Data Visualization (CSCI 627/490)

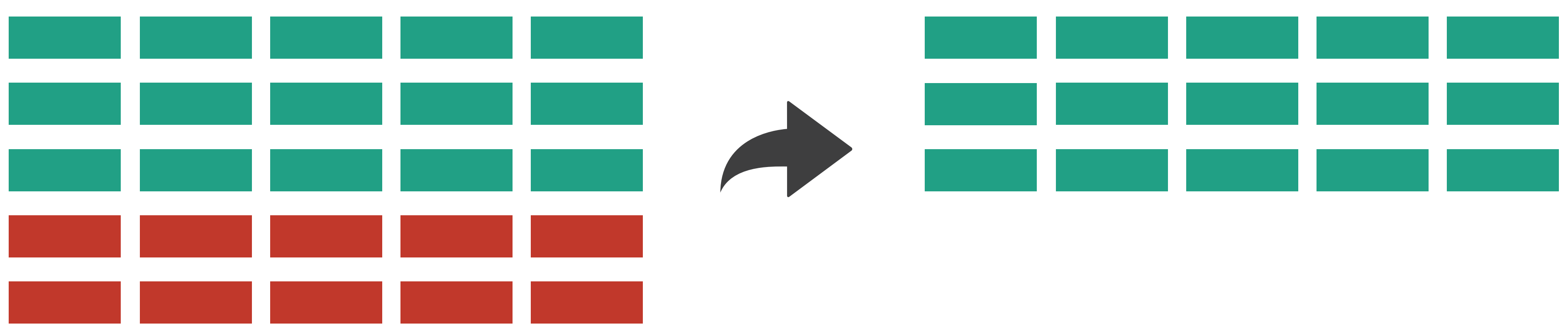
Aggregation & Focus+Context

Dr. David Koop

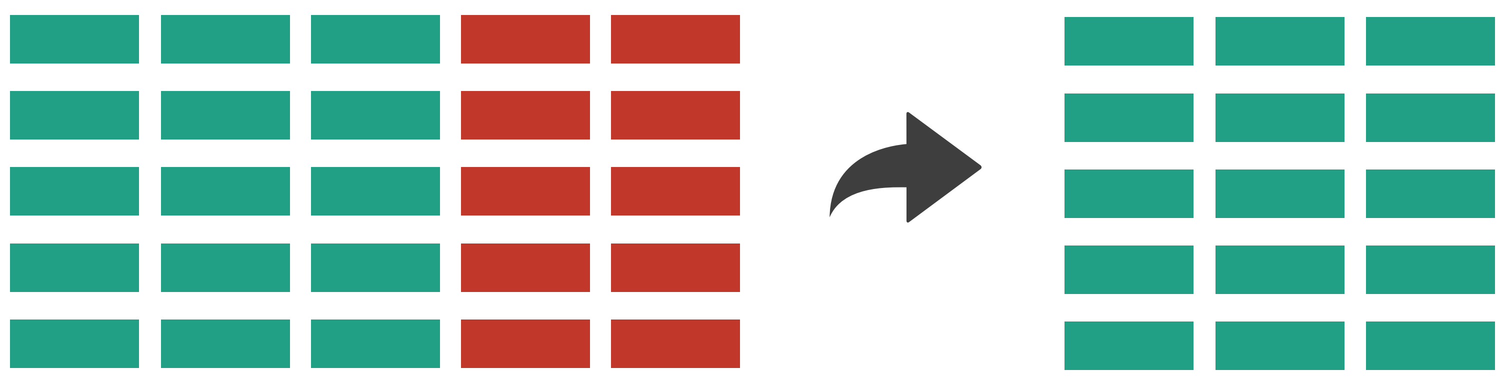
Overview: Reducing Items & Attributes

Filter

- Items

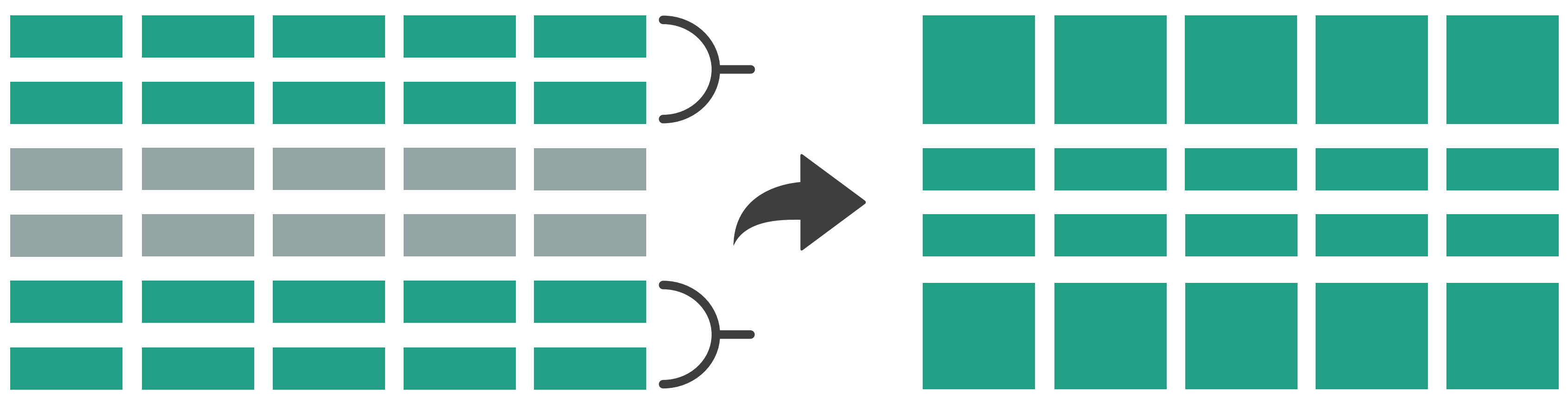


- Attributes

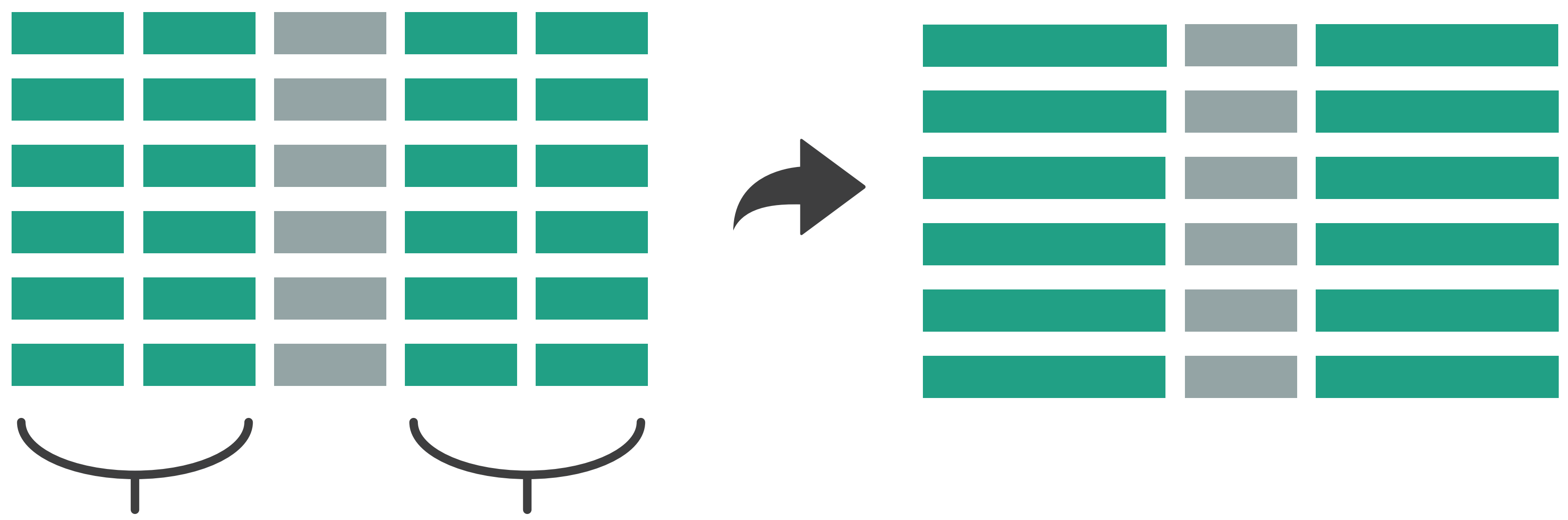


Aggregate

- Items

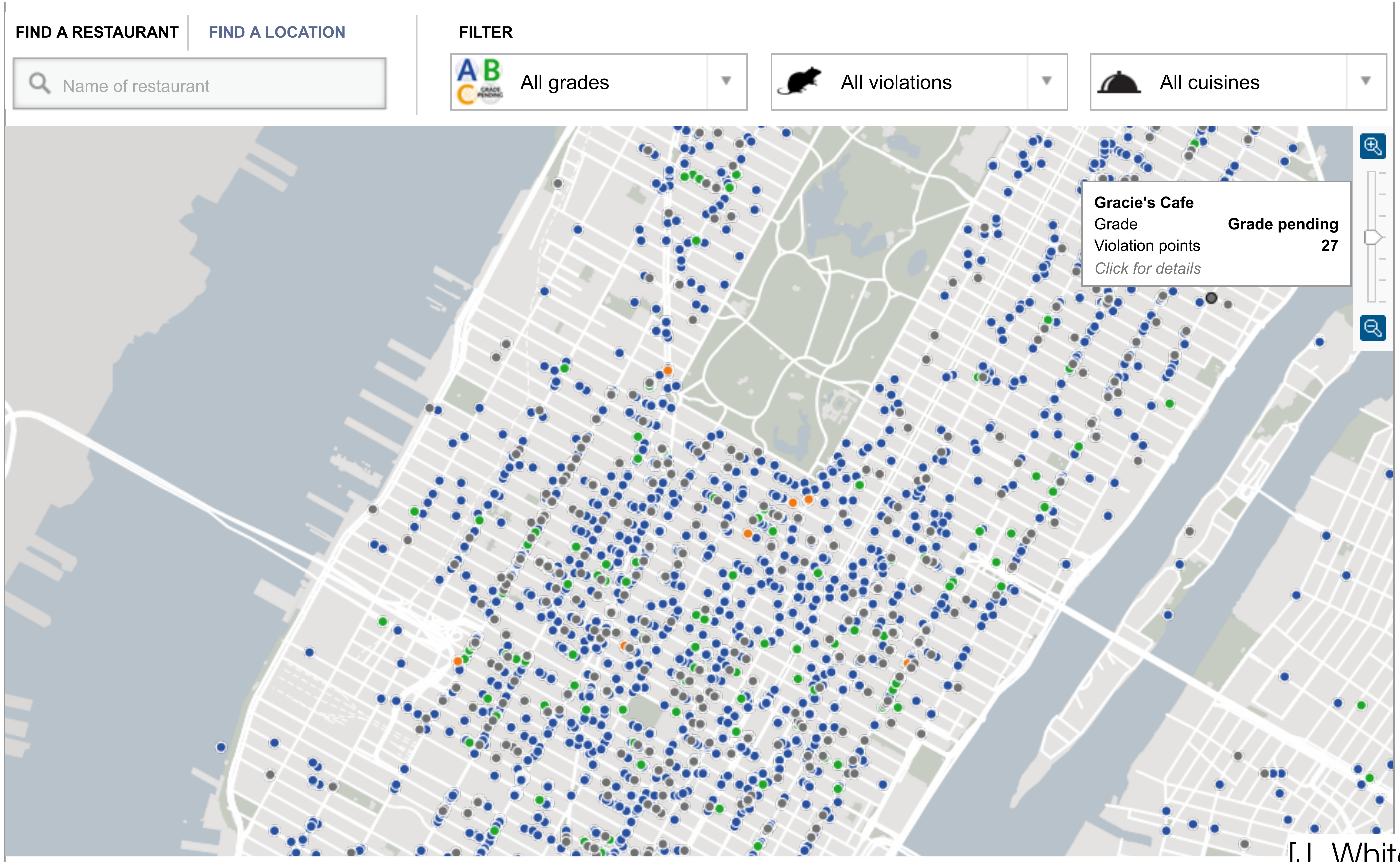


- Attributes



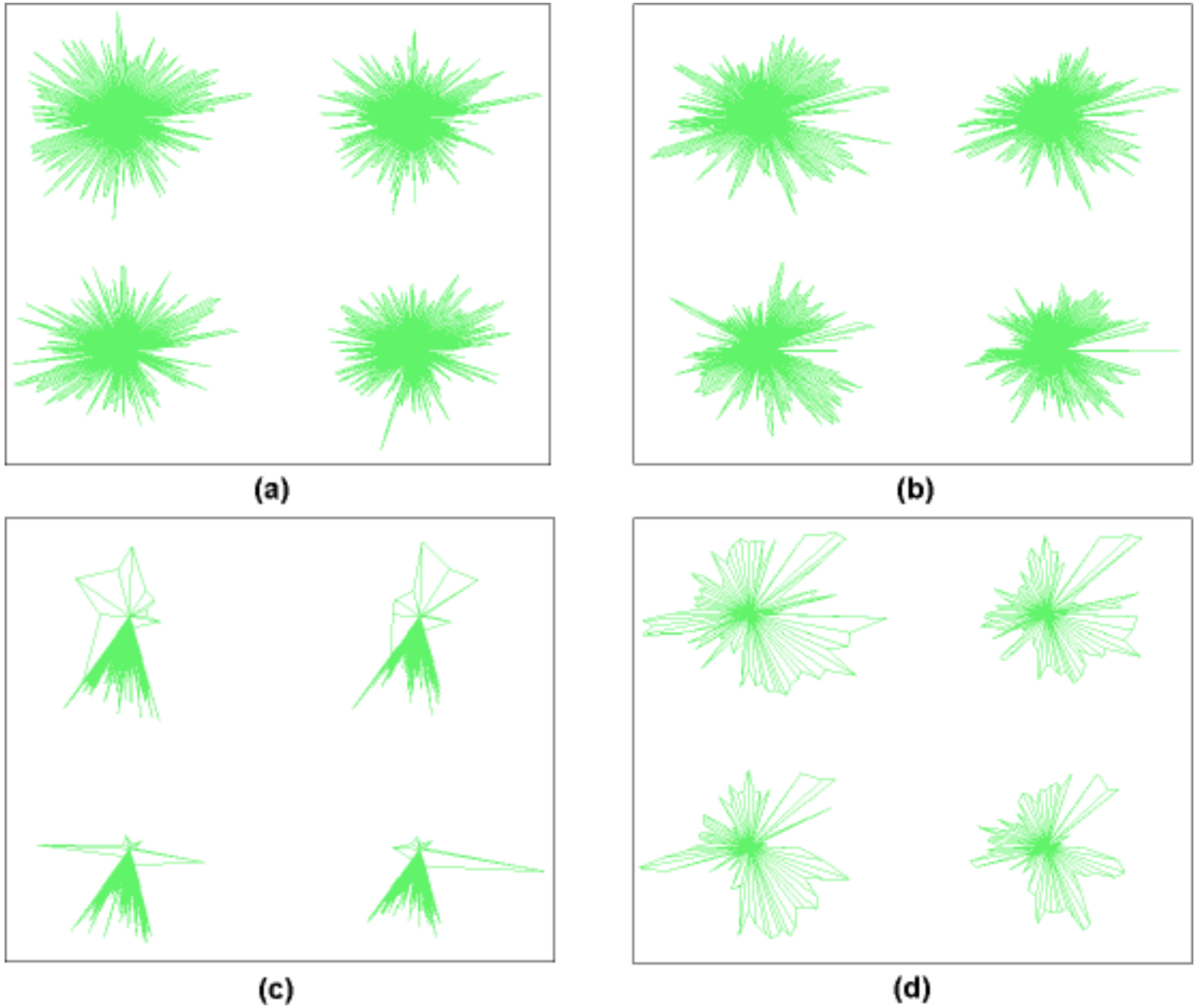
[Munzner (ill. Maguire), 2014]

Item Filtering on Maps



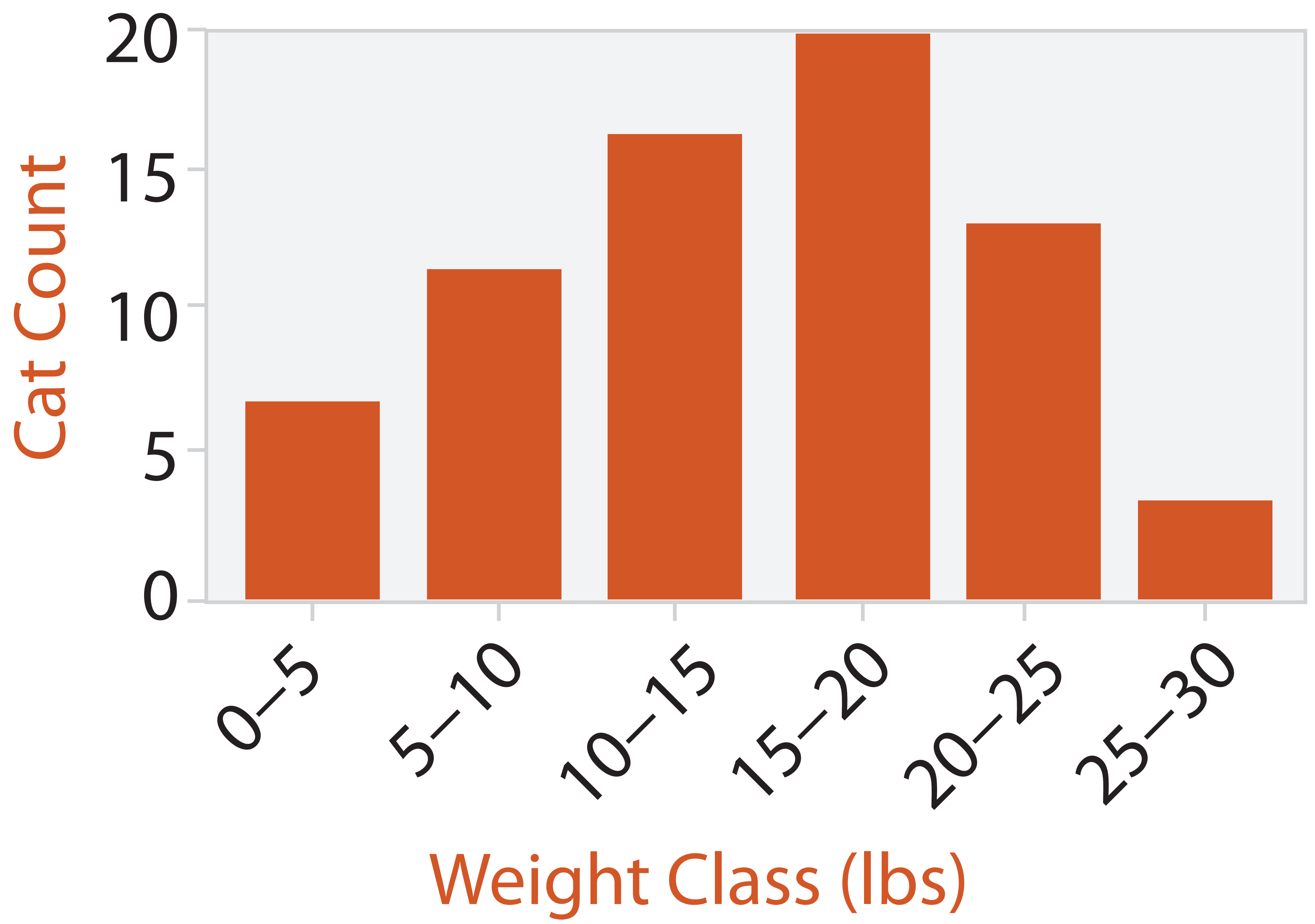
[J. White, New York Times]

Attribute Filtering on Star Plots



[Yang et al., 2003]

