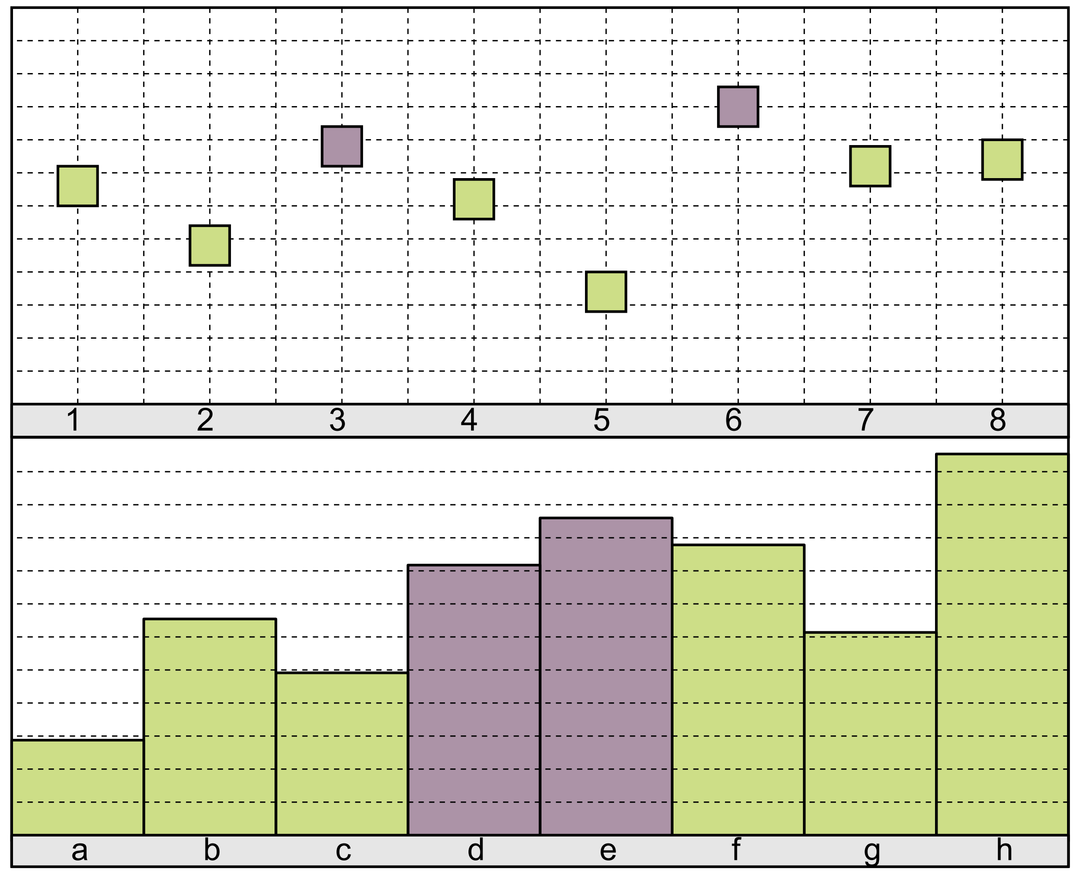
Data Visualization (CSCI 627/490)

