5500):2319–2323, 2000.]

Idiom: **Dimensionality reduction for documents**



Linear dimensionality reduction

- principal components analysis (PCA)
  - finding axes: first with most variance, second with next most, …
  - describe location of each point as linear combination of weights for each axis
- mapping synthesized dims to original dims

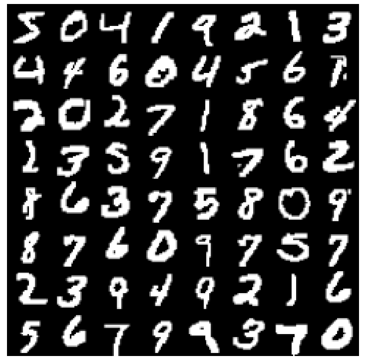


[http://en.wikipedia.org/wiki/File:GaussianScatterPCA.png]

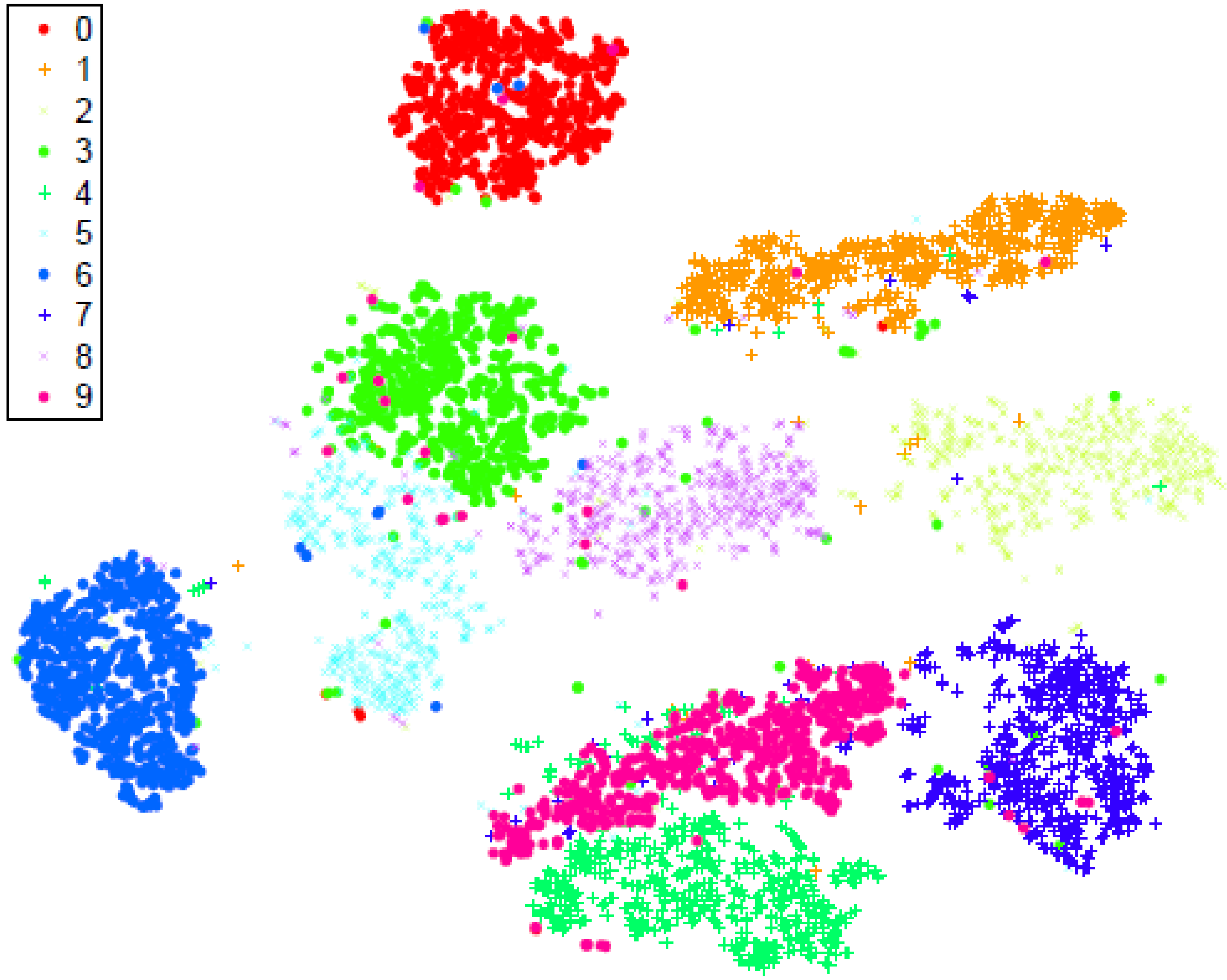
Nonlinear dimensionality reduction

- pro: can handle curved rather than linear structure
- cons: lose all ties to original dims/attribs
  - new dimensions often cannot be easily related to originals
  - mapping synthesized dims to original dims task is difficult
- many techniques proposed
  - many literatures: visualization, machine learning, optimization, psychology, ...
  - techniques: t-SNE, MDS (multidimensional scaling), charting, isomap, LLE,…
    - t-SNE: excellent for clusters
      - but some trickiness remains: http://distill.pub/2016/misread-tsne/
    - MDS: confusingly, entire family of techniques, both linear and nonlinear
      - minimize stress or strain metrics
      - early formulations equivalent to PCA