Aggregation: Histograms

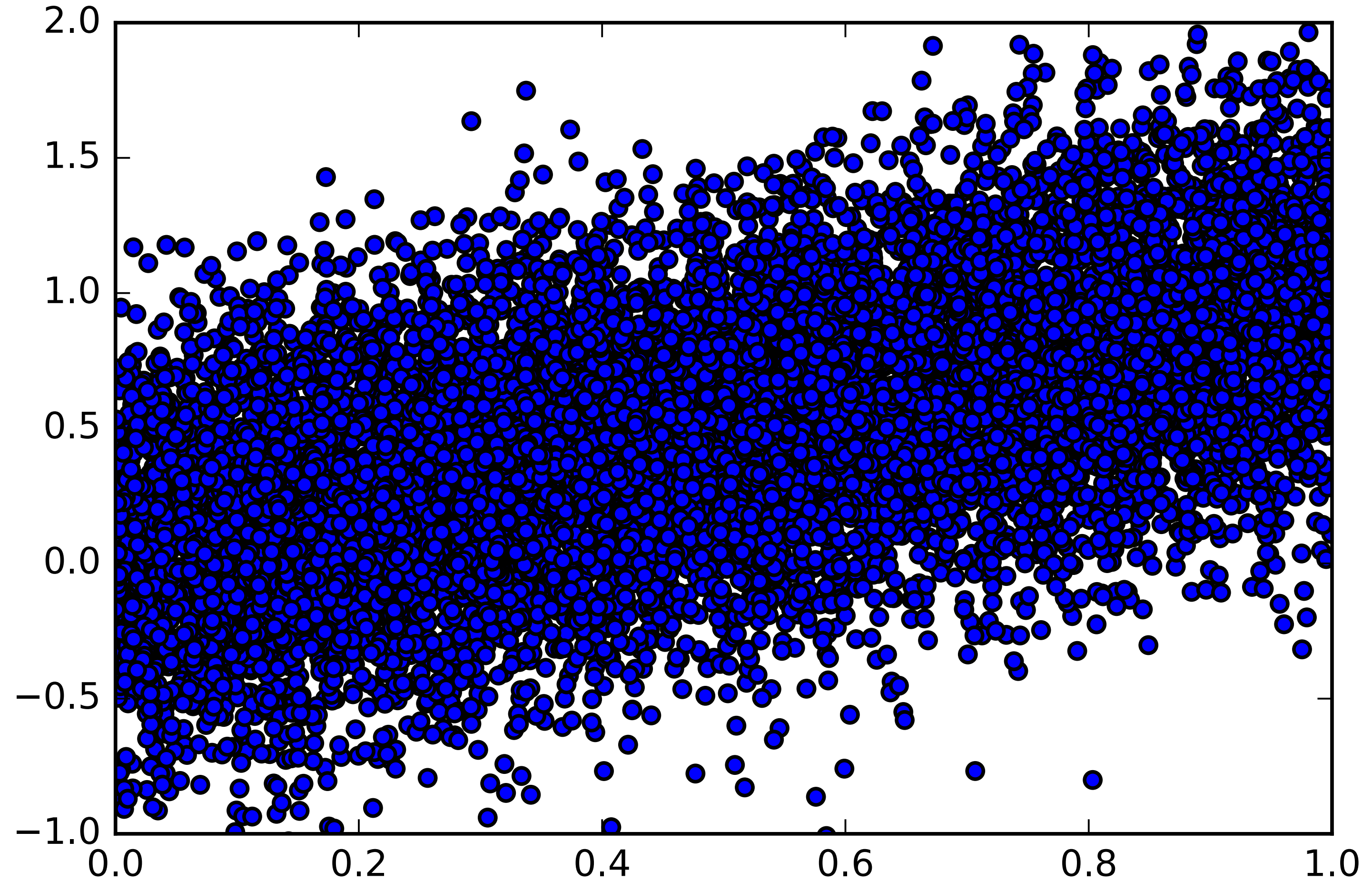


- Very similar to bar charts
- Often shown without space between (continuity)
- Choice of number of bins
  - Important!
  - Viewers may infer different trends based on the layout

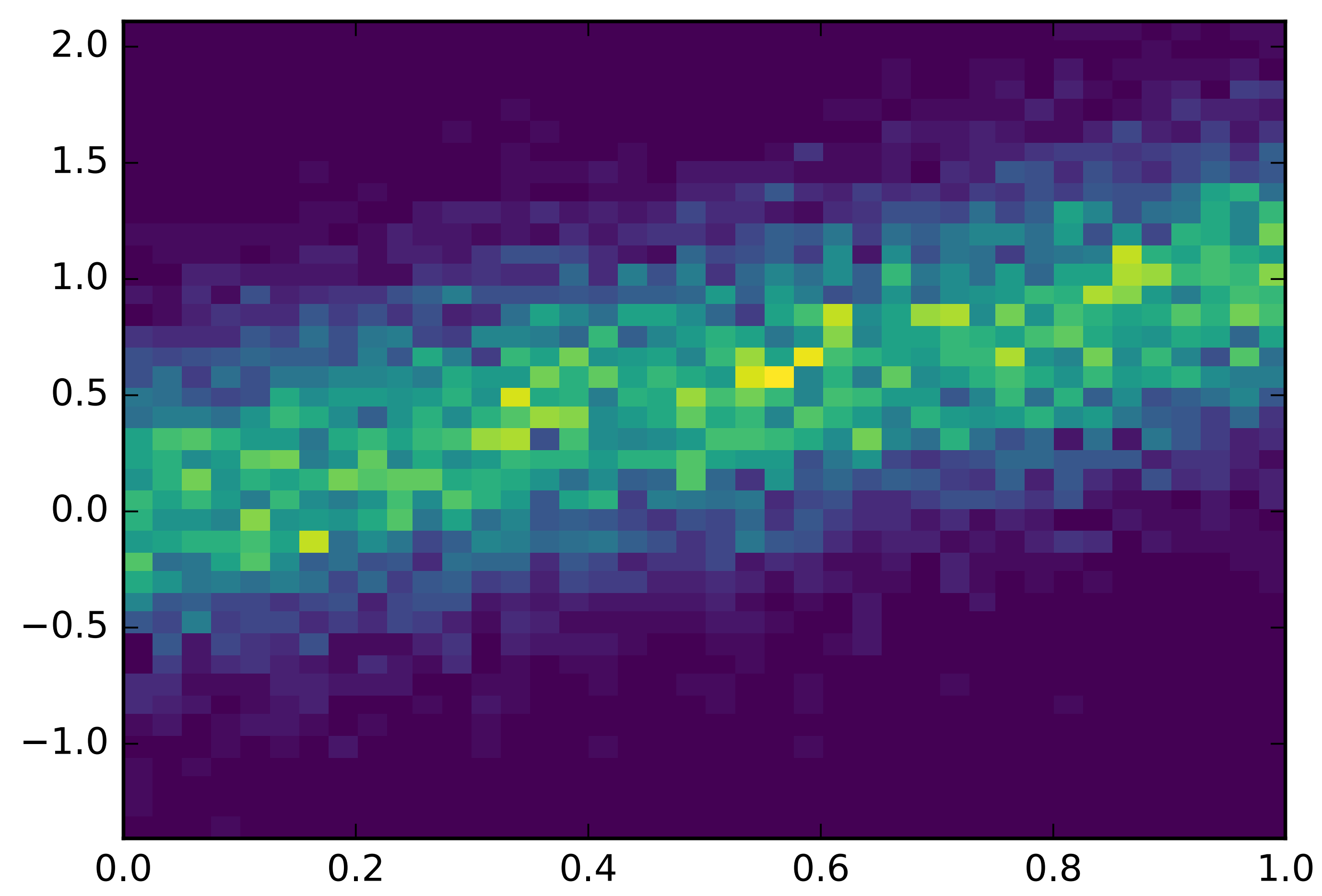
[Munzner (ill. Maguire), 2014]

Binning

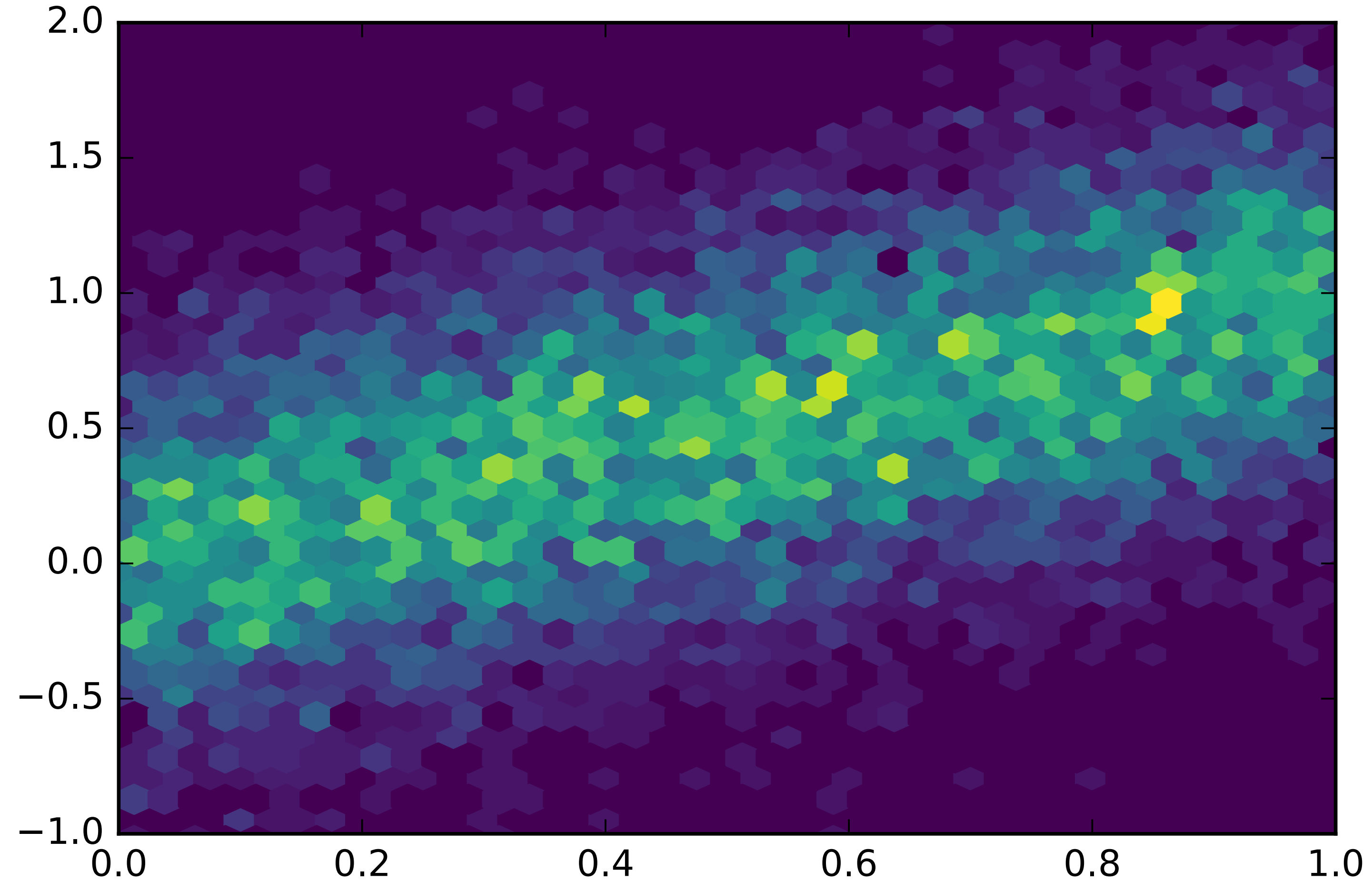
- 2D Histogram is a histogram in 2D encoded using color instead of height
- Hexbin advantages:
  - Bins are more circular so distance to the edge is not as variable
  - More efficient aggregation around the center of the bin



Scatterplot

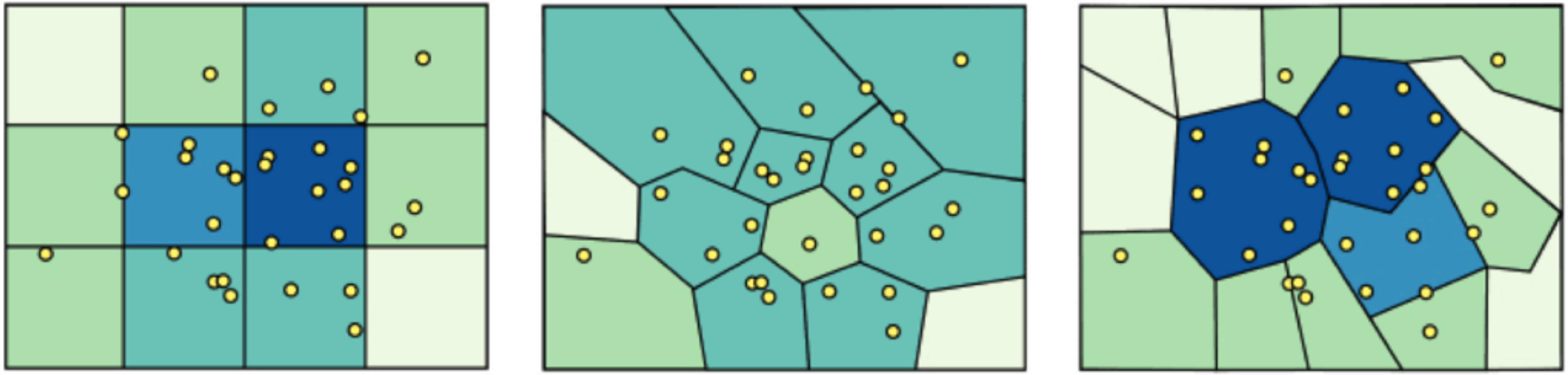


Rectangular Bin



Hexagonal Bin

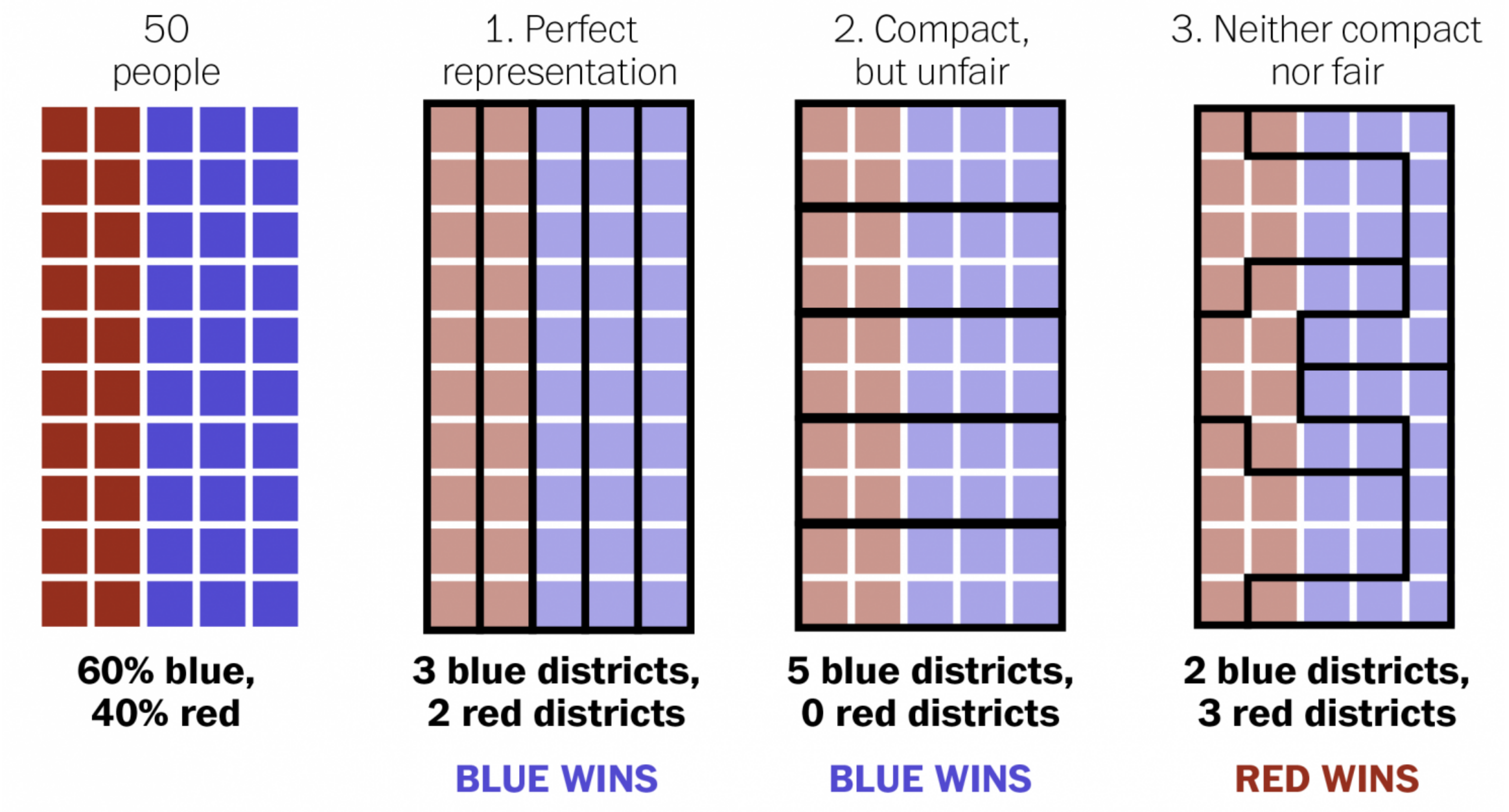
Spatial Aggregation

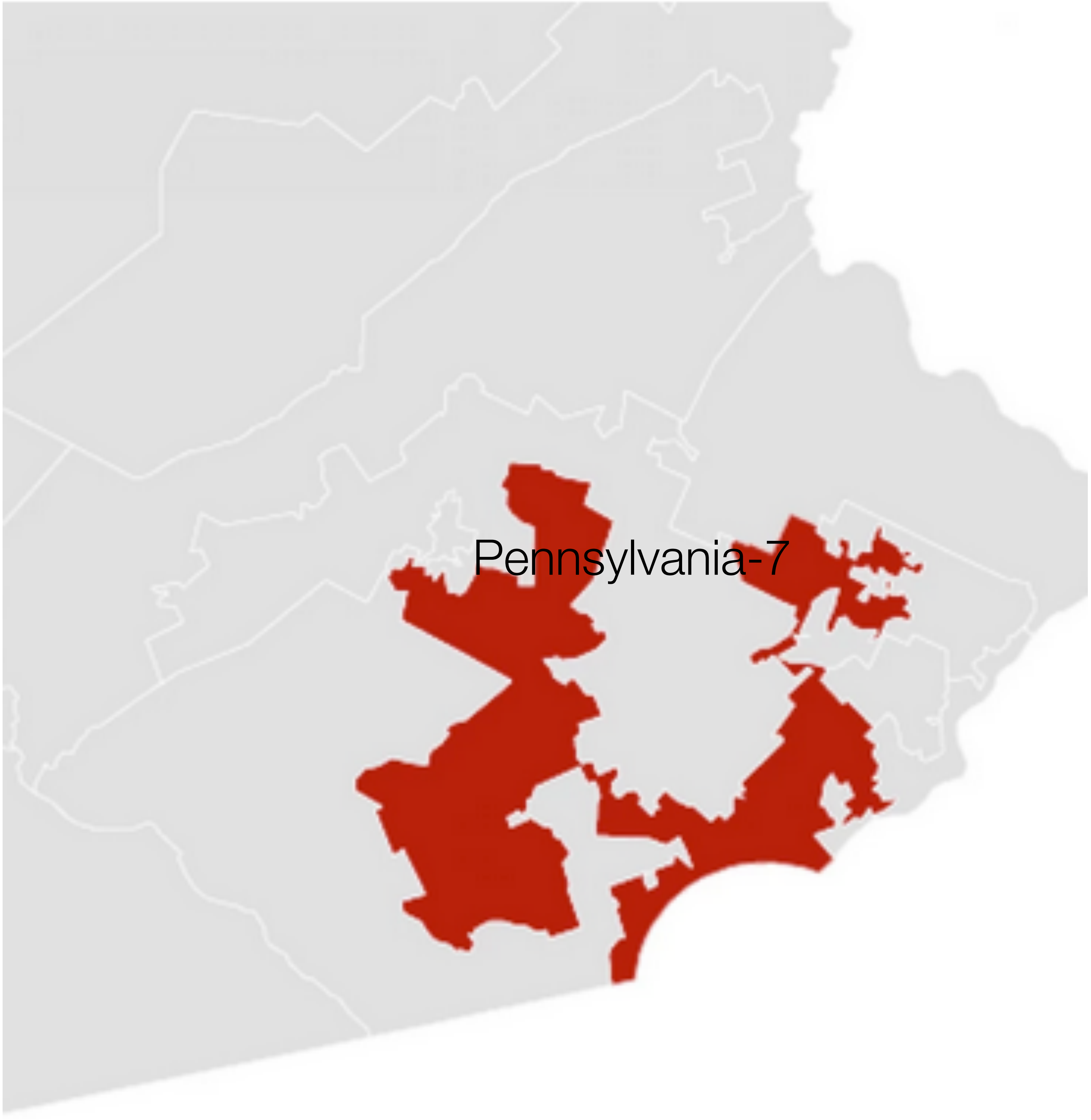


[Penn State, GEOG 486]