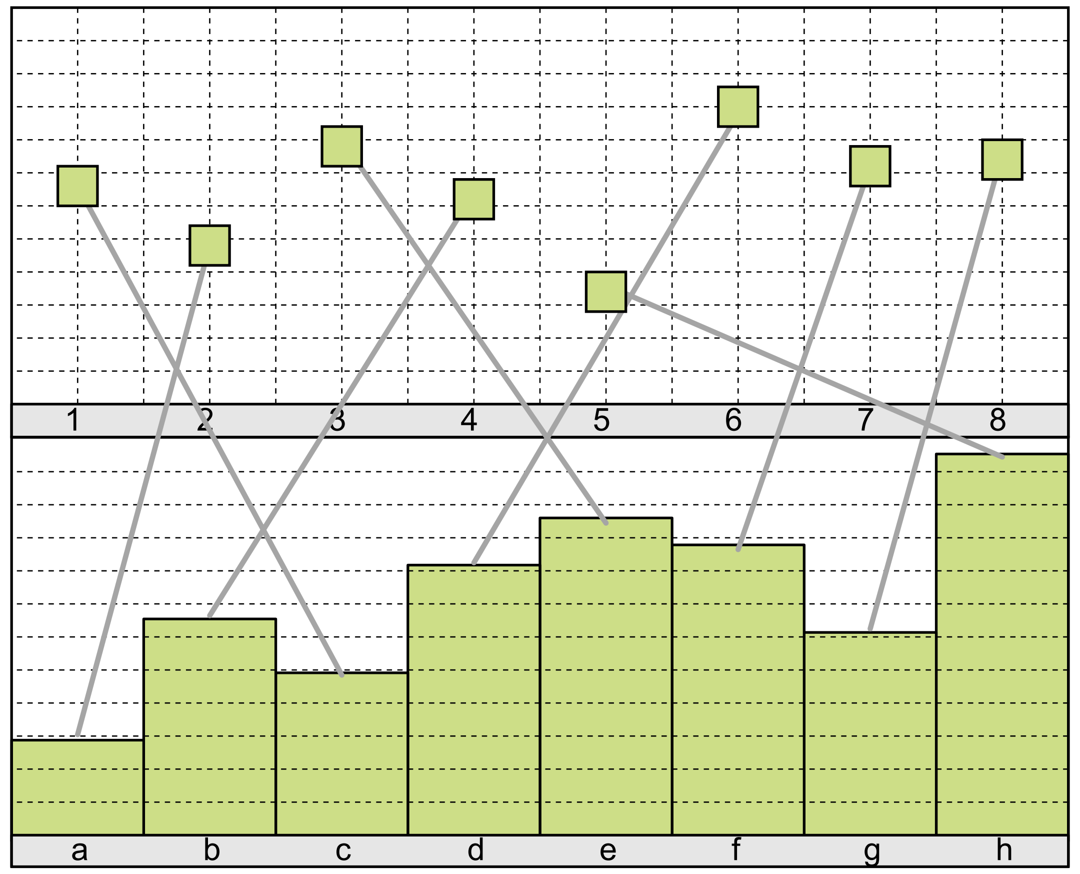
Filtering & Aggregation

Dr. David Koop

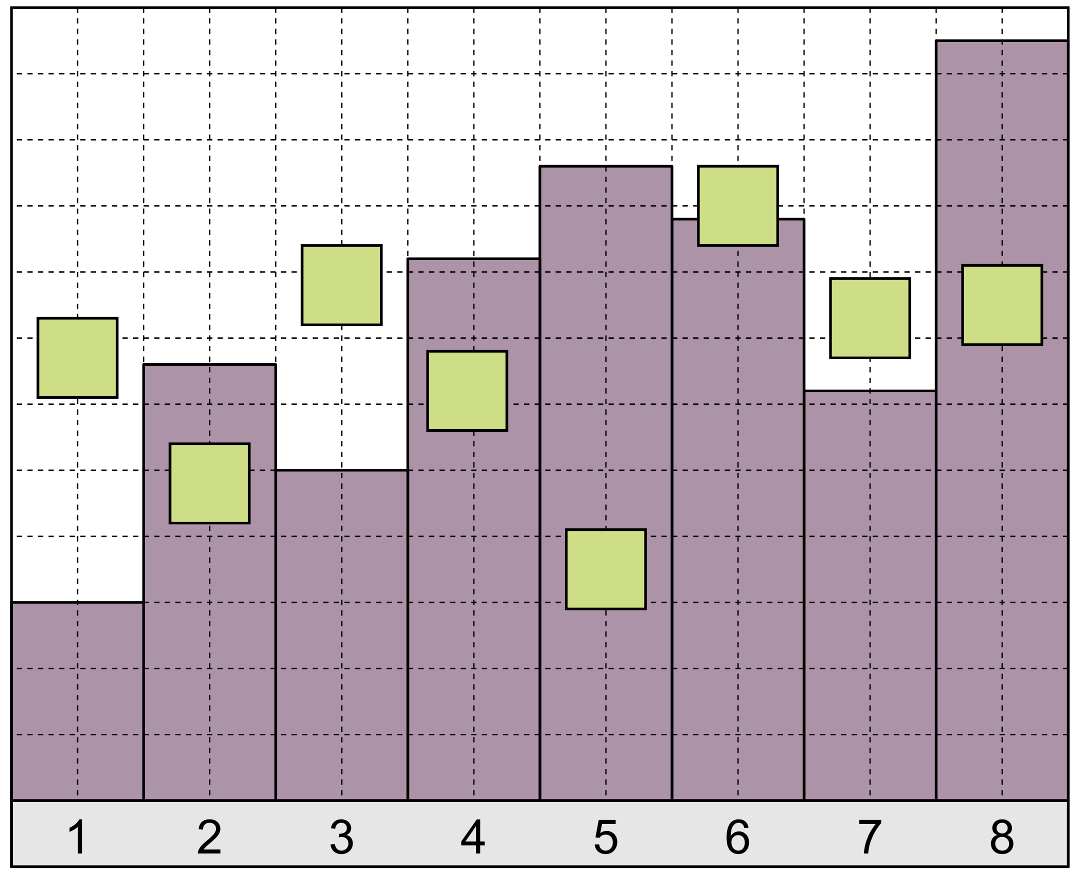
Composite Visualization Techniques