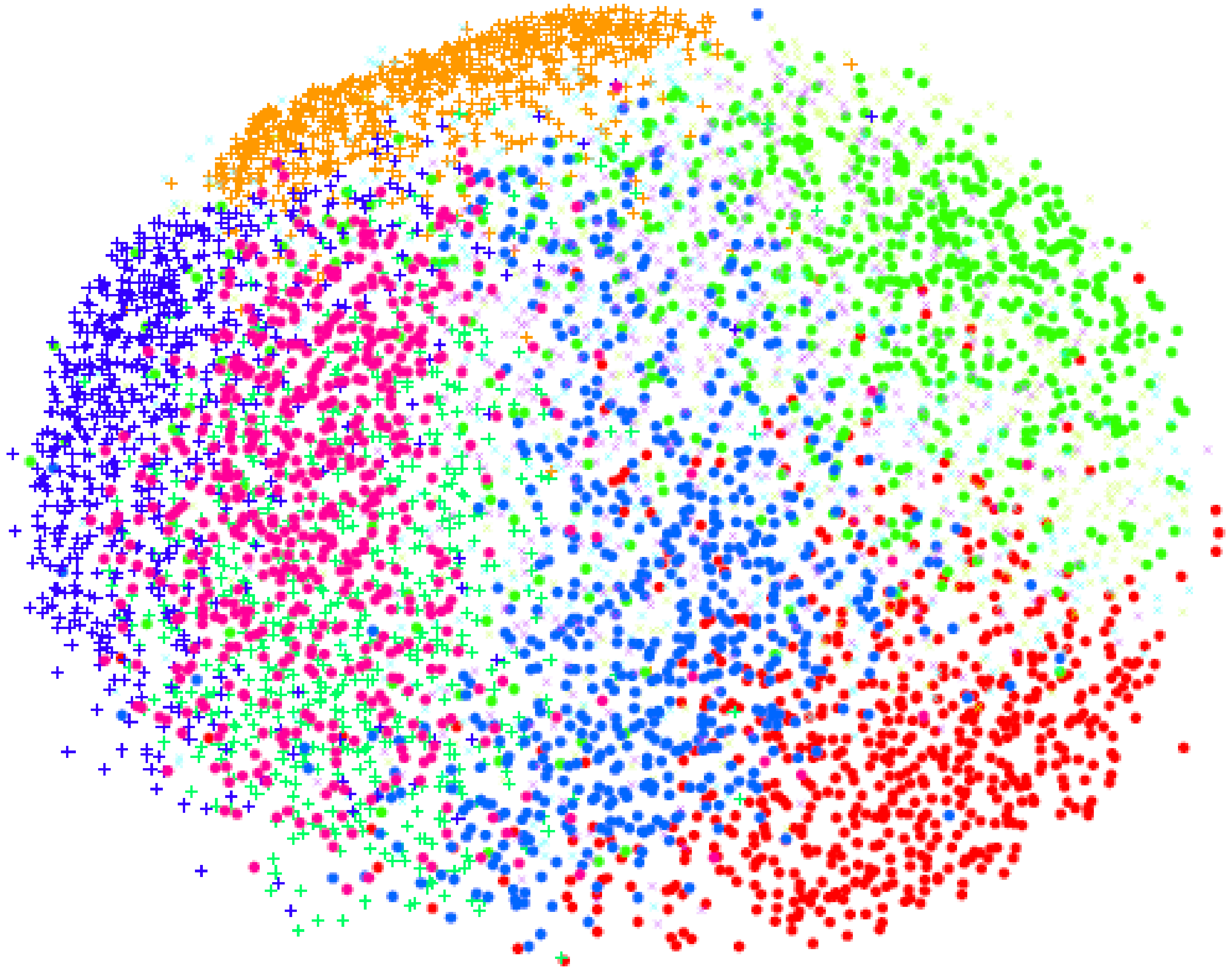
t-distributed stochastic neighbor embedding (t-sne)



MNIST data set



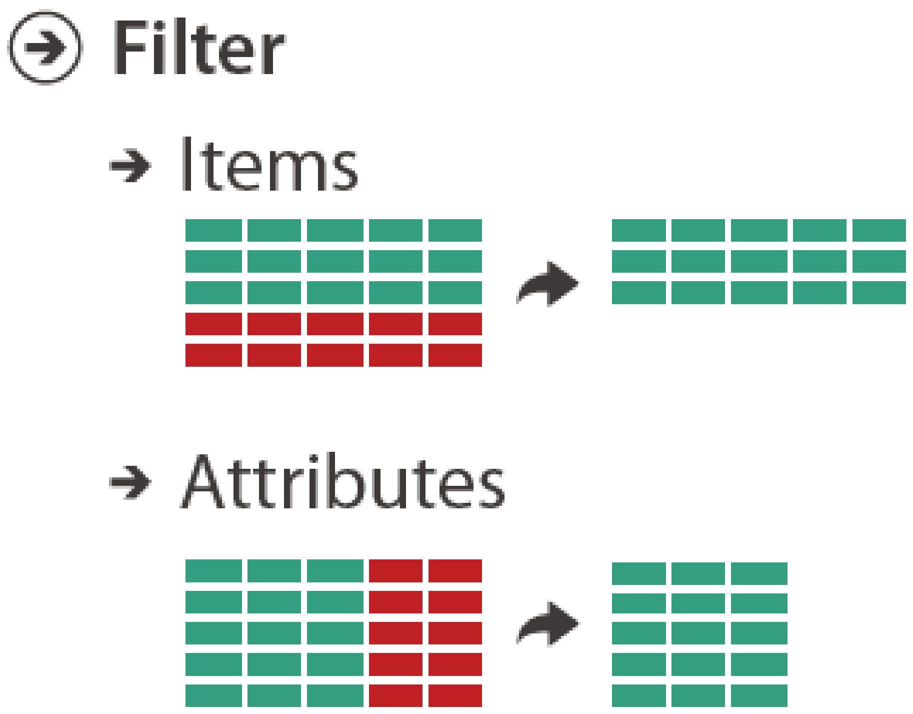
(a) Visualization by t-SNE.

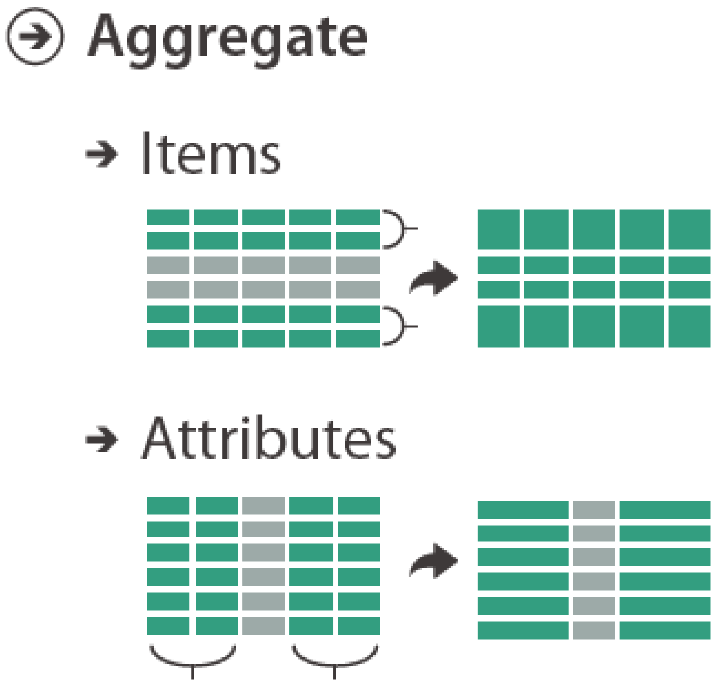


(b) Visualization by Sammon mapping.

t-sne process (video)