Modifiable Areal Unit Problem

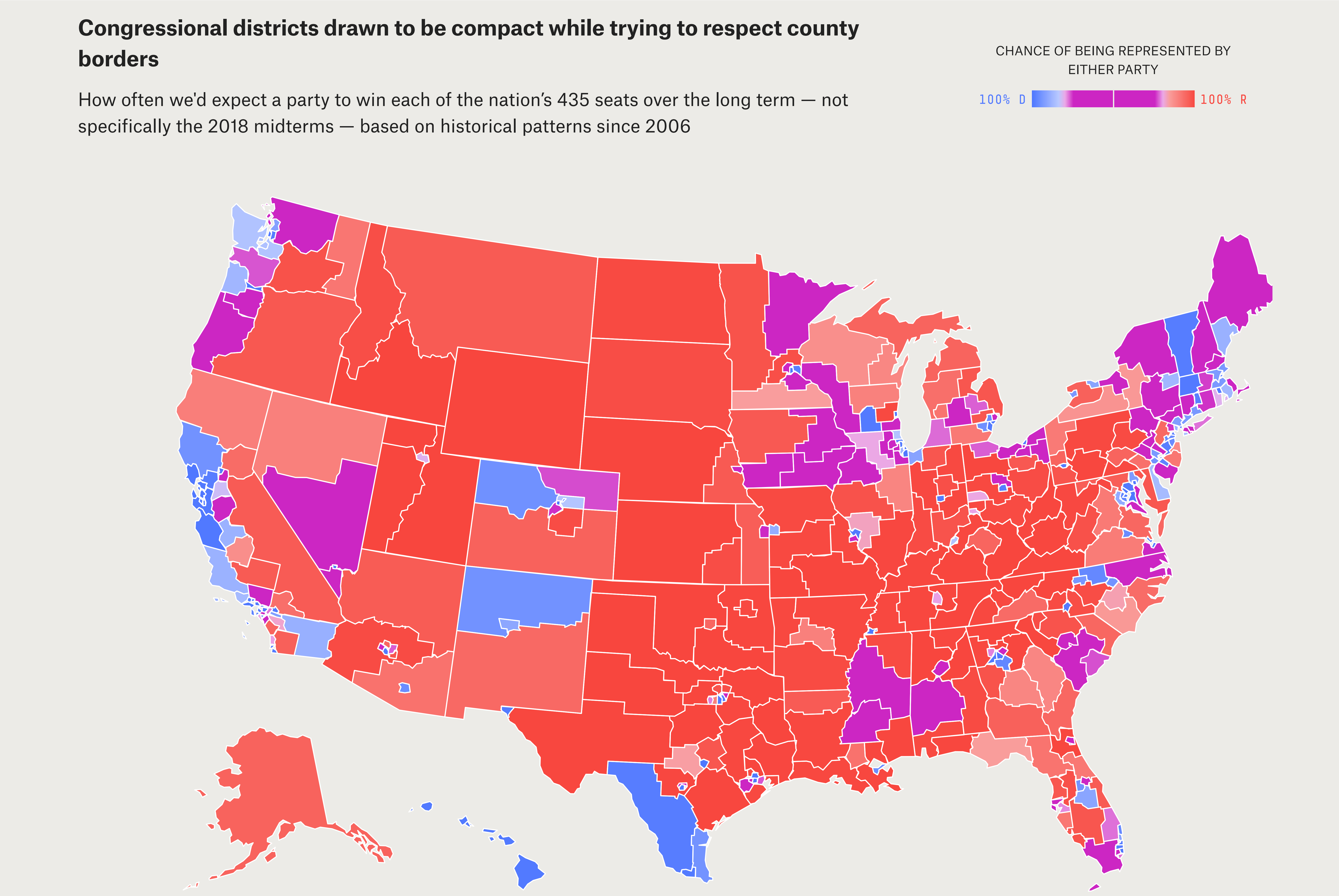
- How you draw boundaries impacts the type of aggregation you get
- Similar to bins in histograms
- Gerrymandering





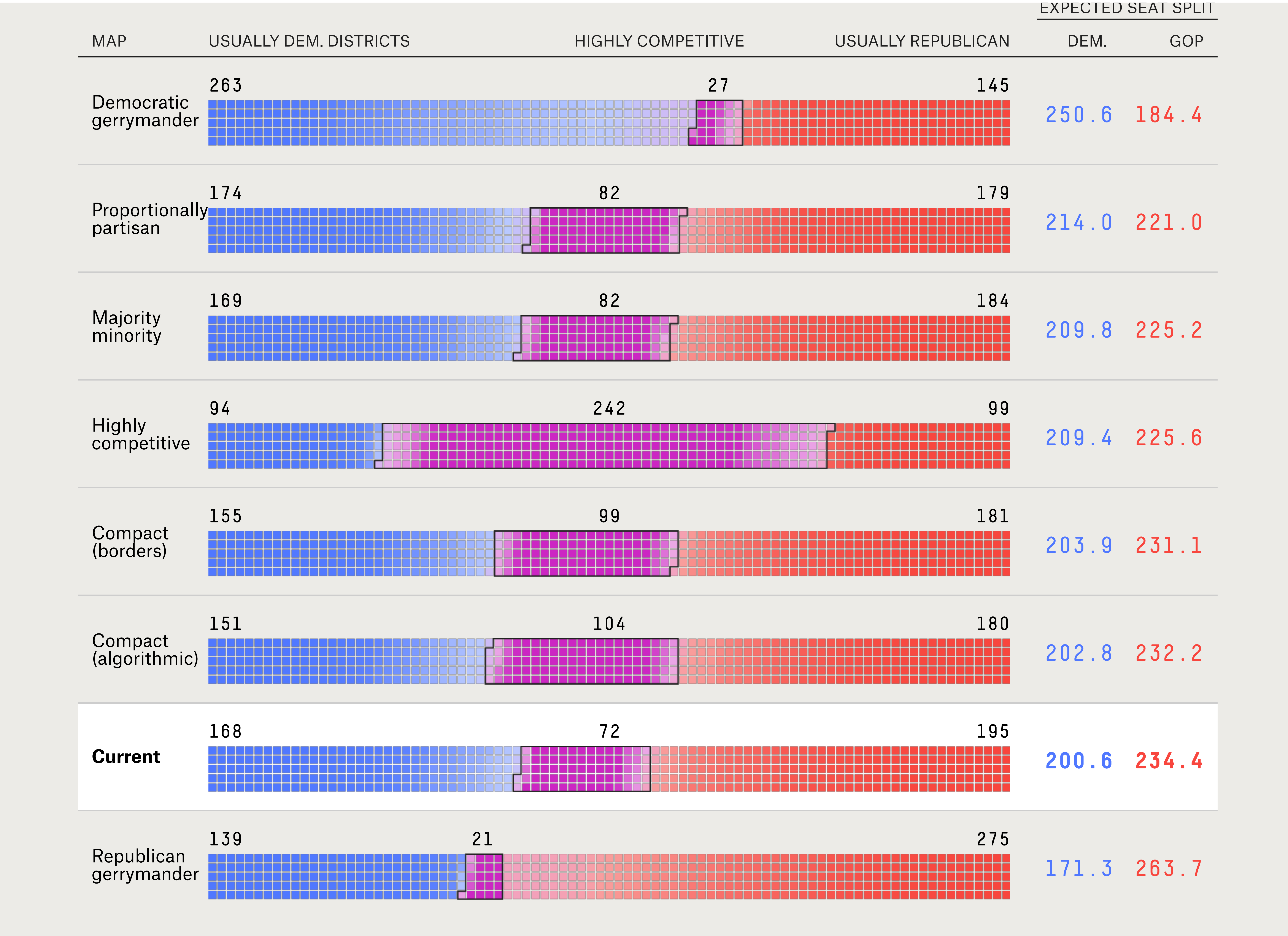
[Wonkblog, Washington Post, Adapted from S. Nass]

Drawing Different Maps: Compactness



[A. Bycoffe et al., 538]

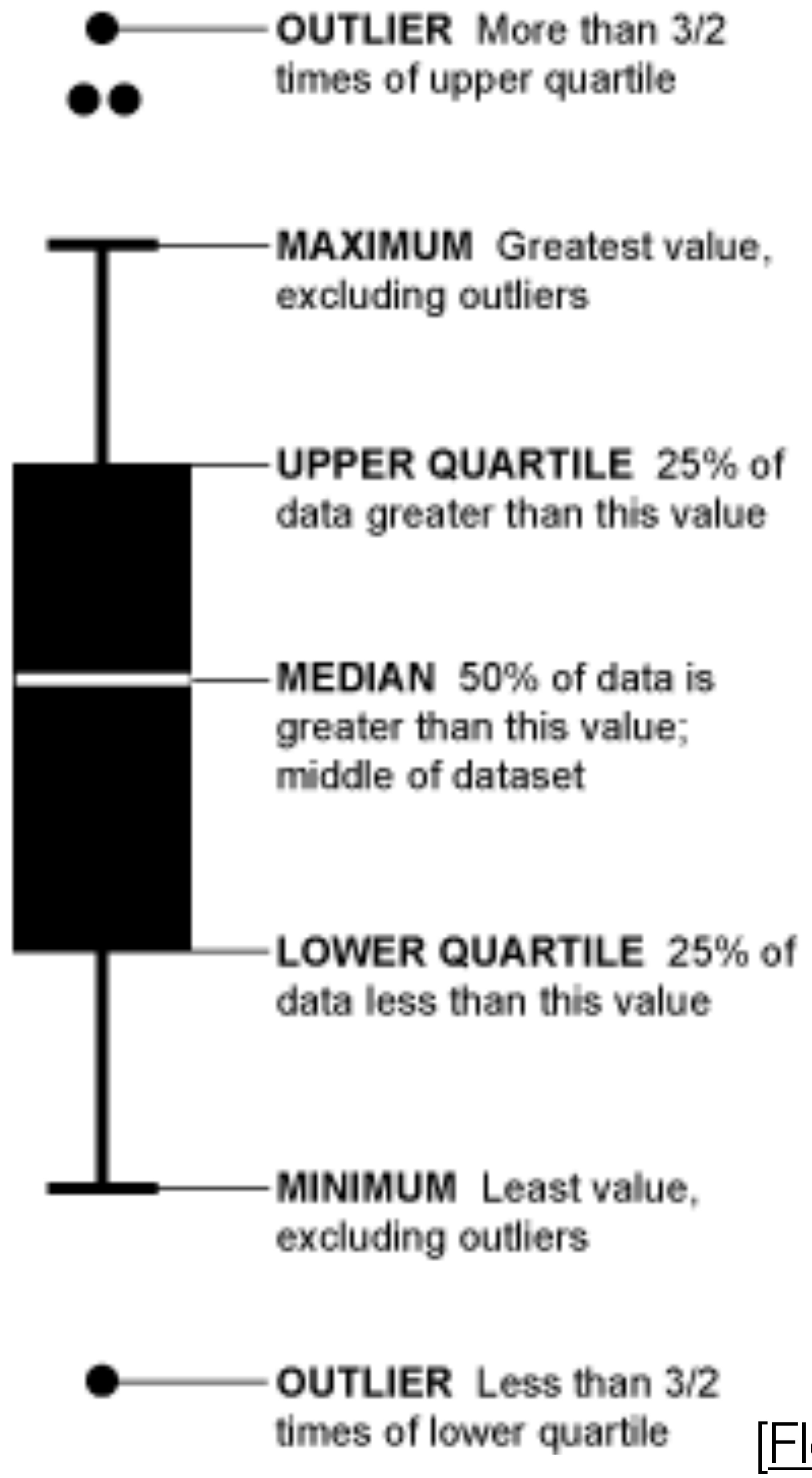
Drawing Different Maps



[A. Bycoffe et al., 538]

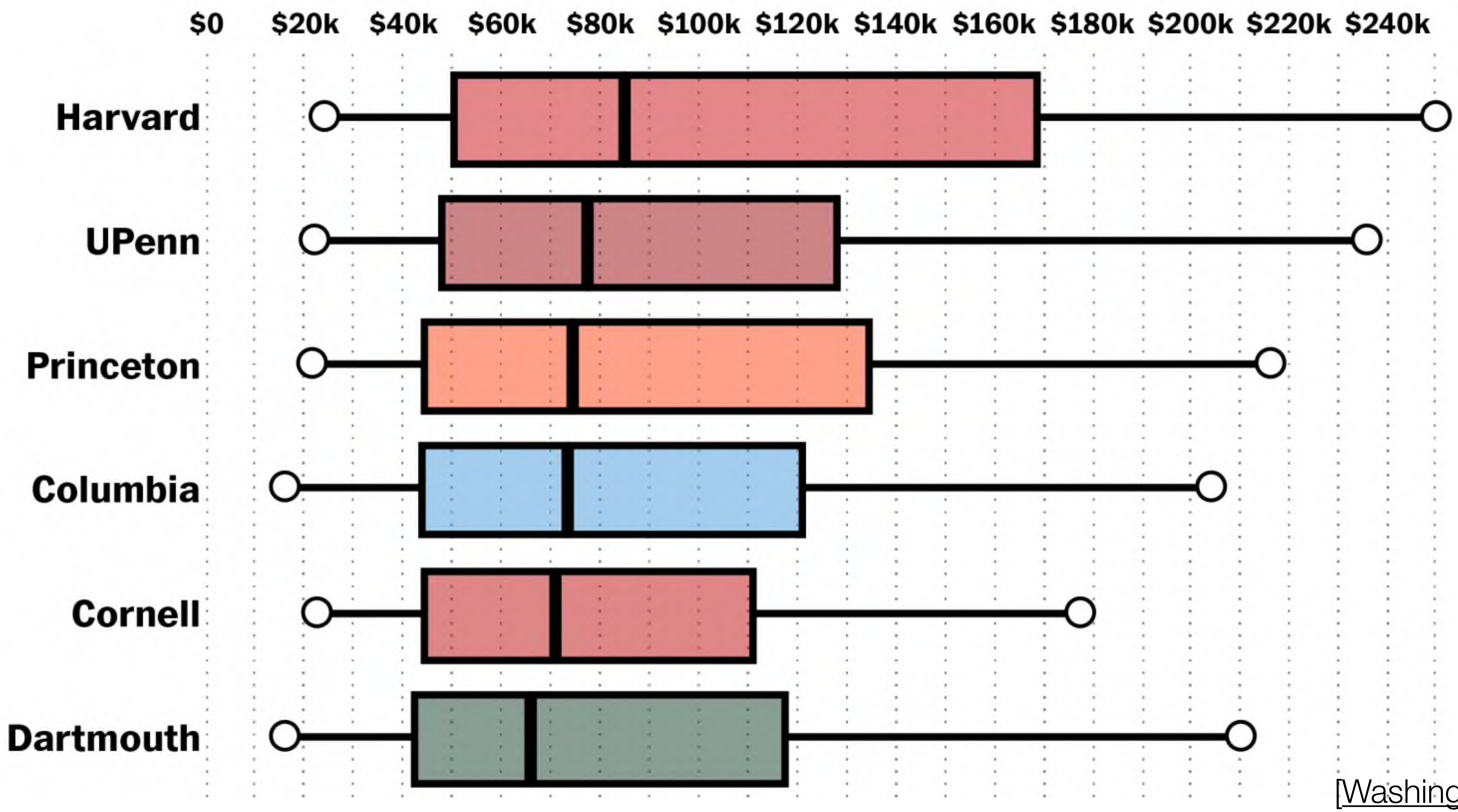
Boxplots

- Show **distribution**
- Single value (e.g. mean, max, min, quartiles) doesn't convey everything
- Created by John Tukey
- Show **spread** and **skew** of data
- Best for **unimodal** data
- Variations like vase plot for multimodal data
- Aggregation here involves many different marks



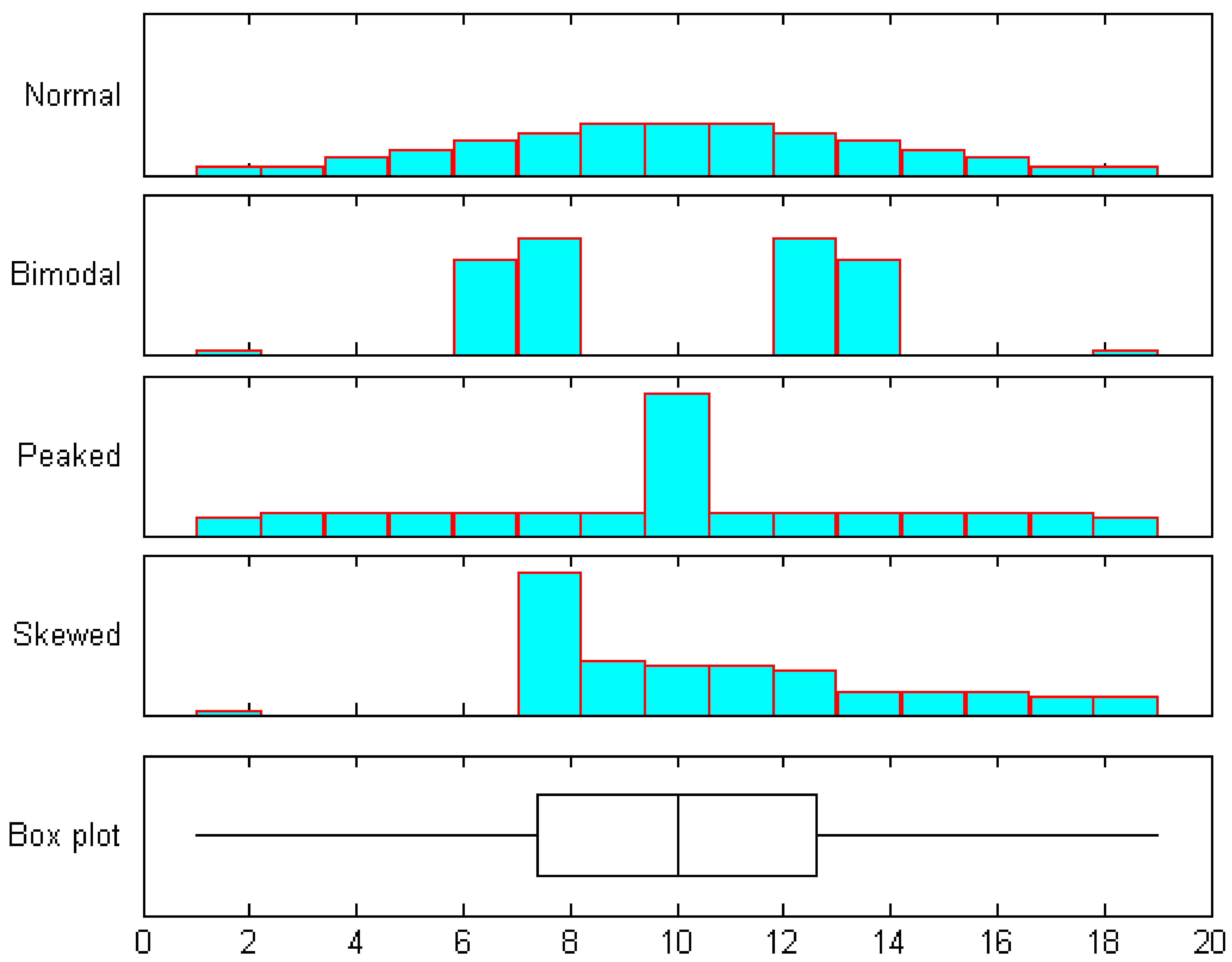
[Flowing Data]

Aggregation: Boxplots



[Washington Post, 2015]

Four Distributions, Same Boxplot…



[C. Choonpradub and D. McNeil, 2005]

Project Design

- Feedback:
  - Data Manipulation?
  - Questions lead, not technique!
  - Be creative! (interaction too) https://xeno.graphics
- Work on turning your visualization ideas into designs
- Turn in:
  - Two Design Sketches (like sheets 2-4 from 5 Sheet Design)
  - One Bad Design Sketch (like sheets 2-4: here, justify why bad)
  - Progress on Implementation
- Due next Wednesday

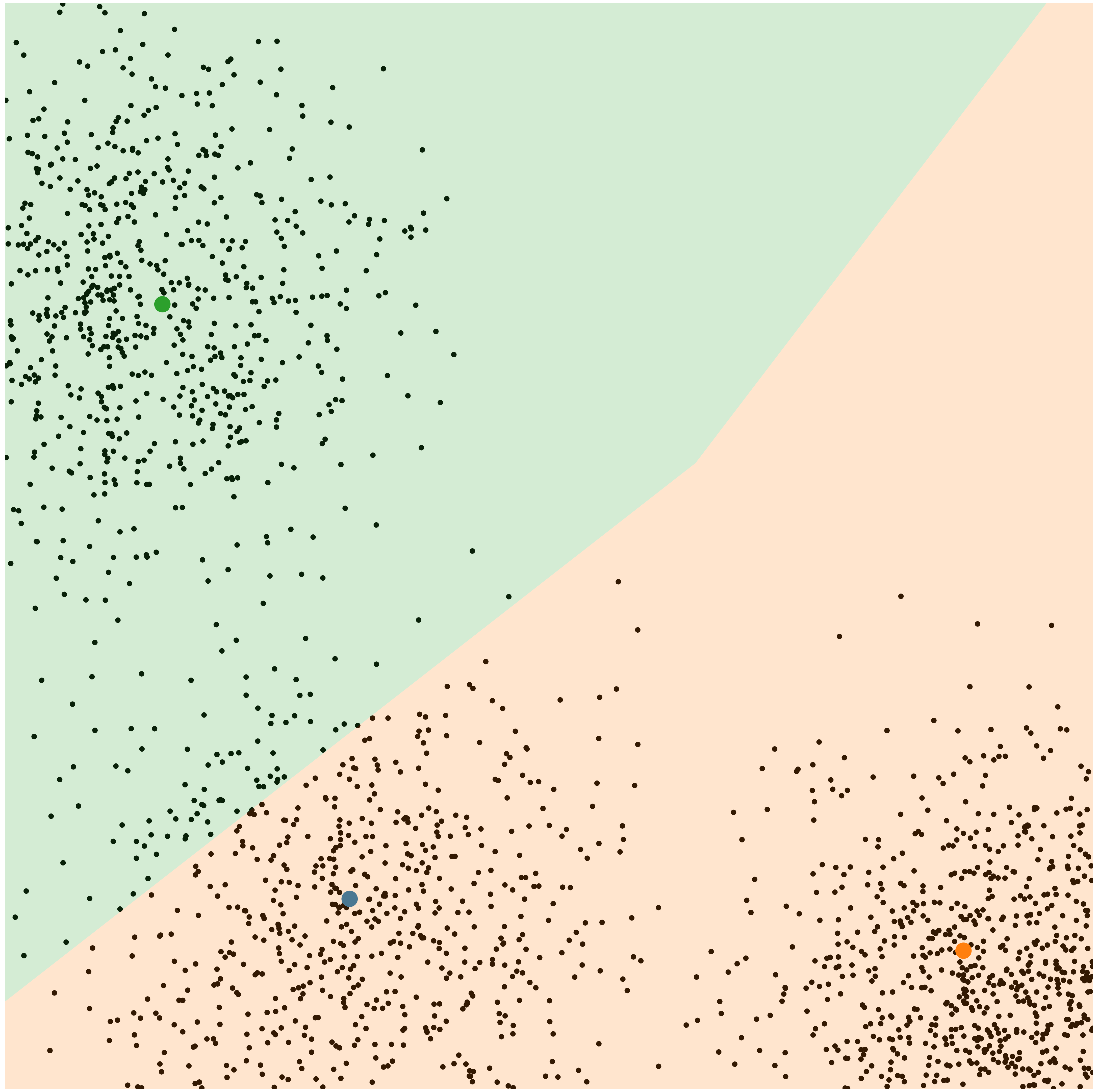
Assignment 5

- Focus on Multiple Views and Interaction
- Soon…

No Class Tuesday

- Vote!