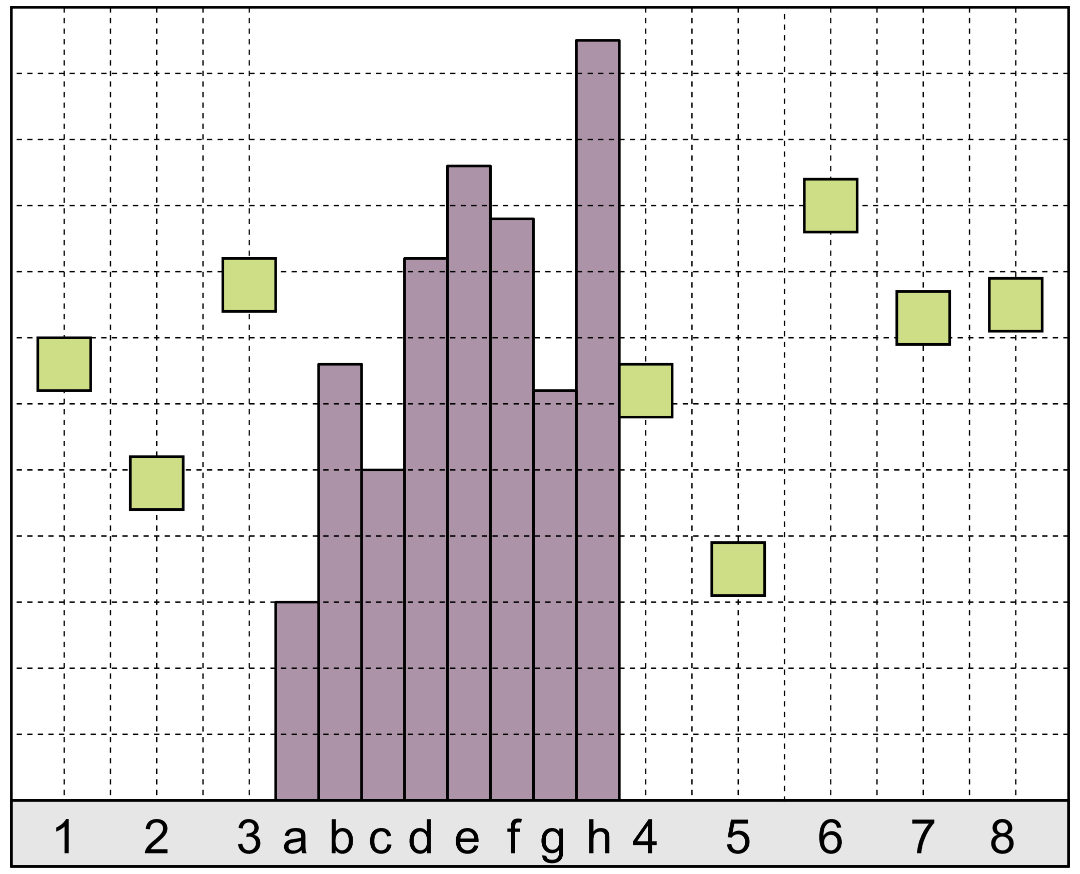
(a) Juxtaposed views.



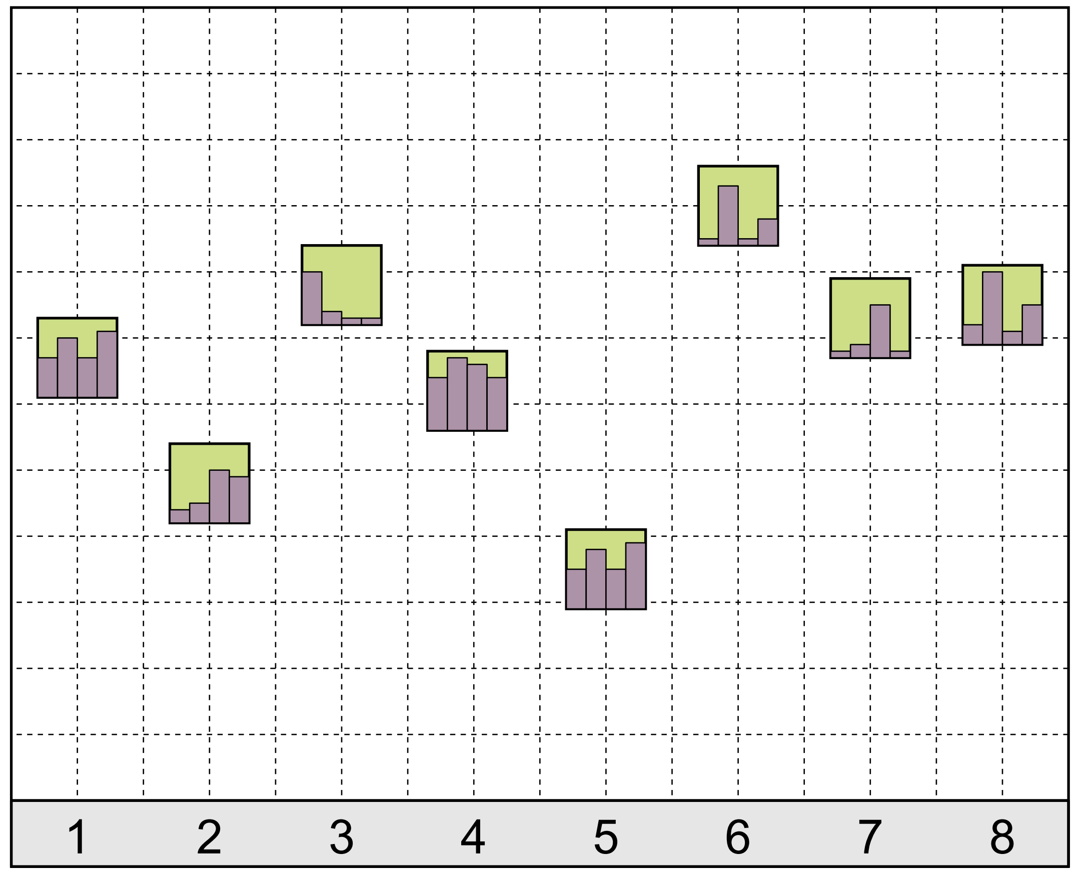
(b) Integrated views.



