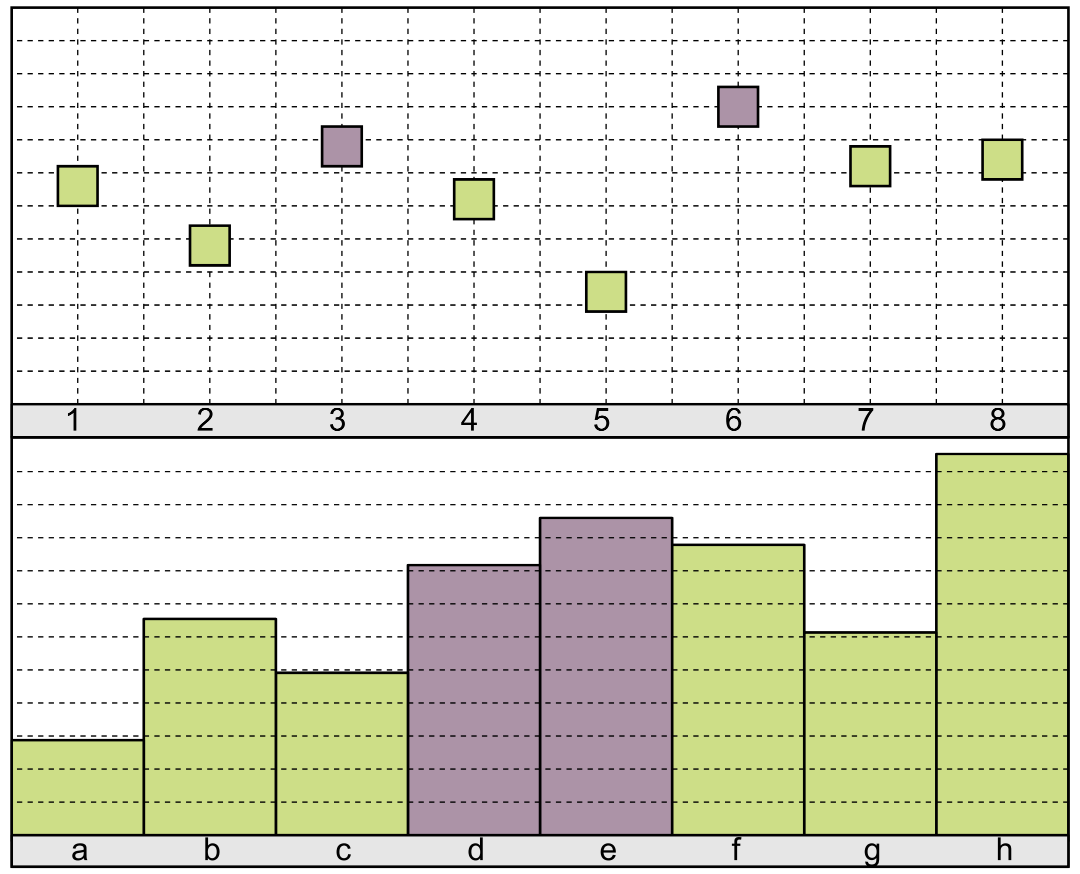
Data Visualization (CSCI 490/680)

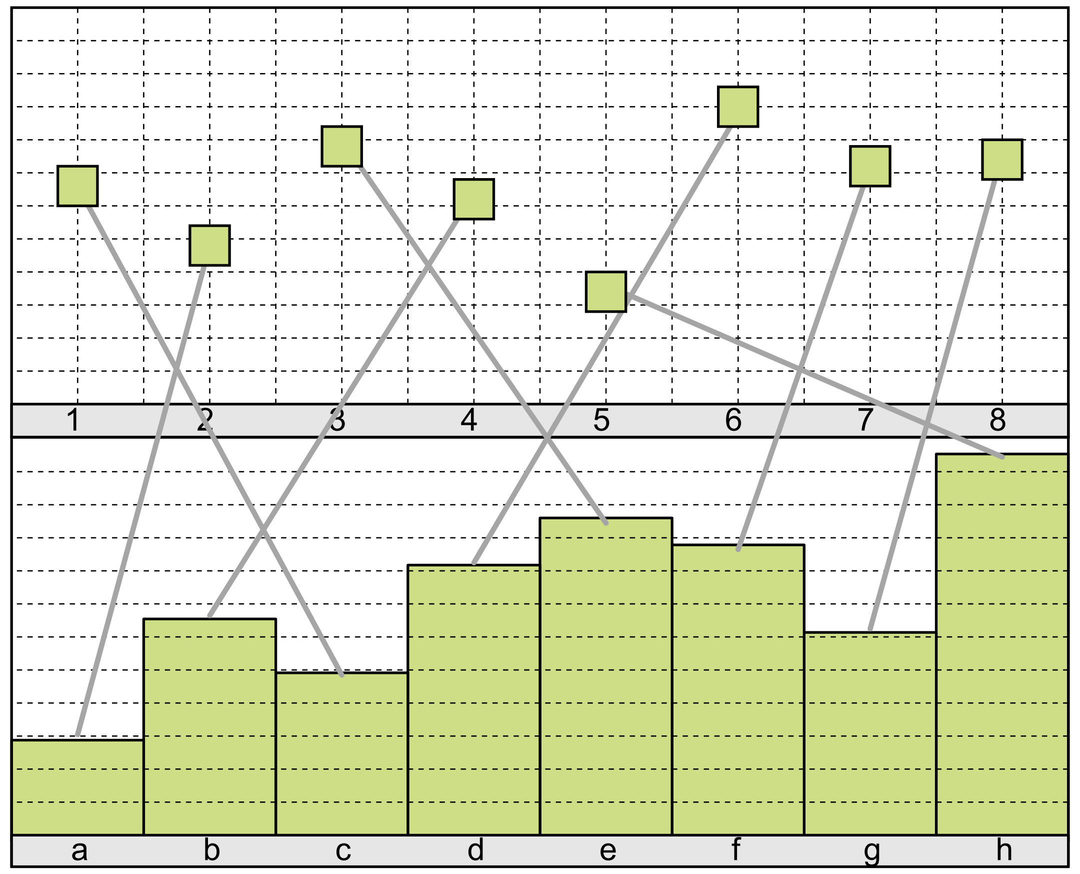
Aggregation

Dr. David Koop

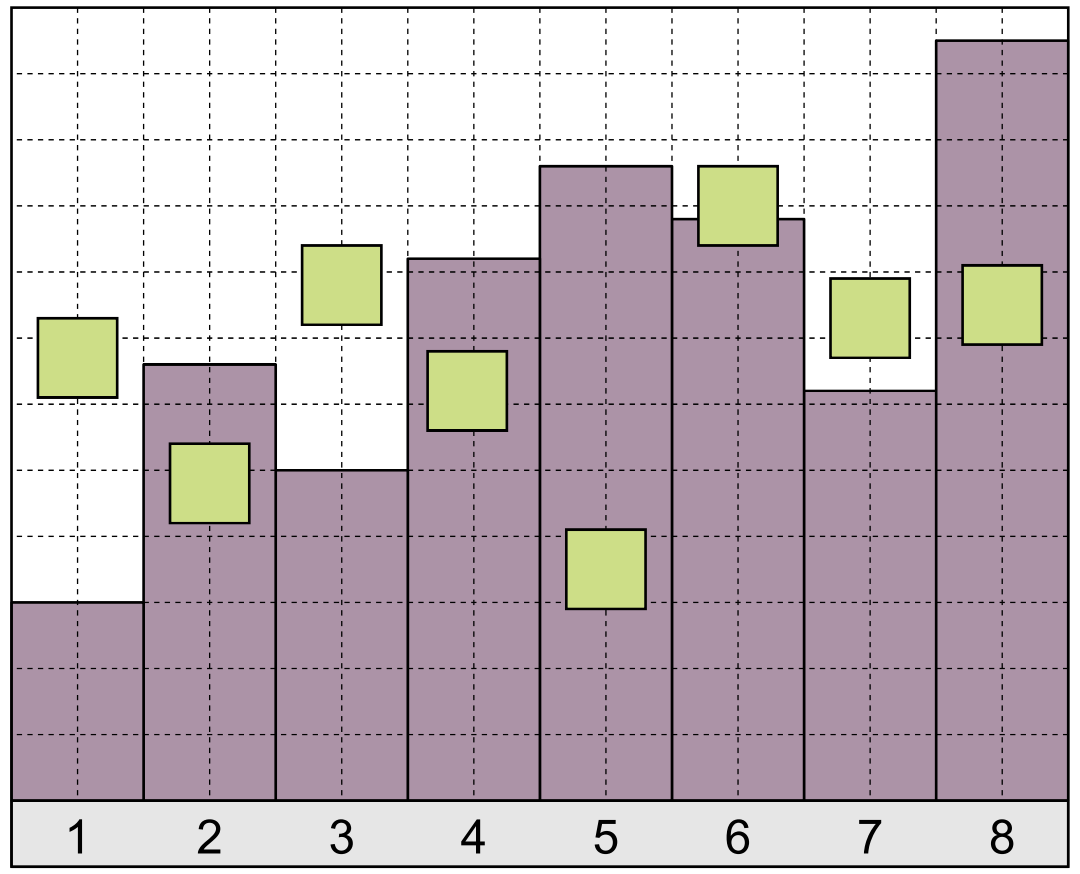
Composite Visualization Techniques



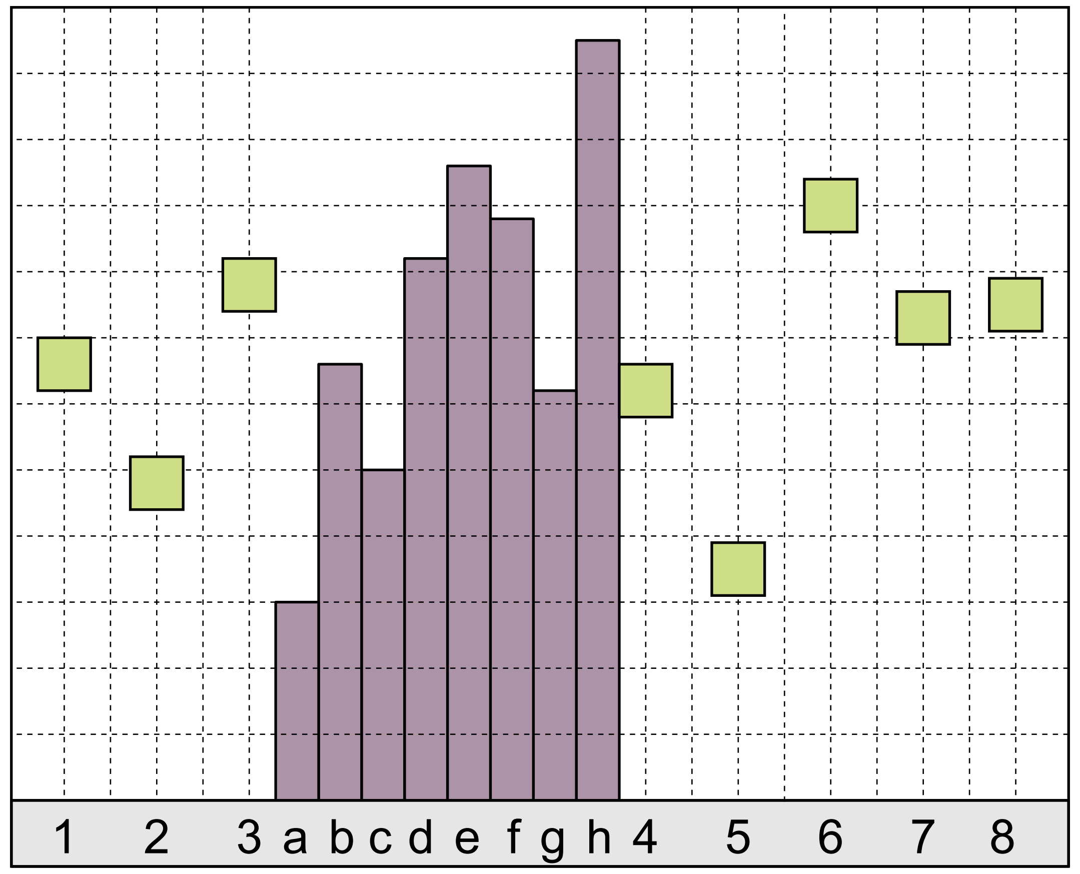
(a) Juxtaposed views.



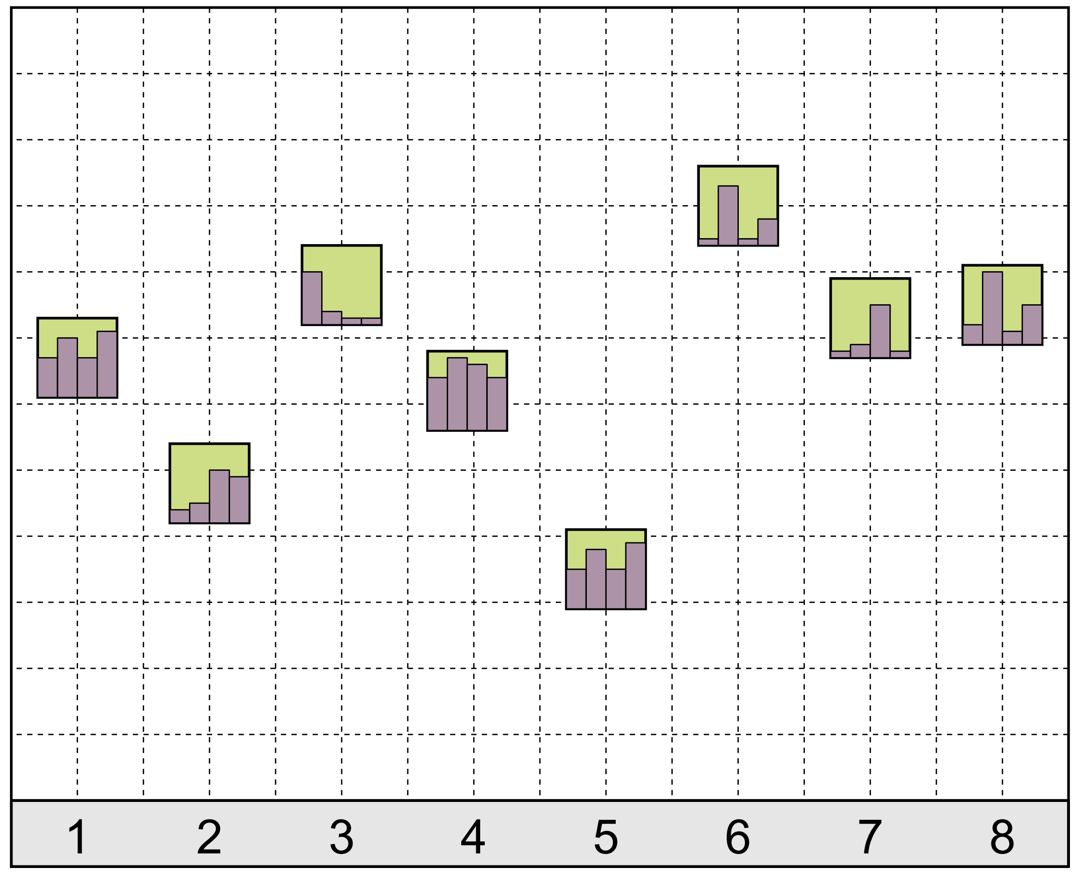
(b) Integrated views.



(c) Superimposed views.



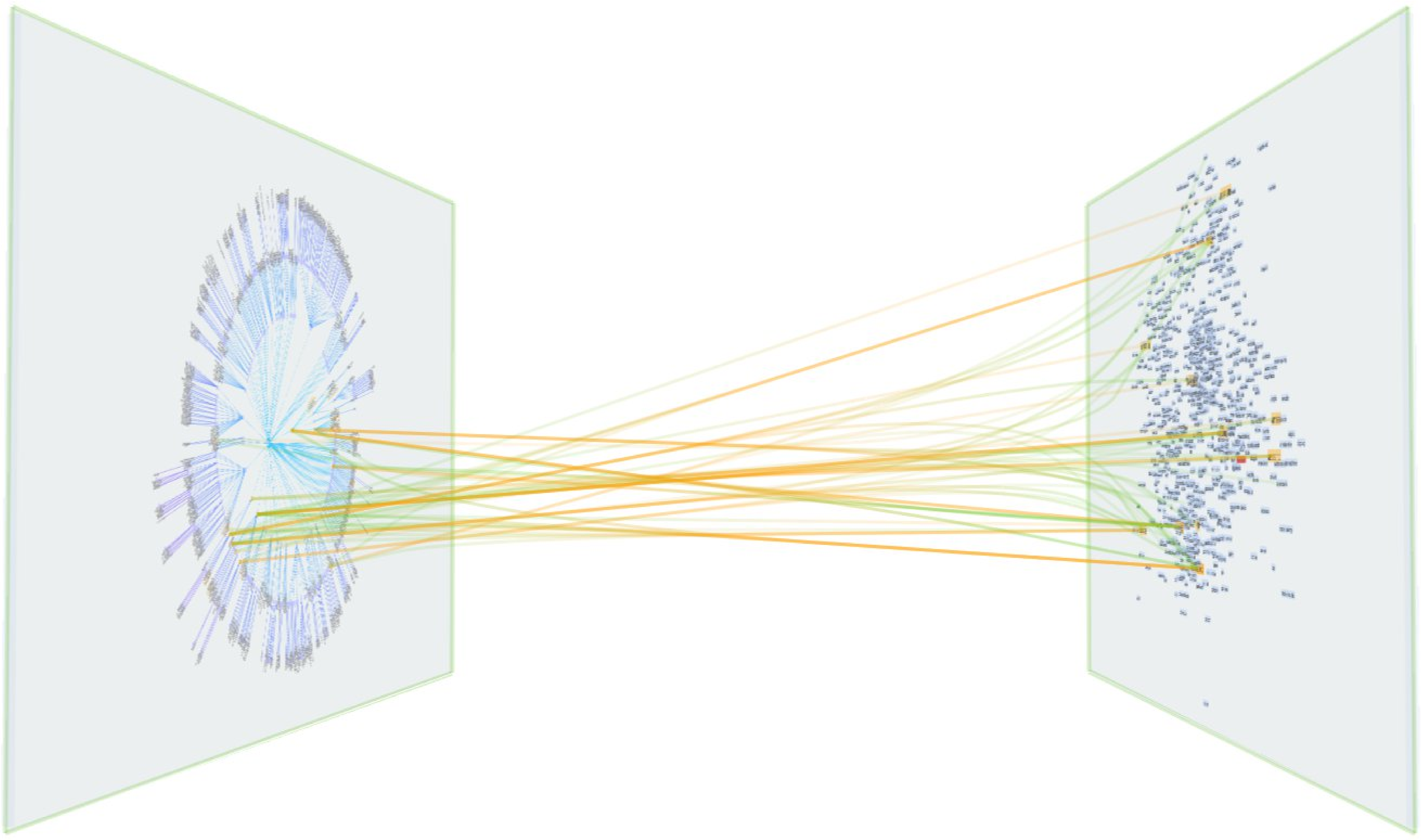
(d) Overloaded views.



(e) Nested views.

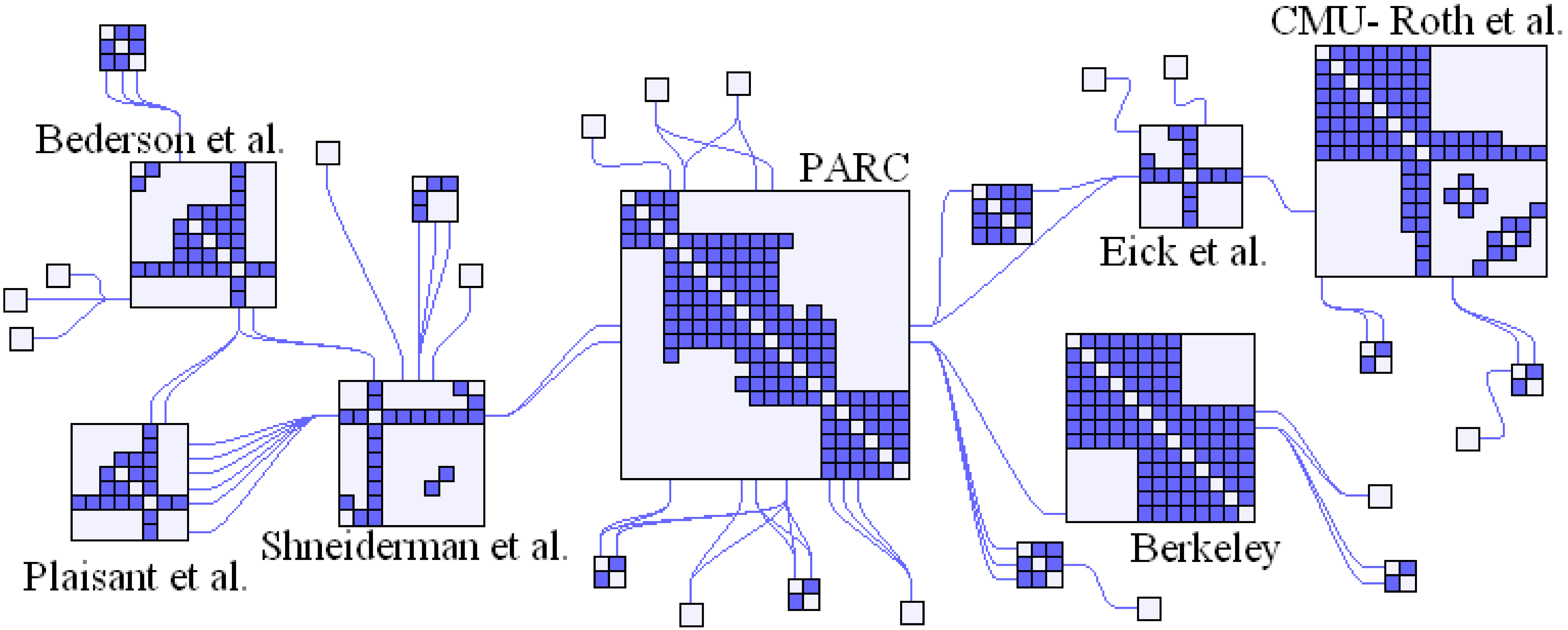
[W. Javed and N. Elmqvist, 2012]

What is this technique?