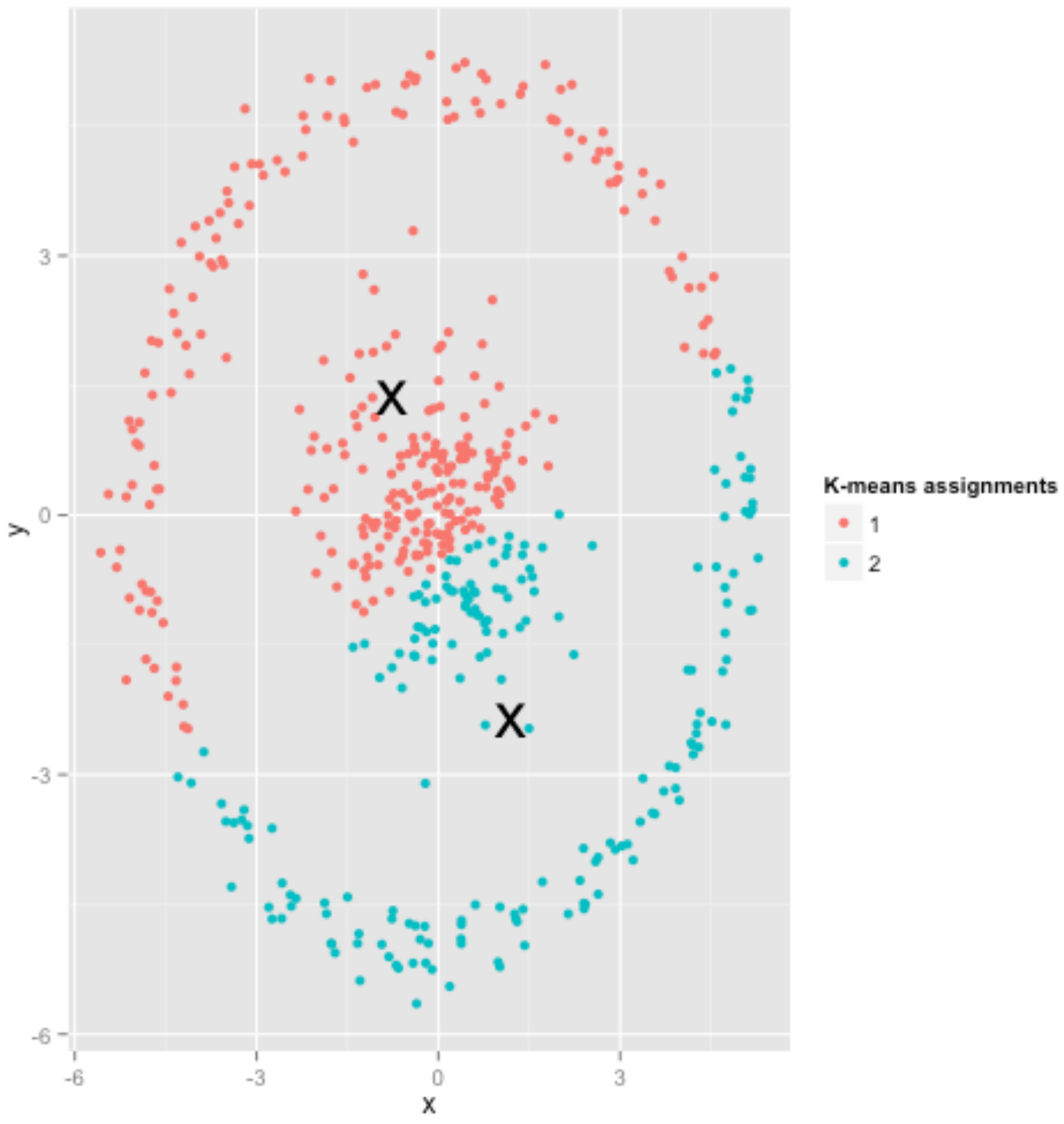
K-Means



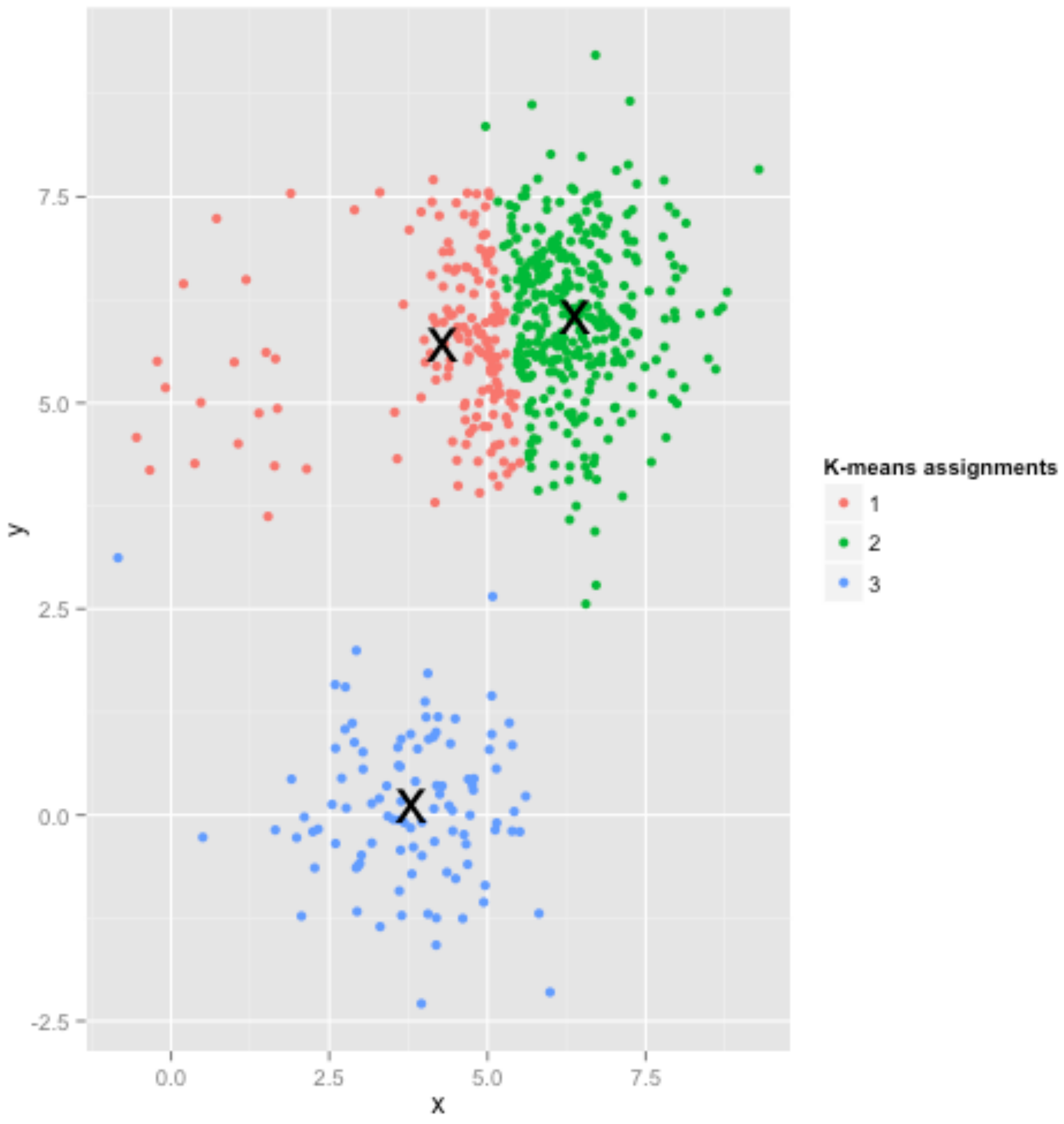
Run

[C. Polis, 2014]

K-Means Issues



Shape



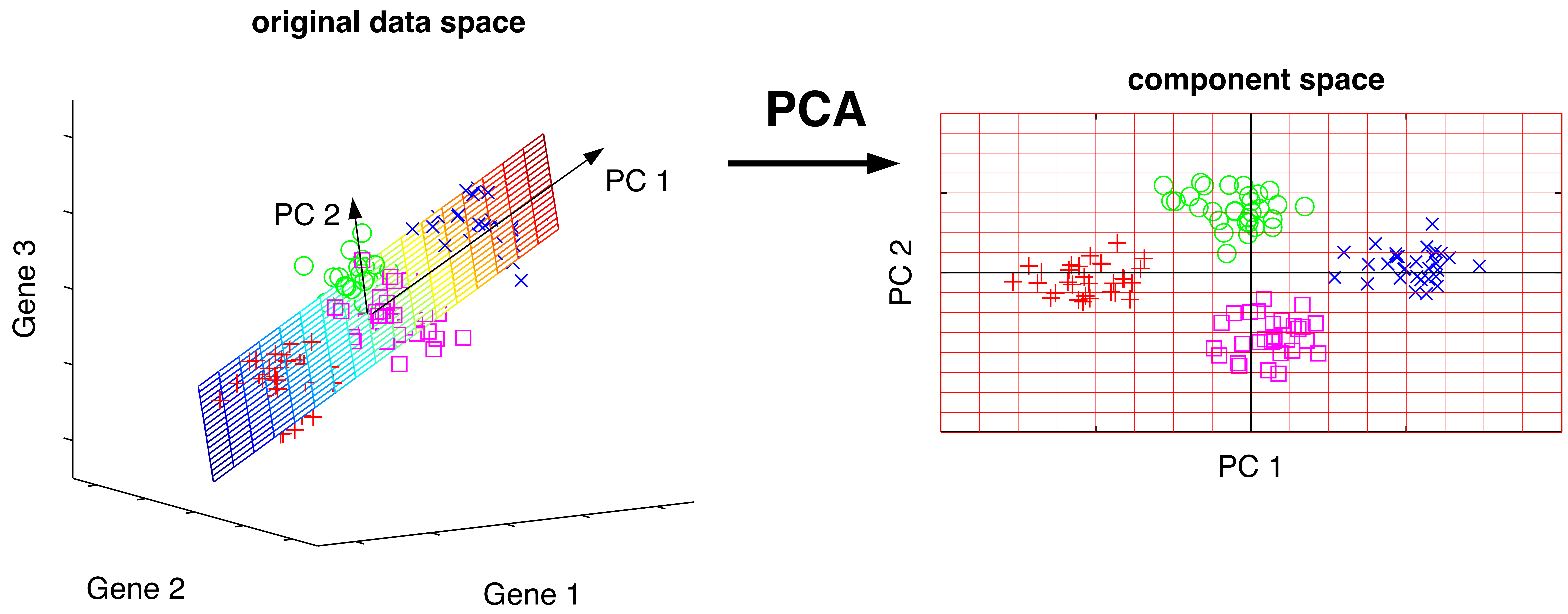
Number of Clusters

[D. Robinson, 2015]

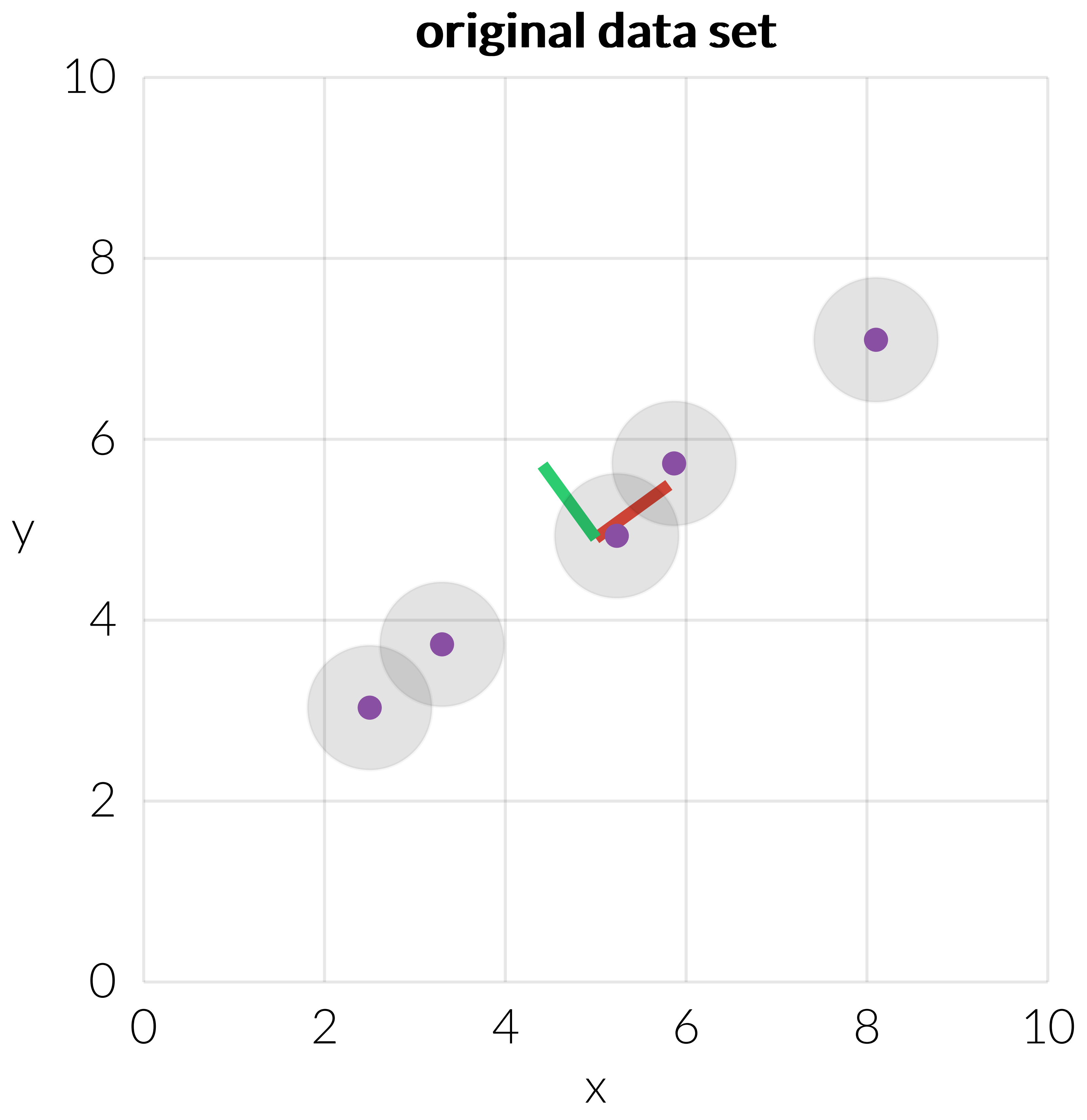
Dimensionality Reduction

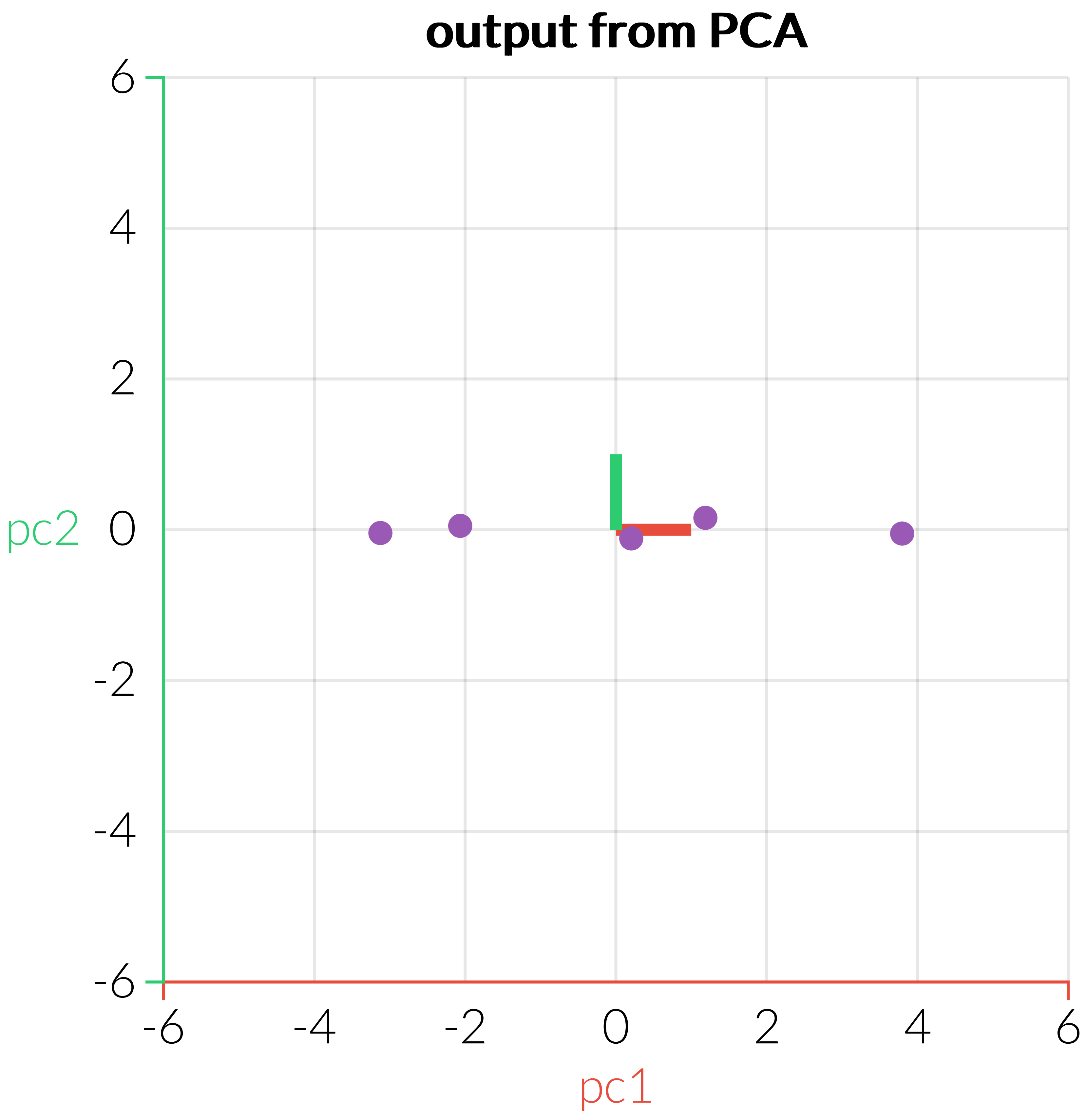
- Attribute Aggregation: Use fewer attributes (dimensions) to represent items
- Combine attributes in a way that is more instructive than examining each individual attribute
- Example: Understanding the language in a collection of books
  - Count the occurrence of each non-common word in each book
  - Huge set of features (attributes), want to represent each with an aggregate feature (e.g. high use of "cowboy", lower use of "city") that allows clustering (e.g. "western")
  - Don't want to have to manually determine such rules
- Techniques: Principle Component Analysis, Multidimensional Scaling family of techniques

Principle Component Analysis (PCA)



PCA

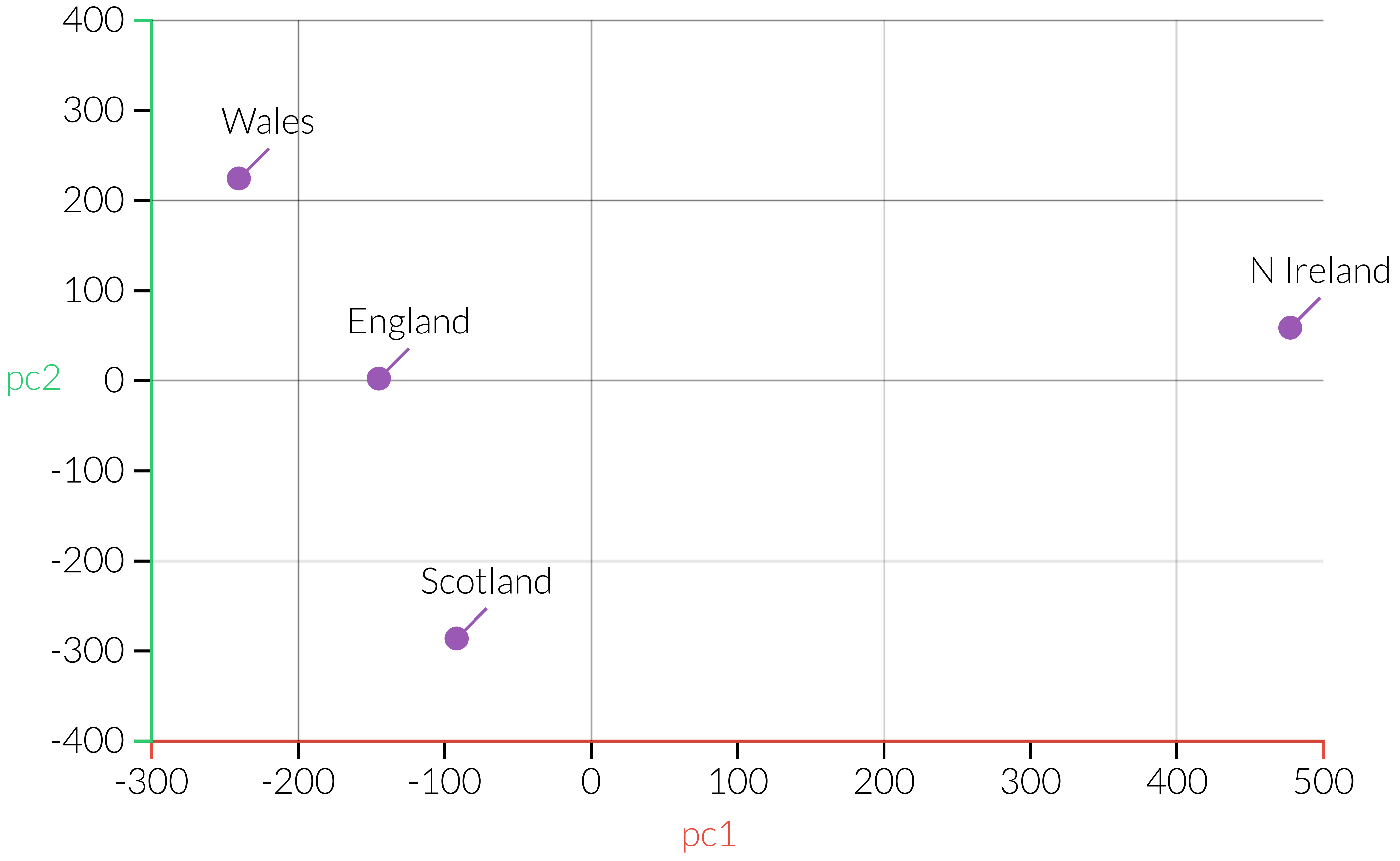




[Principle Component Analysis Explained, Explained Visually, V. Powell & L. Lehe, 2015]