(c) Superimposed views.



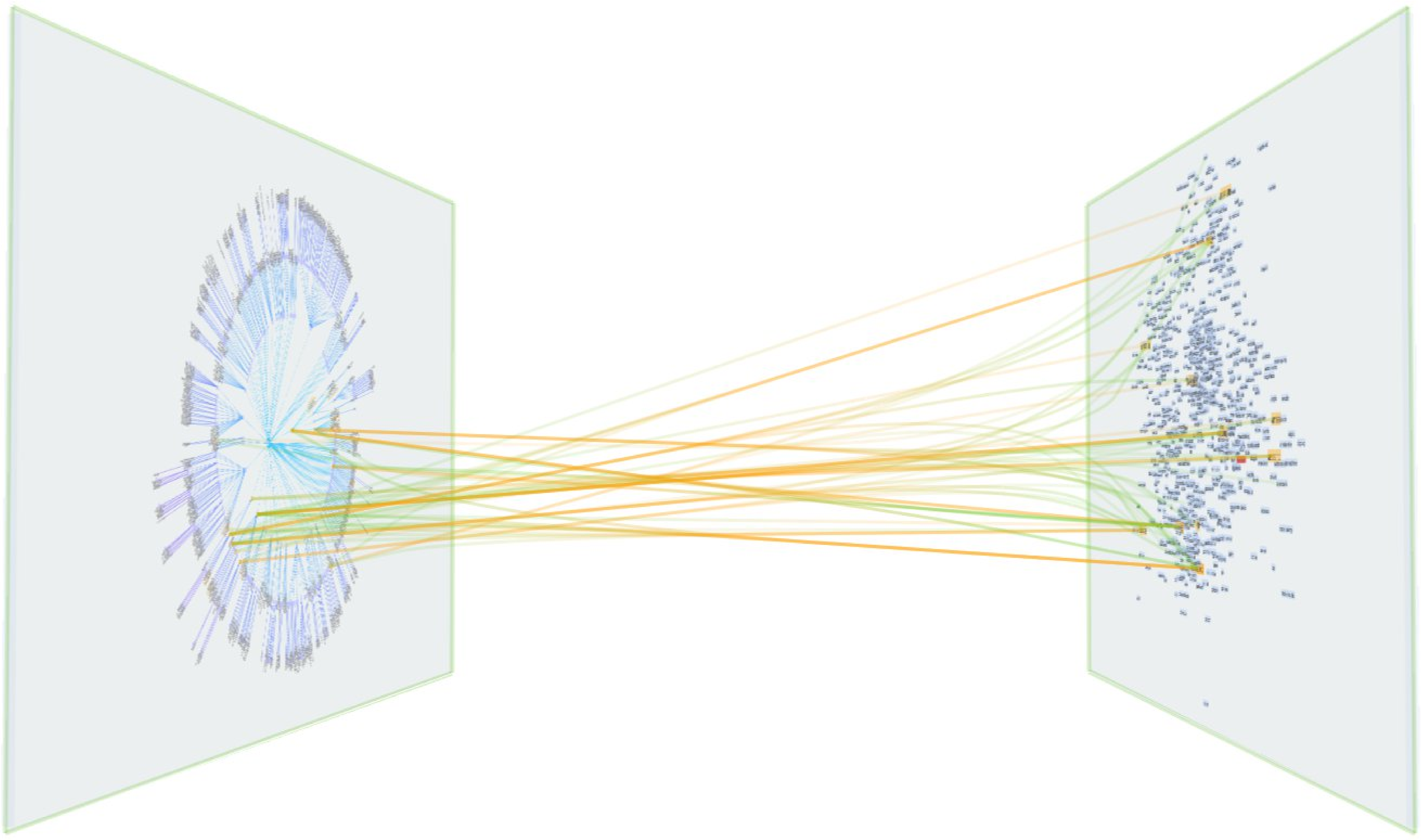
(d) Overloaded views.