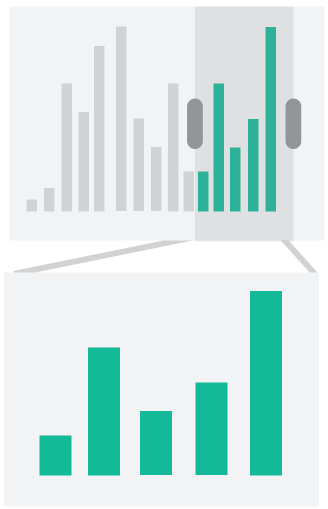
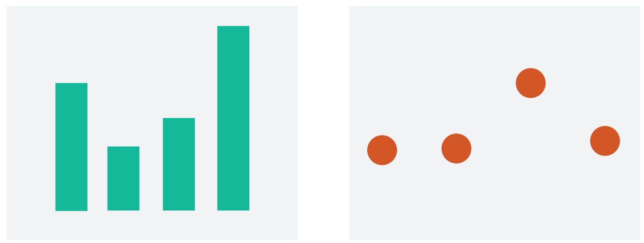
[VisLink, Collins and Carpendale, 2007]

What is this technique?



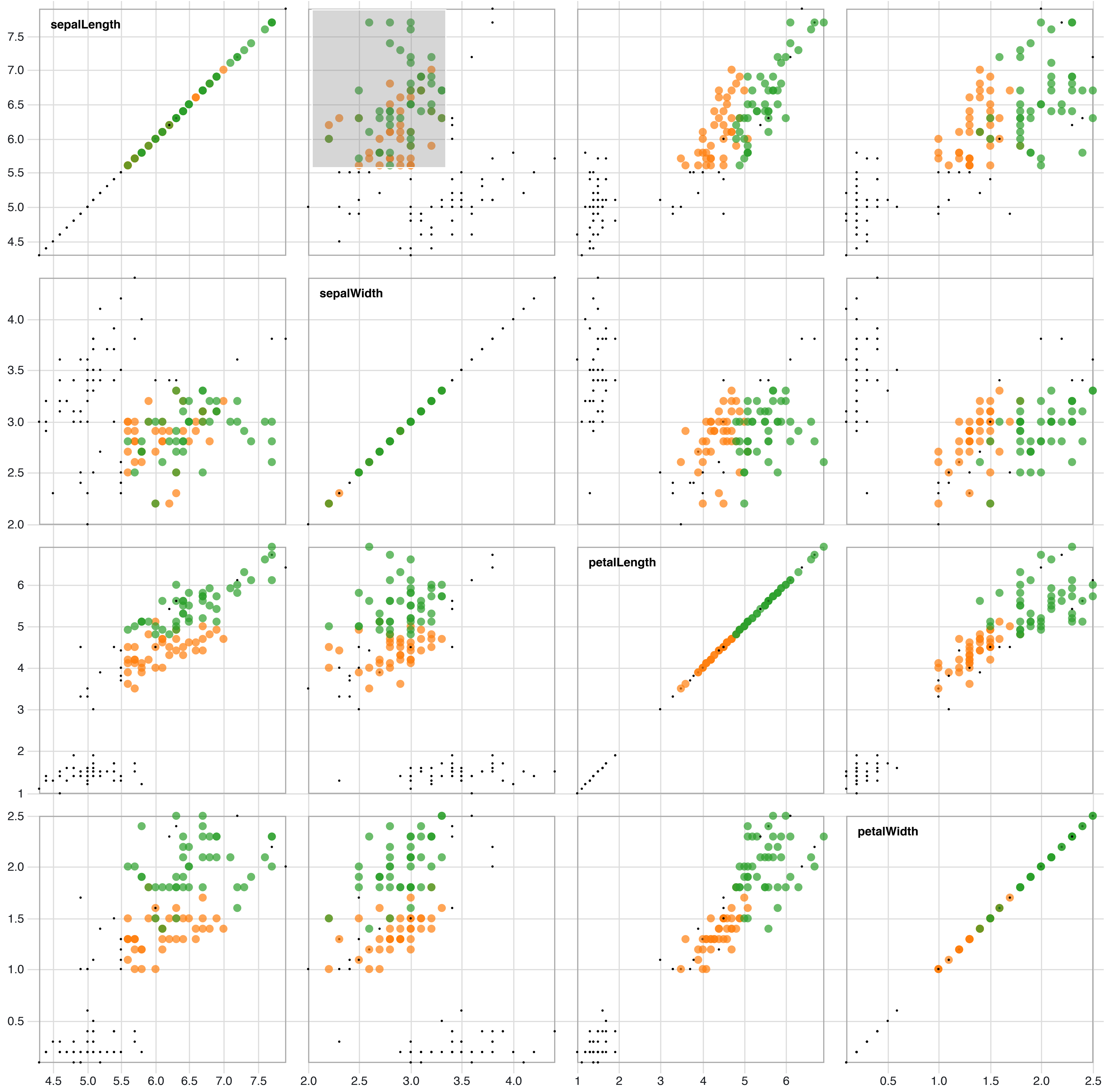
[NodeTrix, N. Henry et al., 2007]

Multiple Views

		Data		
		All	Subset	None
Encoding	Same	Redundant	 Overview/ Detail	 Small Multiples
	Different	 Multiform	 Multiform, Overview/ Detail	No Linkage

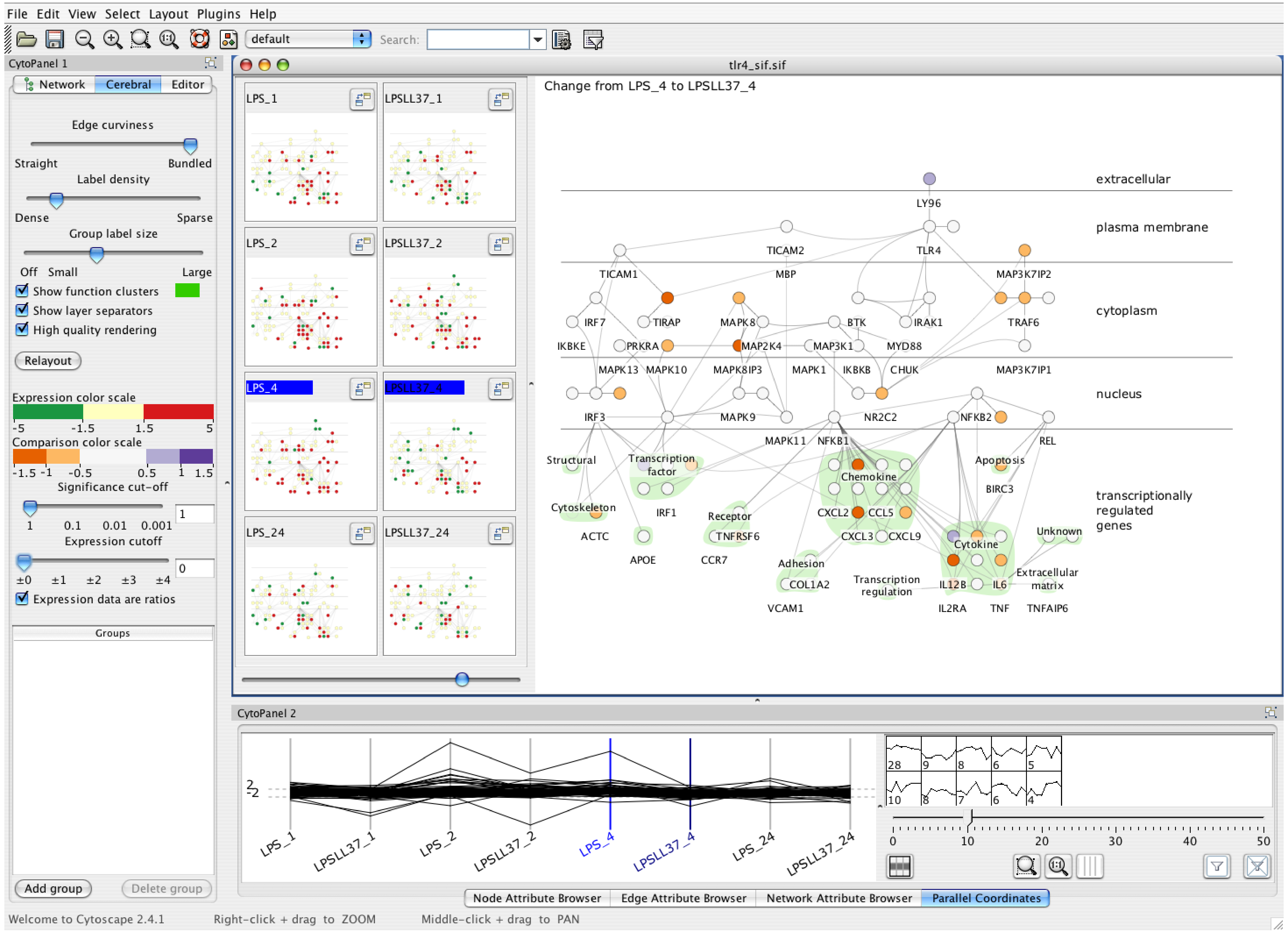
[Munzner (ill. Maguire), 2014]

Brushing



[M. Bostock]

Multiform & Small Multiples

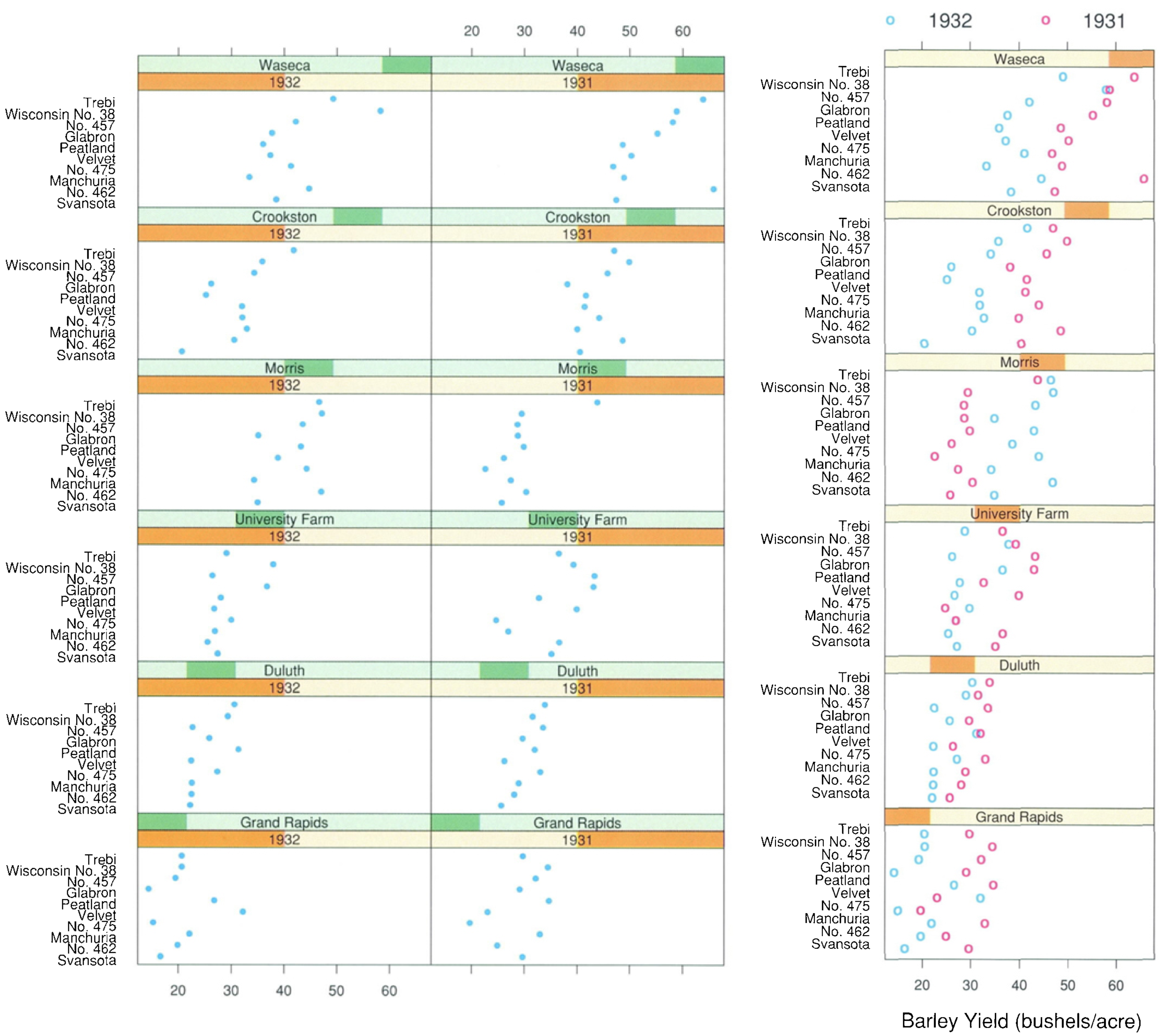


[Barsky et al., 2008]

Partitioned Views

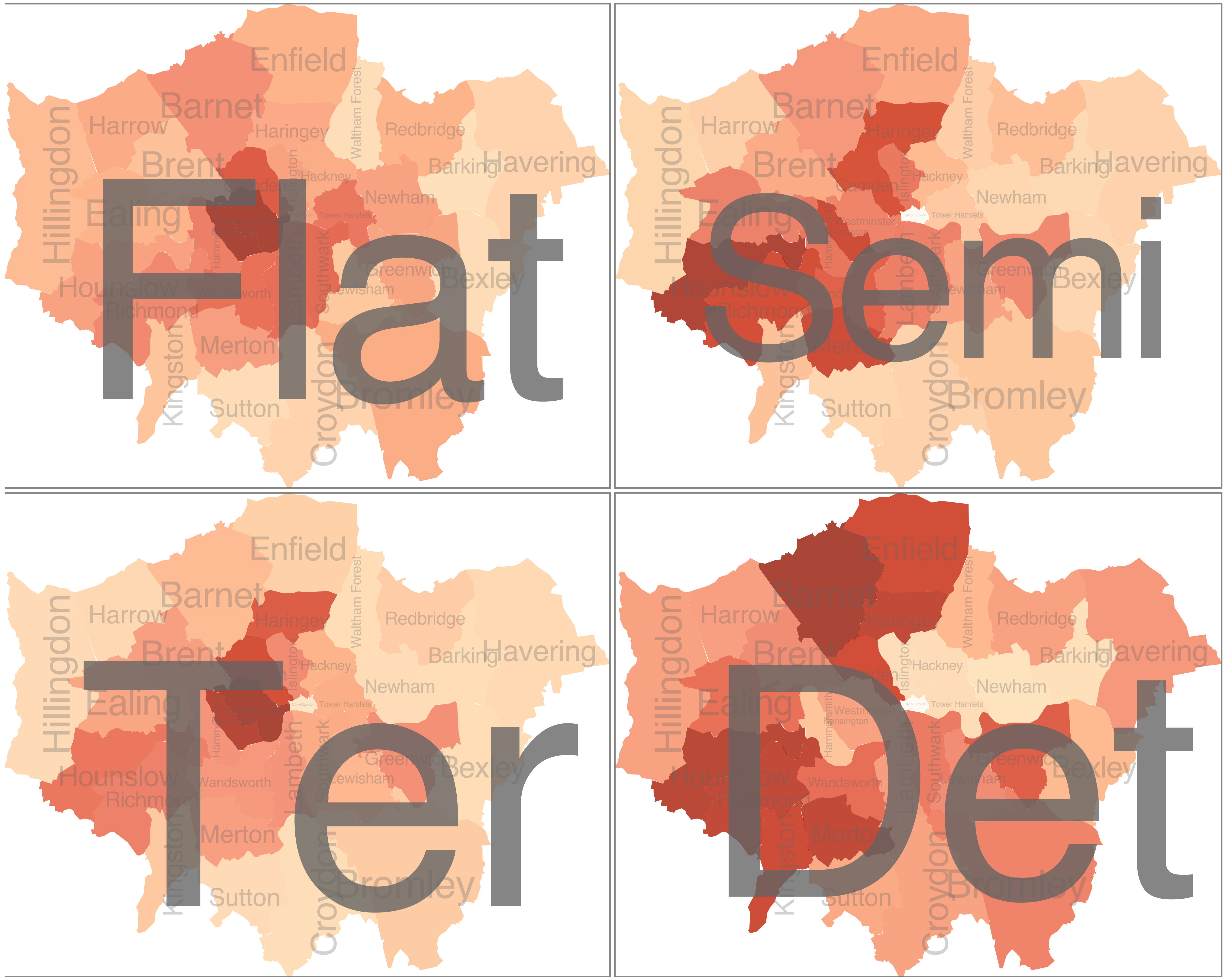
- Split dataset into groups and visualize each group
- Extremes: one item per group, one group for all items
- Can be a hierarchy
  - Order: which splits are more "related"?
  - Which attributes are used to split? usually categorical

Partitioned Views: Trellis Matrix Alignment



[Becker et al., 1996]

Recursive Subdivision: HiVE System

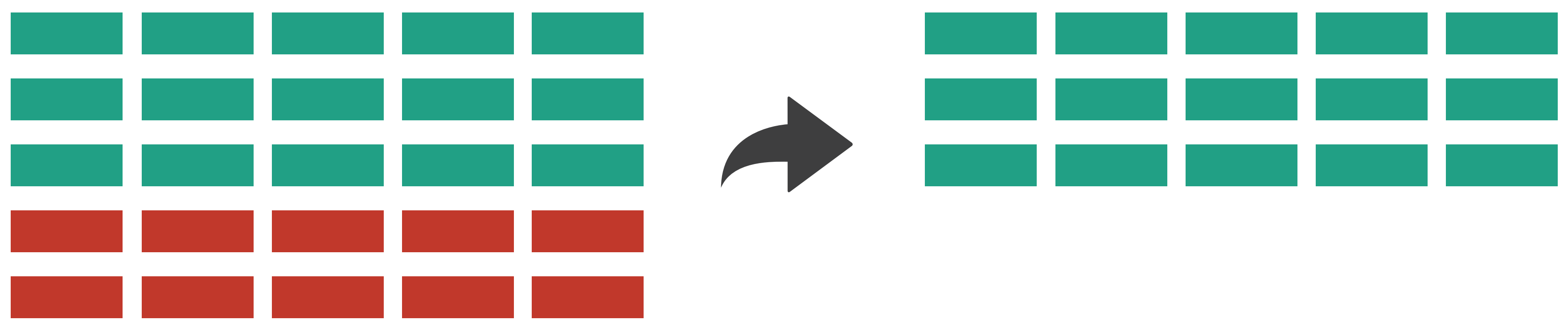


[Slingsby et al., 2009]