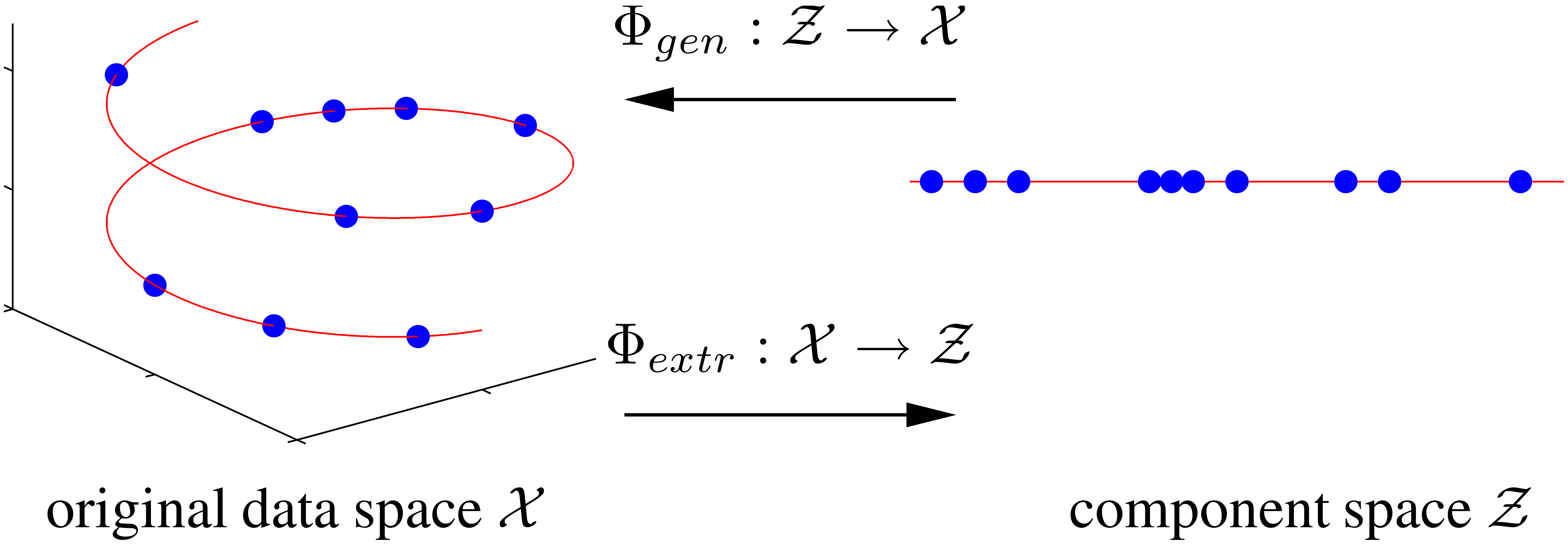
17 dimensions to 2

	England	N Ireland	Scotland	Wales
Alcoholic drinks	375	135	458	475
Beverages	57	47	53	73
Carcase meat	245	267	242	227
Cereals	1472	1494	1462	1582
Cheese	105	66	103	103
Confectionery	54	41	62	64
Fats and oils	193	209	184	235
Fish	147	93	122	160
Fresh fruit	1102	674	957	1137
Fresh potatoes	720	1033	566	874
Fresh Veg	253	143	171	265
Other meat	685	586	750	803
Other Veg	488	355	418	570
Processed potatoes	198	187	220	203
Processed Veg	360	334	337	365
Soft drinks	1374	1506	1572	1256
Sugars	156	139	147	175

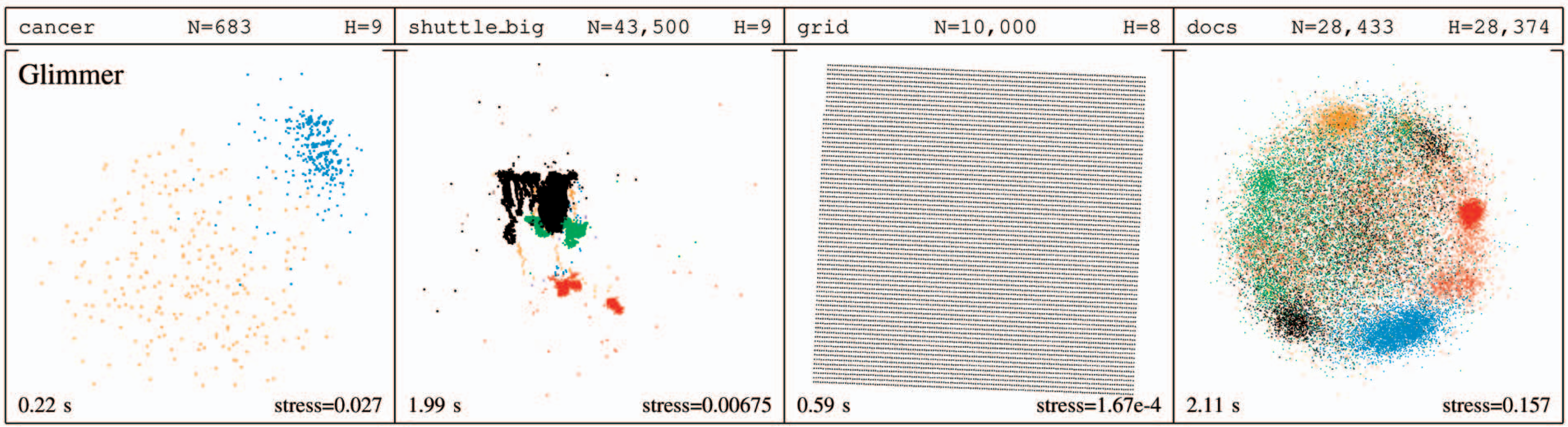


[Principle Component Analysis Explained, Explained Visually, V. Powell & L. Lehe, 2015]

Non-linear Dimensionality Reduction

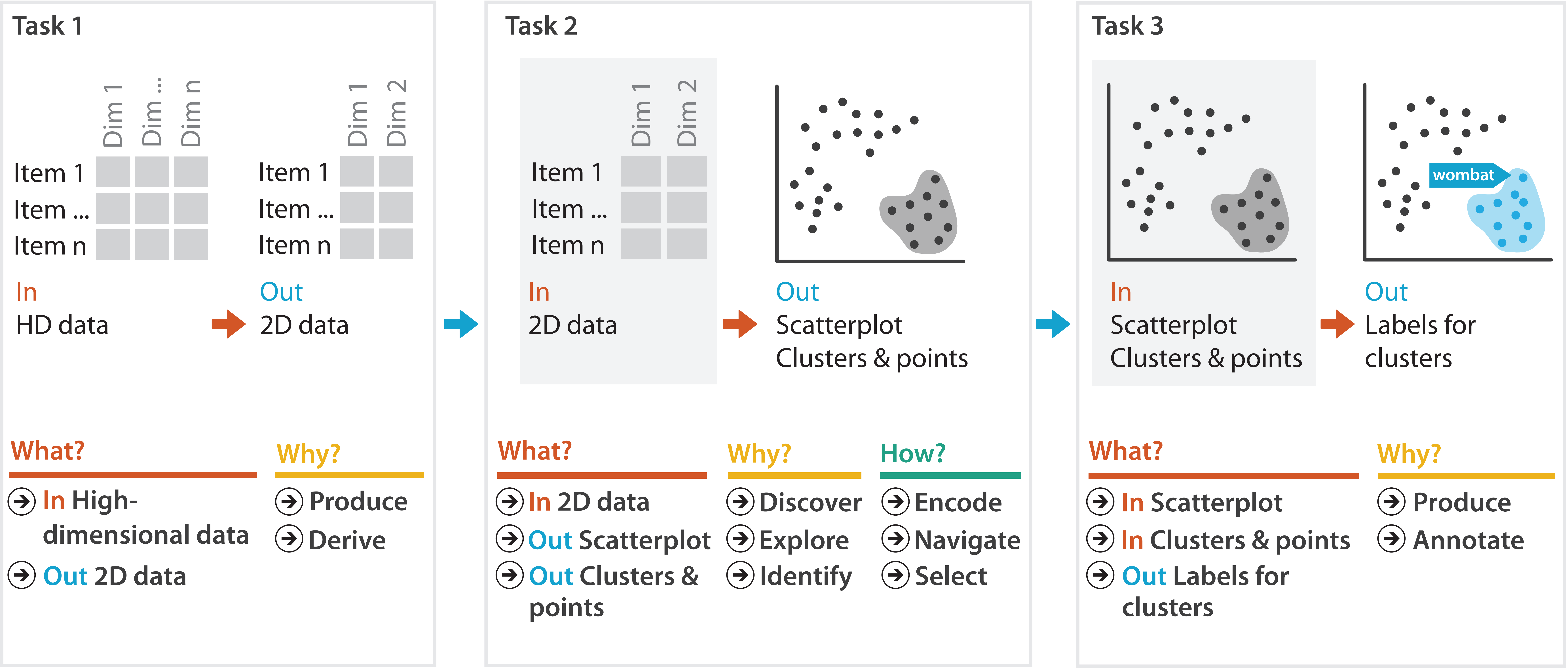


Dimensionality Reduction in Visualization



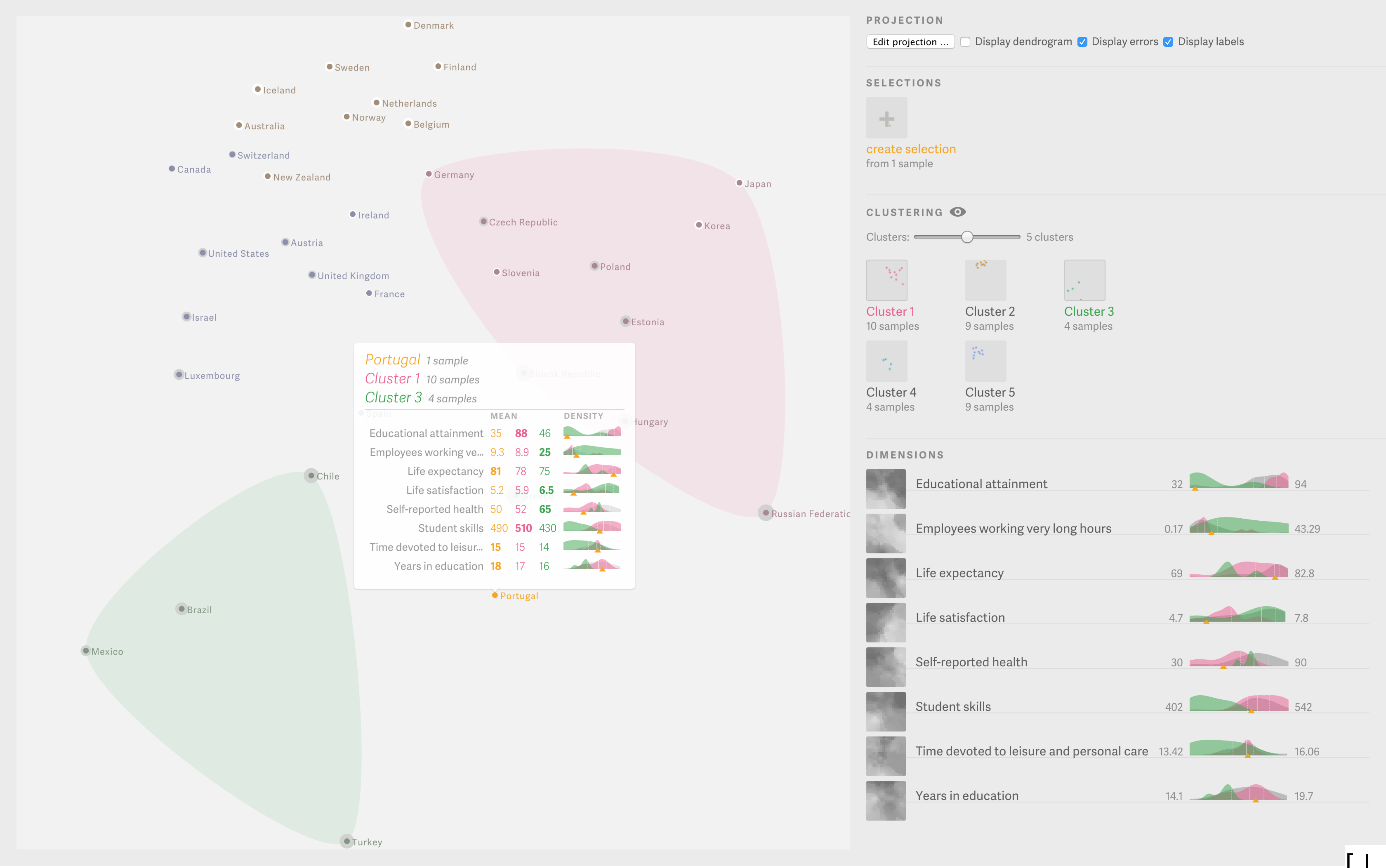
[Glimmer, Ingram et al., 2009]

Tasks in Understanding High-Dim. Data



[Munzner (ill. Maguire), 2014]

Probing Projections



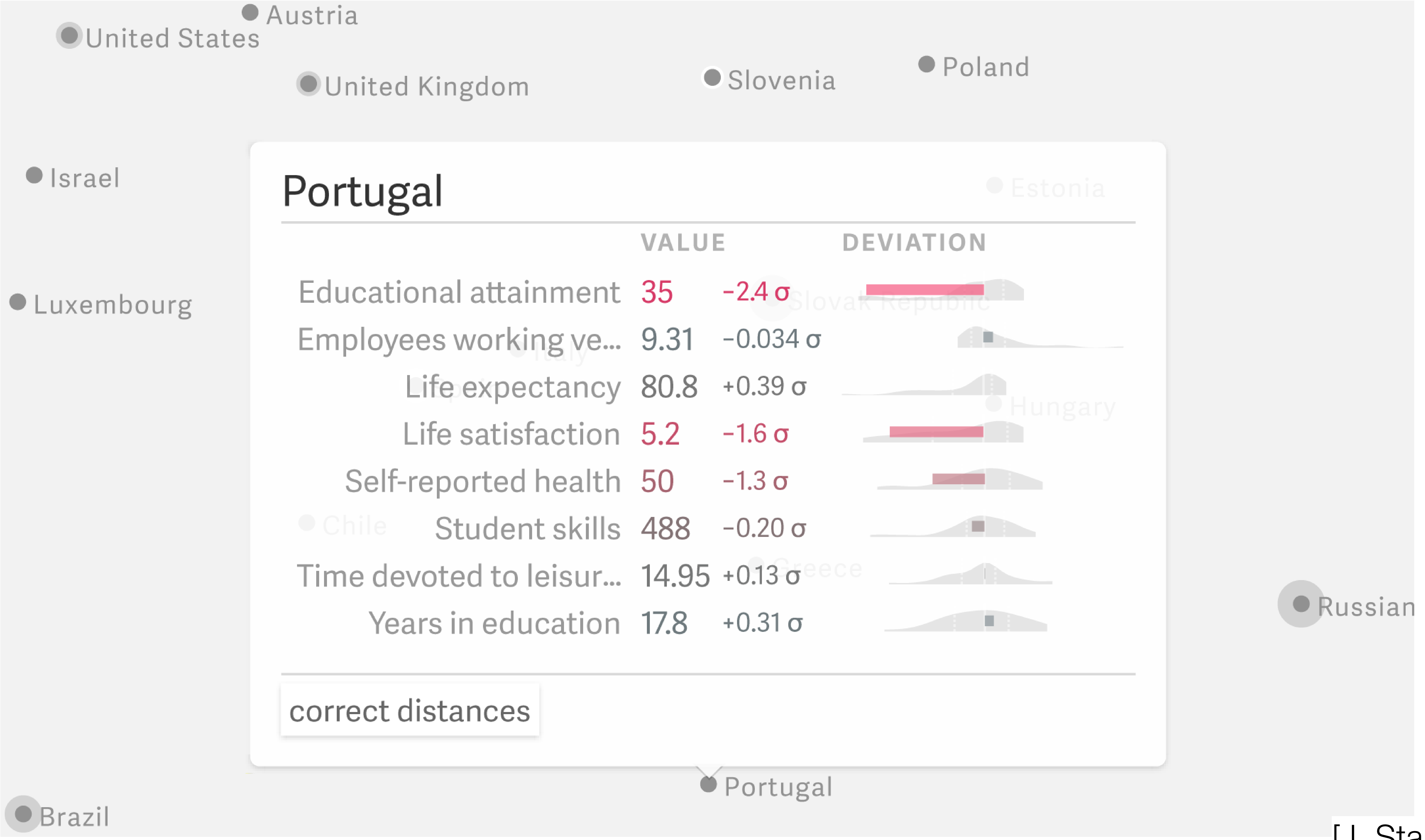
[J. Stahnke et al., 2015]

Probing Projection Goals

- Examining the Projection
- Exploring the Data
- Design Goals:
  - Show and correct approximation errors
  - Allow for multi-level comparisons
  - Spatial orientation
  - Consistent design
- Allow **grouping** of samples
  - Selections
  - Classes
  - Clusters

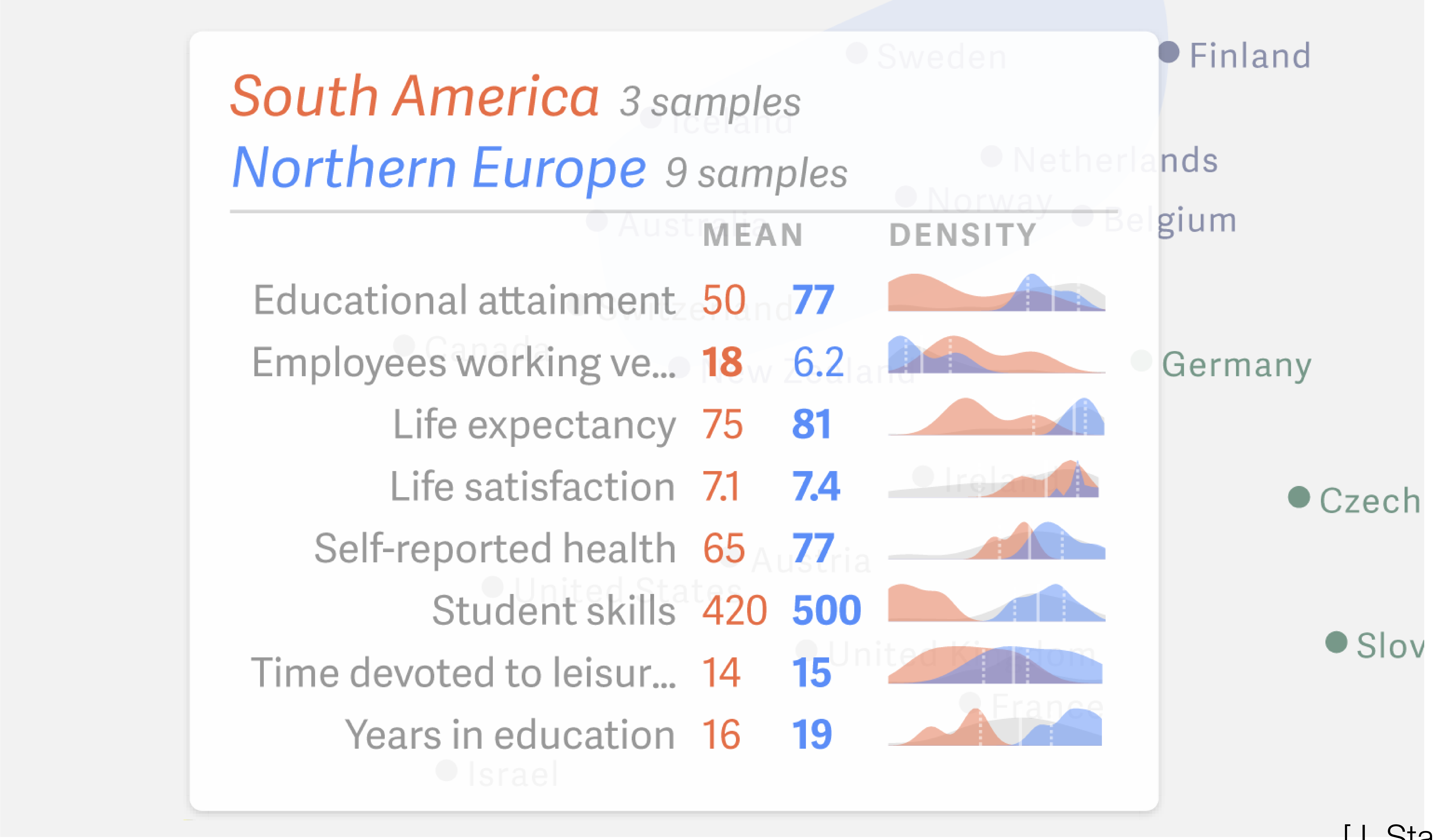
[J. Stahnke et al., 2015]