(e) Nested views.

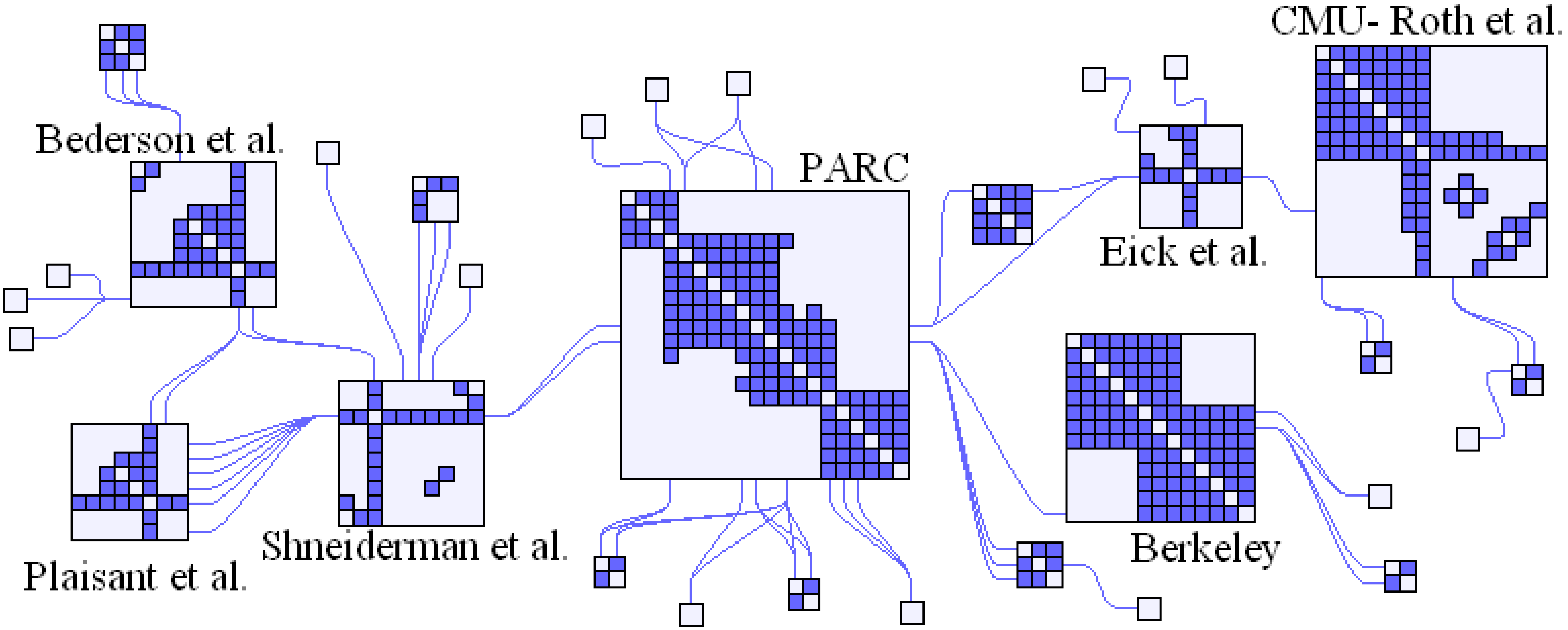
[W. Javed and N. Elmqvist, 2012]

What is this technique?