Reduce items and attributes





Further reading

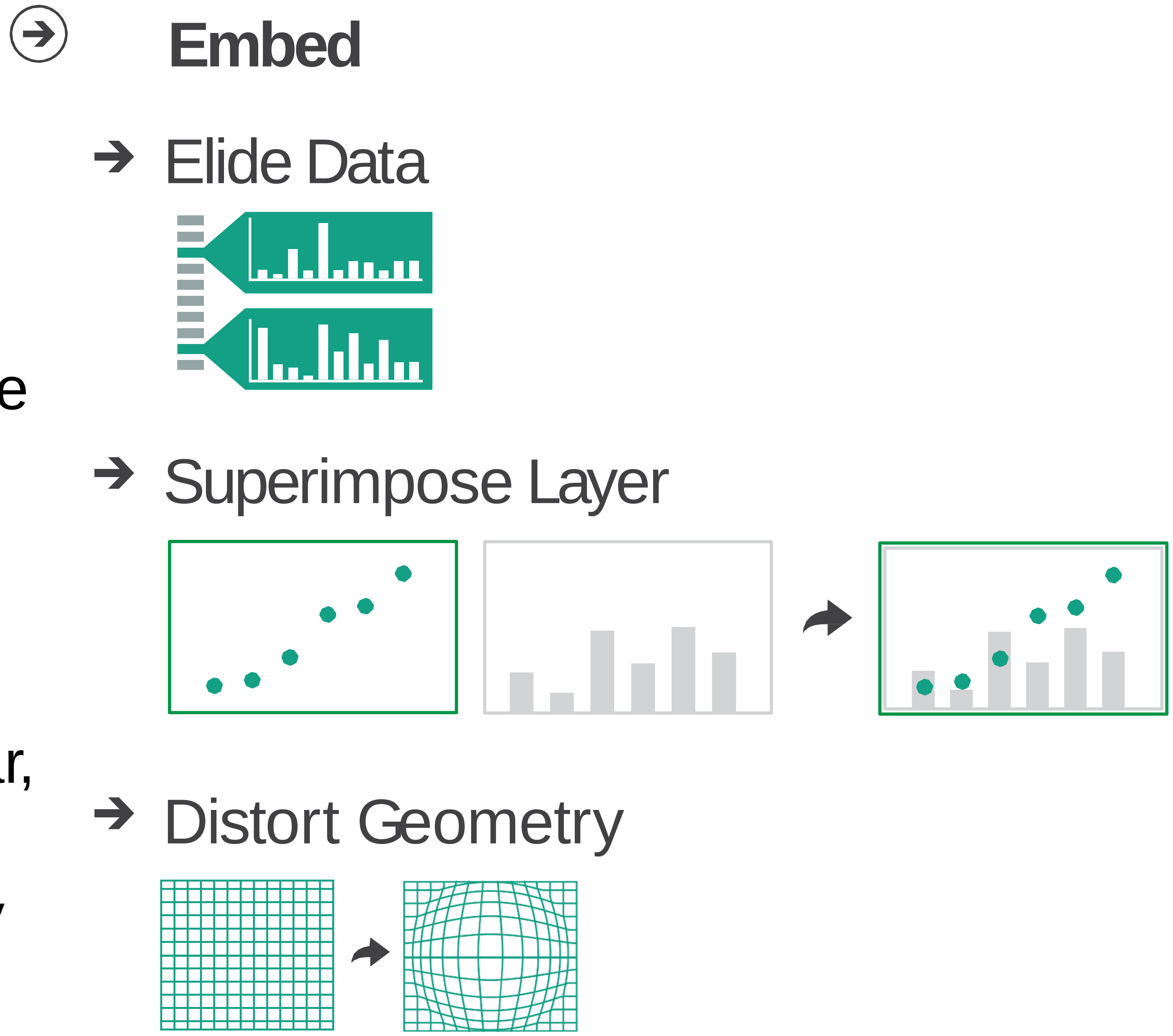
- Visualization Analysis and Design. Munzner. AK Peters Visualization Series, CRC Press, 2014.
  - *Chap 13: Reduce Items and Attributes*

- *Hierarchical Aggregation for Information Visualization: Overview, Techniques and Design Guidelines.* Elmqvist and Fekete. IEEE Transactions on Visualization and Computer Graphics 16:3 (2010), 439–454.
- *A Review of Overview+Detail, Zooming, and Focus+Context Interfaces.* Cockburn, Karlson, and Bederson. ACM Computing Surveys 41:1 (2008), 1–31.
- *A Guide to Visual Multi-Level Interface Design From Synthesis of Empirical Study Evidence.* Lam and Munzner. Synthesis Lectures on Visualization Series, Morgan Claypool, 2010.