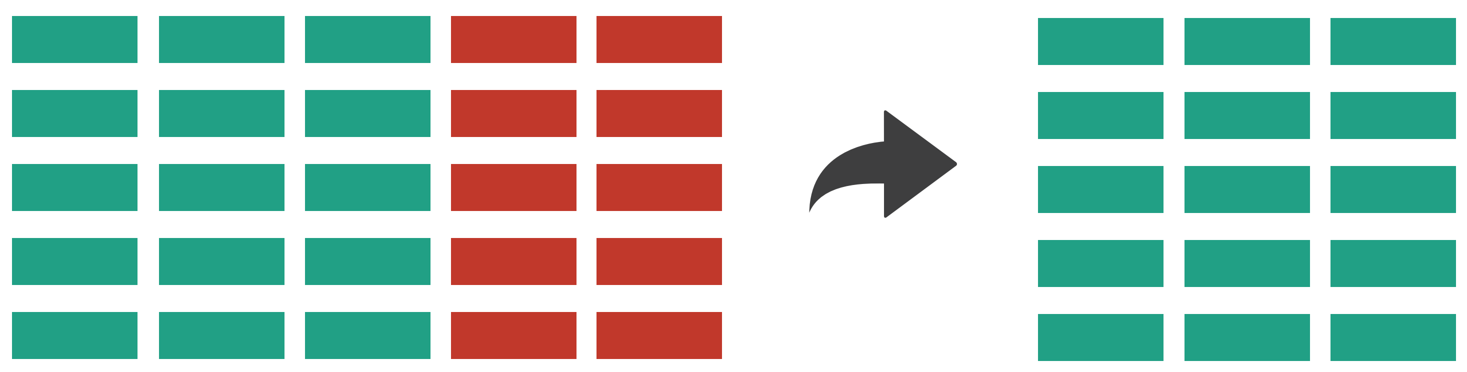
Overview: Reducing Items & Attributes

Filter

- Items

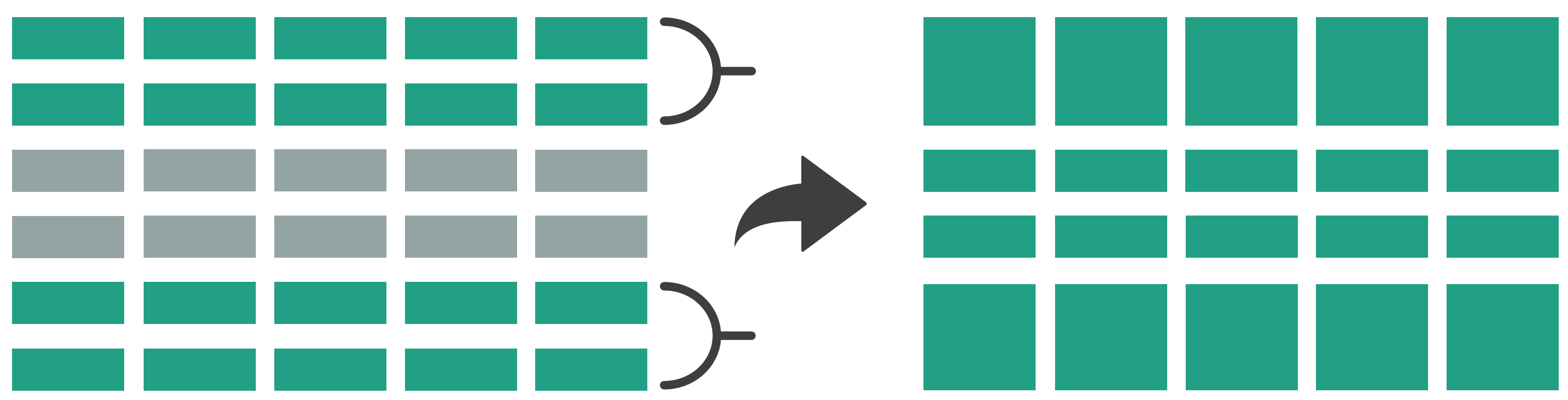


- Attributes

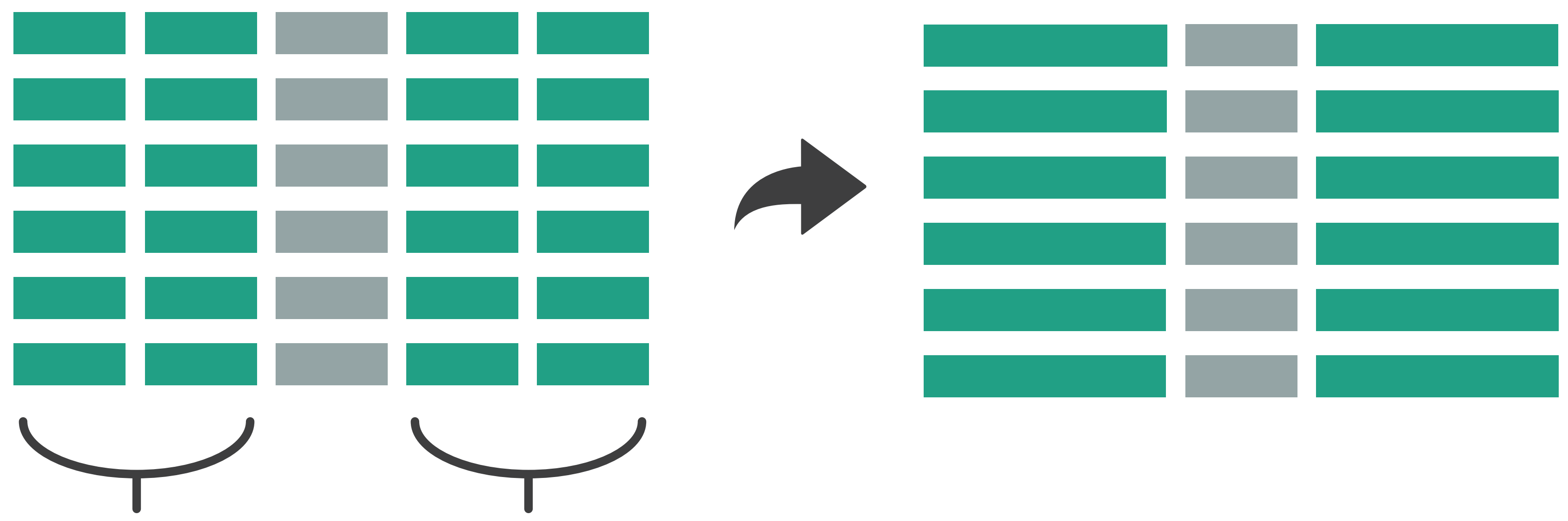


Aggregate

- Items

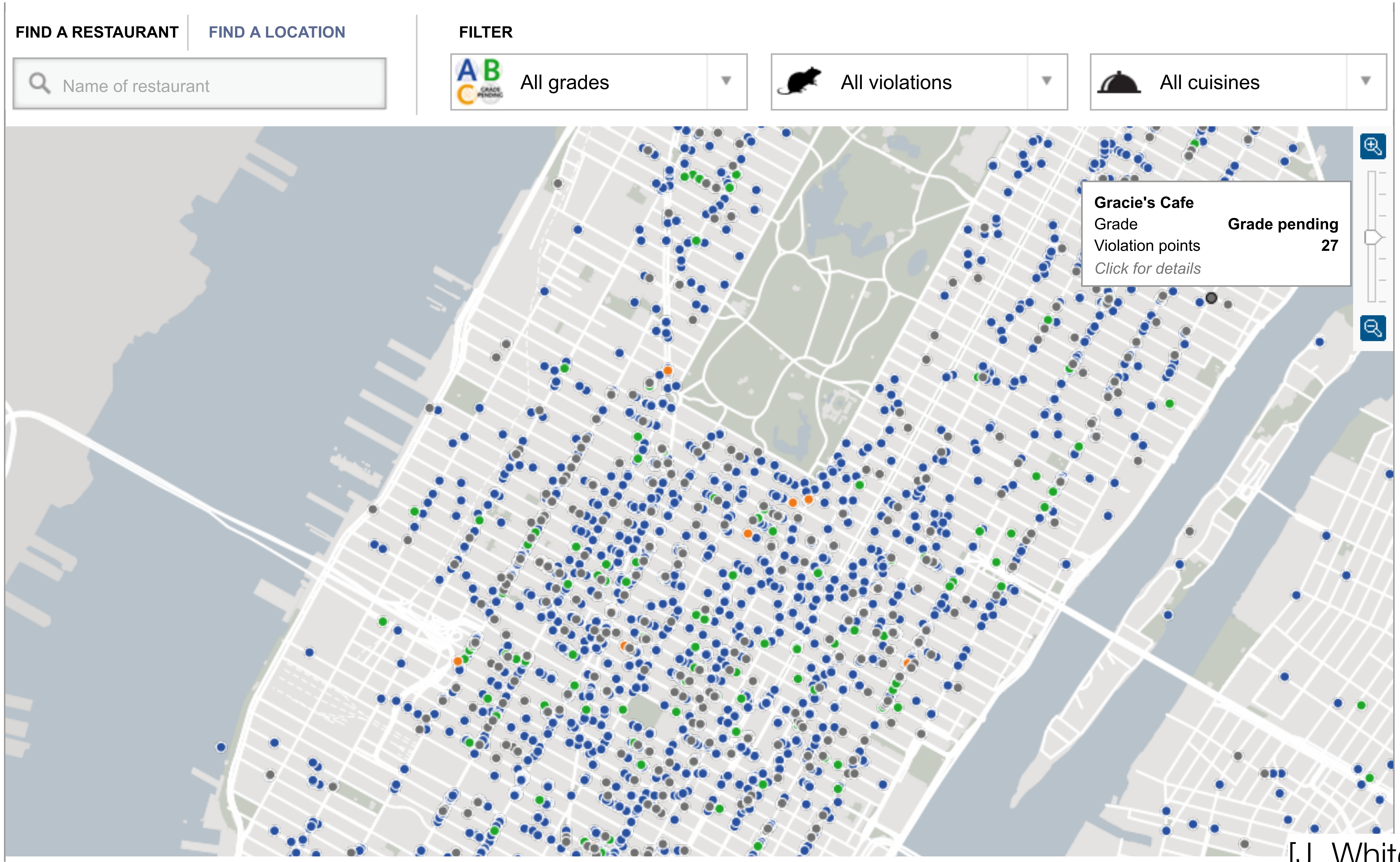


- Attributes



[Munzner (ill. Maguire), 2014]

Example: NYC Health Dept. Restaurant Ratings



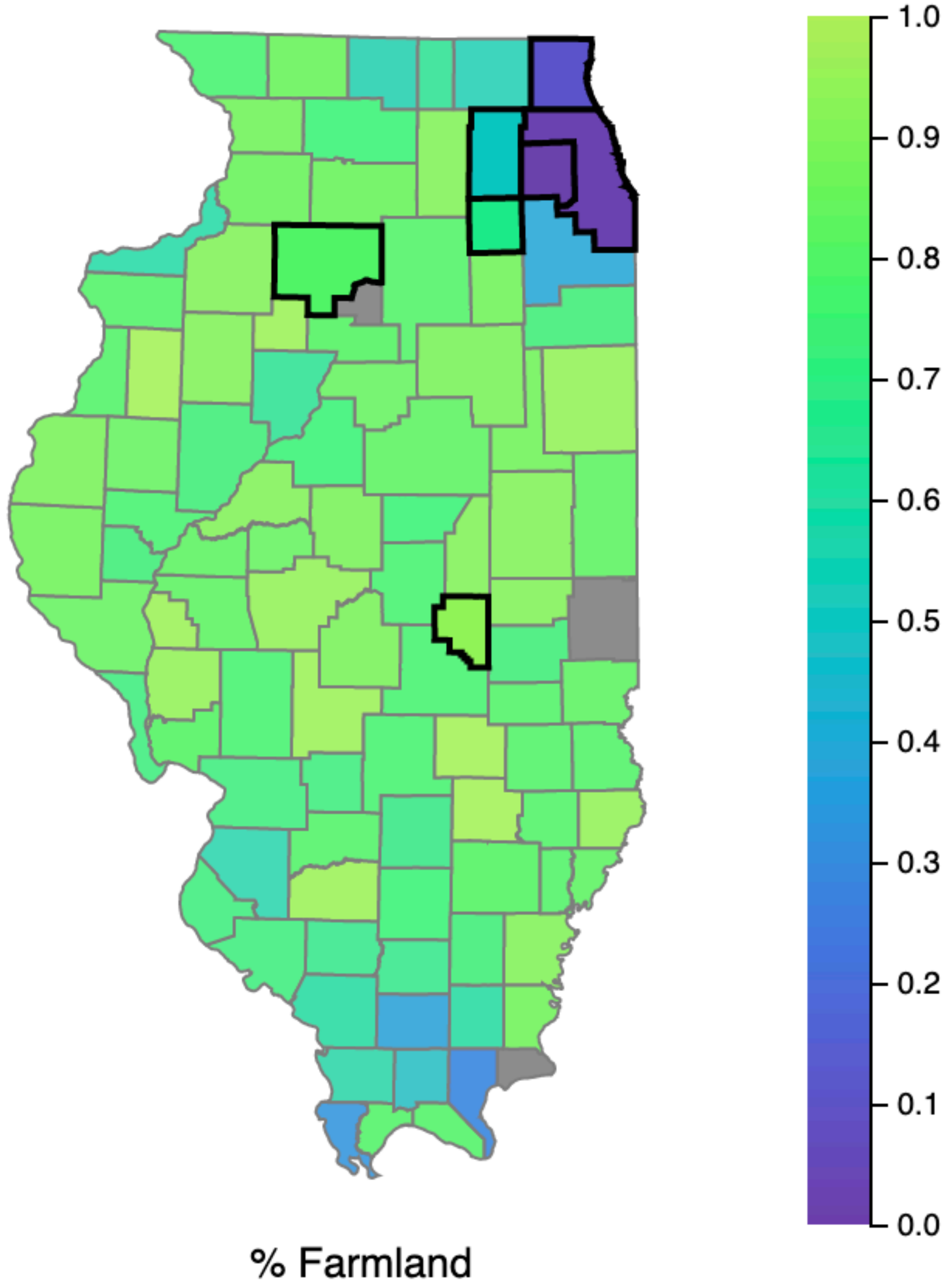
[J. White, New York Times]

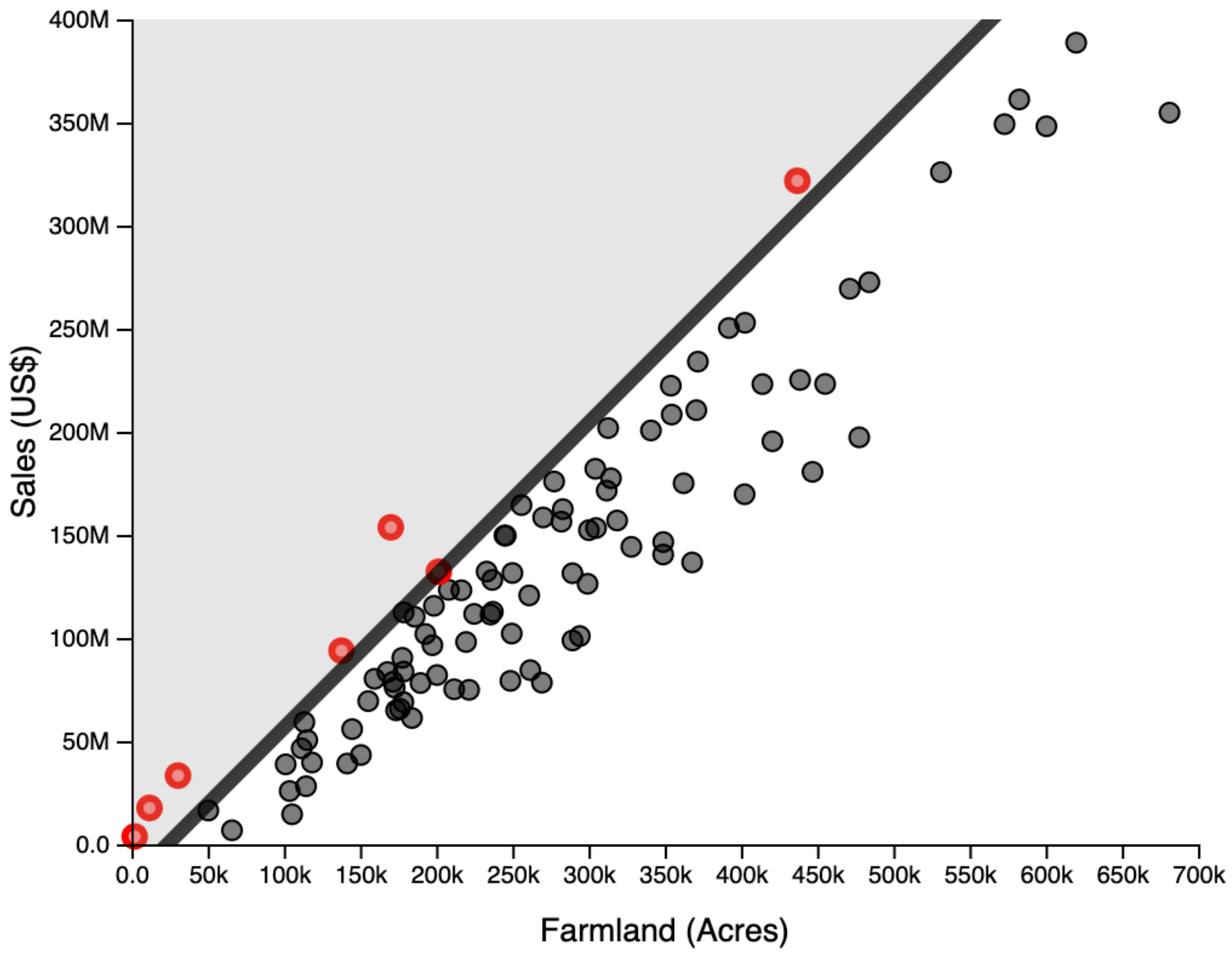
Project Design

- Work on turning your visualization ideas into designs
- Turn in:
  - Three Designs Sketches
  - Progress on Implementation
- Options:
  - Try vastly different options
  - Refine an initial idea
- Due Monday, Nov. 11

Assignment 5

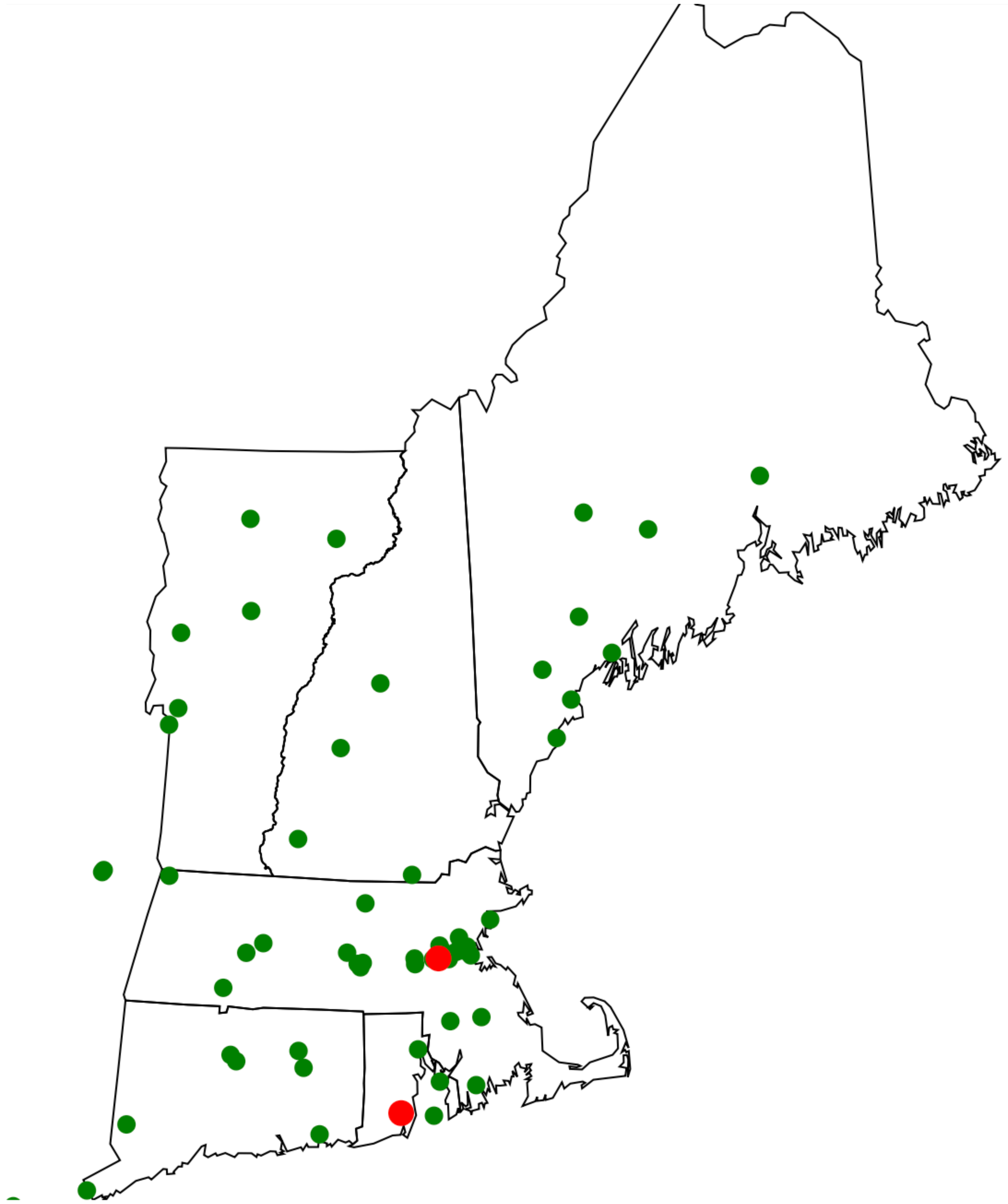
- Multiple Views and Interaction using Linked Highlighting
- Due November 22





Linked Highlighting Example

- https://codepen.io/dakoop/pen/oQxxmx



Nov. 15 Williams	Rhode Island College	59-47
Nov. 15 Wheaton (Mass.)	Plymouth State	67-48
Nov. 15 Tufts	Keene State	79-48
Nov. 16 Western Connecticut	Juniata	59-52
Nov. 16 Yeshiva	Plymouth State	43-81
Nov. 16 UMass Dartmouth	Colby-Sawyer	84-62
Nov. 16 Southern Maine	Bridgewater State	67-49
Nov. 16 Babson	Rhode Island College	95-56
Nov. 16 Eastern Connecticut	Framingham State	74-47
Nov. 16 Wellesley	Keene State	35-74
Nov. 18 Rivier	Plymouth State	41-46
Nov. 18 Keene State	Amherst	67-84
Nov. 18 Emmanuel (Mass.)	Eastern Connecticut	63-76
Nov. 19 Southern Maine	University of New England	73-77
Nov. 19 Western Connecticut	Sage	51-47
Nov. 19 WPI	UMass Dartmouth	71-73
Nov. 20 Rhode Island College	Emmanuel (Mass.)	56-72
Nov. 20 Eastern Connecticut	Trinity (Conn.)	60-27
Nov. 21 Lasell	UMass Boston	53-82
Nov. 22 UMass Dartmouth	Coast Guard	68-58
Nov. 22 Lyndon St.	Plymouth State	56-62
Nov. 22 Trinity (Conn.)	Western Connecticut	57-86
Nov. 22 Smith	Keene State	63-77
Nov. 22 Rhode Island College	WPI	35-56
Nov. 22 Salem State	UMass Boston	41-63
Nov. 25 Keene State	Trinity (Conn.)	66-46
Nov. 25 Bates	Southern Maine	61-60
Nov. 25 Fitchburg State	UMass Boston	60-57
Nov. 25 Castleton	Plymouth State	67-52
Nov. 25 Rhode Island College	Bridgewater State	67-68
Nov. 25 Salve Regina	UMass Dartmouth	50-73

Aggregation

- Usually involves **derived** attributes
- Examples: mean, median, mode, min, max, count, sum
- Remember expressiveness principle: still want to avoid implying trends or similarities based on aggregation

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

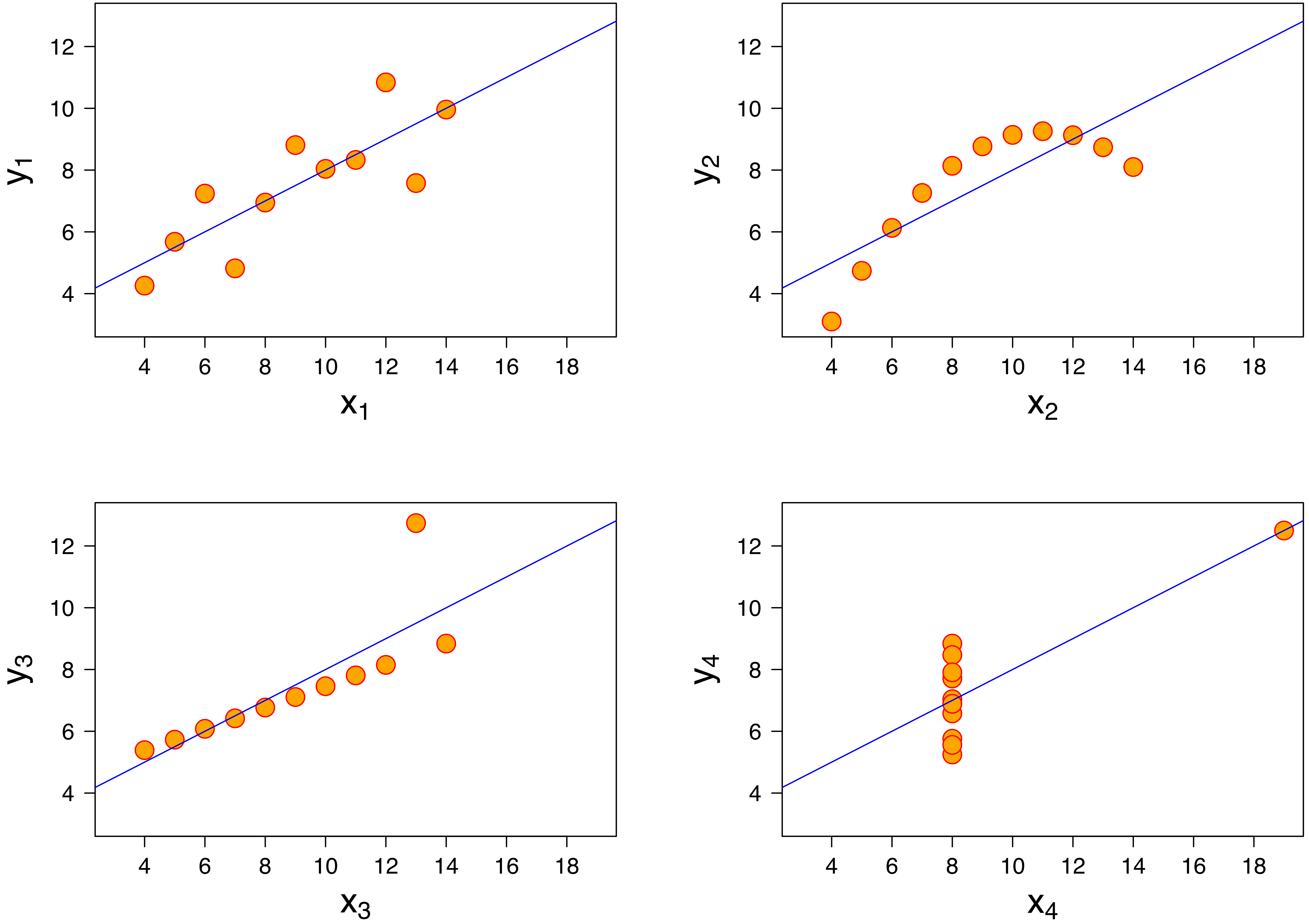
Aggregation

- Usually involves **derived** attributes
- Examples: mean, median, mode, min, max, count, sum
- Remember expressiveness principle: still want to avoid implying trends or similarities based on aggregation