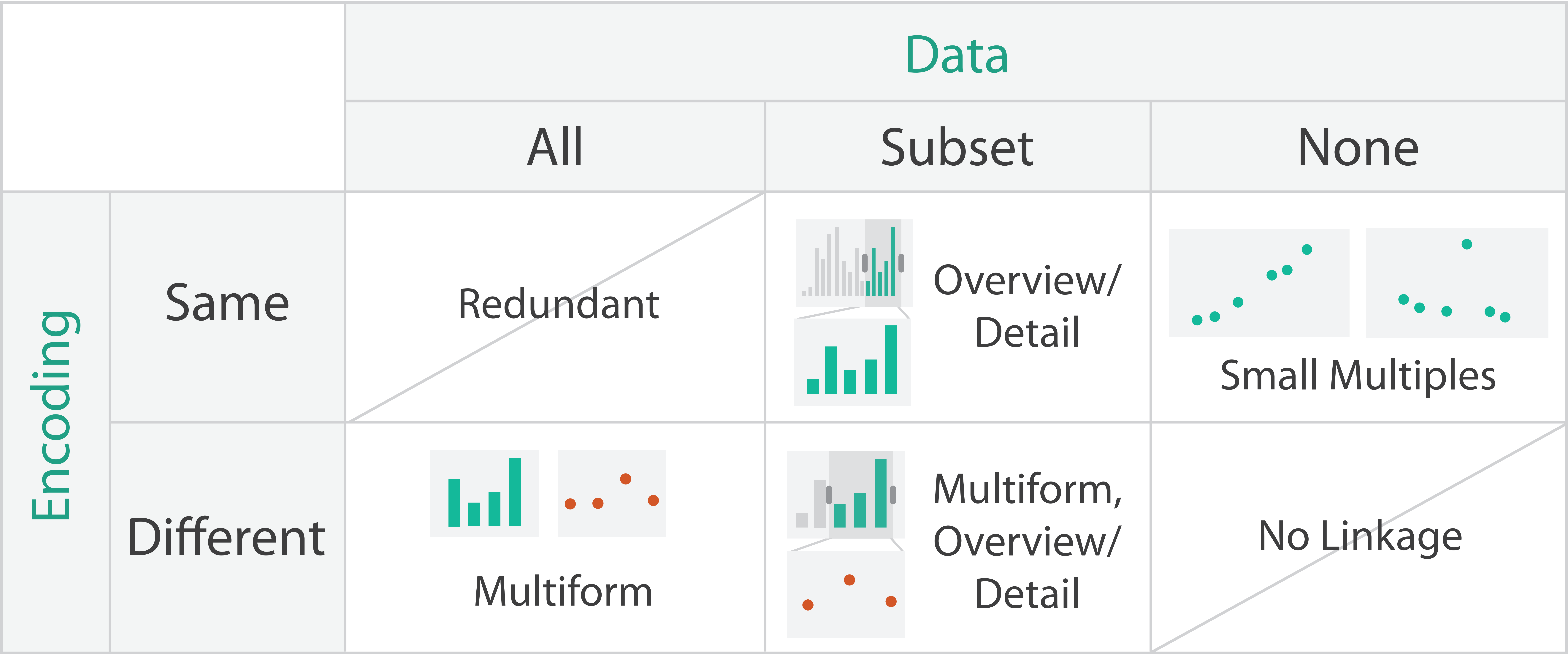
[VisLink, Collins and Carpendale, 2007]

What is this technique?



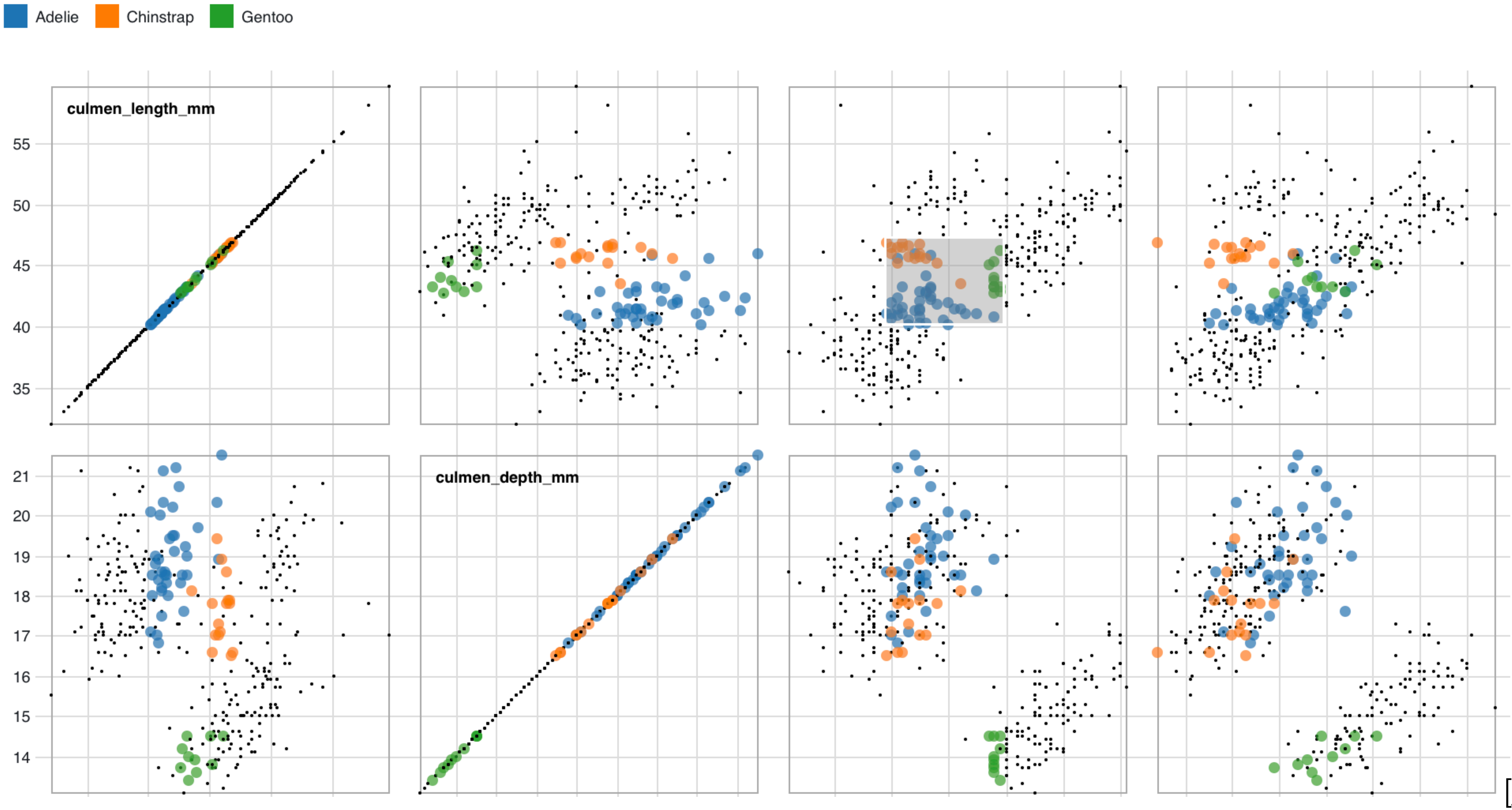
[NodeTrix, N. Henry et al., 2007]