Embed

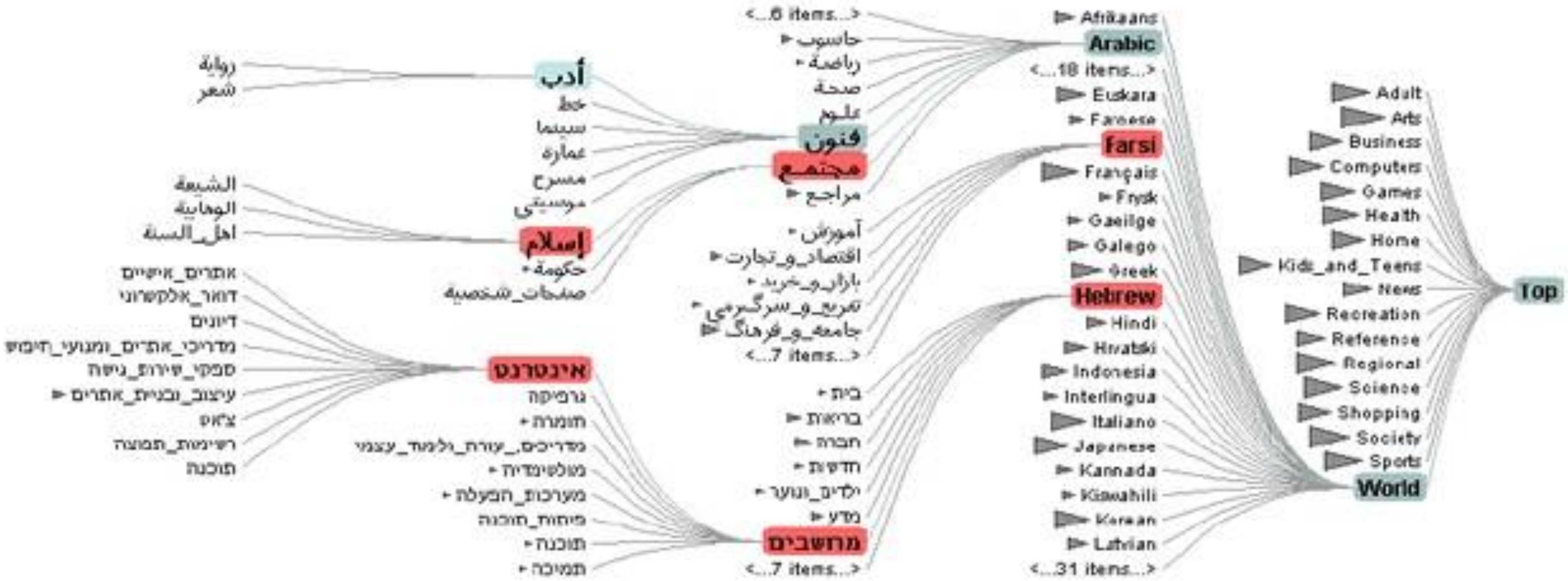
Embed: Focus+Context

- combine information within single view
- elide
  - selectively filter and aggregate
- superimpose layer
  - local lens
- distortion design choices
  - region shape: radial, rectilinear, complex
  - how many regions: one, many
  - region extent: local, global
  - interaction metaphor



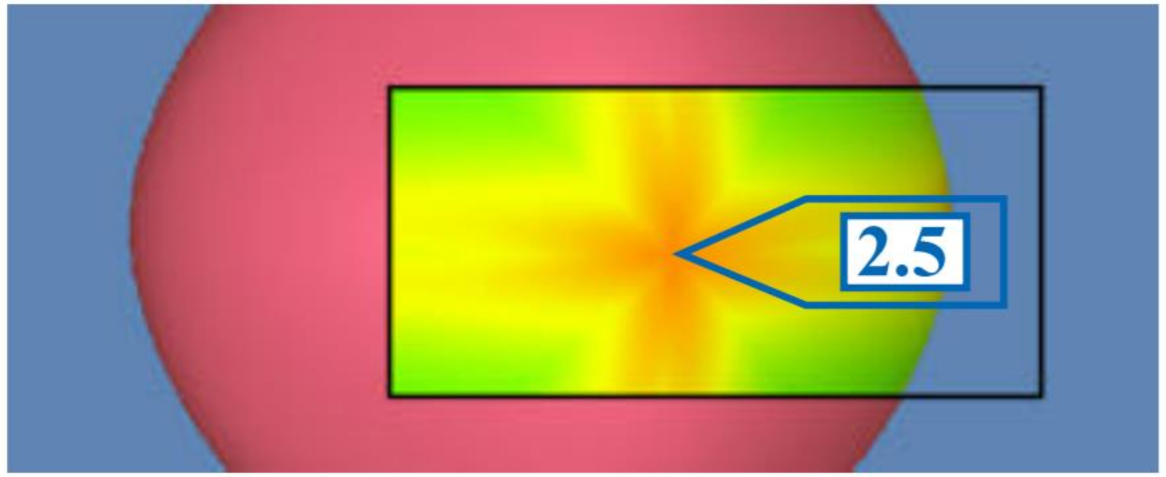
Idiom: **DOITrees Revisited**

- elide
  - some items dynamically filtered out
  - some items dynamically aggregated together
  - some items shown in detail



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

Superimpose

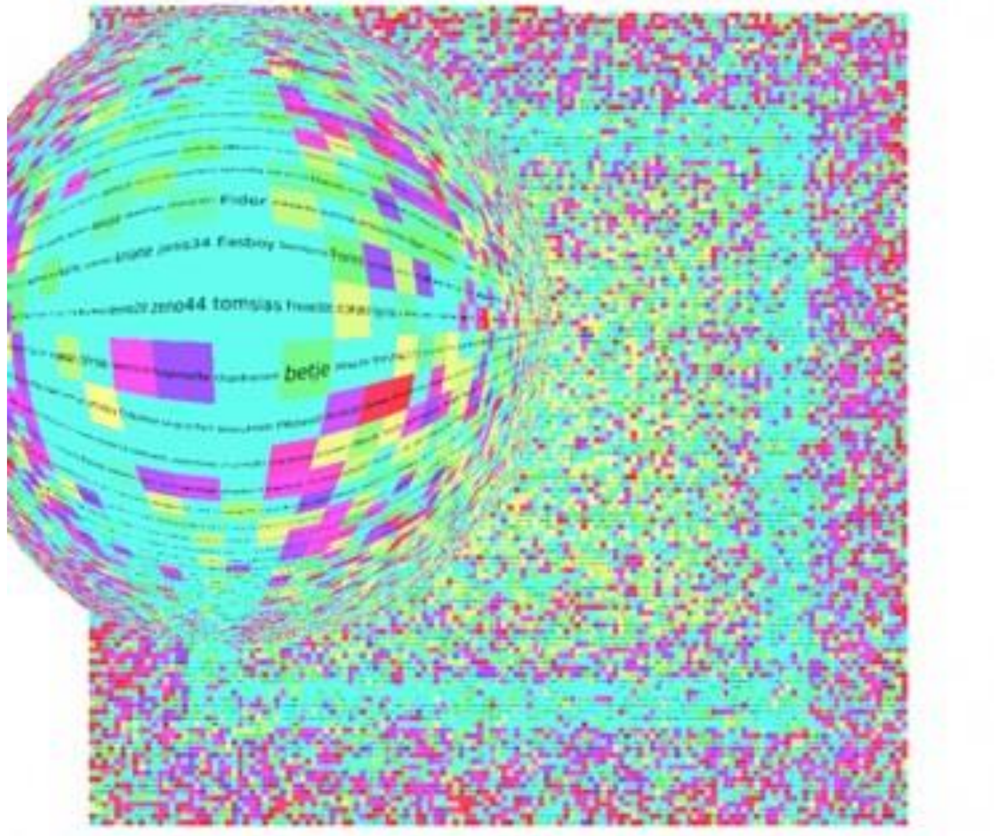


Gaussian curvature pseudo-color lens with overlaid tool to read the numeric value of the curvature.