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

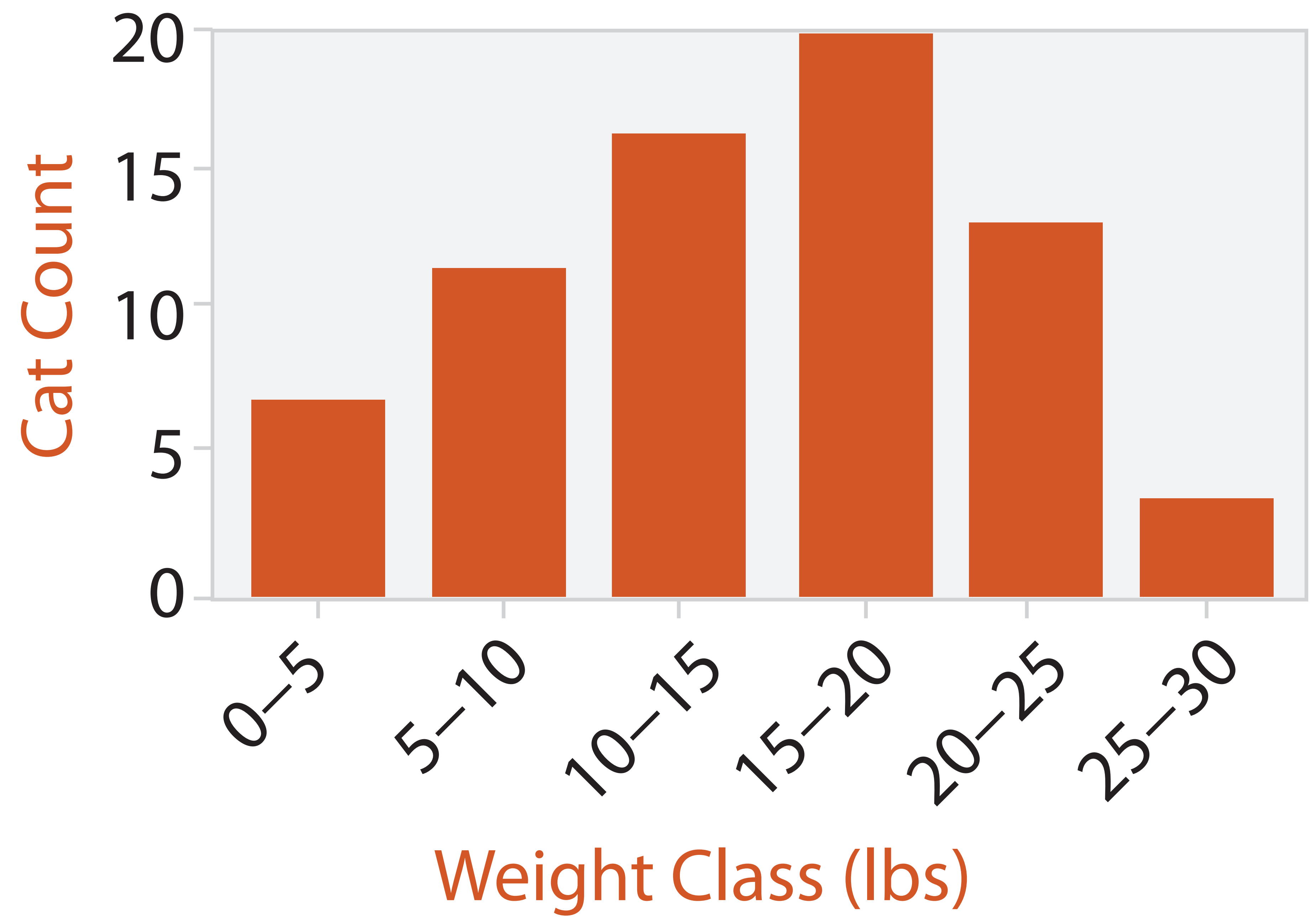
Mean of x	9
Variance of x	11
Mean of y	7.50
Variance of y	4.122
Correlation	0.816

Anscombe's Quartet



[F. J. Anscombe]

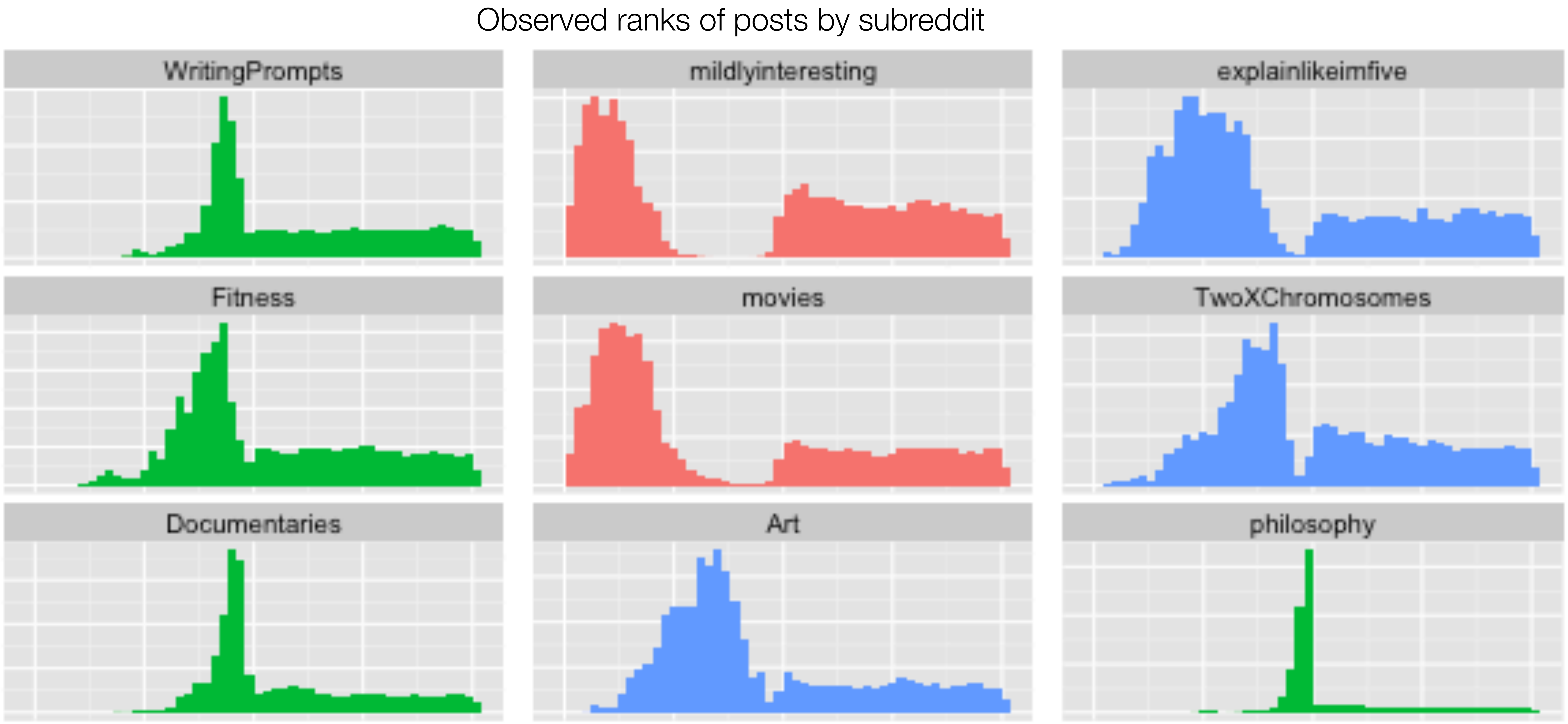
Aggregation: Histograms



- 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

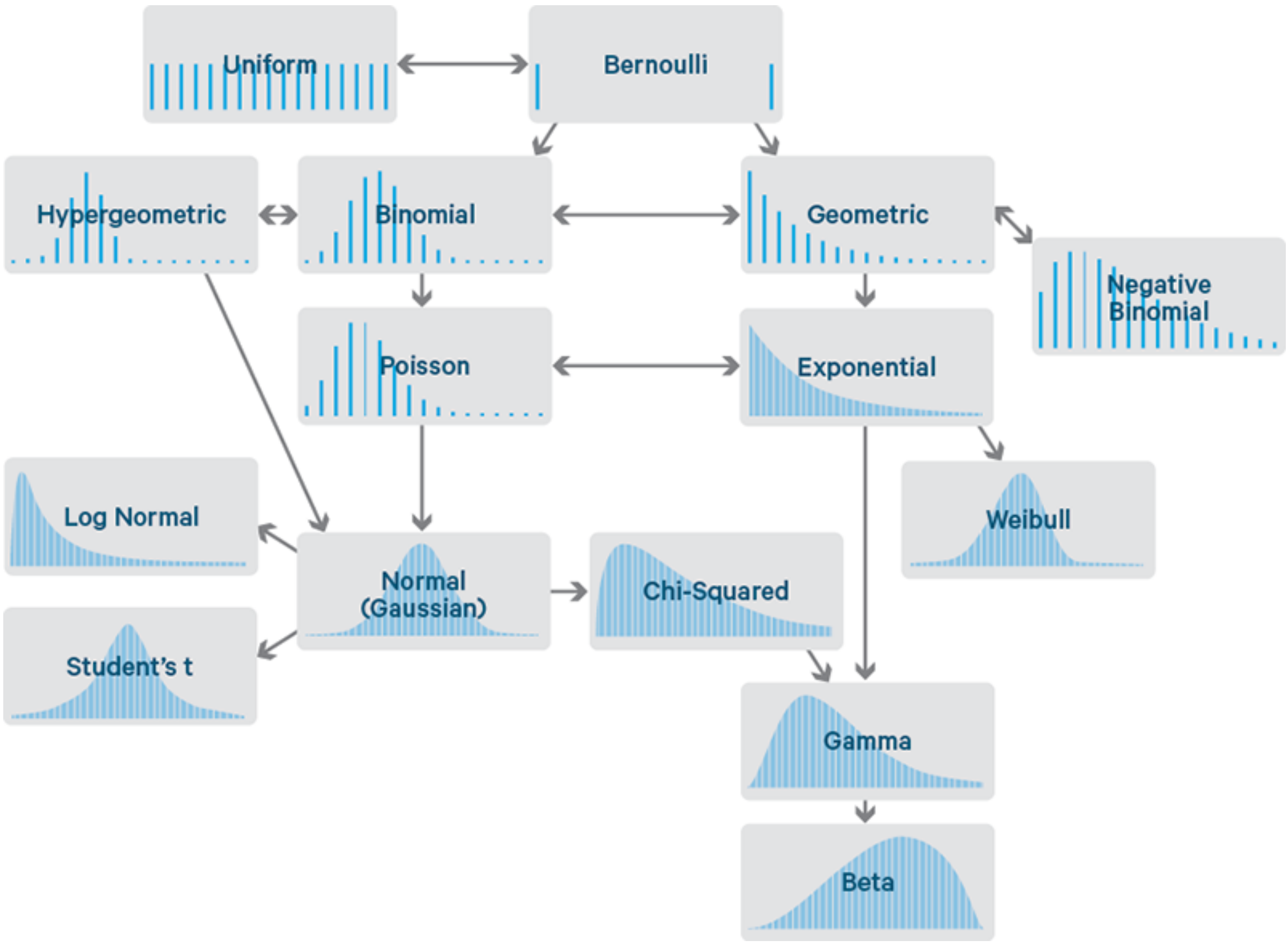
[Munzner (ill. Maguire), 2014]

Aggregation: Histograms



["The reddit Front Page is Not a Meritocracy", T. W. Schneider]

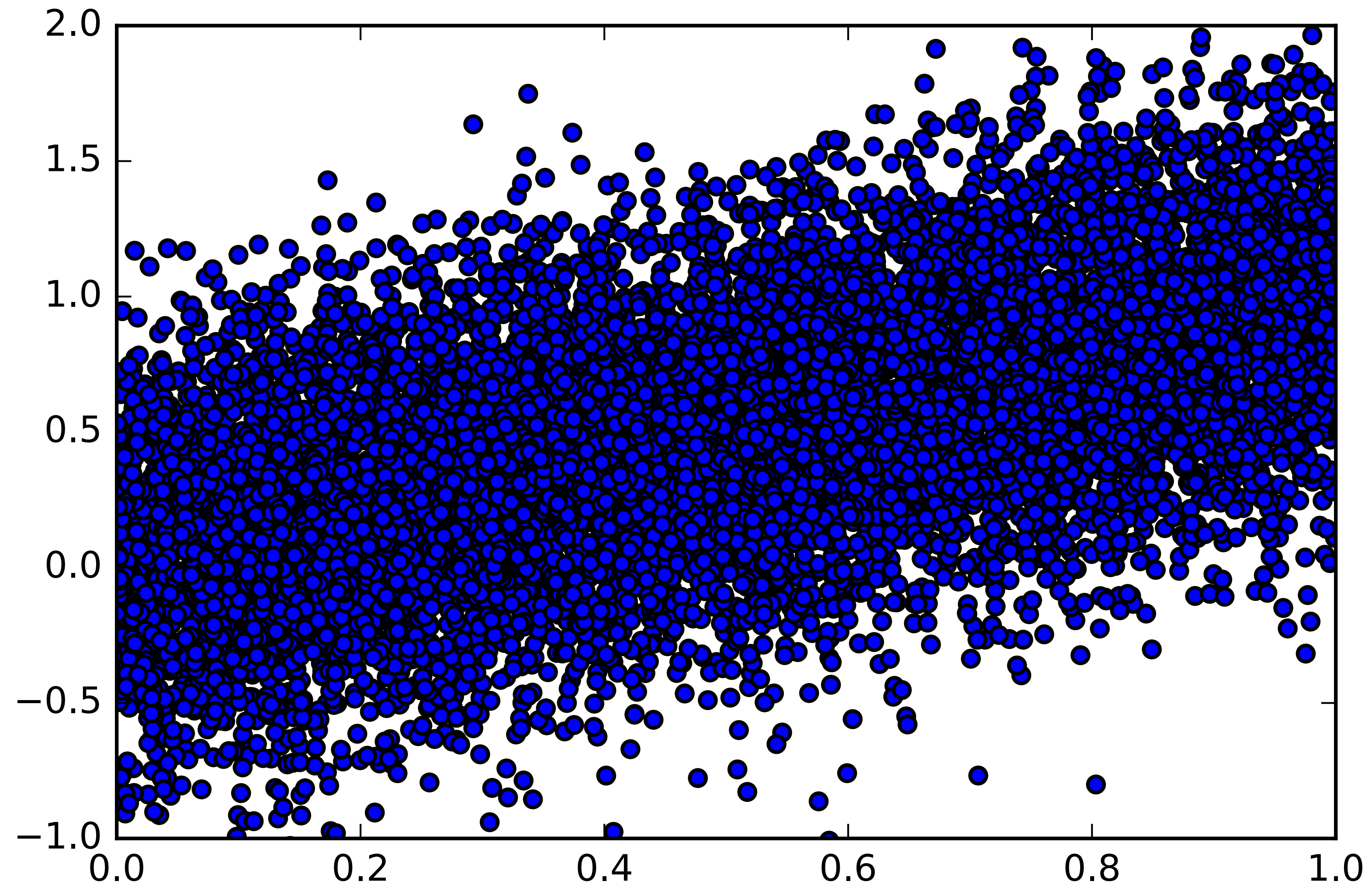
Common Distributions

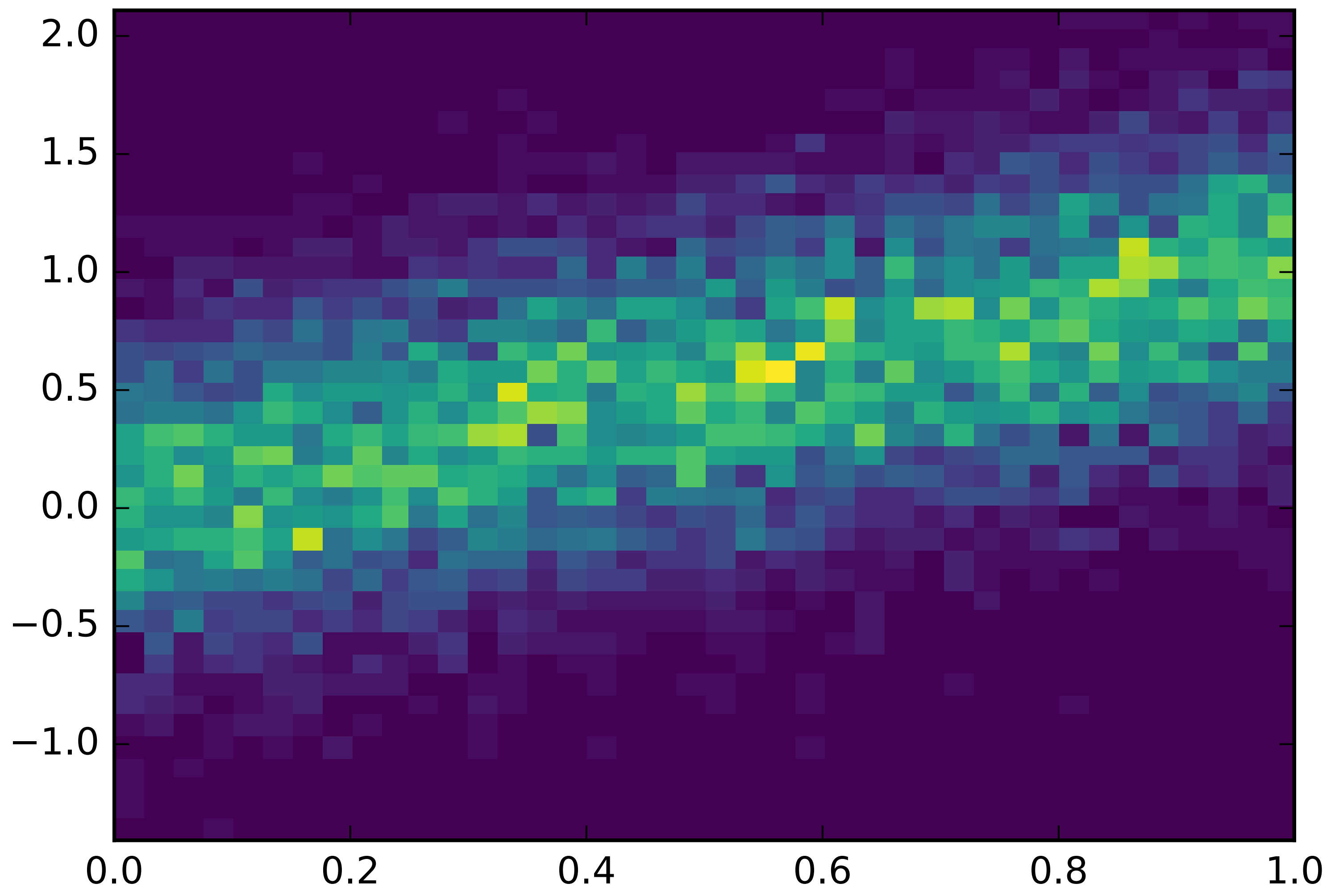


[Cloudera]

Binning Scatterplots

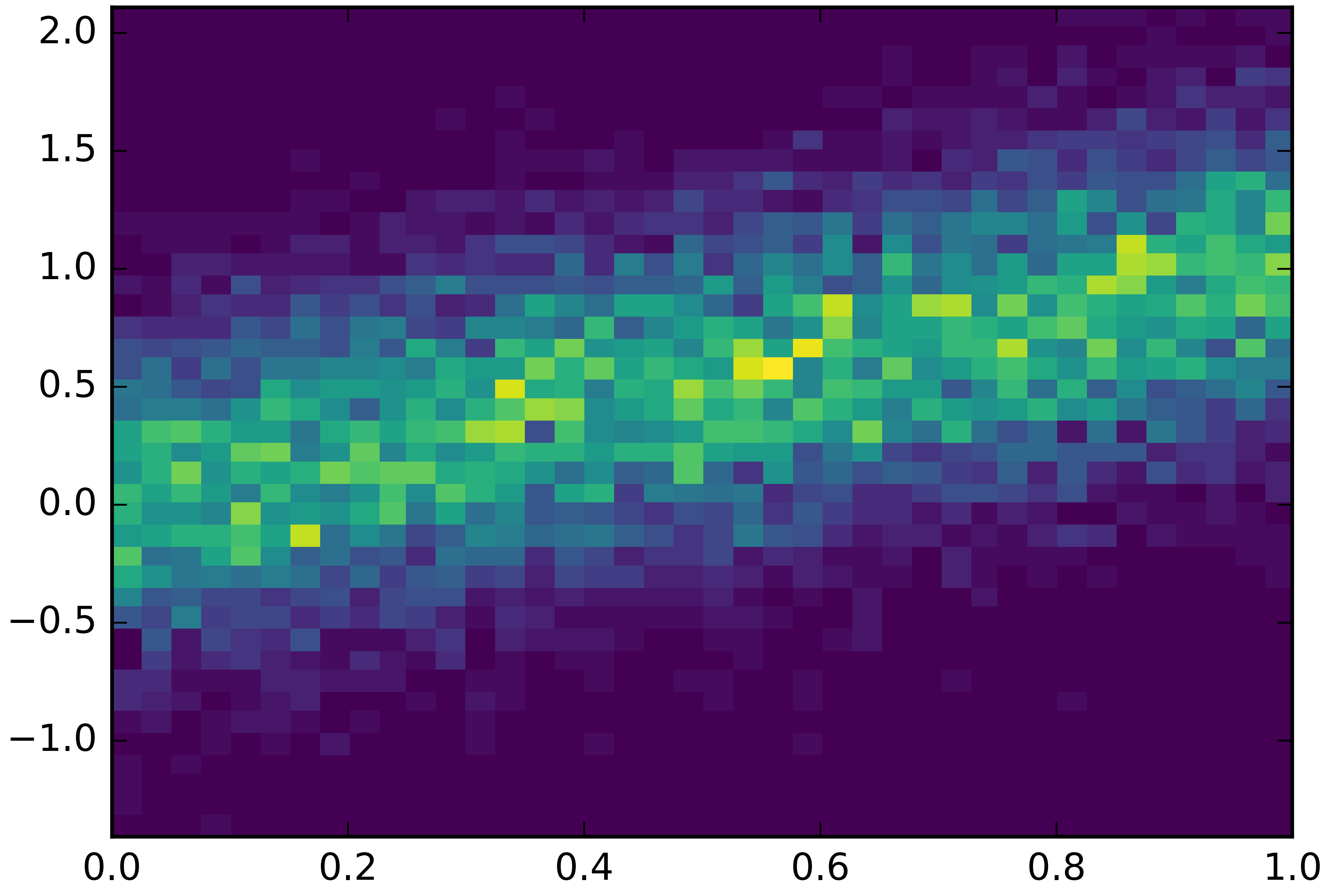
- At some point, cannot see density
- Blobs on top of blobs
- 2D Histogram is a histogram in 2D encoded using color instead of height
- Each region is aggregated