Multiple Views



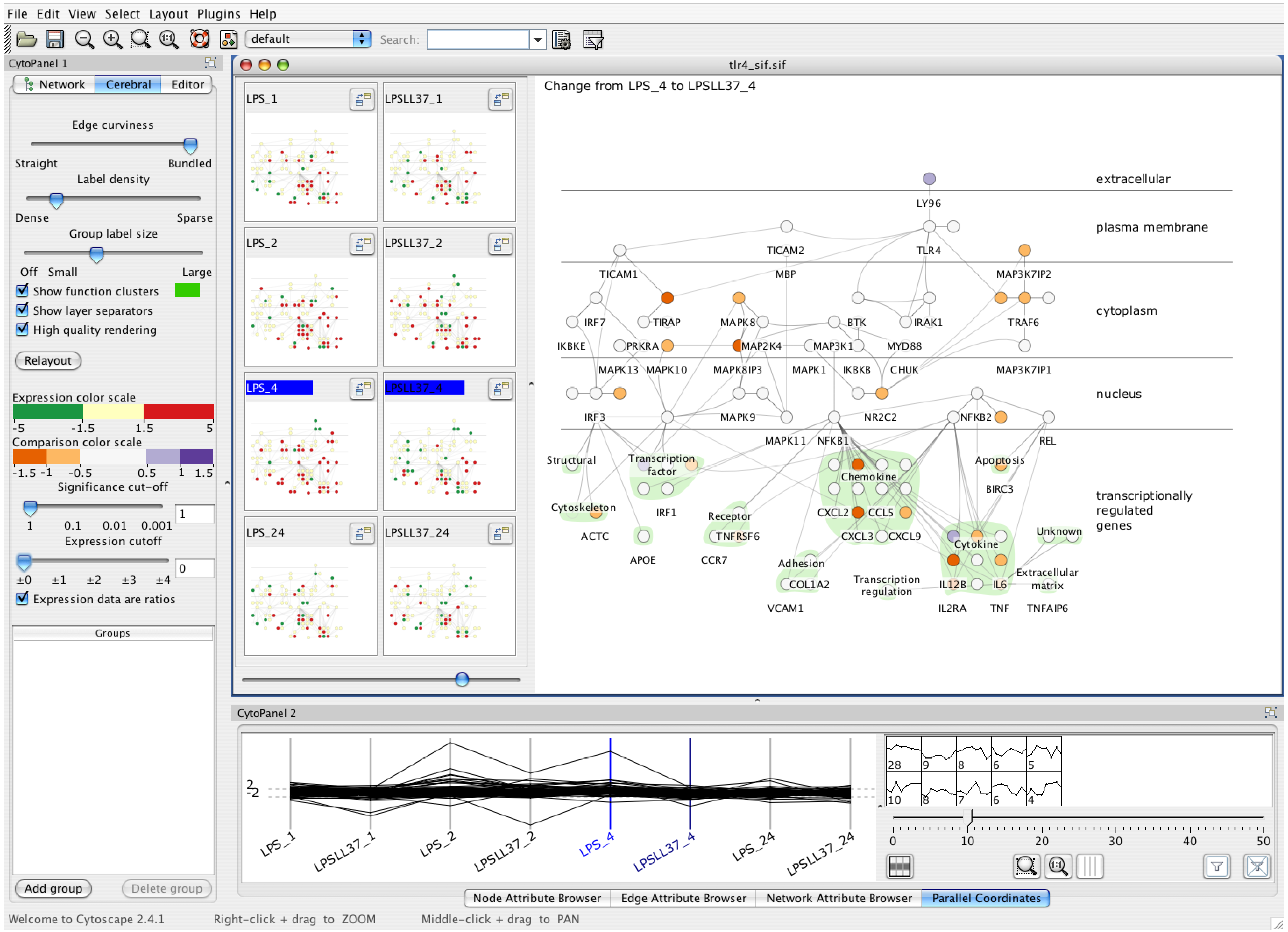
[Munzner (ill. Maguire), 2014]