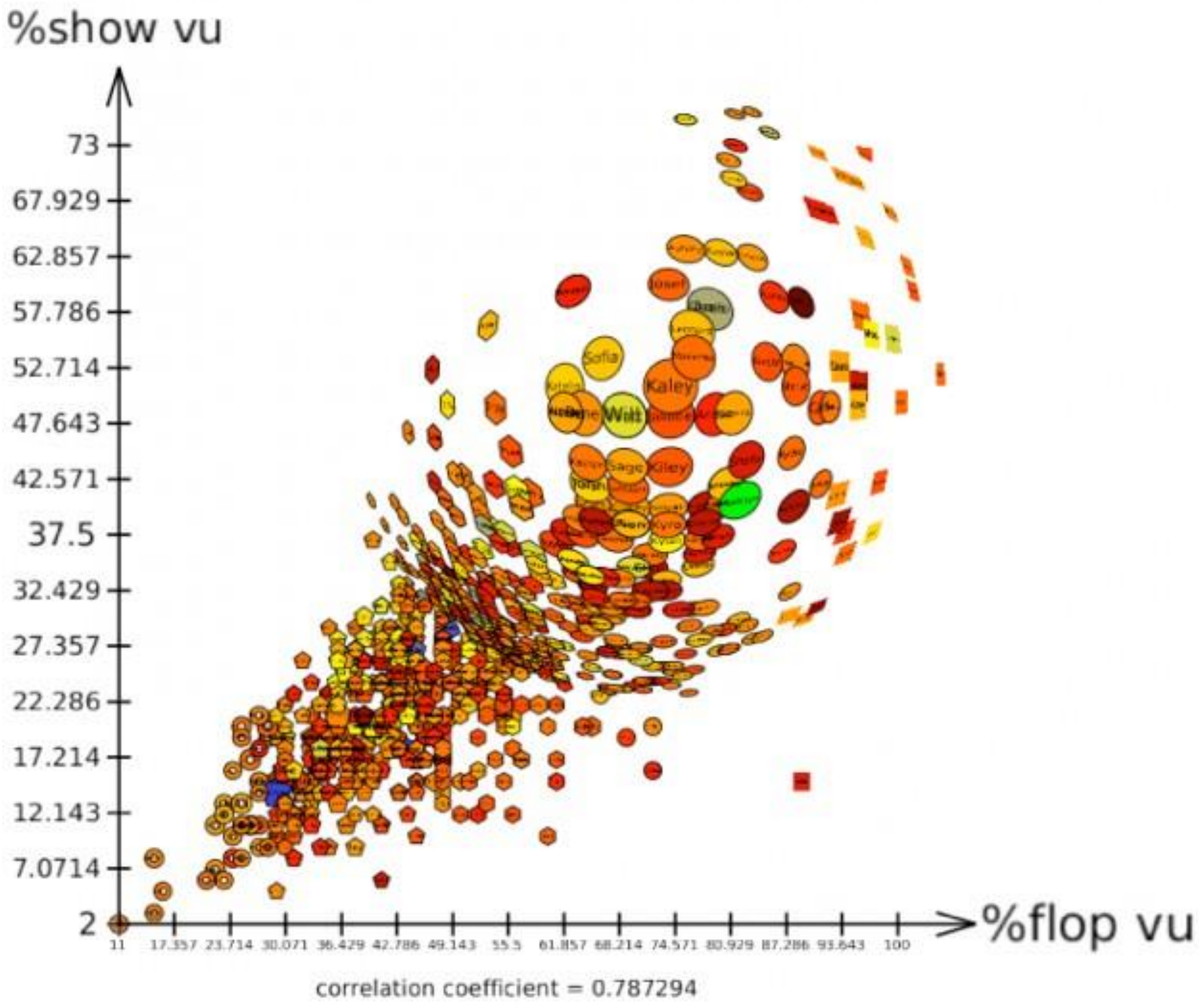
Toolglass and magic lenses: the see-through interface. 1993

Idiom: **Fisheye Lens**

- distort geometry
  - shape: radial
  - focus: single extent
  - extent: local
  - metaphor: draggable



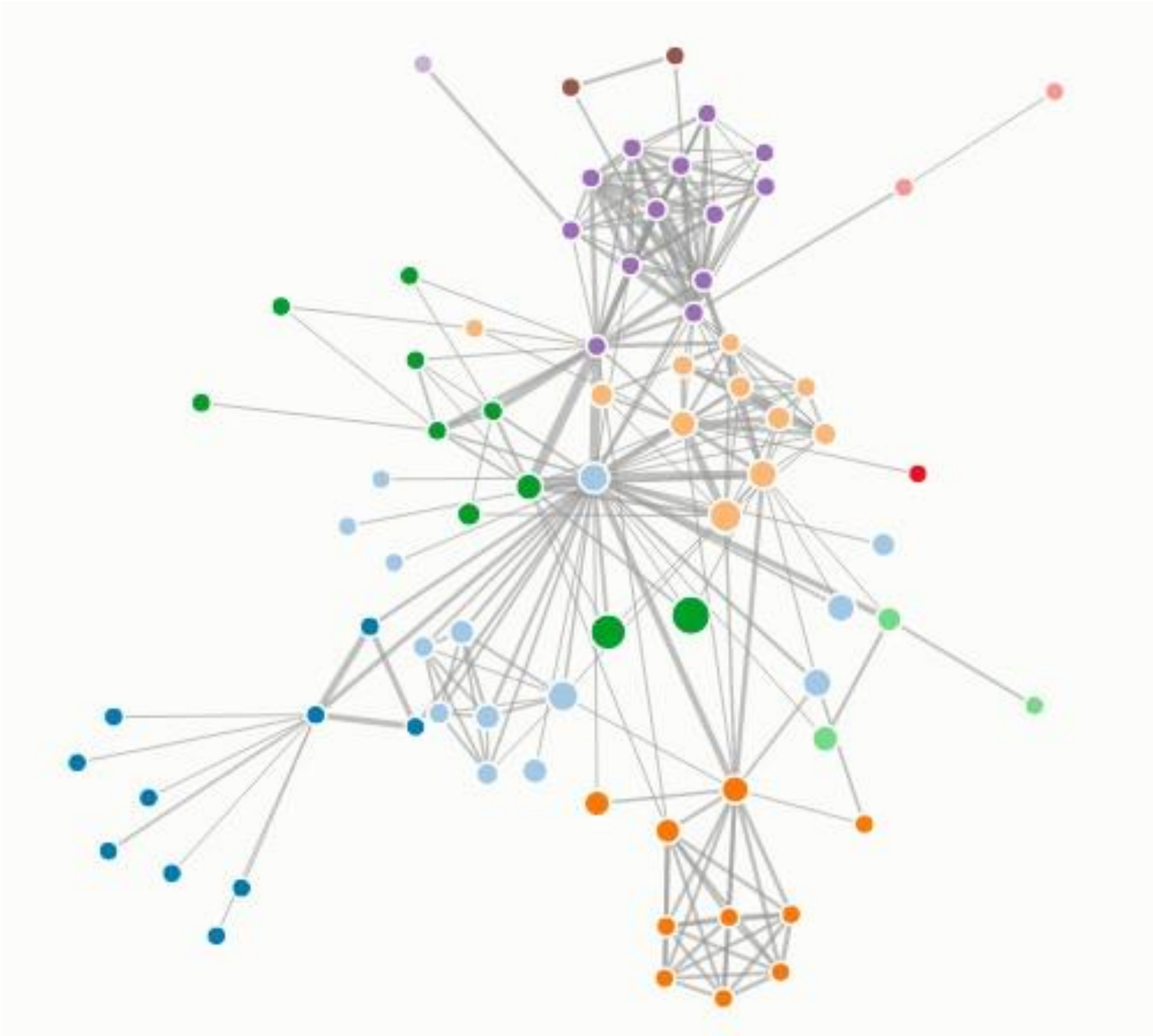
https://tulip.labri.fr/TulipDrupal/?q=node/371