Tooltips with statistics



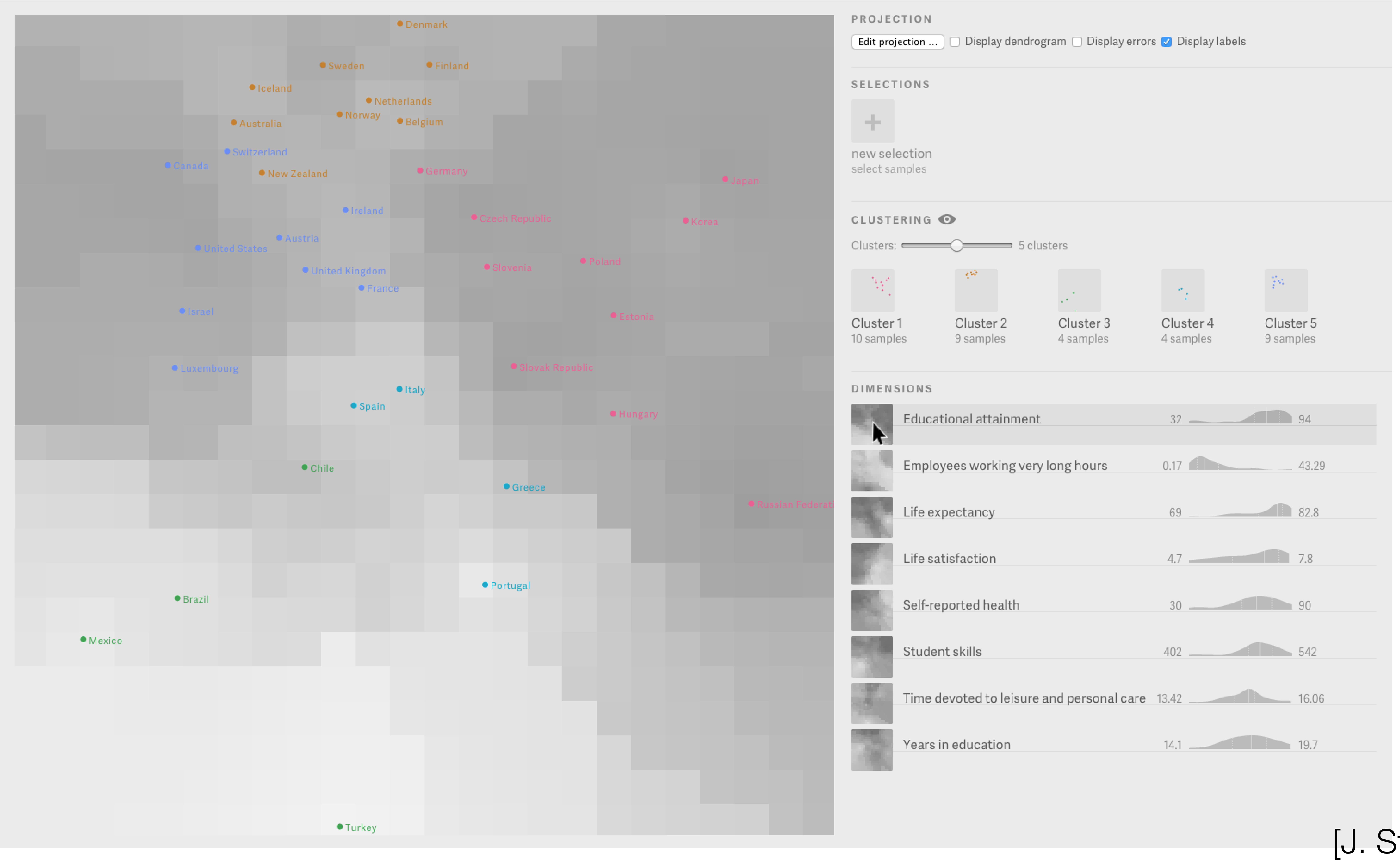
[J. Stahnke et al., 2015]

Comparing Two Groups



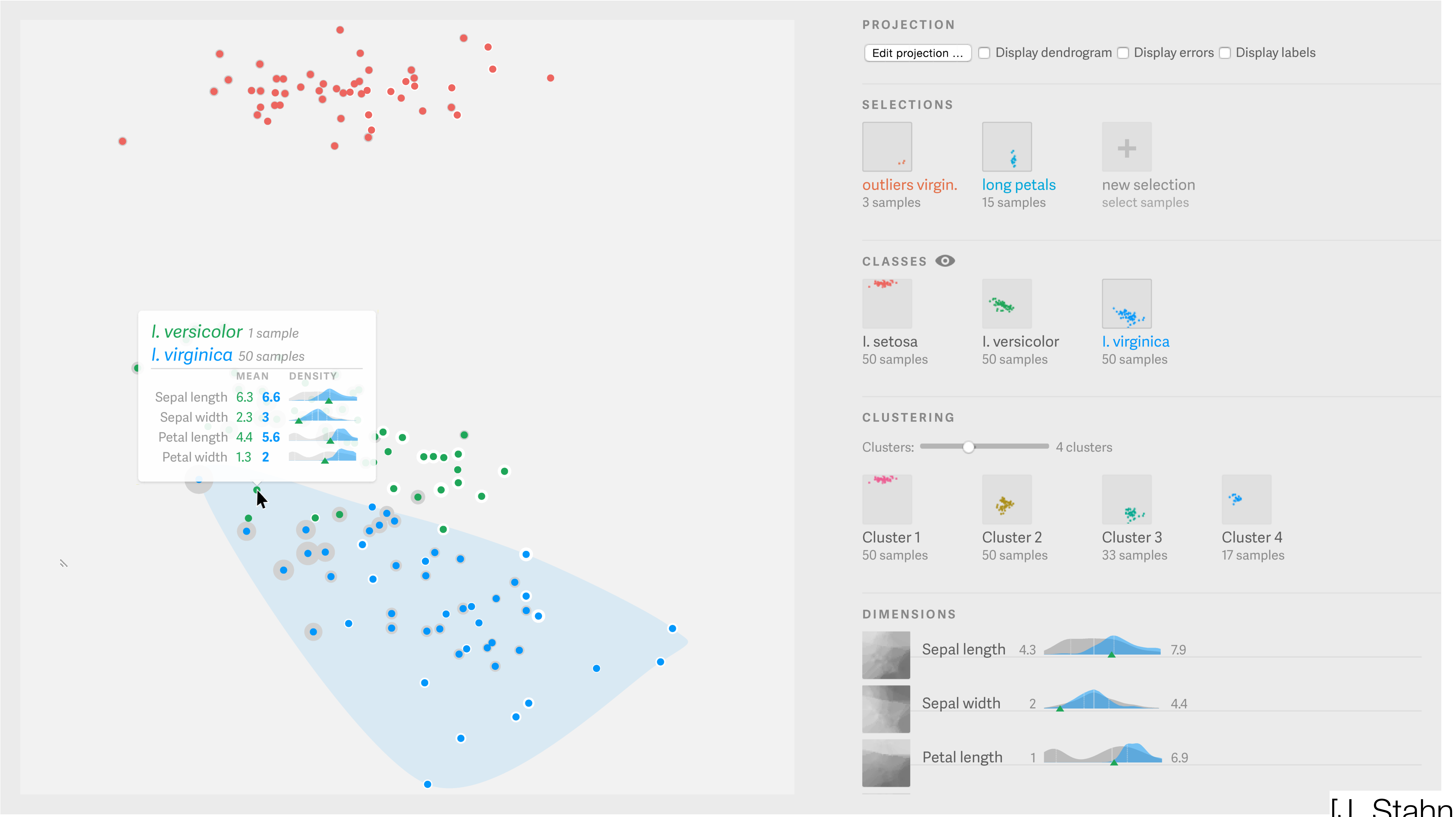
[J. Stahnke et al., 2015]

Heatmap from Dimension Hover



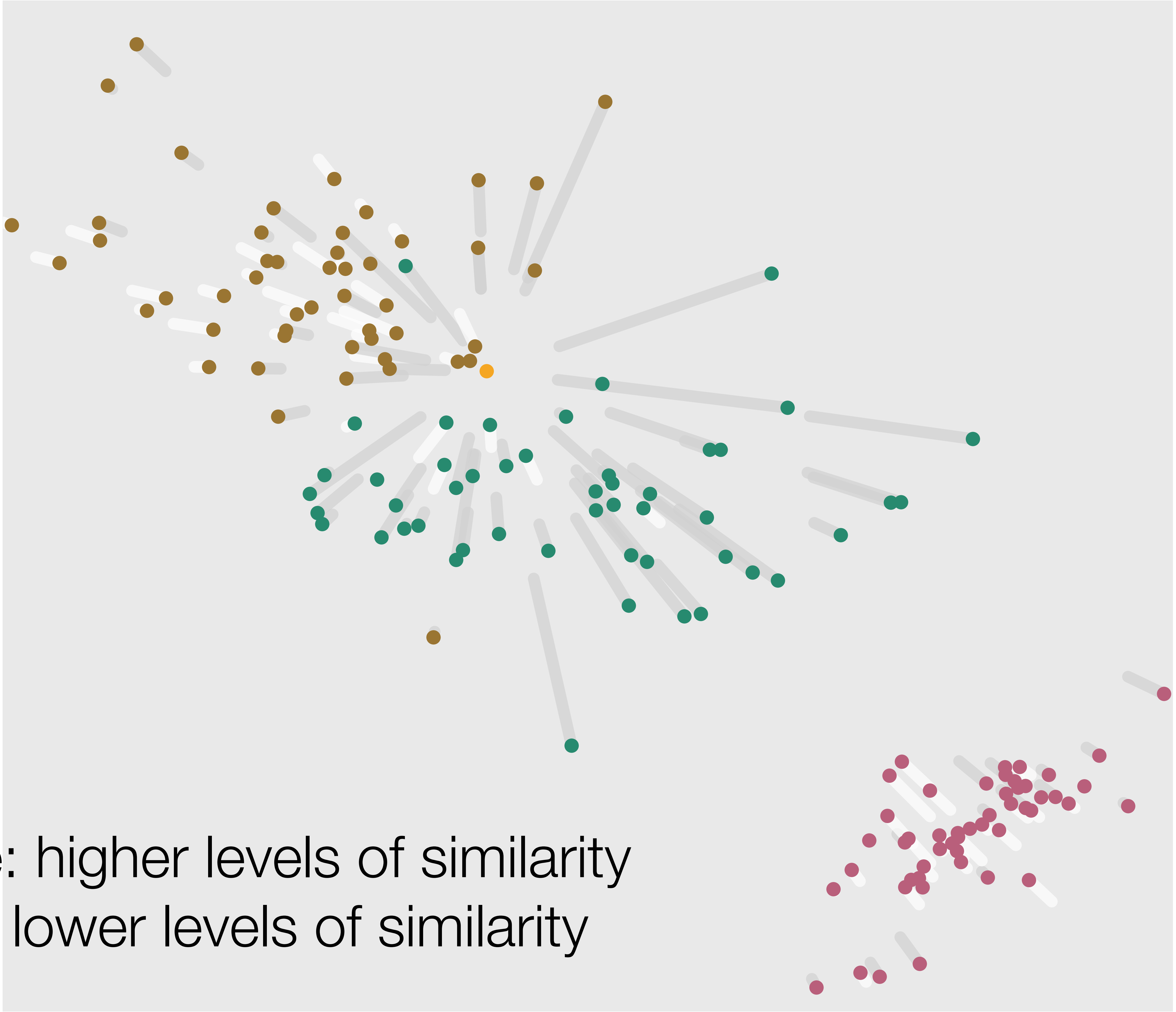
[J. Stahnke et al., 2015]

Showing Error via Sample-centric Halos



[J. Stahnke et al., 2015]

Showing Projection Errors



White: higher levels of similarity
Gray: lower levels of similarity

[J. Stahnke et al., 2015]

User Study & Results

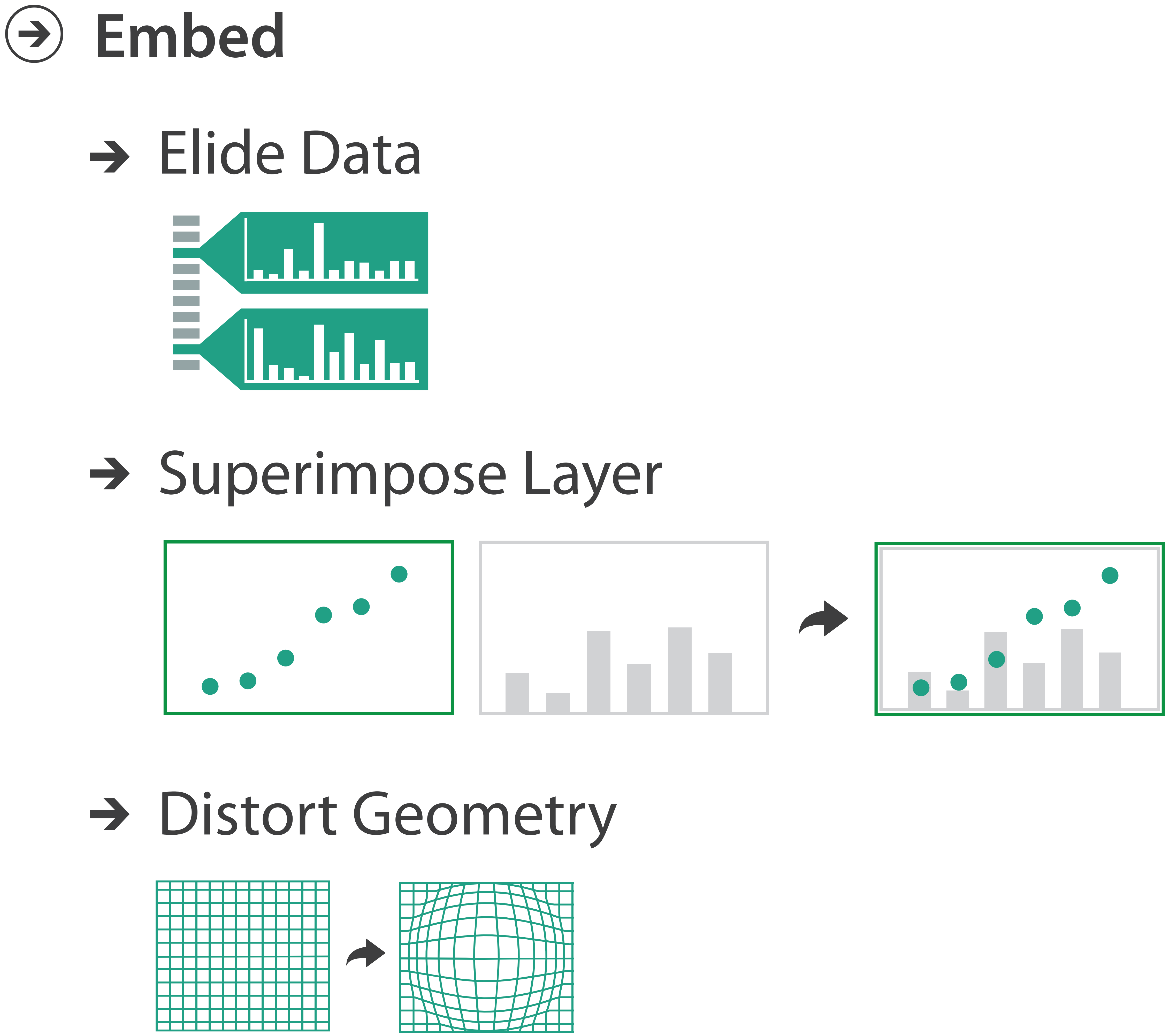
- Types of Questions:
  - How would you try to characterize the type X?
  - In what way are X and Y different in their properties?
  - Are the projections of X and Y correct or do they deviate? How do you interpret this?
  - Can you discover which parts of the cluster combinations are A, B, and C?
- Discussion:
  - Learnability: need more effective mechanisms for grasping the concepts behind dimensionality reduction
  - Manipulation: What happens with results?
  - Large data: What about text corpora?

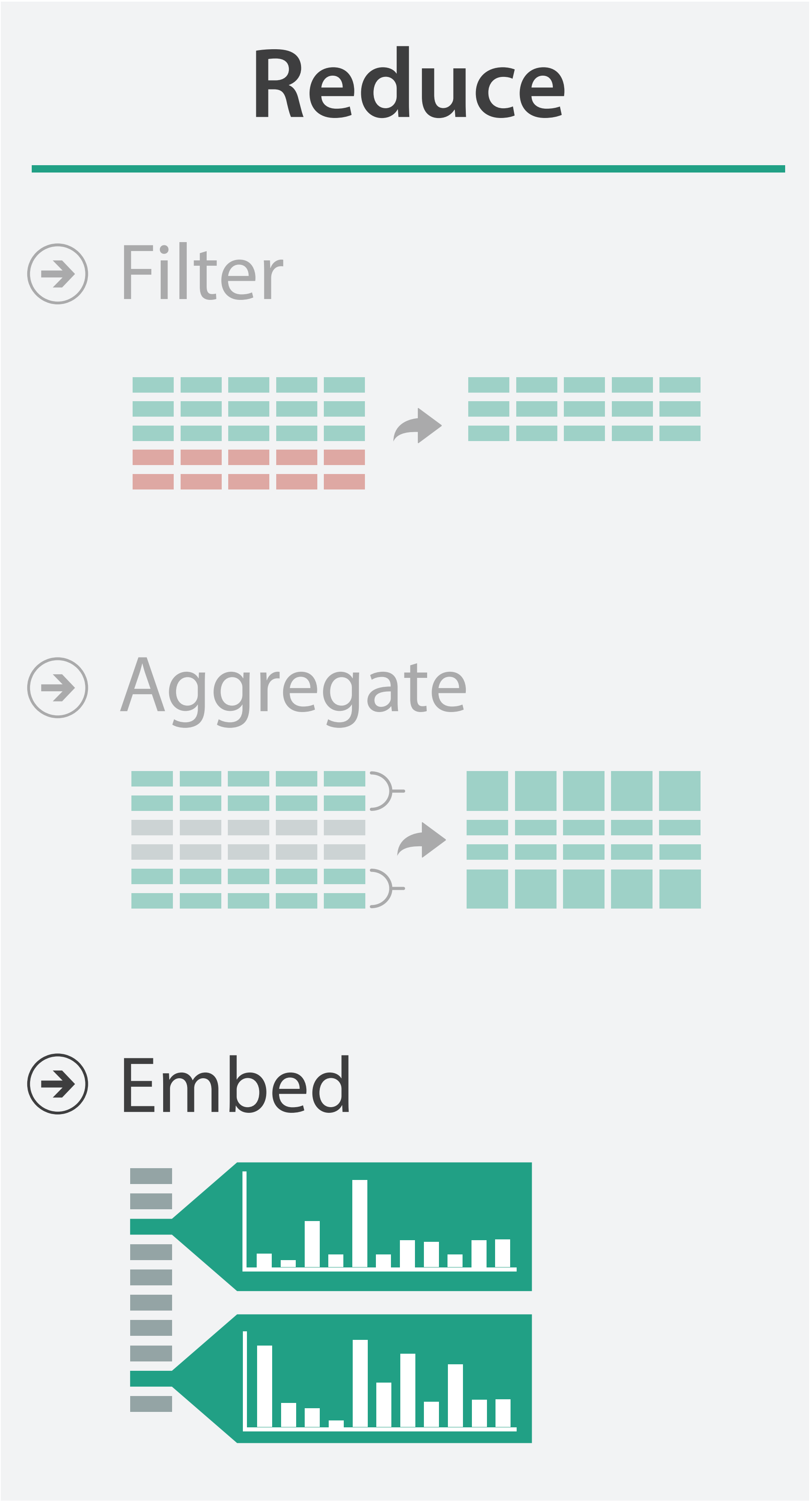
[J. Stahnke et al., 2015]

Focus+Context

- Show everything at once but compress regions that are not the current focus
  - User shouldn't lose sight of the overall picture
  - May involve some aggregation in non-focused regions
  - "Nonliteral navigation" like semantic zooming
- Elision
- Superimposition: more directly tied than with layers
- Distortion

Focus+Content Overview





[Munzner (ill. Maguire), 2014]