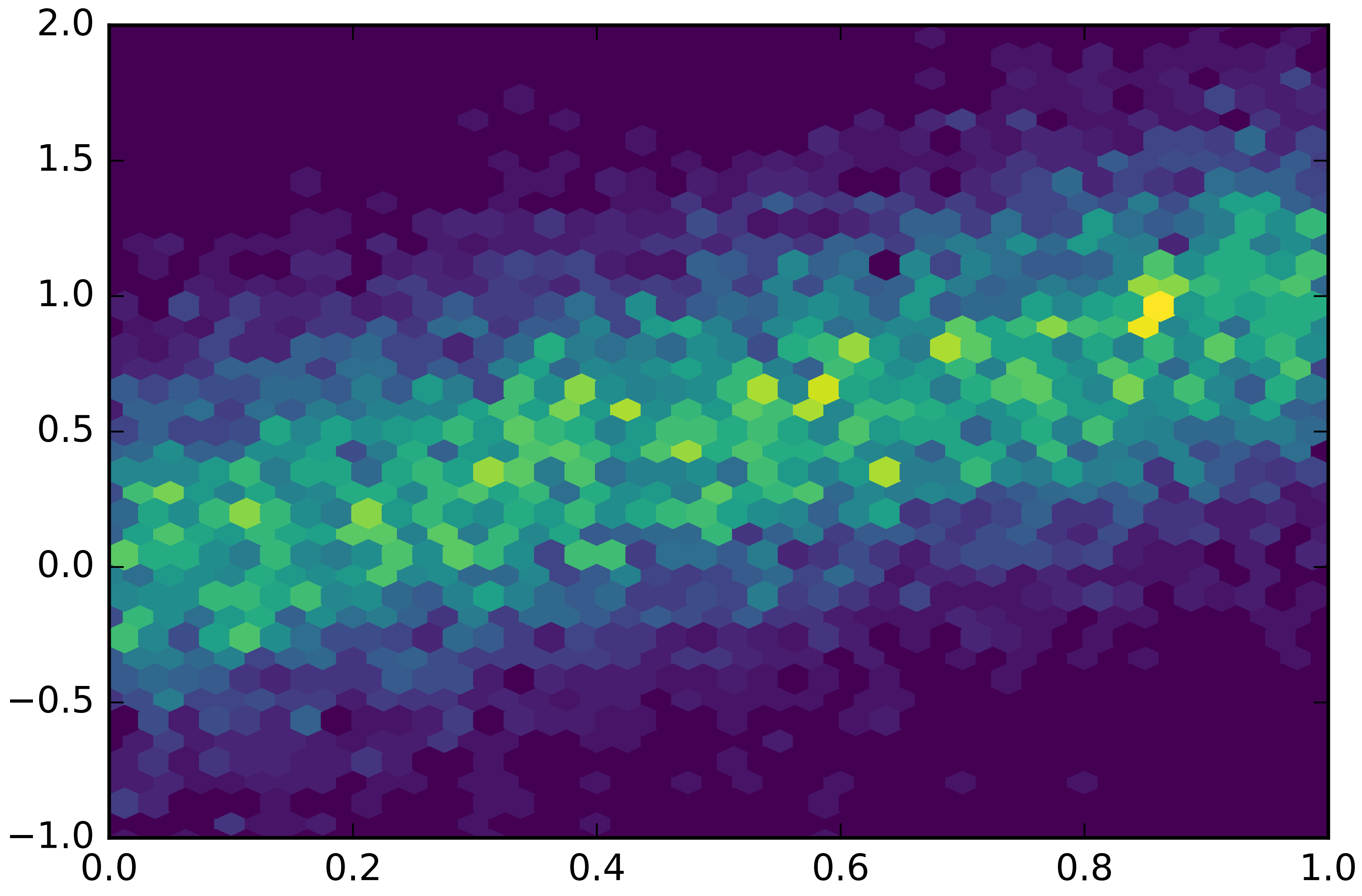




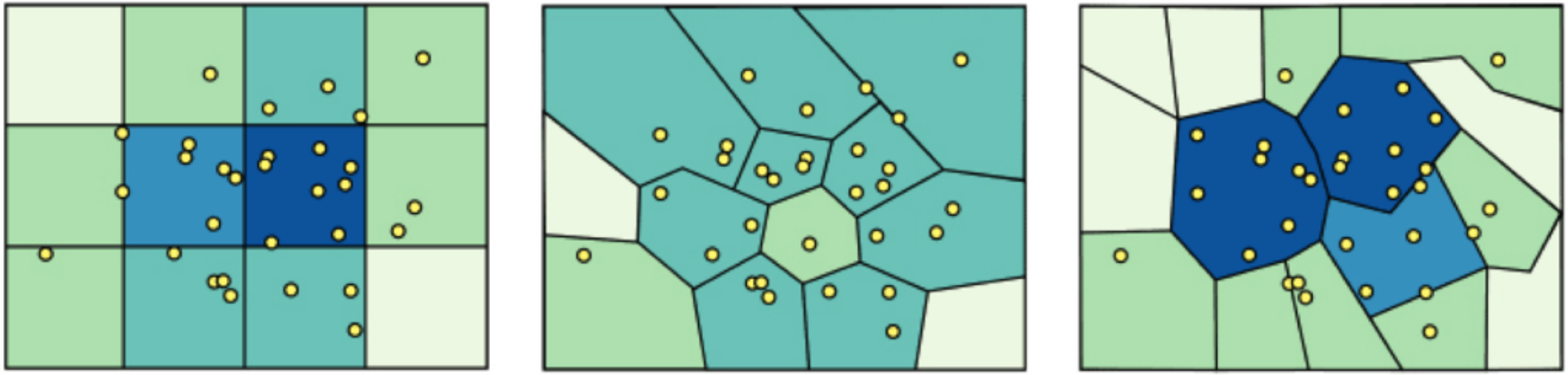
Binning

- Hexagonal bins are more circular
- Distance to the edge is not as variable
- More efficient aggregation around the center of the bin





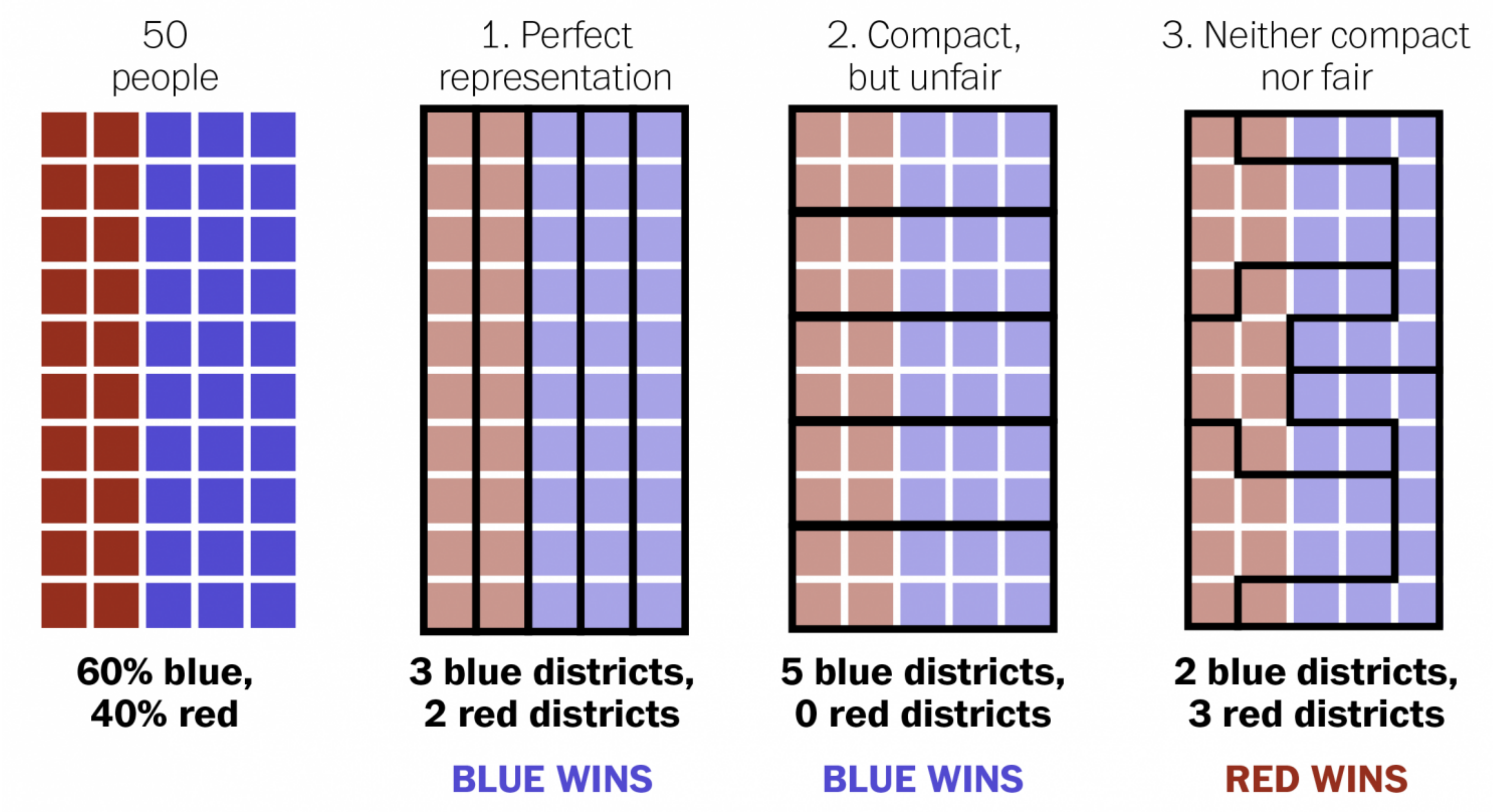
Spatial Aggregation

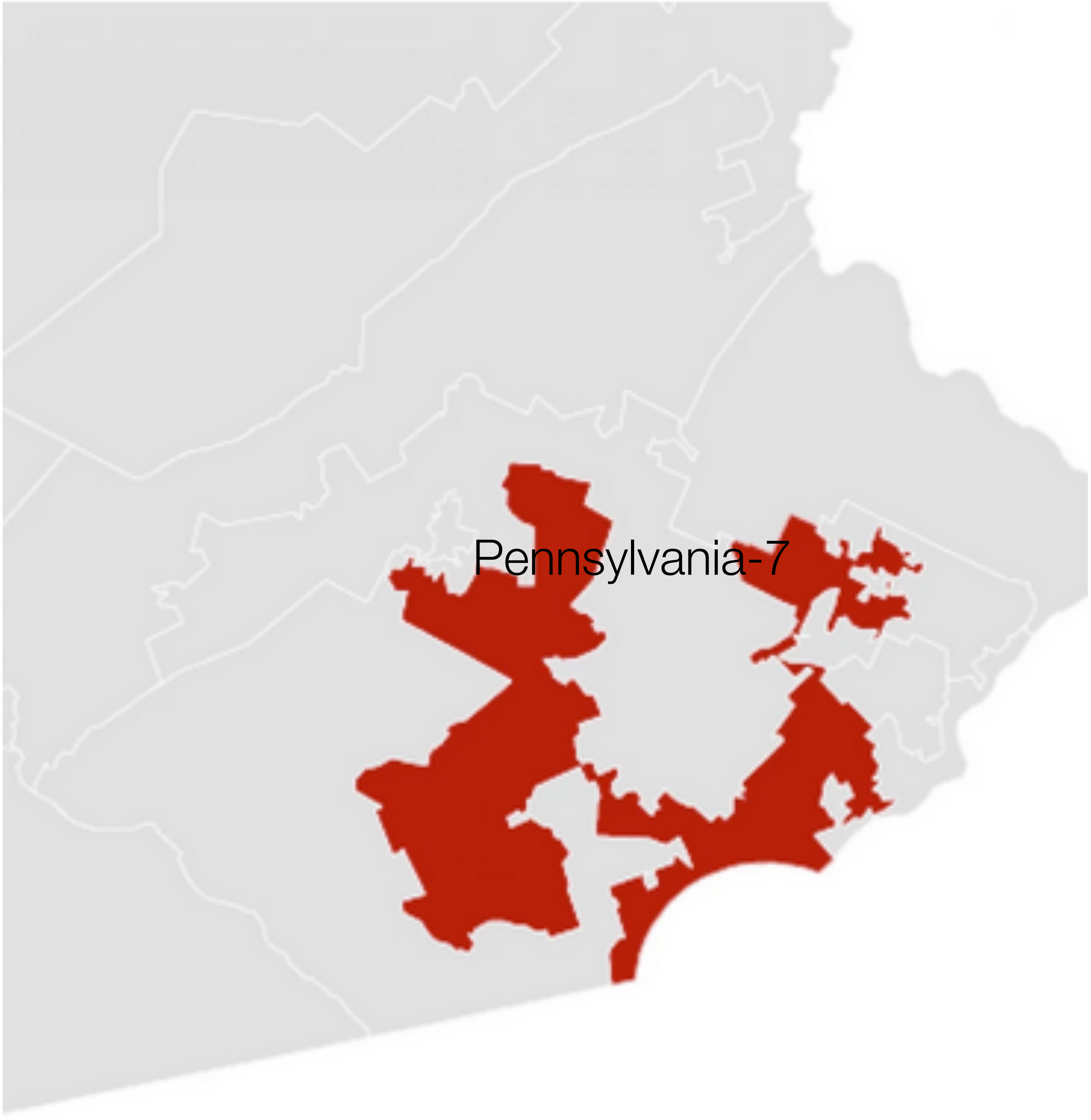


[Penn State, GEOG 486]

Modifiable Areal Unit Problem

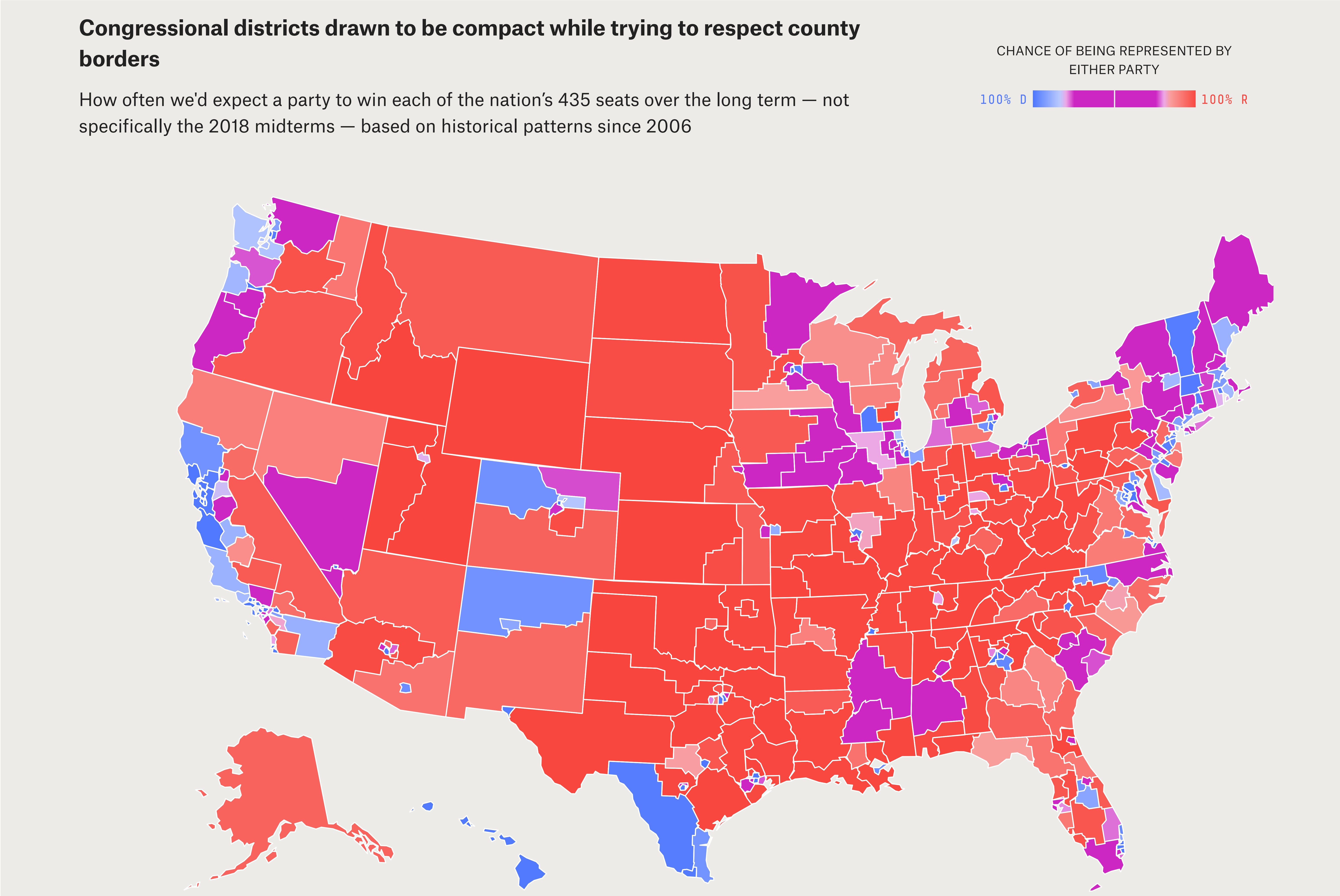
- How you draw boundaries impacts the type of aggregation you get
- Similar to bins in histograms
- Gerrymandering





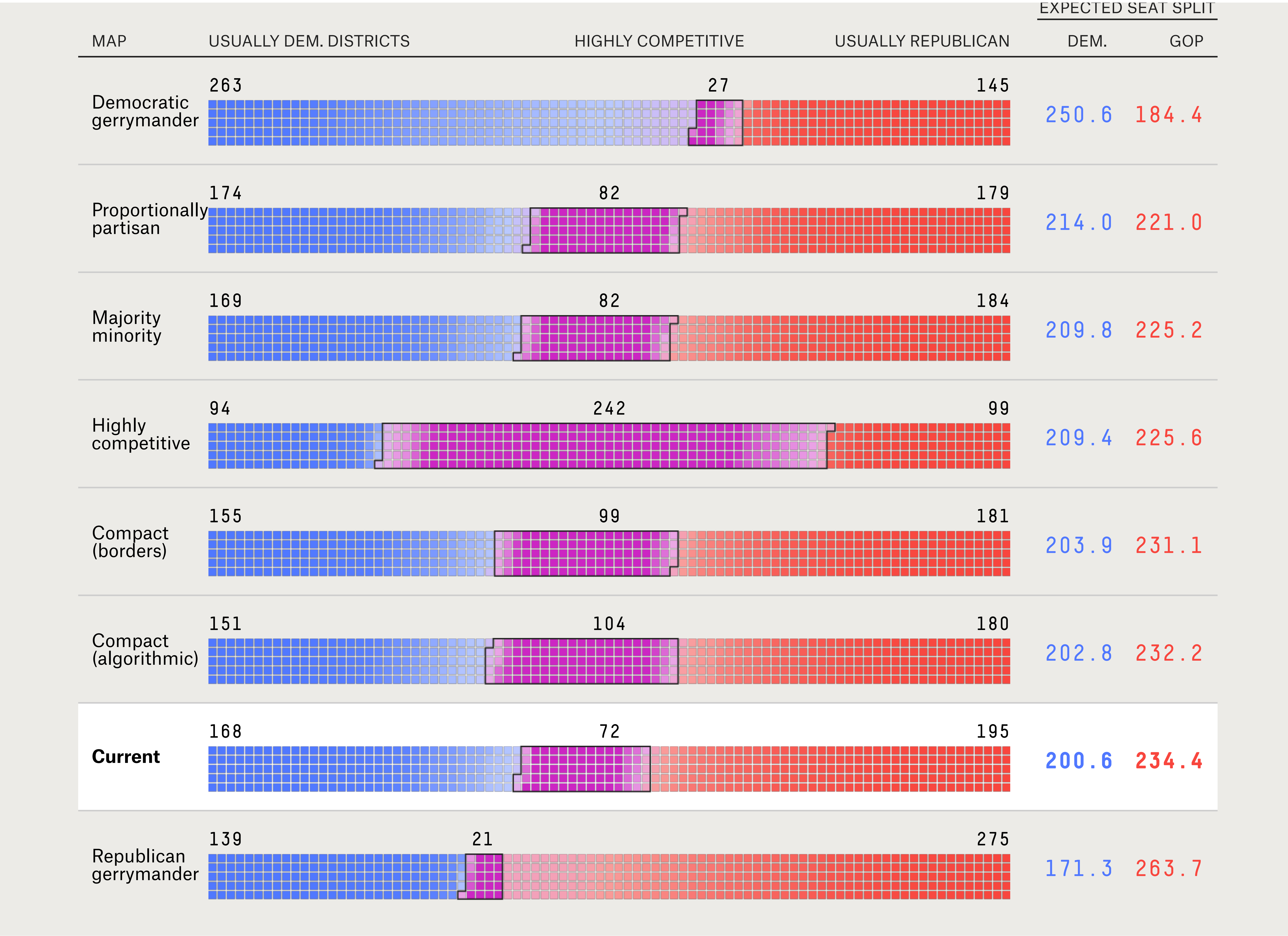
[Wonkblog, Washington Post, Adapted from S. Nass]

Drawing Different Maps: Compactness



[A. Bycoffe et al., 538]