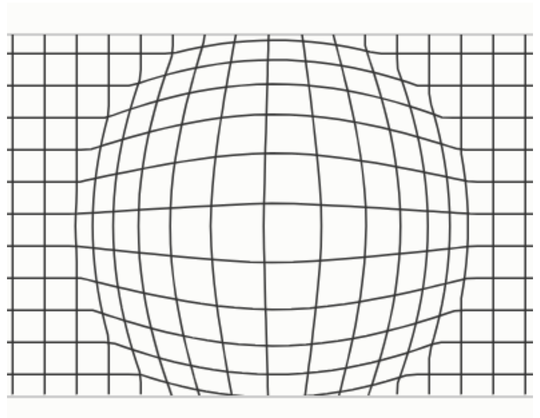


https://tulip.labri.fr/TulipDrupal/?q=node/351

Idiom: **Fisheye Lens**

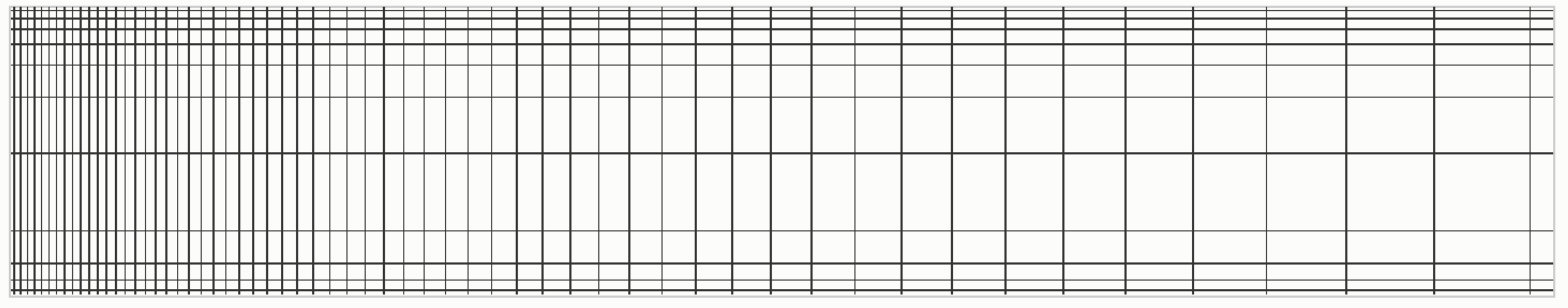
System:
D3

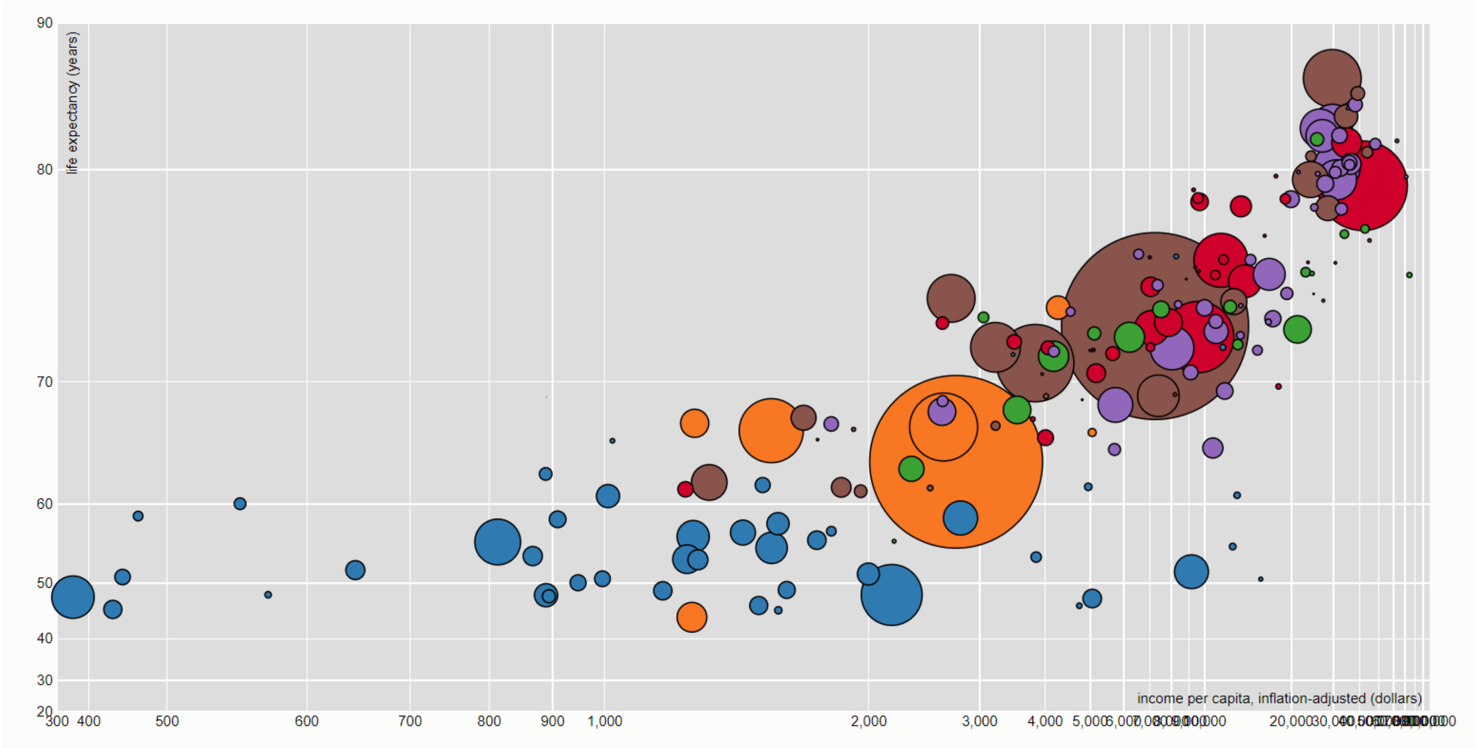


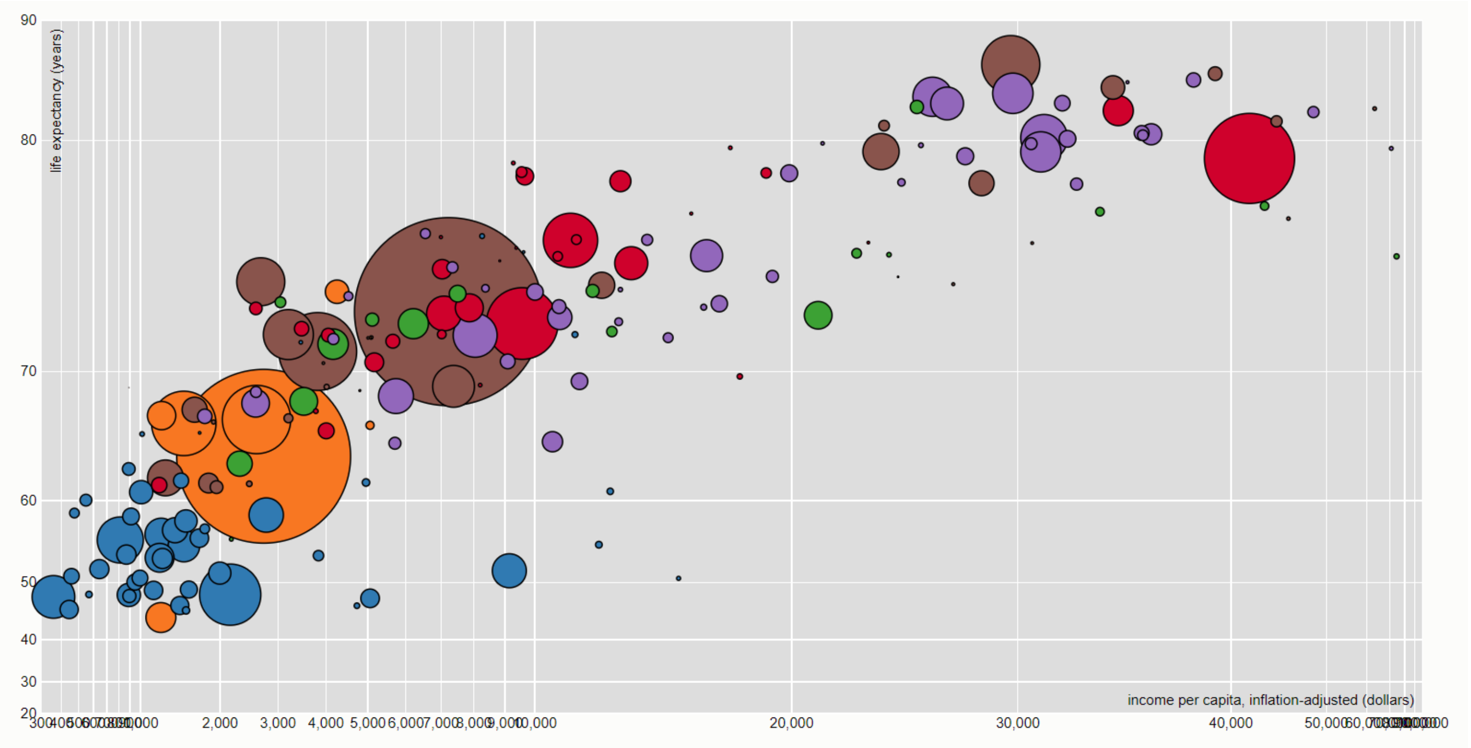


[D3 Fisheye Lens]
(https://bost.ocks.org/mike/fisheye/)

Cartesian Distortion