Elision

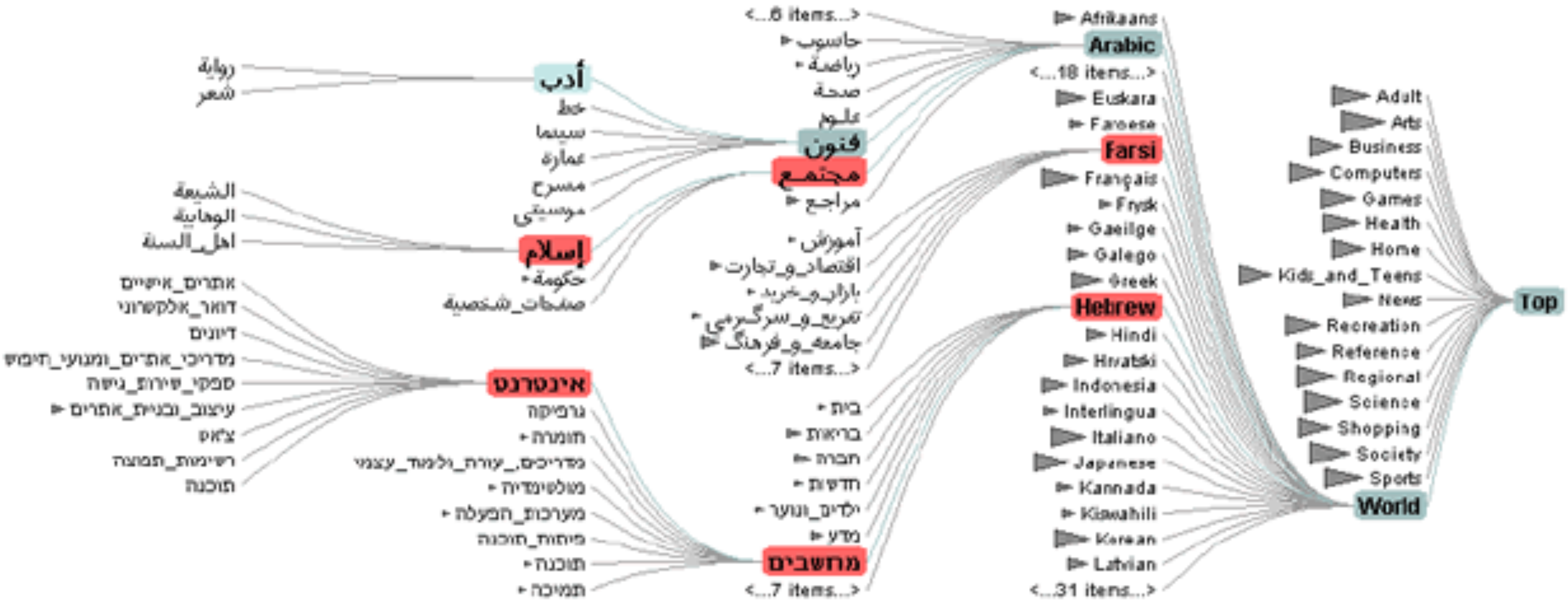
- There are a number of examples of elision including in text , DOITrees, …
- Includes both filtering and aggregation but goal is to give overall view of the data
- In visualization, usually correlated with focus regions

Degree of Interest Function

- DOI = I(x) - D(x,y)
  - I: interest function
  - D: distance (semantic or spatial)
  - x: location of item
  - y: current focus point (could be more than one)
- Interactive: y changes

Elision: DOITrees

- Example: 600,000 node tree
  - Multiple foci (from search results or via user selection)
  - Distance computed topologically (levels, not geometric)

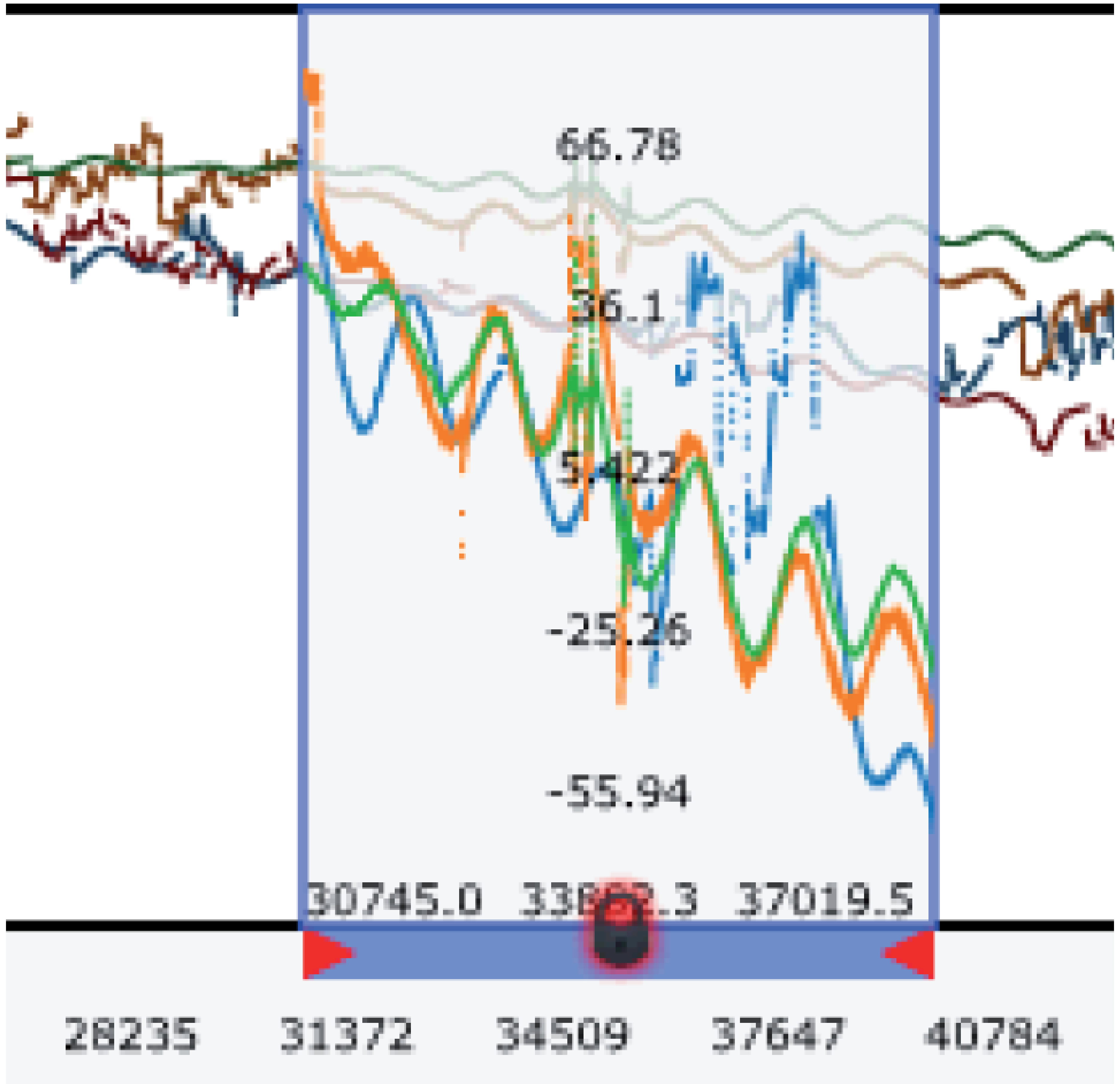


[Heer and Card, 2004]

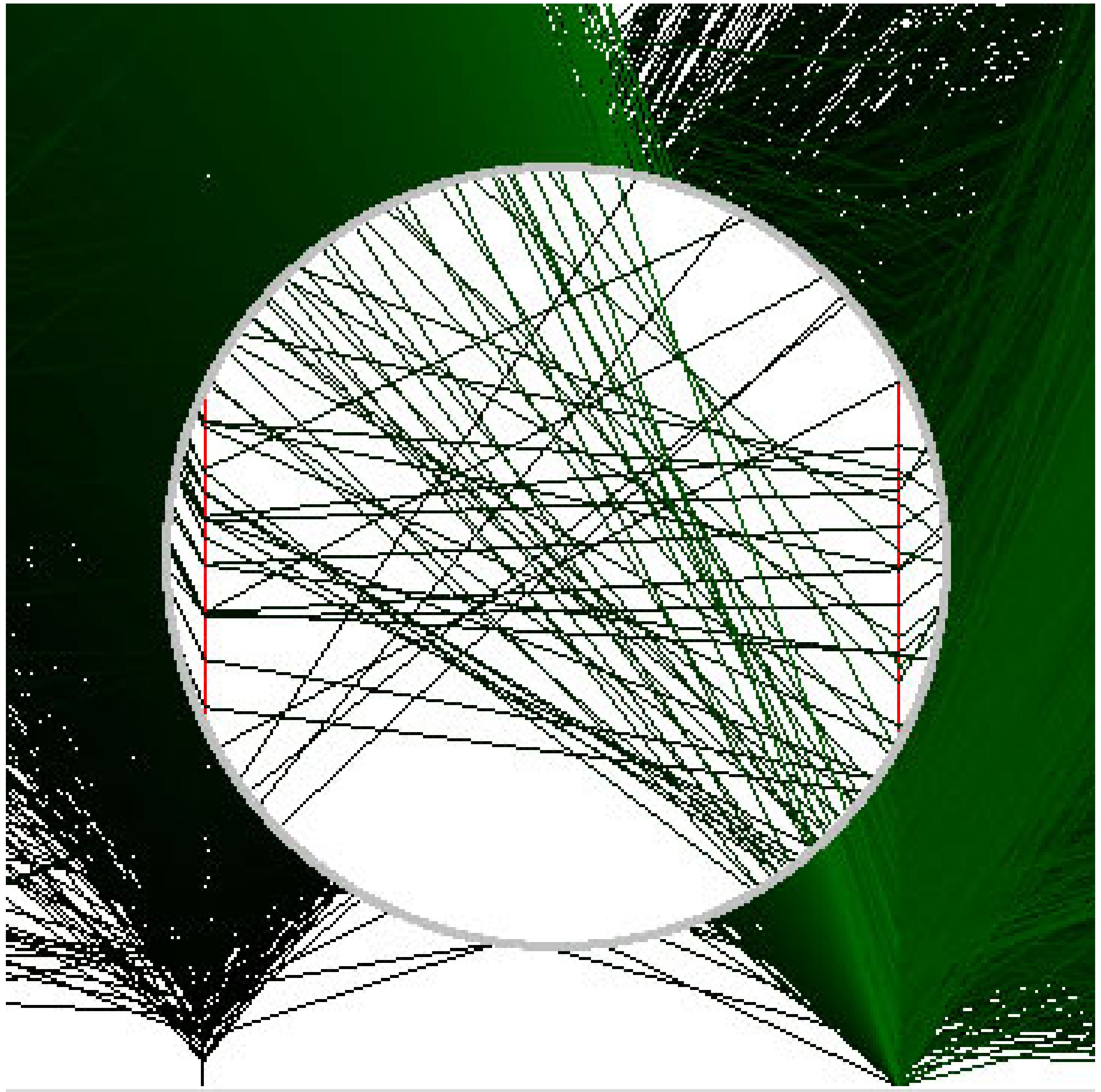
Superimposition

- Different from layers because this is restricted to a particular region
  - For Focus+Context, superimposition is **not global**
  - More like overloading
- Lens may occlude the layer below

Superimposition with Interactive Lenses



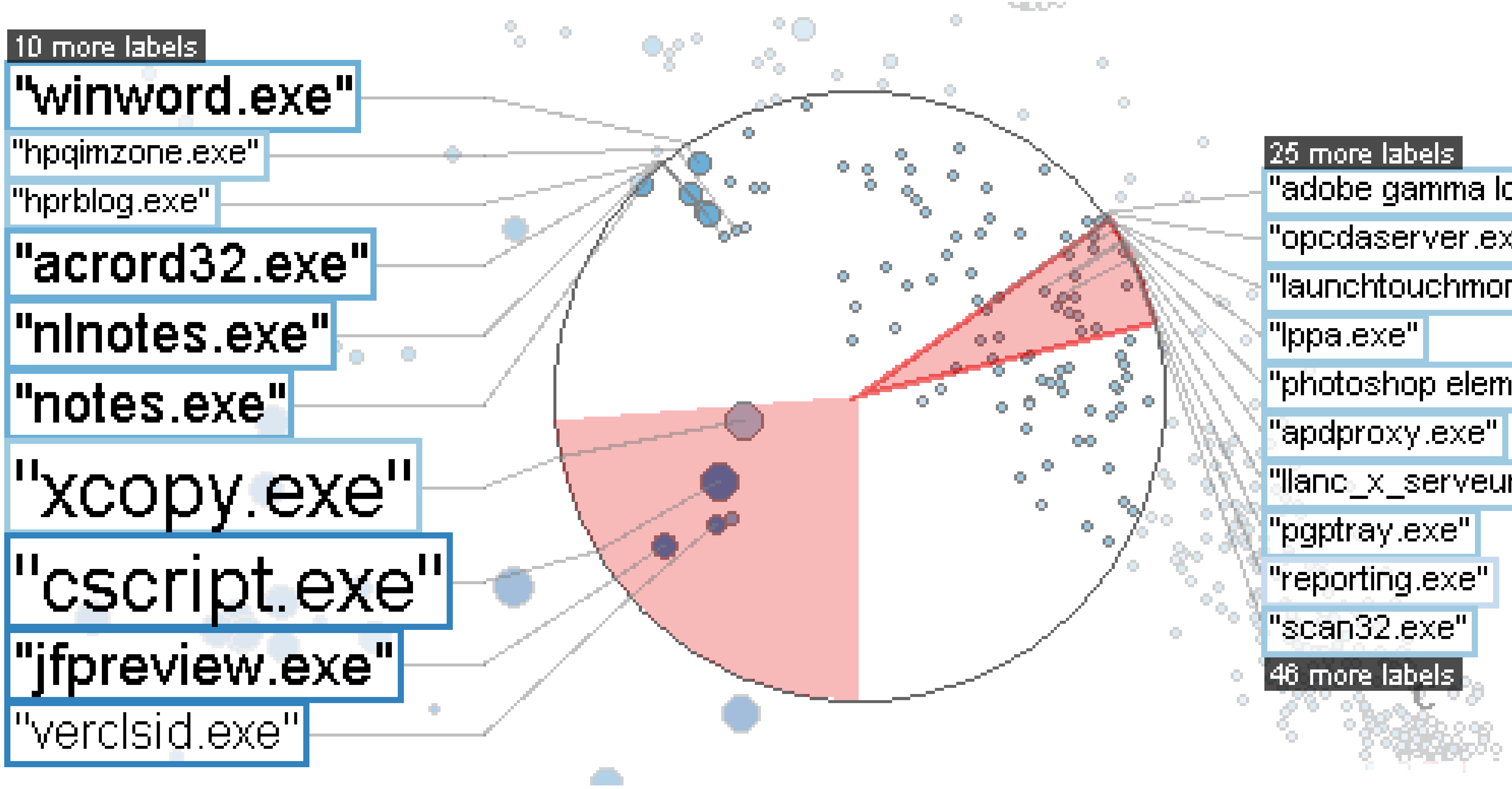
(a) Alteration



(b) Suppression

[ChronoLenses and Sampling Lens in Tominski et al., 2014]

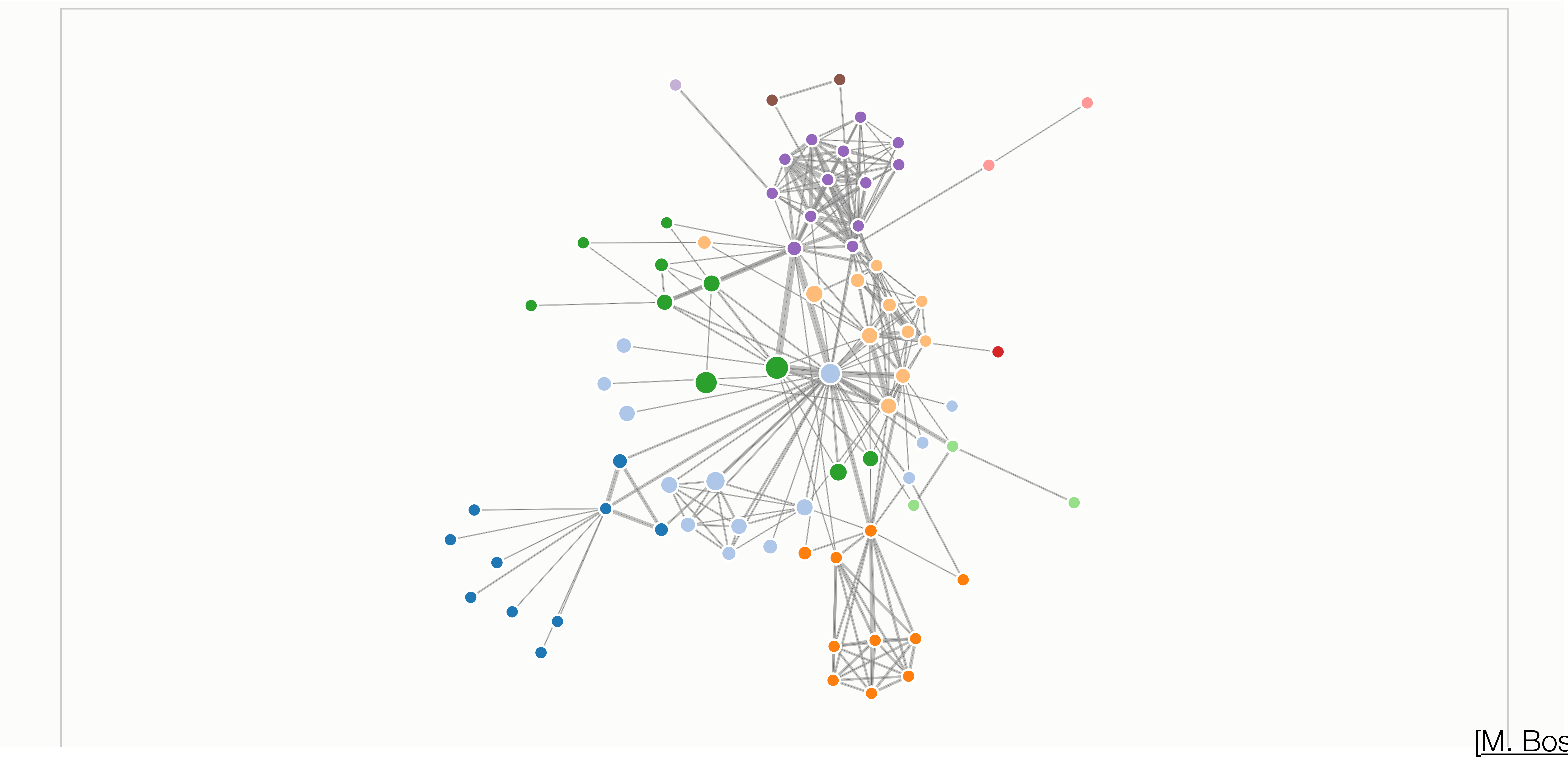
Superimposition with Interactive



(c) Enrichment

[Extended Lens in Tominski et al., 2014]

Distortion

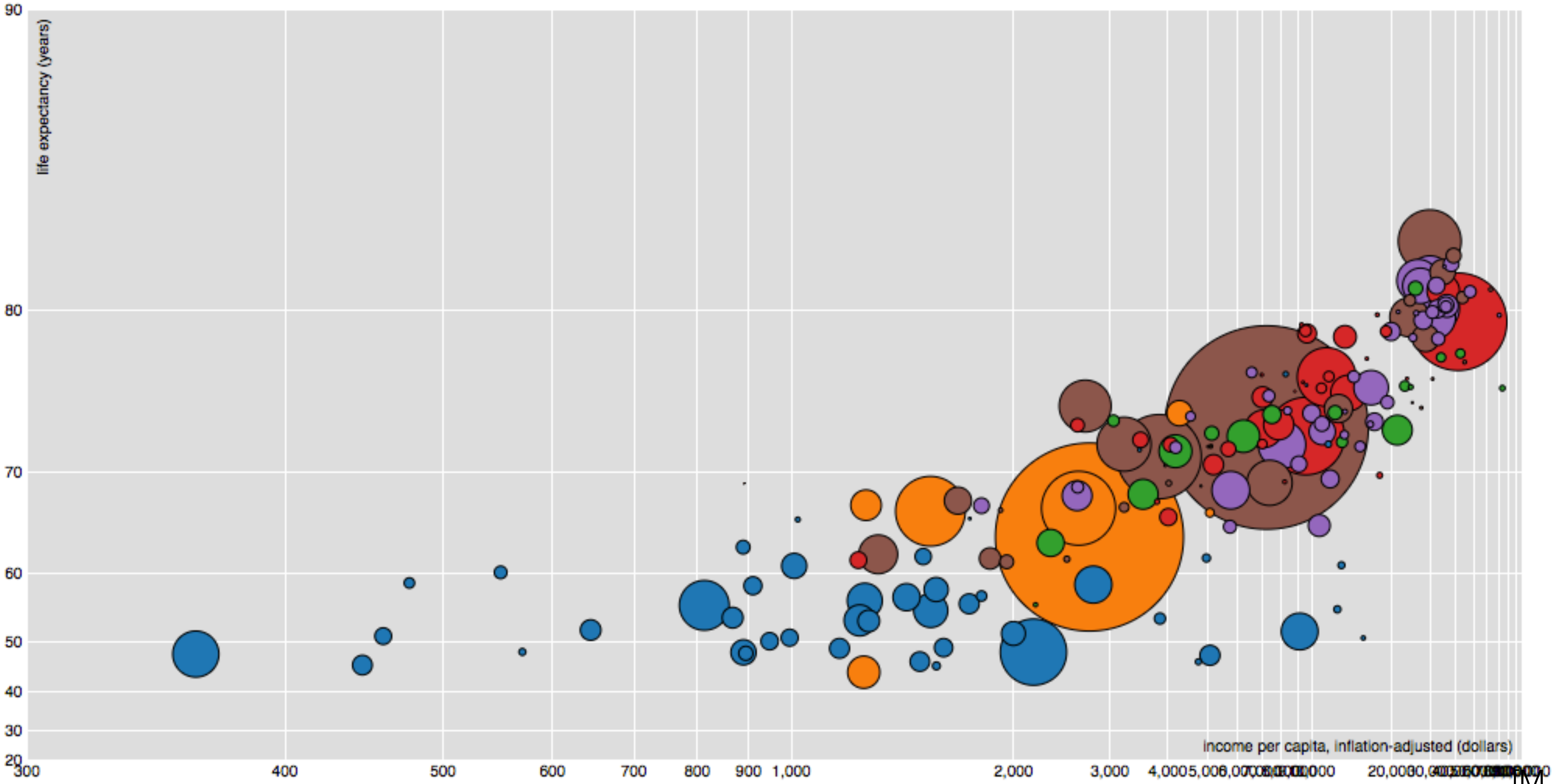


[M. Bostock]