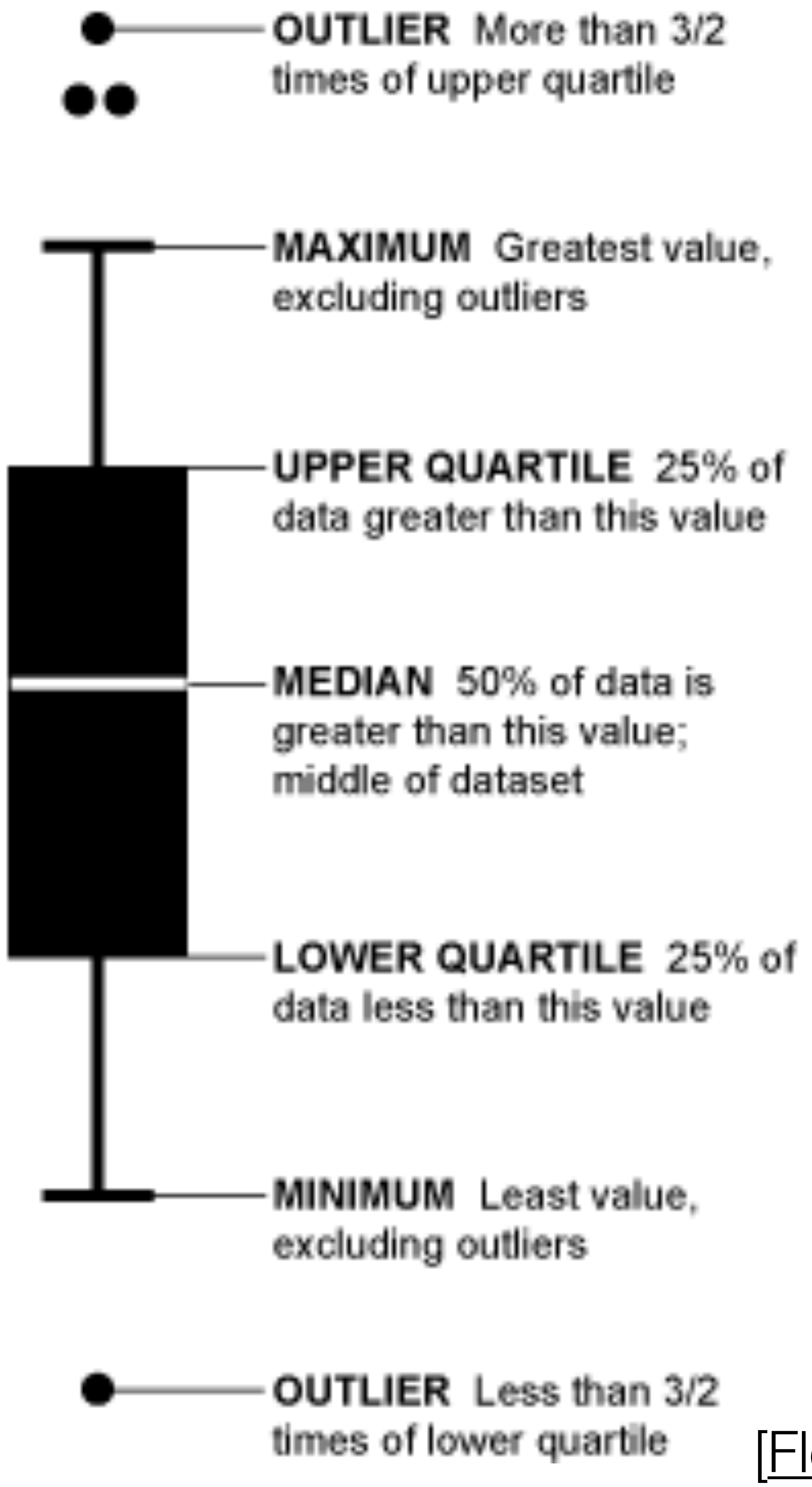
Drawing Different Maps



[A. Bycoffe et al., 538]

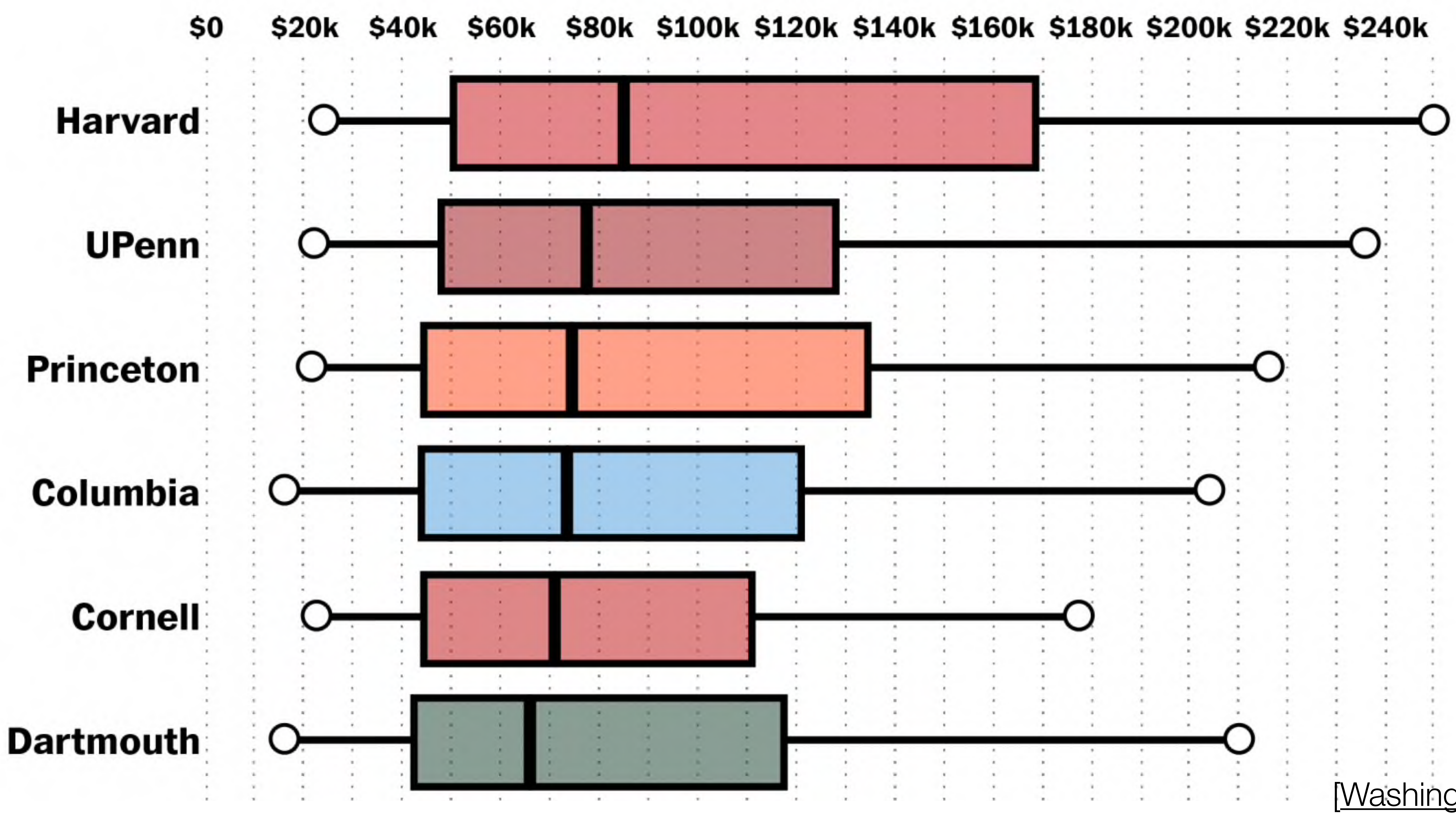
Boxplots

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



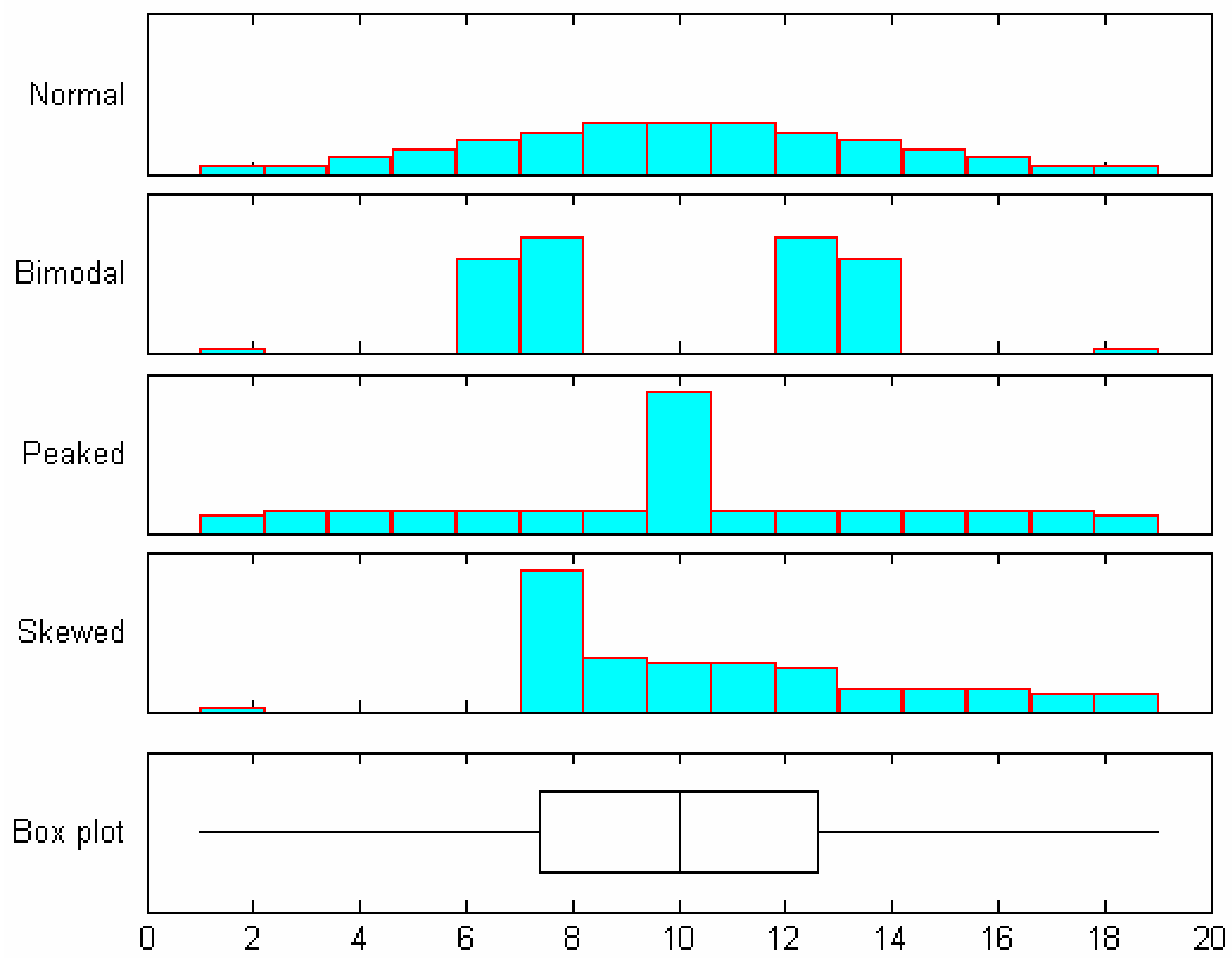
[Flowing Data]

Aggregation: Boxplots



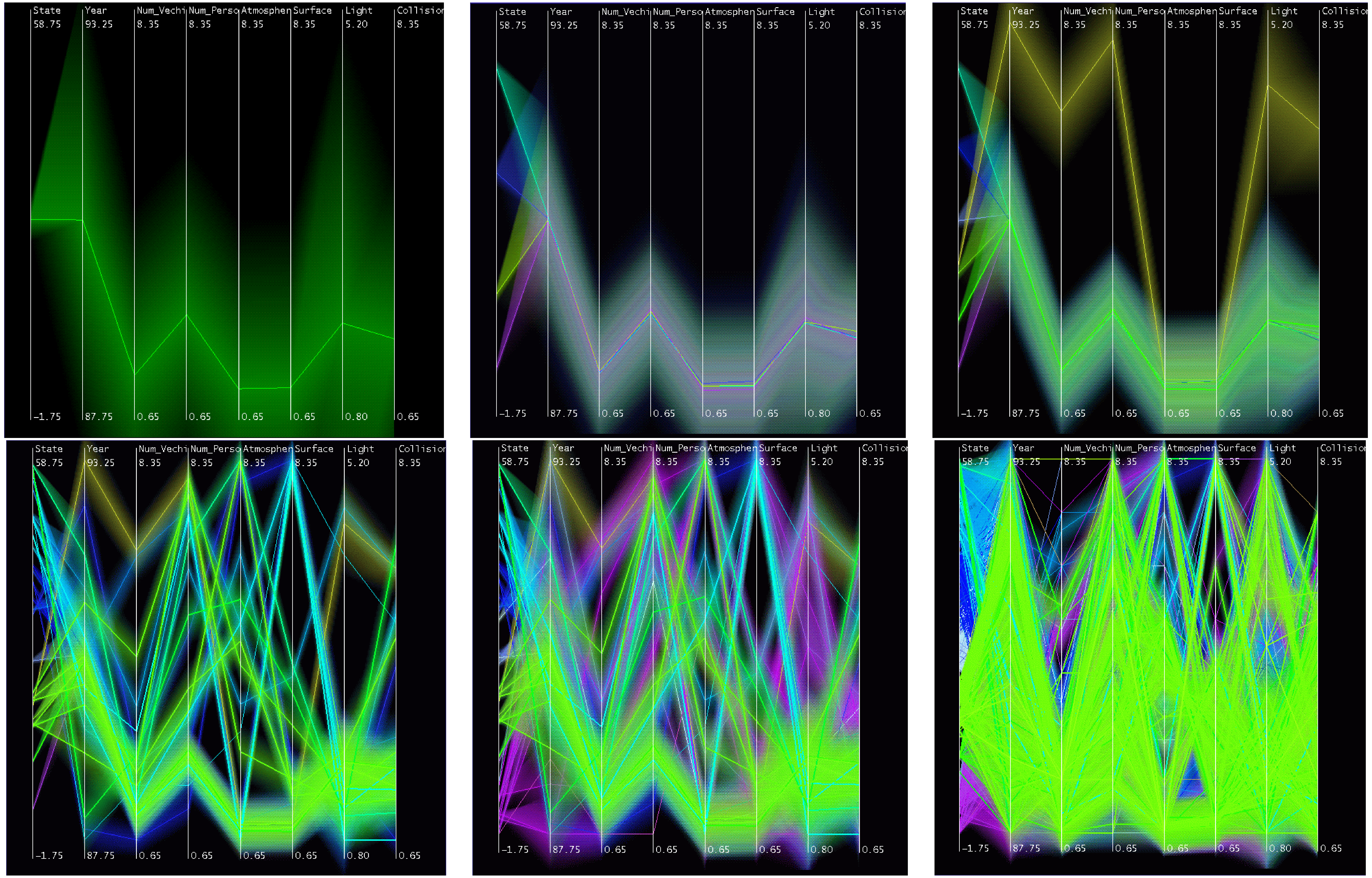
[Washington Post, 2015]

Four Distributions, Same Boxplot…



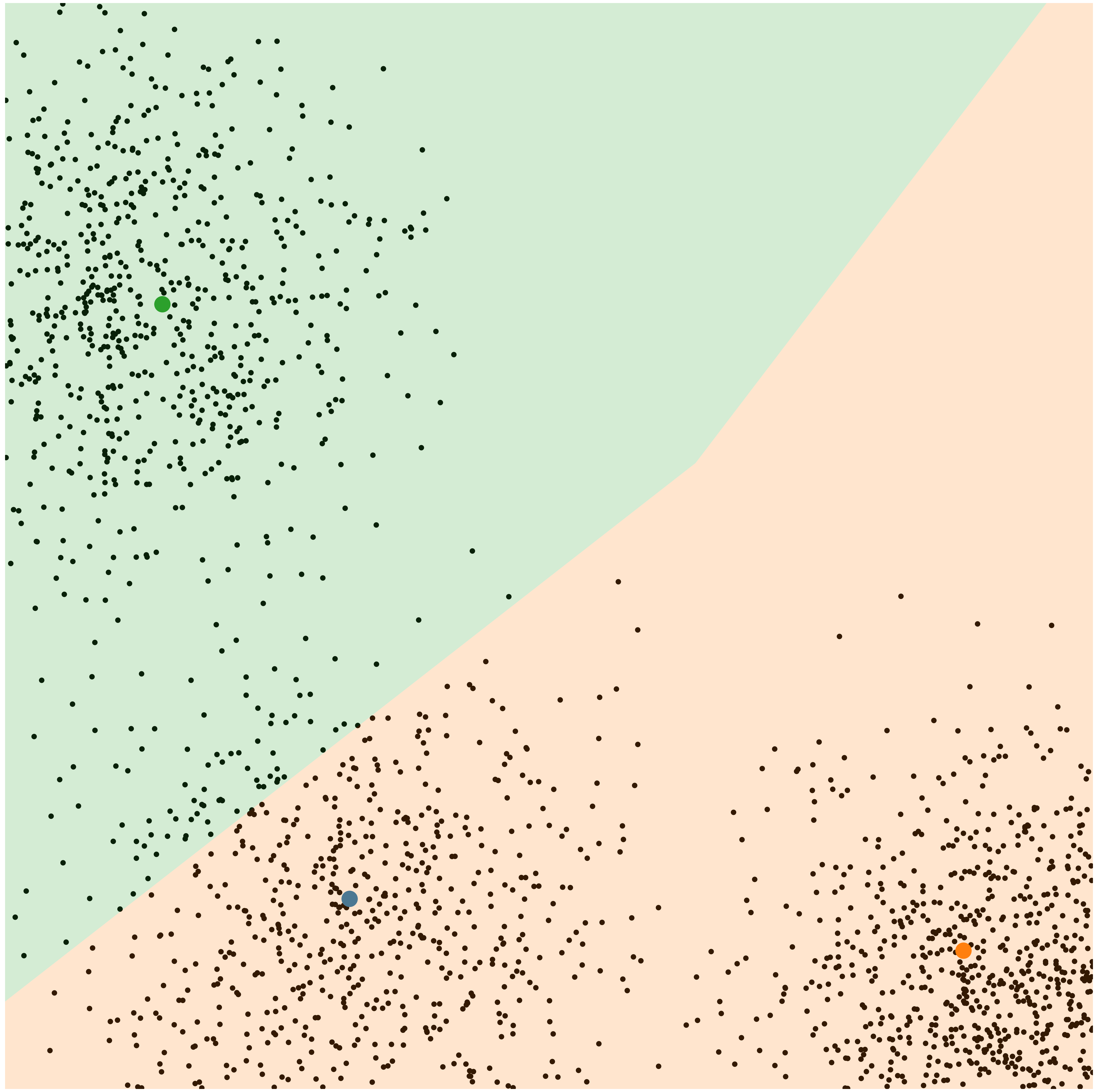
[C. Choonpradub and D. McNeil, 2005]

Hierarchical Parallel Coordinates



[Fua et al., 1999]