Distortion costs and benefits

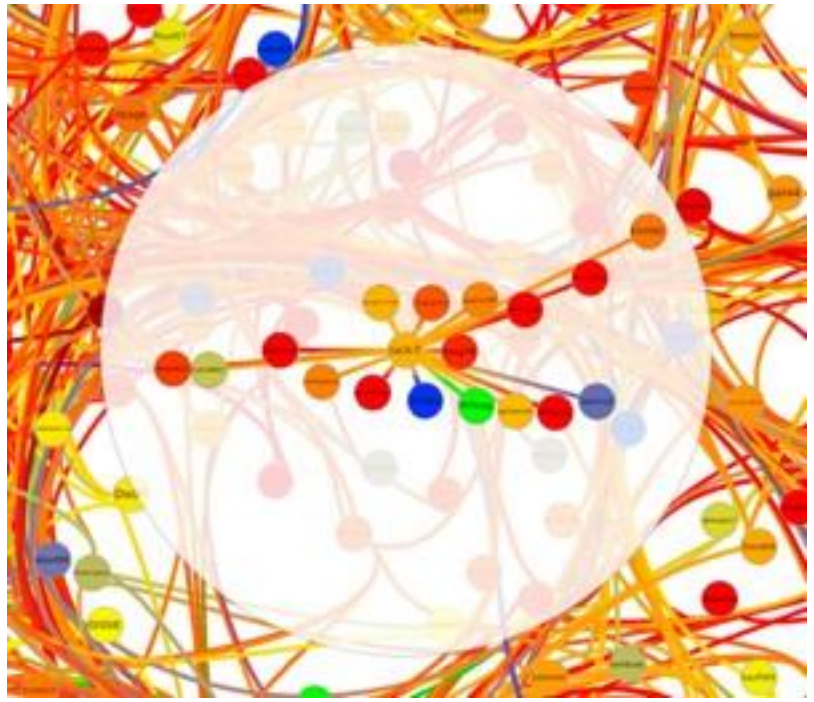
- benefits
  - combine focus and context information in single view
- costs
  - length comparisons impaired
    - network/tree topology comparisons unaffected: connection, containment
  - effects of distortion unclear if original structure unfamiliar
  - object constancy/tracking maybe impaired