Distortion Choices

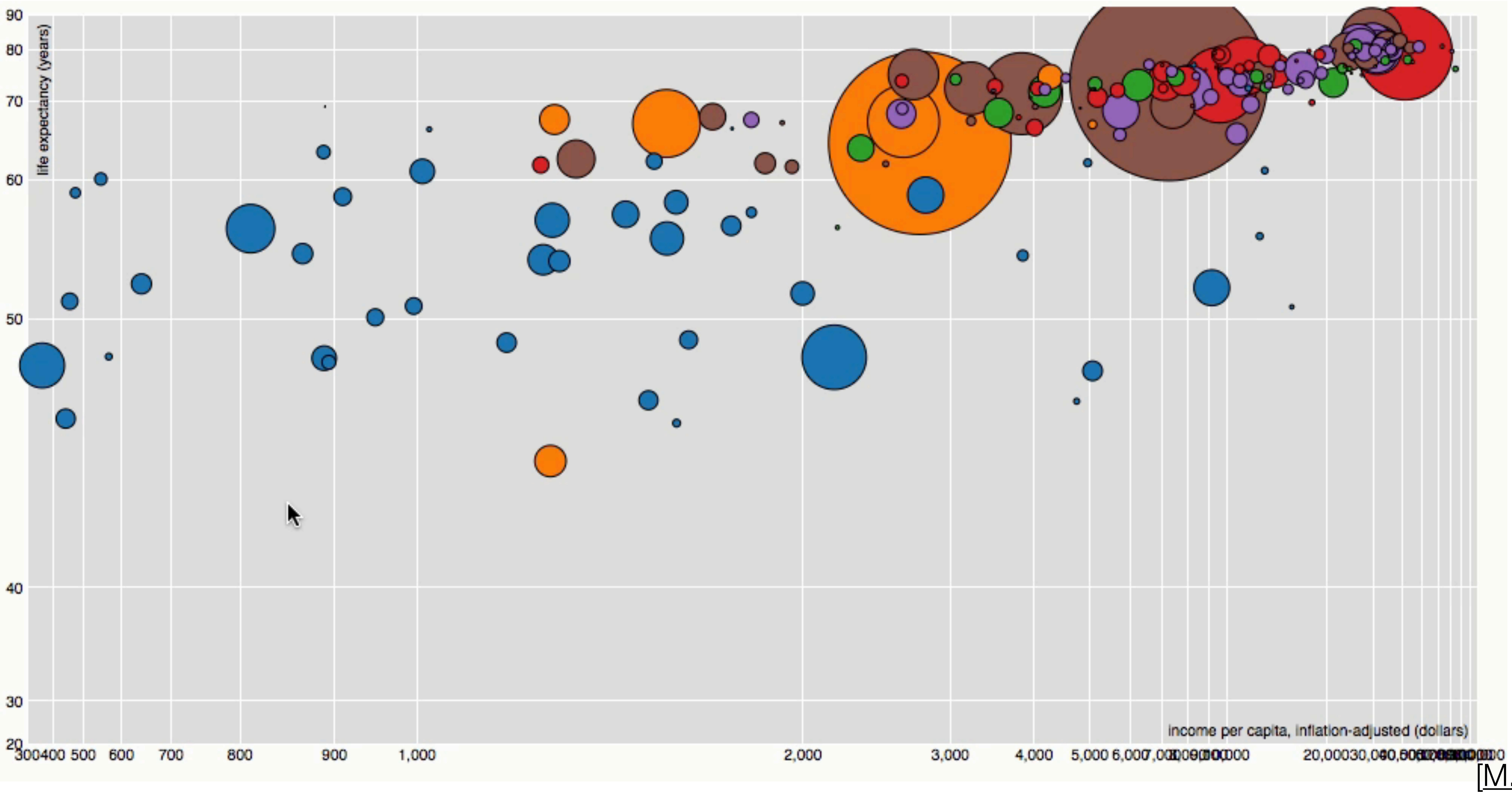
- How many focus regions? One or Multiple
- Shape of the focus?
  - Radial
  - Rectangular
  - Other
- Extent of the focus
  - Constrained similar to magic lenses
  - Entire view changes
- Type of interaction: Geometric, moveable lenses, rubber sheet

Overplotting



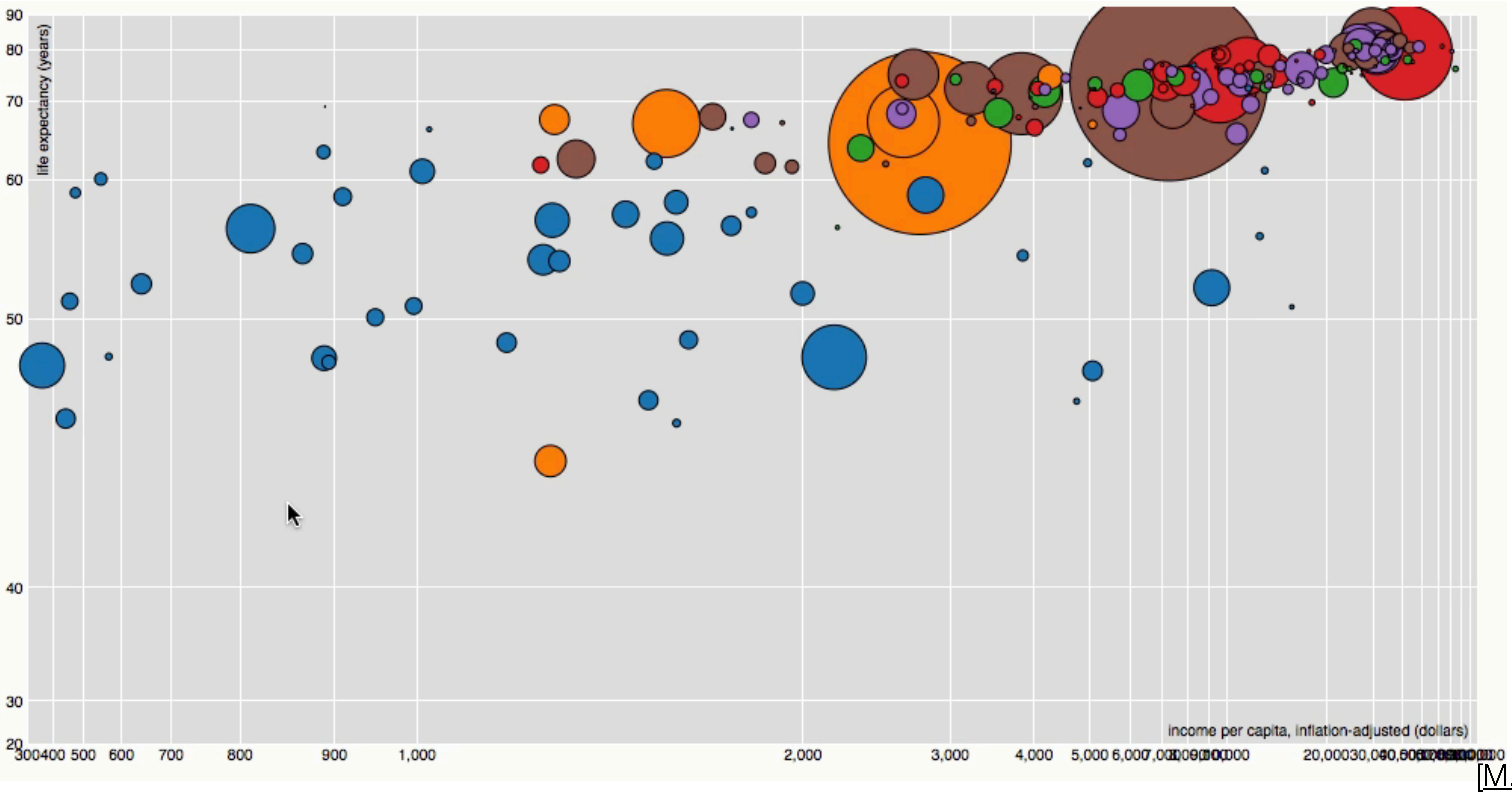
[M. Bostock]

Cartesian Distortion



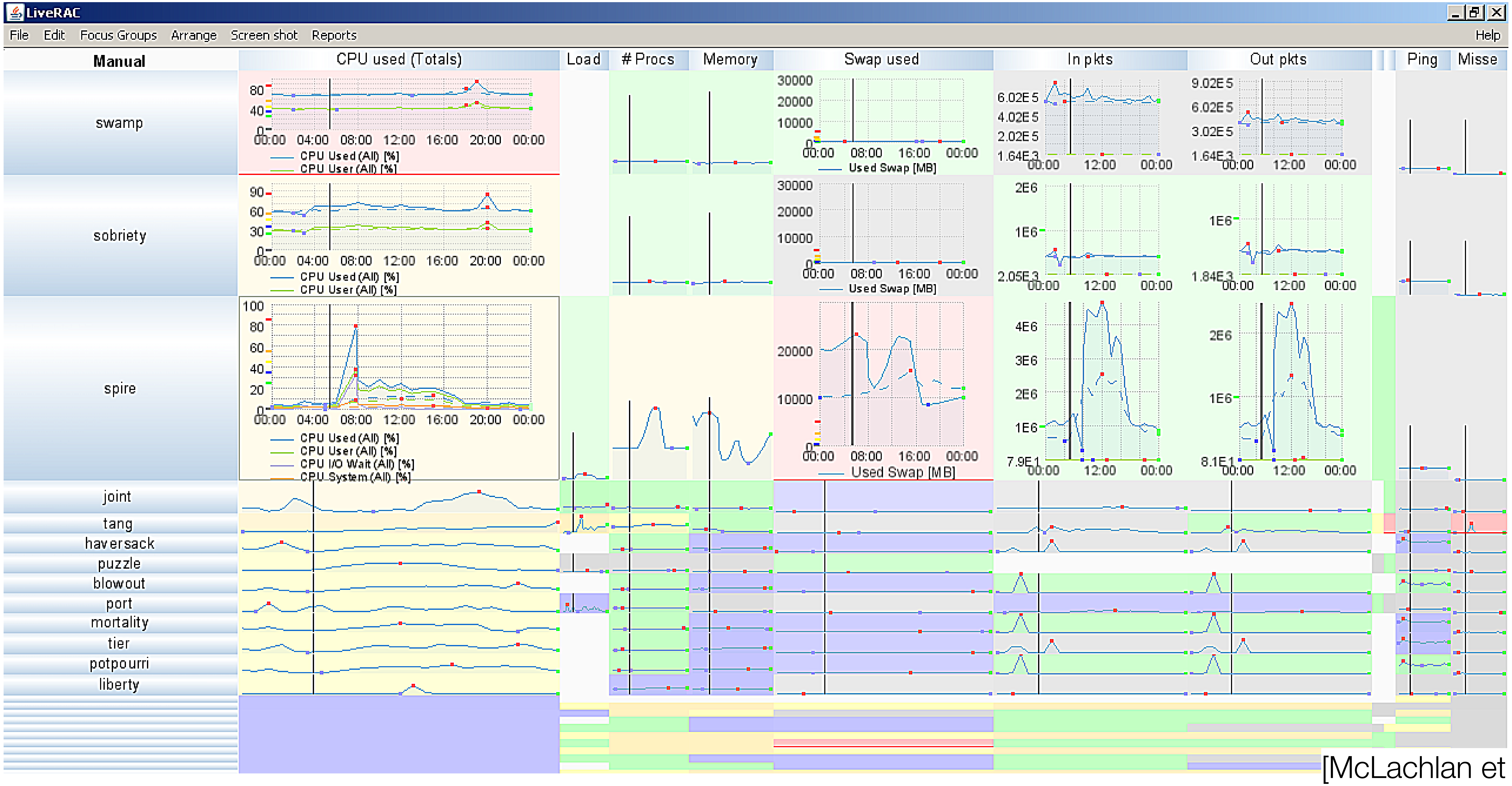
[M. Bostock]

Cartesian Distortion



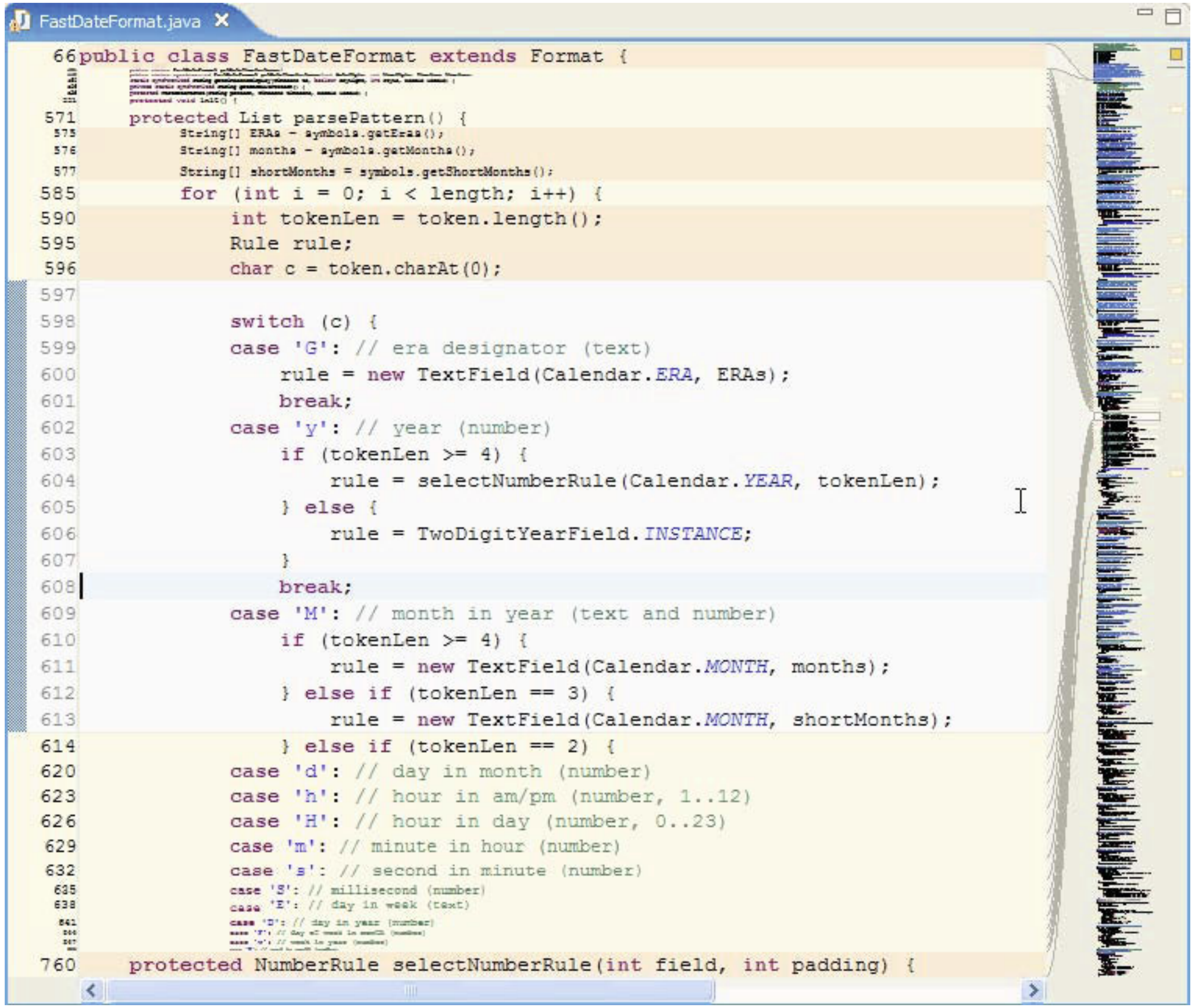
[M. Bostock]

Stretch and Squish Navigation



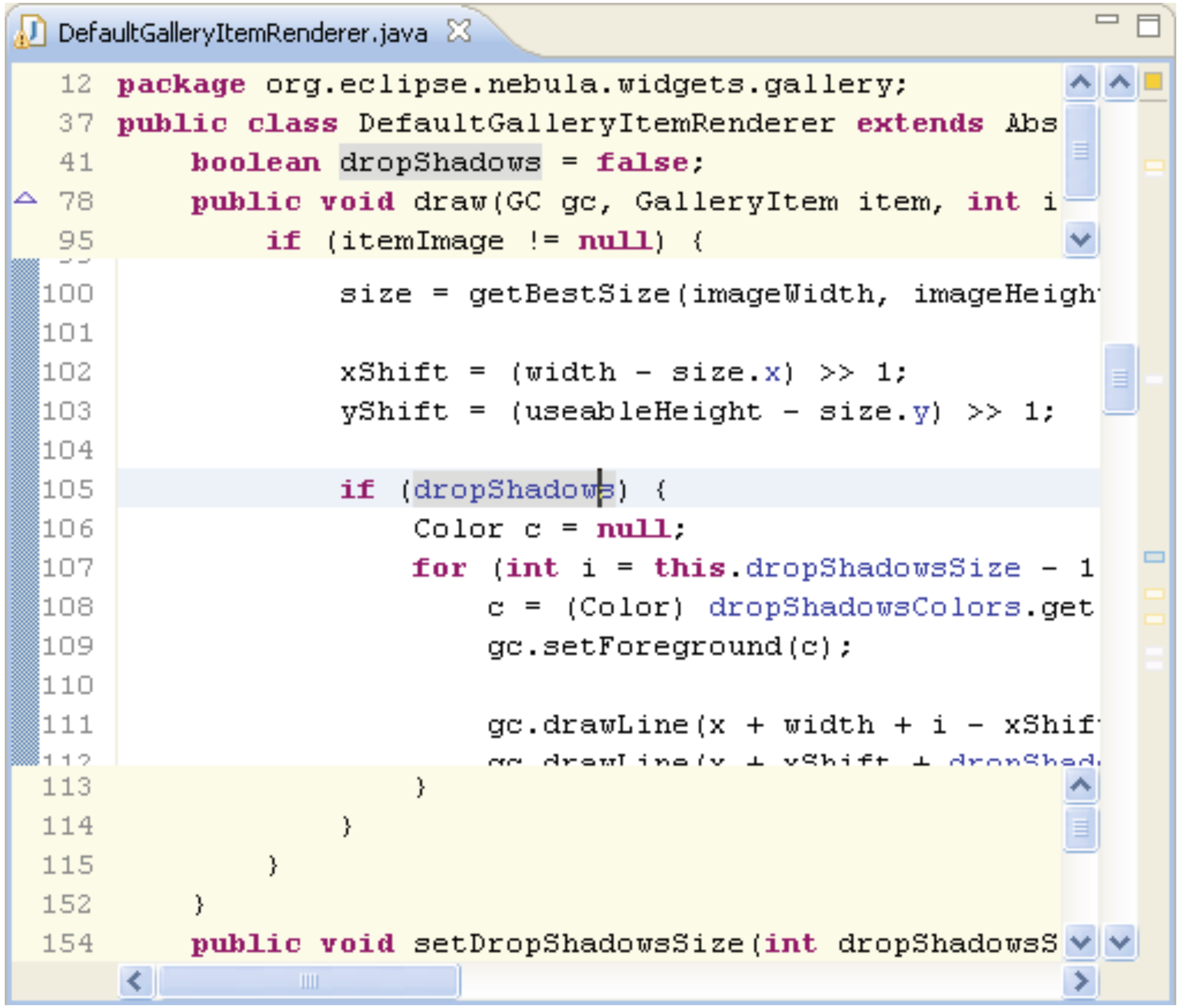
[McLachlan et al., 2008]

Fisheye Distortion in Programming



[Jakobsen and Hornbaek, 2011]

Distortion vs. Hide



[Jakobsen and Hornbaek, 2011]

Research Questions

- Is a priori importance useful (and for what)?
- What does the user focus on?
  - predictability of view changes when focus changes
  - how direct user control is
  - task & context
- What interesting information should be displayed
  - degree of interest function may produce varied result sizes
- Do fisheye views integrate or disintegrate?
  - interference with other interactions; allow on-demand use?
- Are fisheye views suitable for large displays?

[Jakobsen and Hornbaek, 2011]

Distortion Concerns

- Distance and length judgments are **harder**
  - Example: Mac OS X Dock with Magnification
  - Spatial position of items changes as the focus changes
- Node-link diagrams not an issue… why?
- Users have to be made aware of distortion
  - Back to scatterplot with distortion example
  - Lenses or shading give clues to users
- **Object constancy**: understanding when two views show the same object
  - What happens under distortion?
  - 3D Perspective is distortion… but we are well-trained for that
- Think about **what** is being shown (filtering) and method (fisheye)