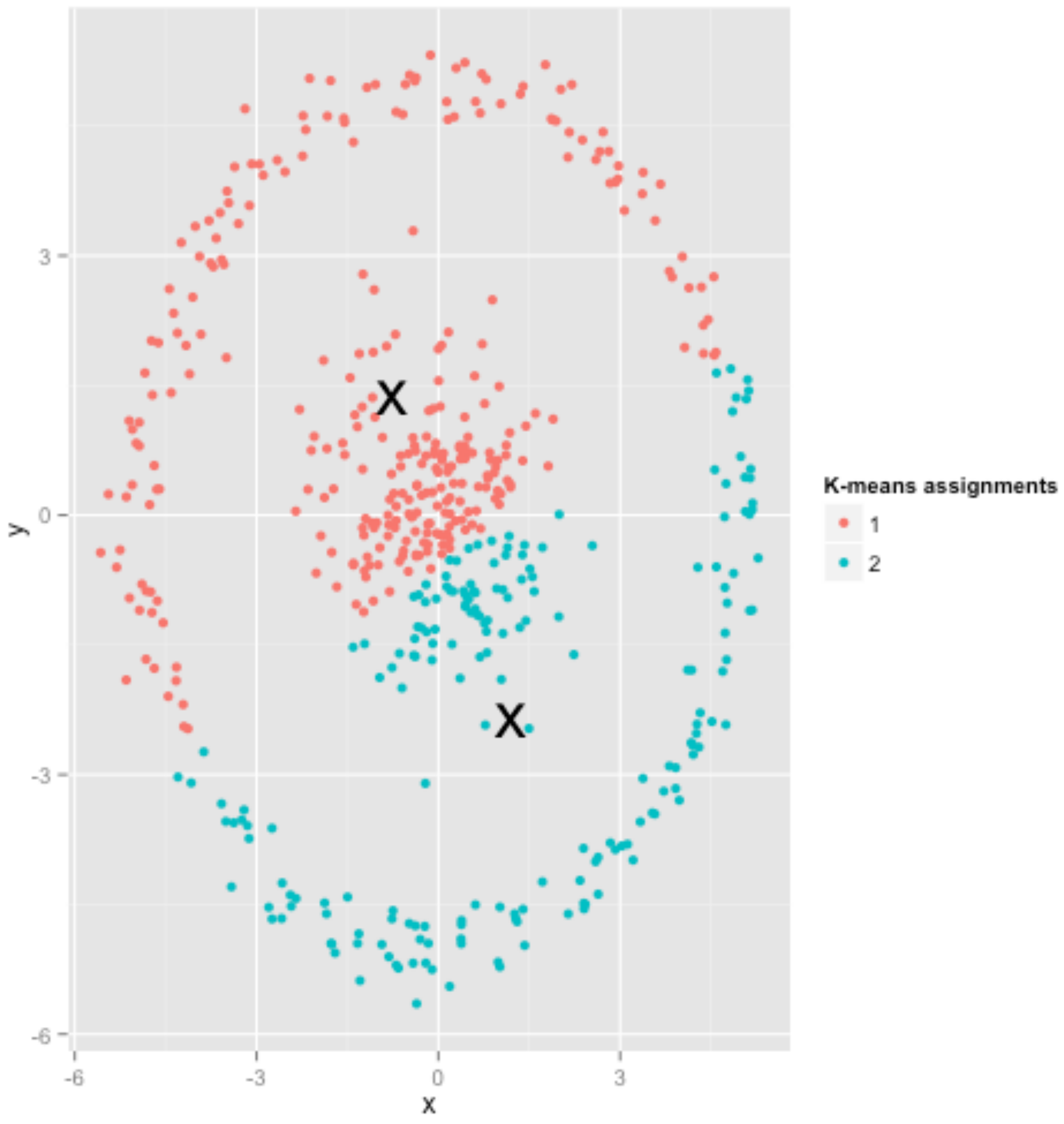
K-Means



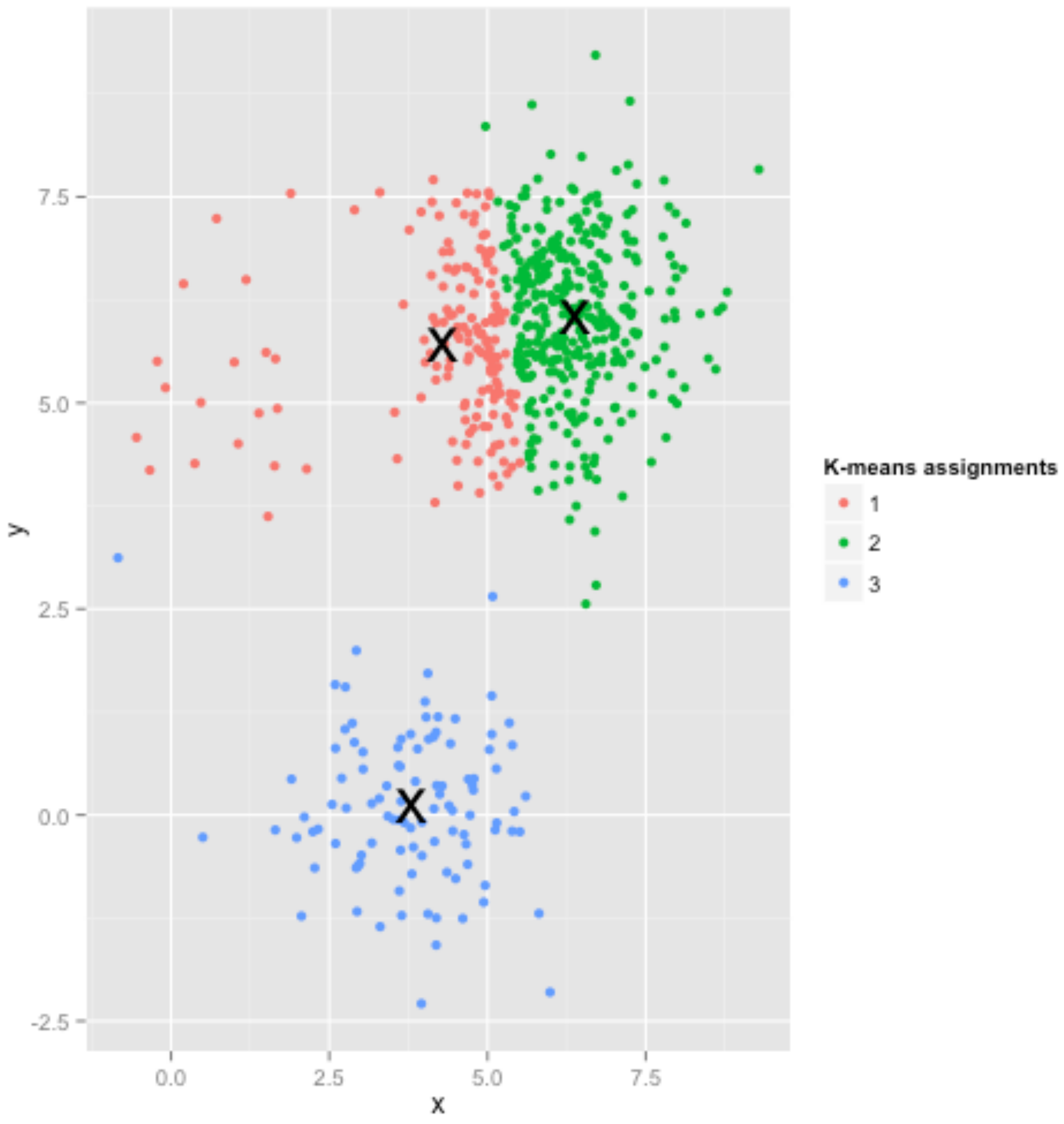
Run

[C. Polis, 2014]

K-Means Issues



Shape



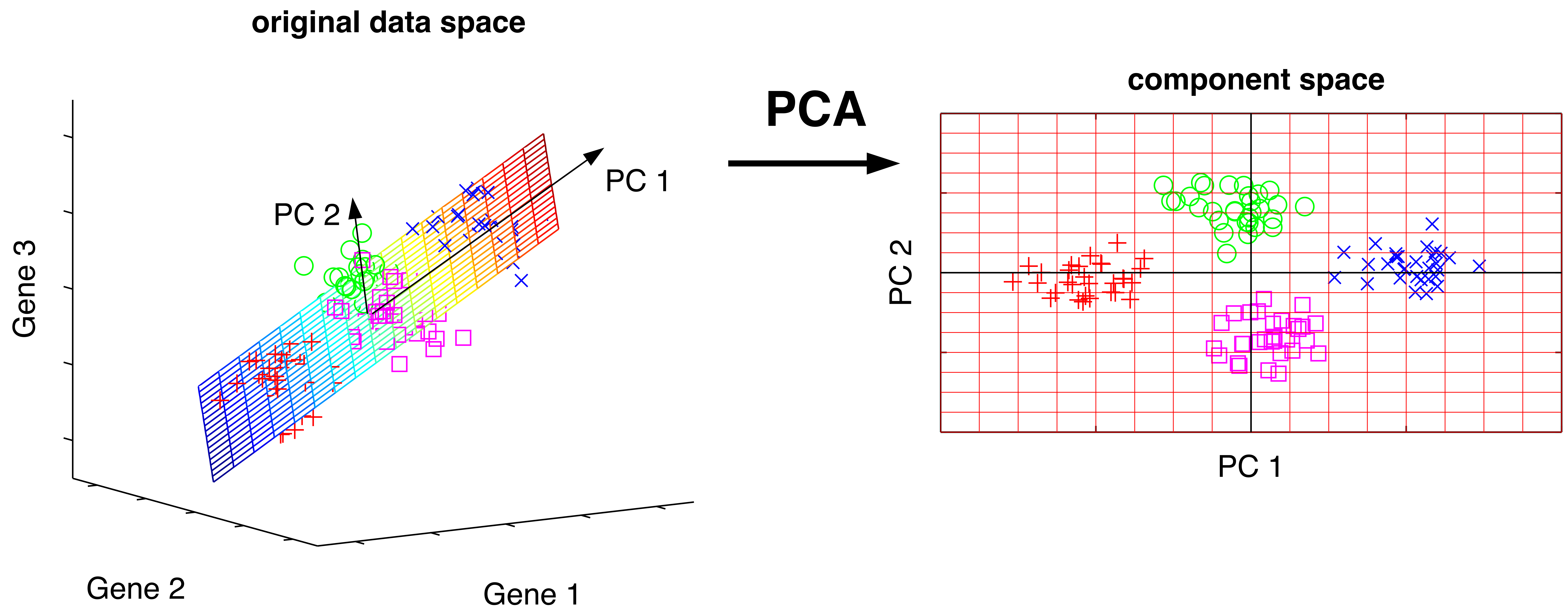
Number of Clusters

[D. Robinson, 2015]

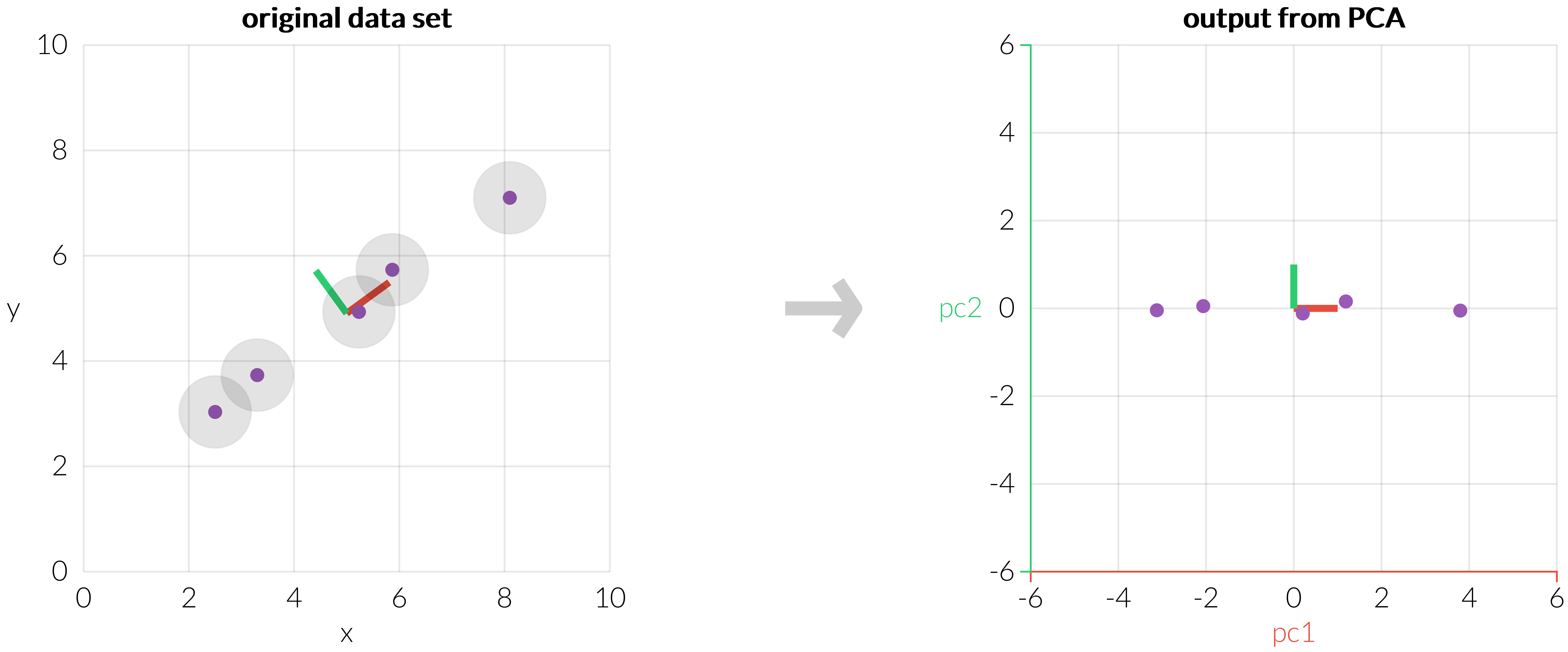
Dimensionality Reduction

- 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)



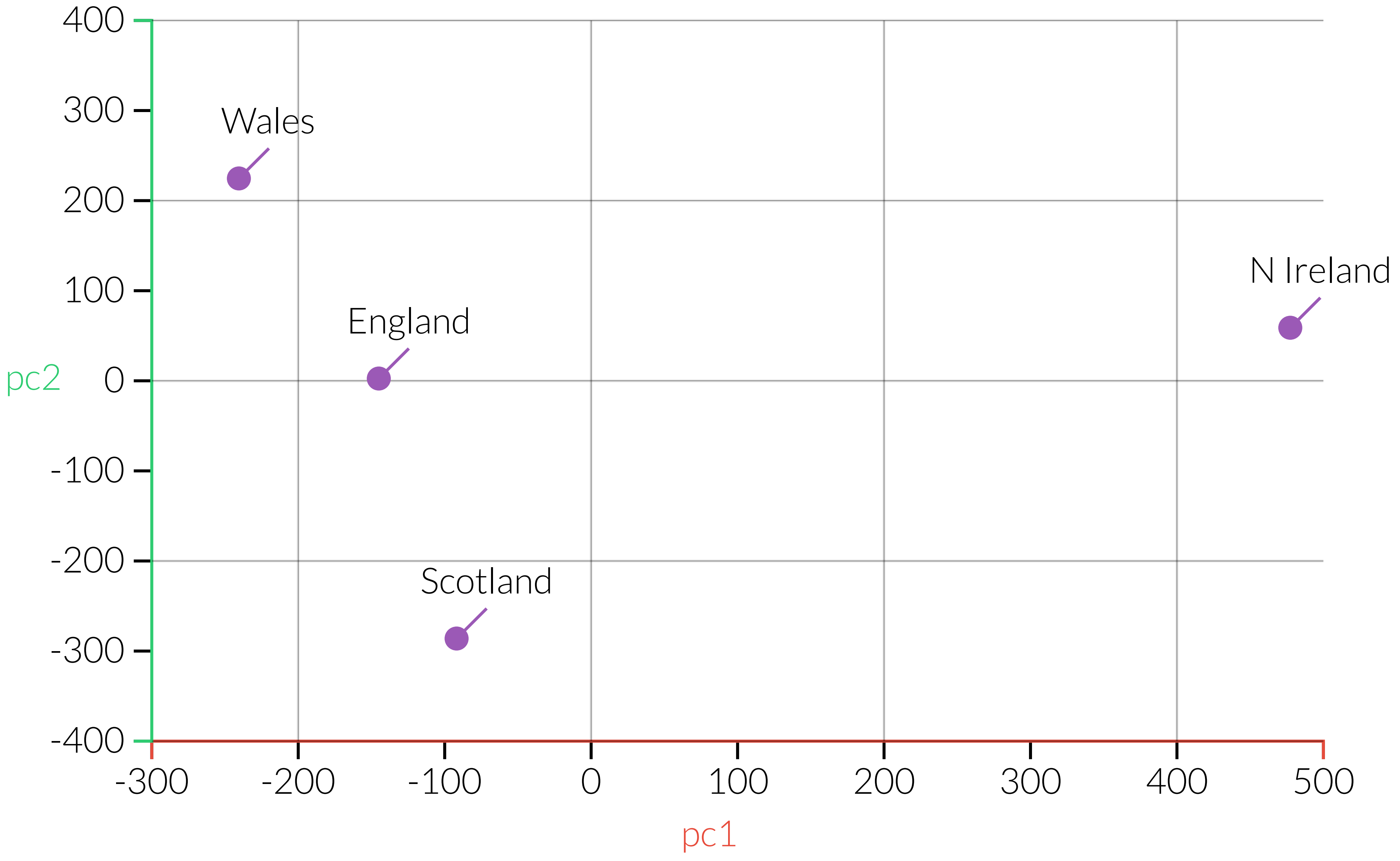
PCA



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

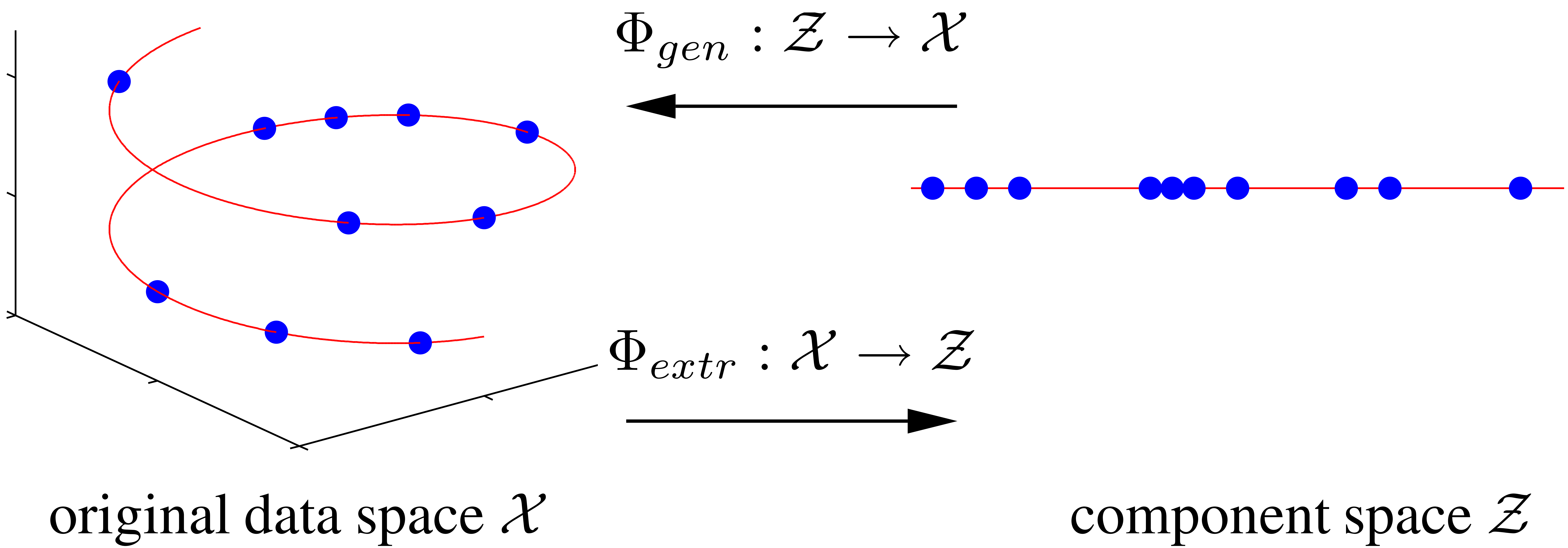
17 dimensions to 2

	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

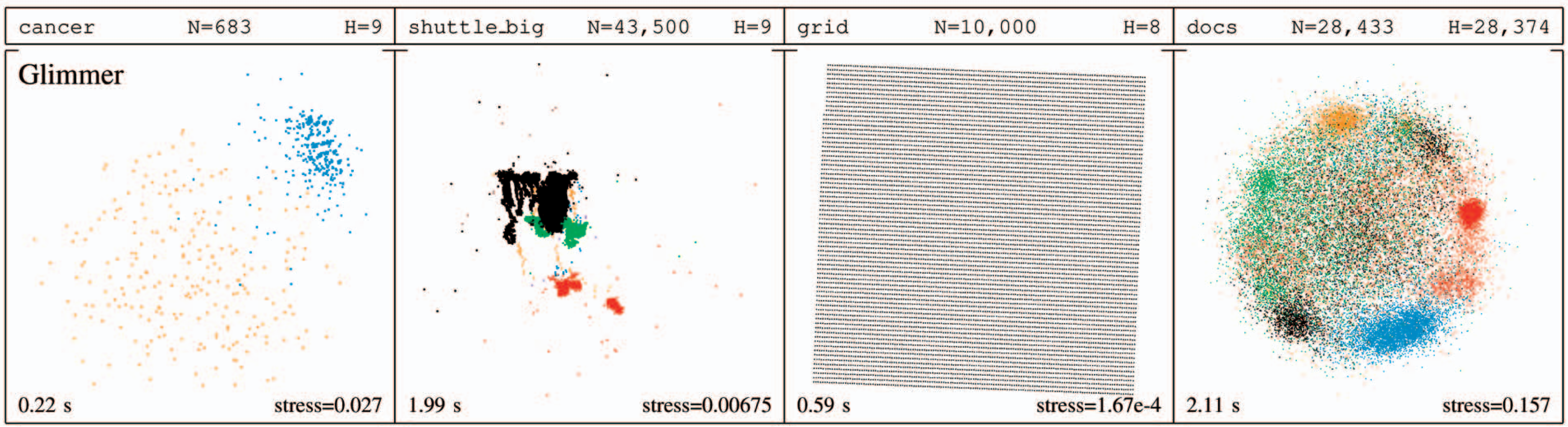


[Principle Component Analysis Explained, Explained Visually, V. Powell & L. Lehe, 2015]