fisheye lens

magnifying lens

neighborhood layering

Bring and Go

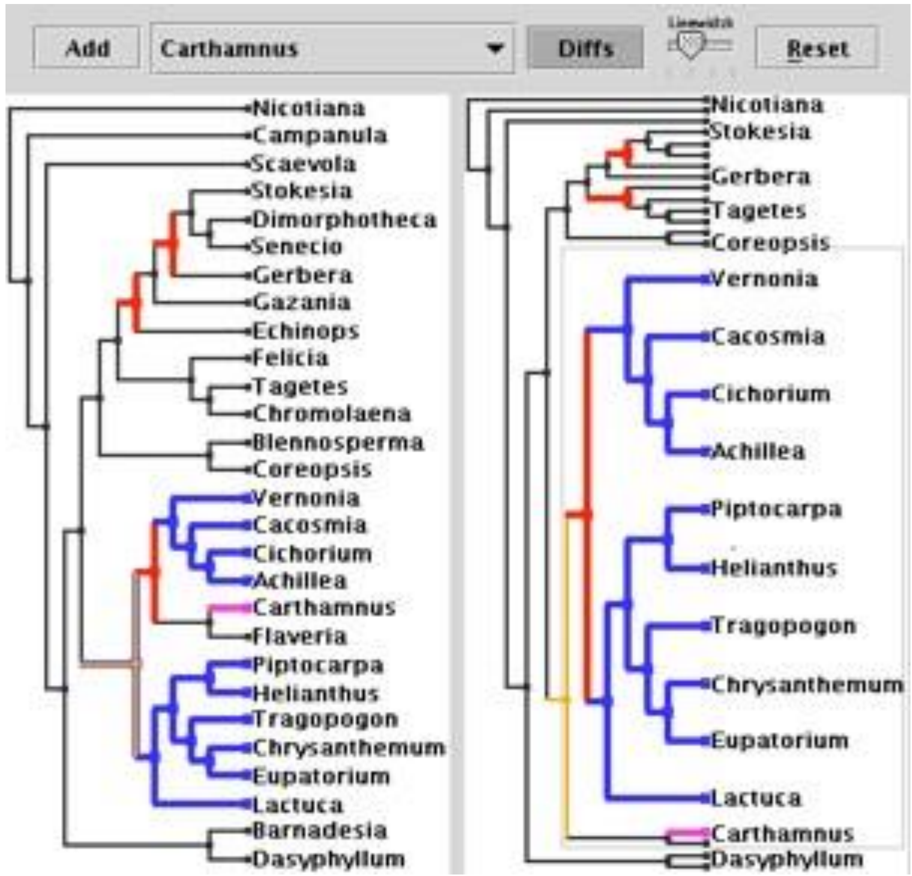


[Living Flows: Enhanced Exploration of Edge-Bundled Graphs Based on GPU-Intensive Edge Rendering. Lambert, Auber, and Melançon. Proc. Intl. Conf. Information Visualisation (IV), pp. 523–530, 2010.]

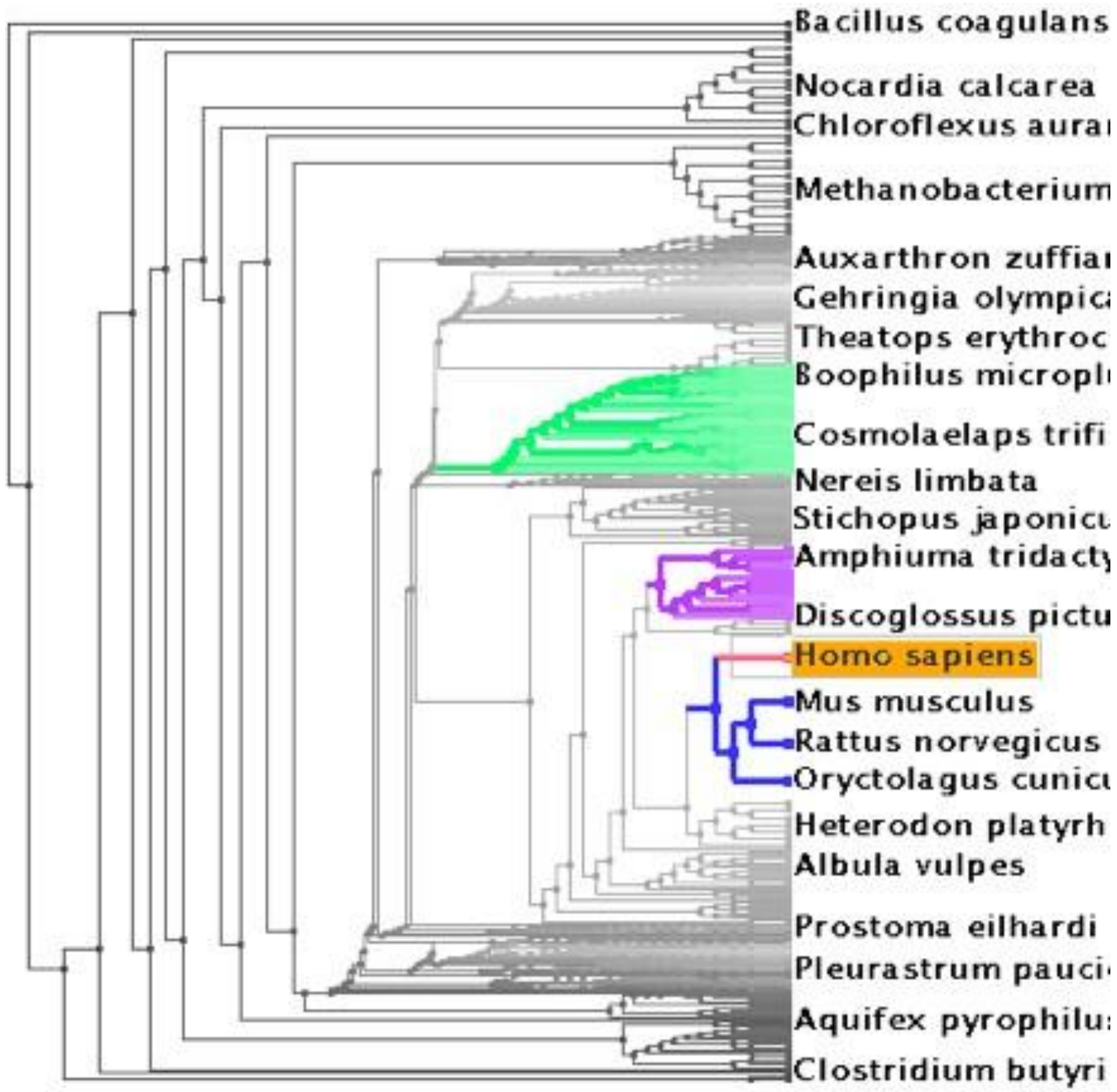
Idiom: **Stretch and Squish Navigation**

- distort geometry
  - shape: rectilinear
  - foci: multiple
  - impact: global
  - metaphor: stretch and squish, borders fixed

System: **TreeJuxtaposer**



[video](#)



[TreeJuxtaposer: Scalable Tree Comparison Using Focus+Context With Guaranteed Visibility. Munzner, Guimbretiere, Tasiran, Zhang, and Zhou. ACM Transactions on Graphics (Proc. SIGGRAPH) 22:3 (2003), 453– 462.]

Focus+Contest Visualization with distortion minimization

Drawing Road Networks with Focus Regions

Edited from Jan-Henrik Haunert's slides

Focus+Context Metro Maps

Can you see the station names?

Yes, you can!