Non-linear Dimensionality Reduction

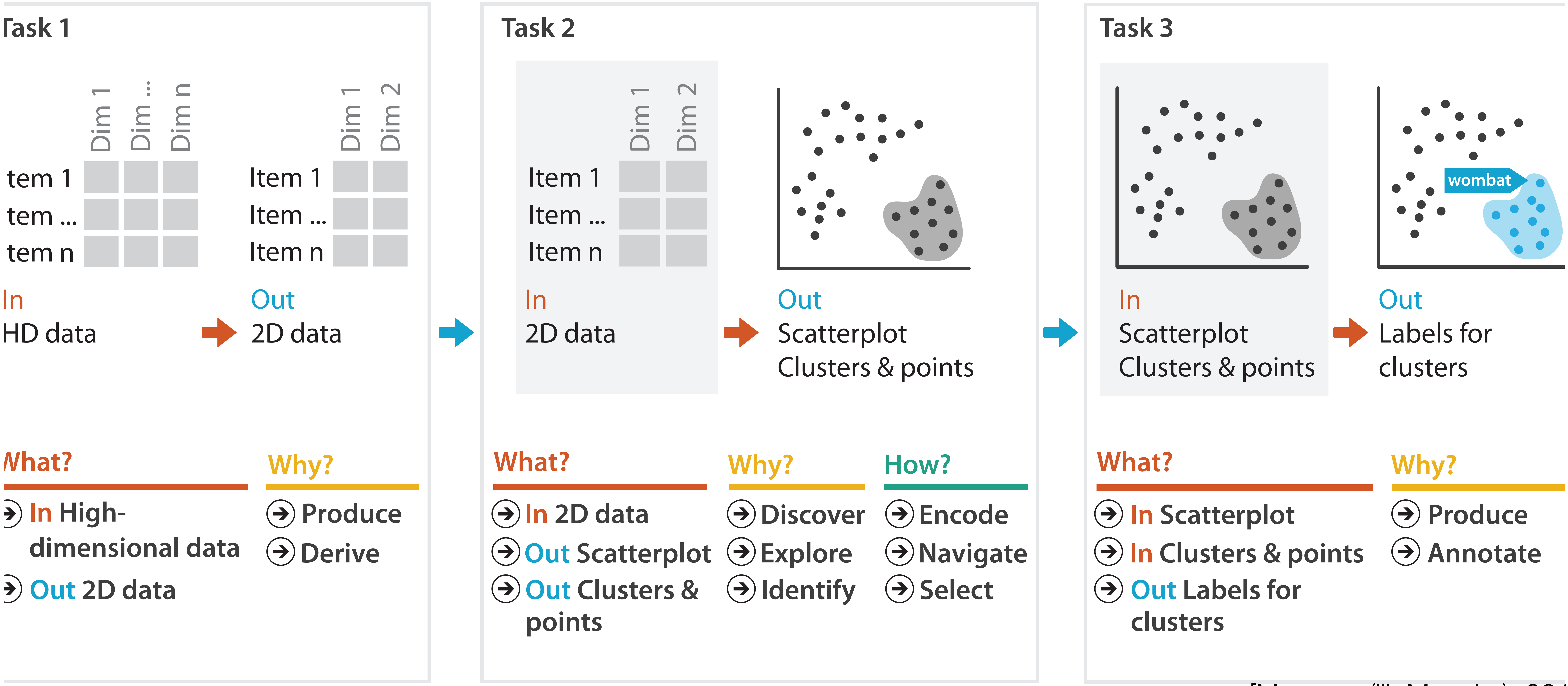


Dimensionality Reduction in Visualization



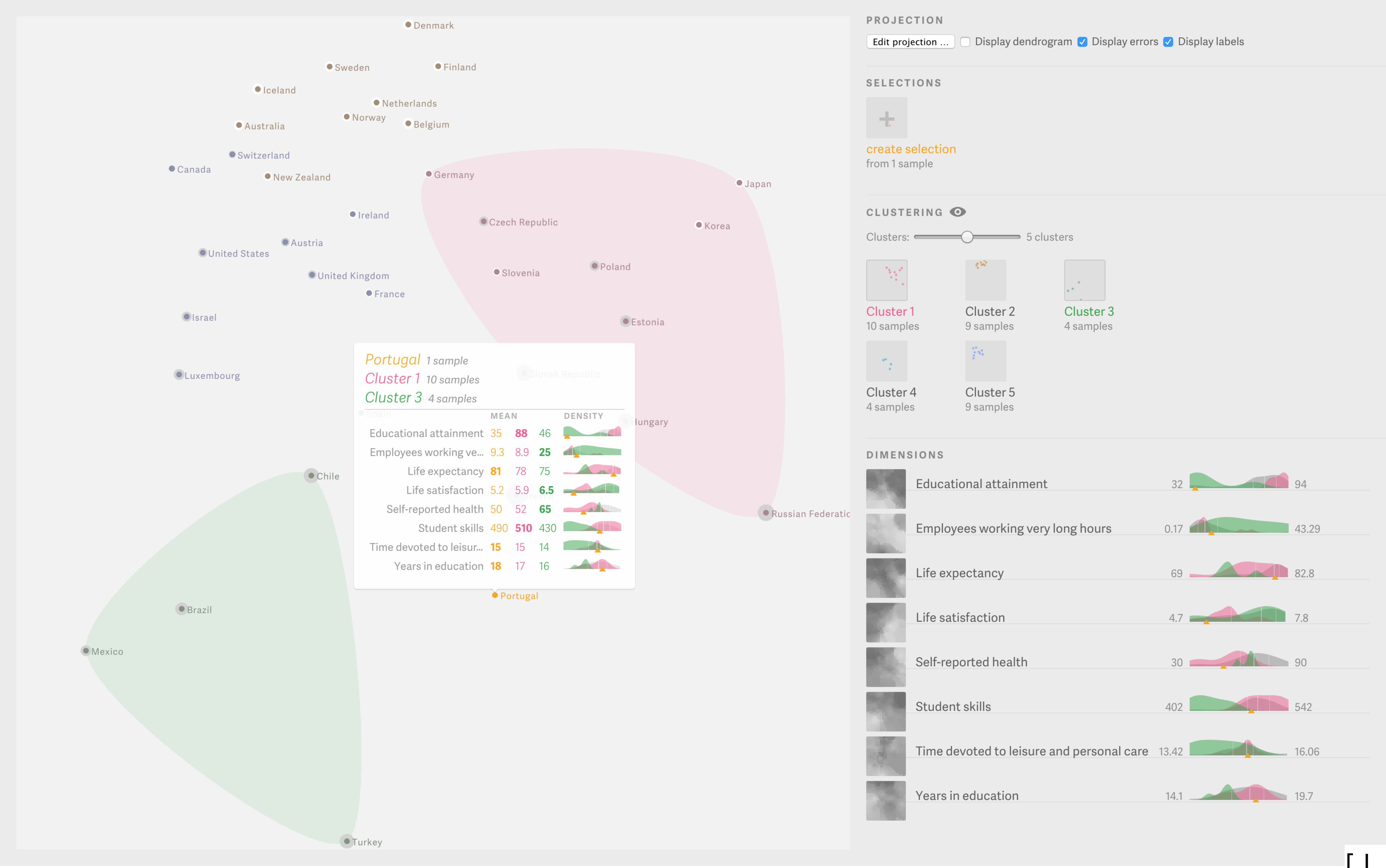
[Glimmer, Ingram et al., 2009]

Tasks in Understanding High-Dim. Data



[Munzner (ill. Maguire), 2014]

Probing Projections



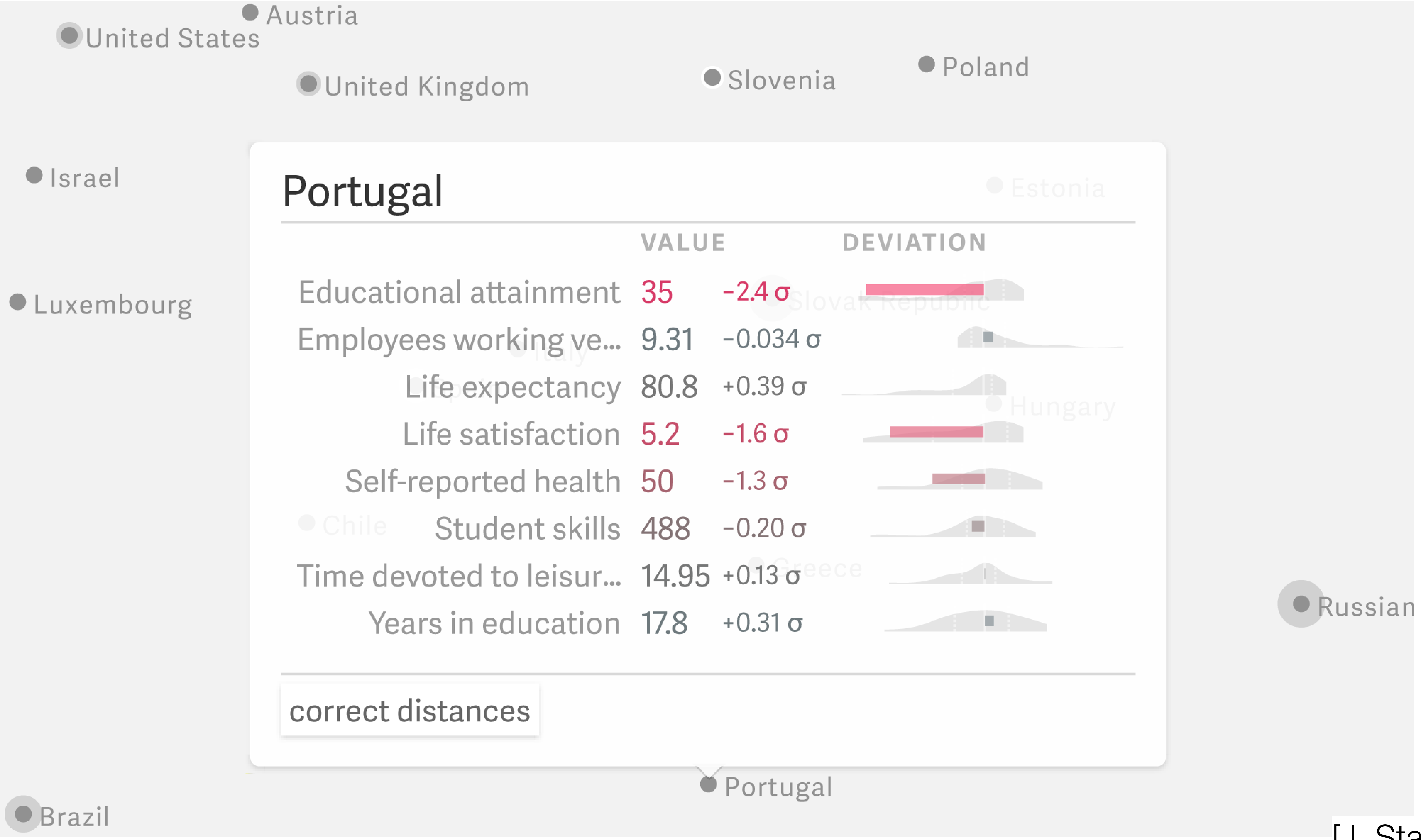
[J. Stahnke et al., 2015]

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

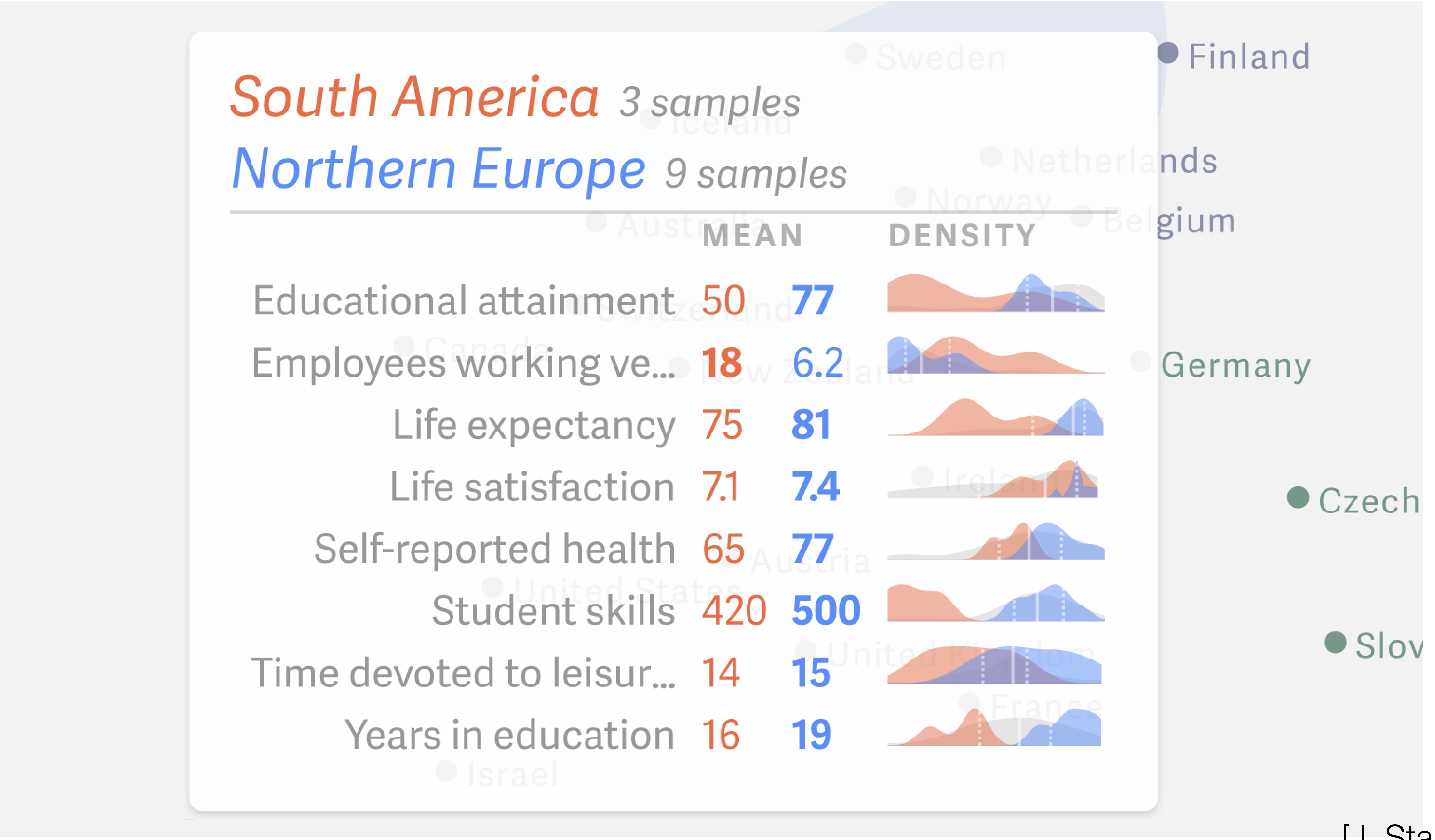
[J. Stahnke et al., 2015]

Tooltips with statistics



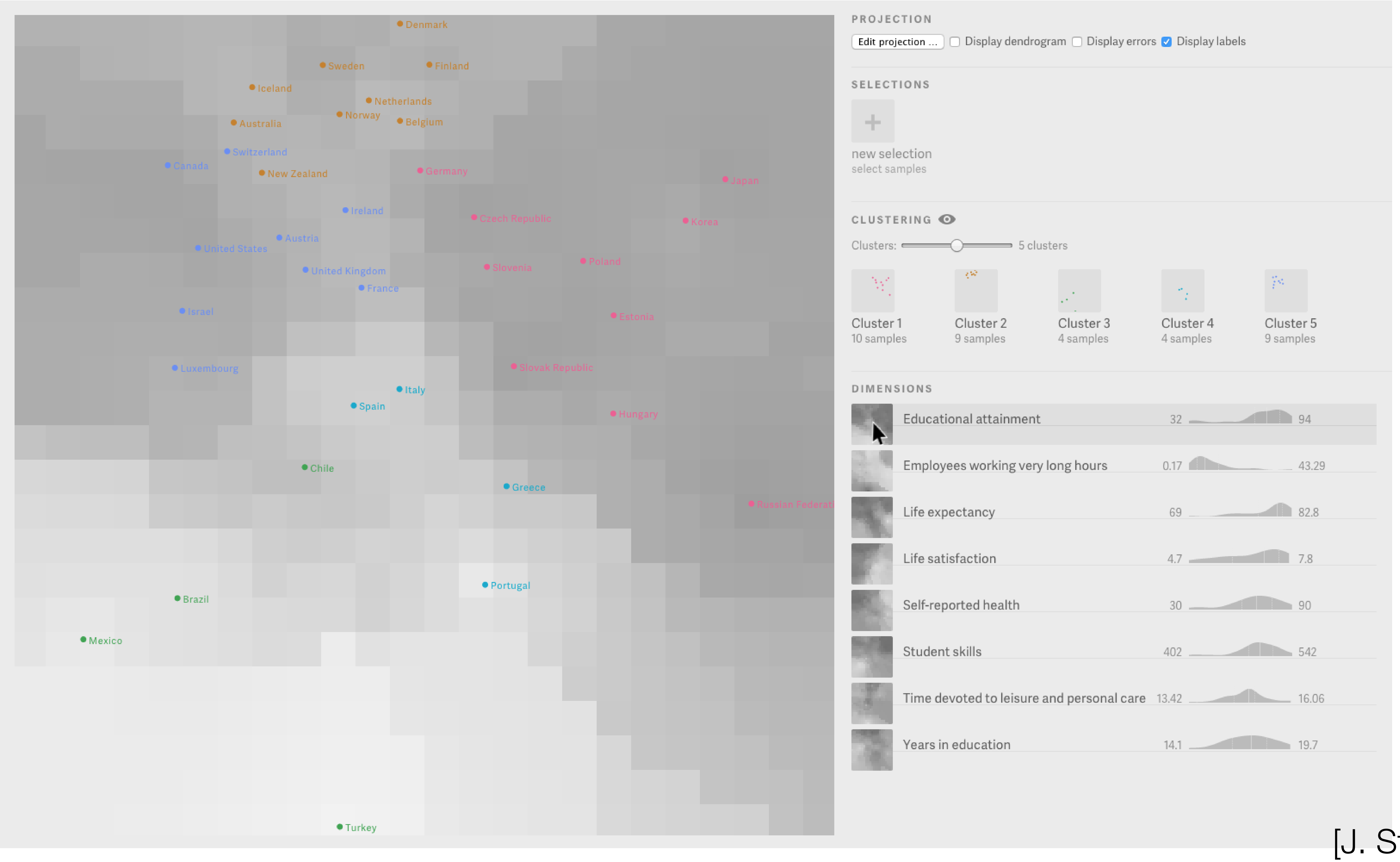
[J. Stahnke et al., 2015]

Comparing Two Groups



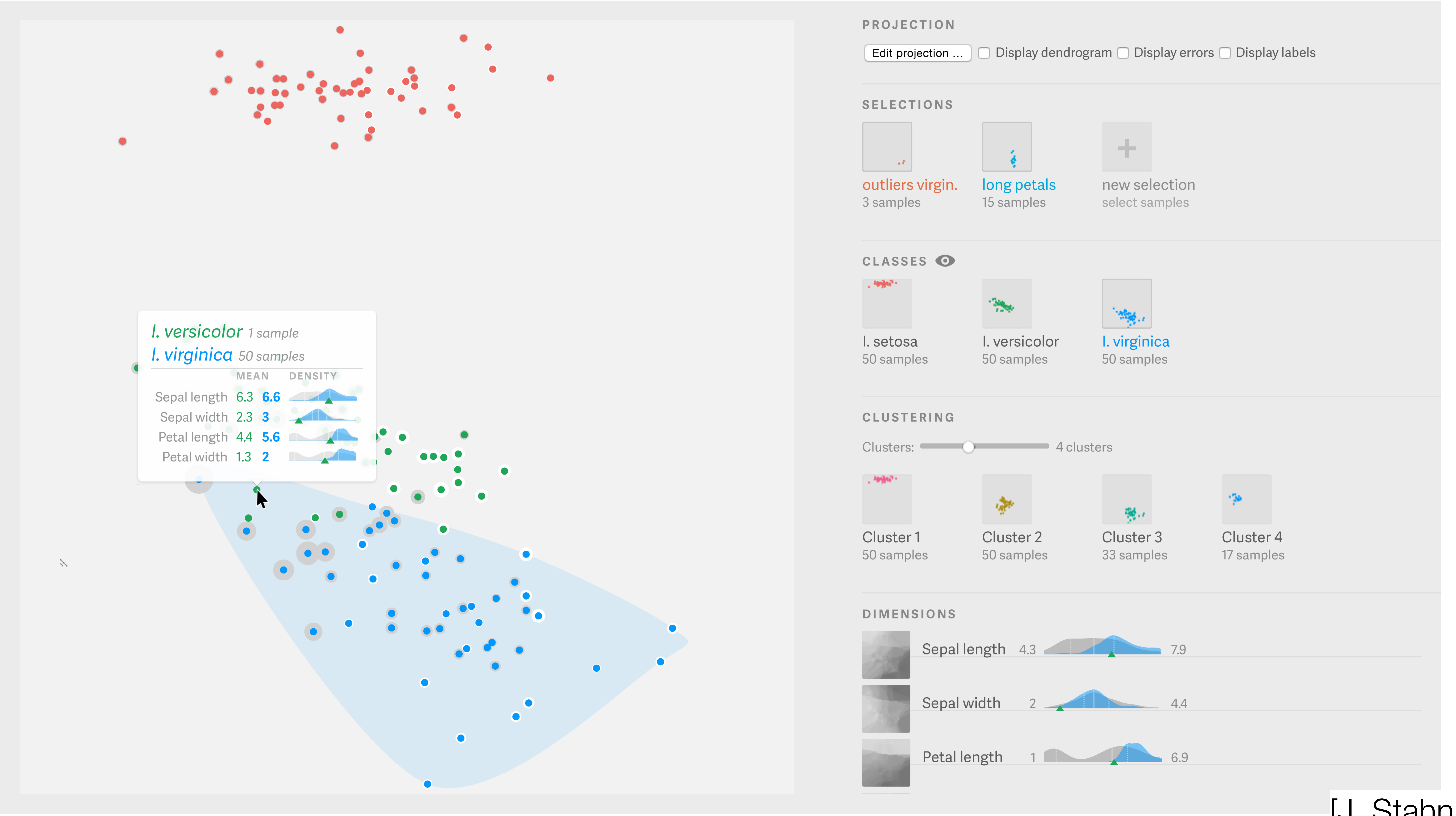
[J. Stahnke et al., 2015]

Heatmap from Dimension Hover



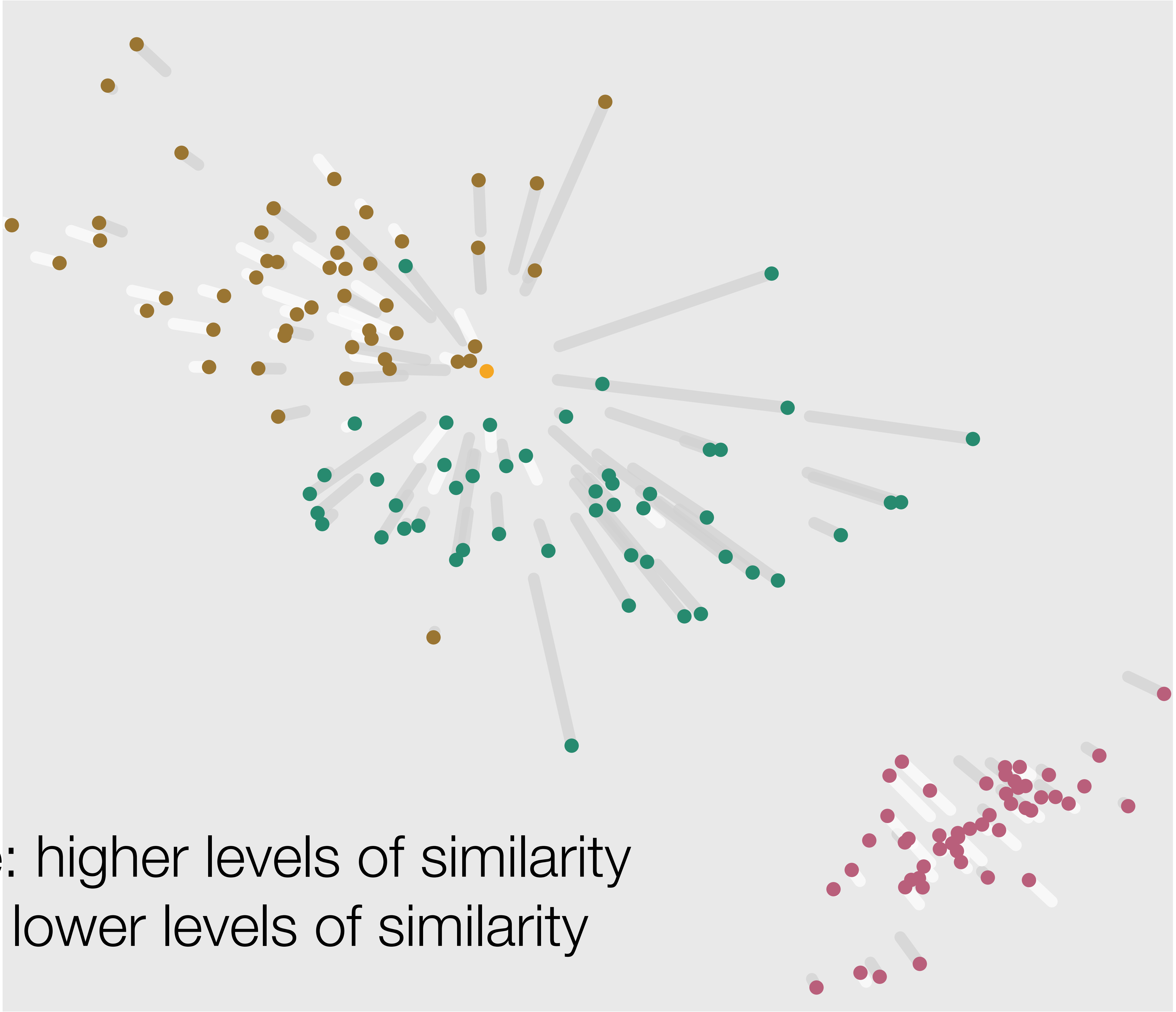
[J. Stahnke et al., 2015]

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]