Brushing



[M. Bostock]

Multiform & Small Multiples

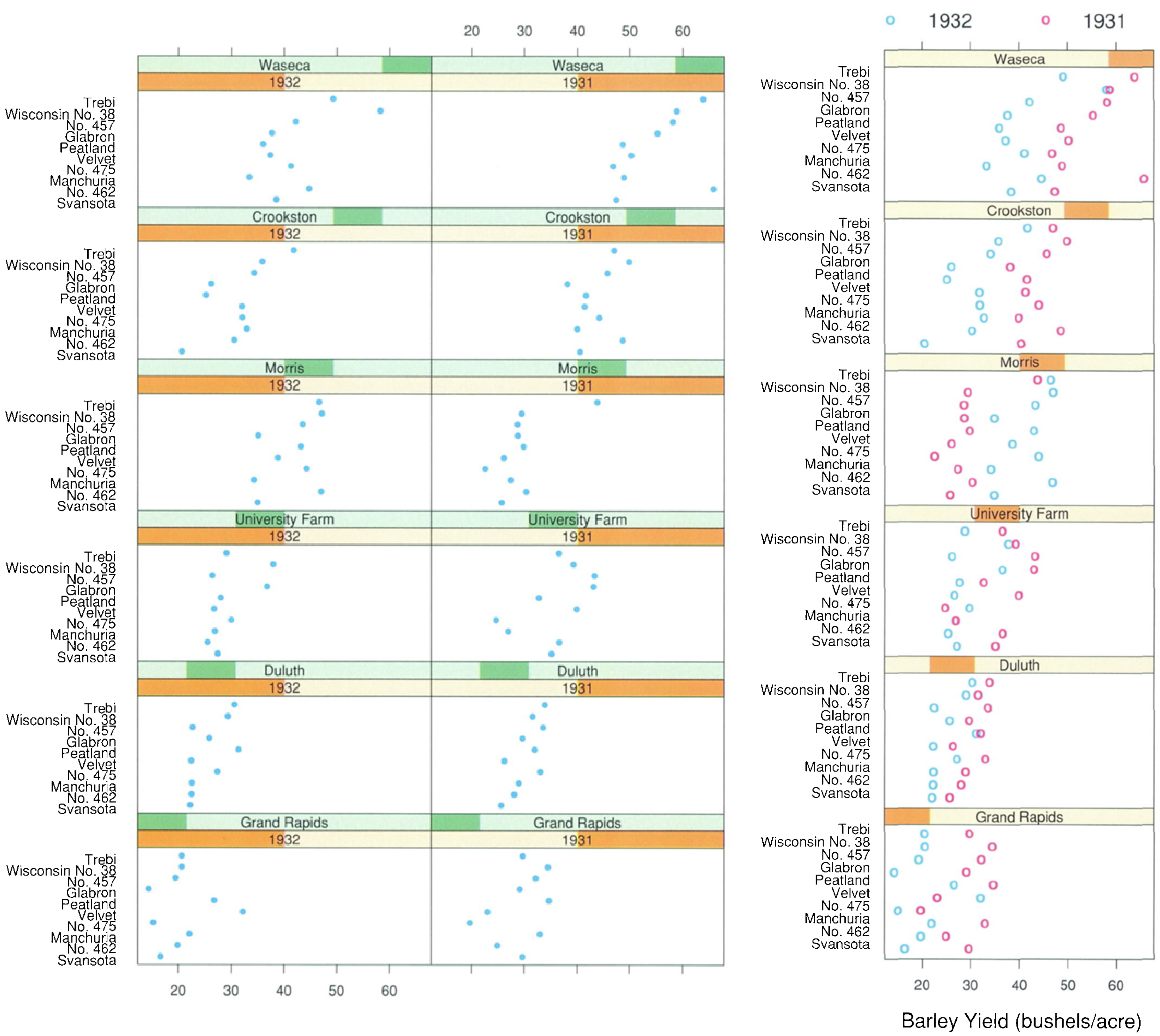


[Barsky et al., 2008]

Partitioned Views

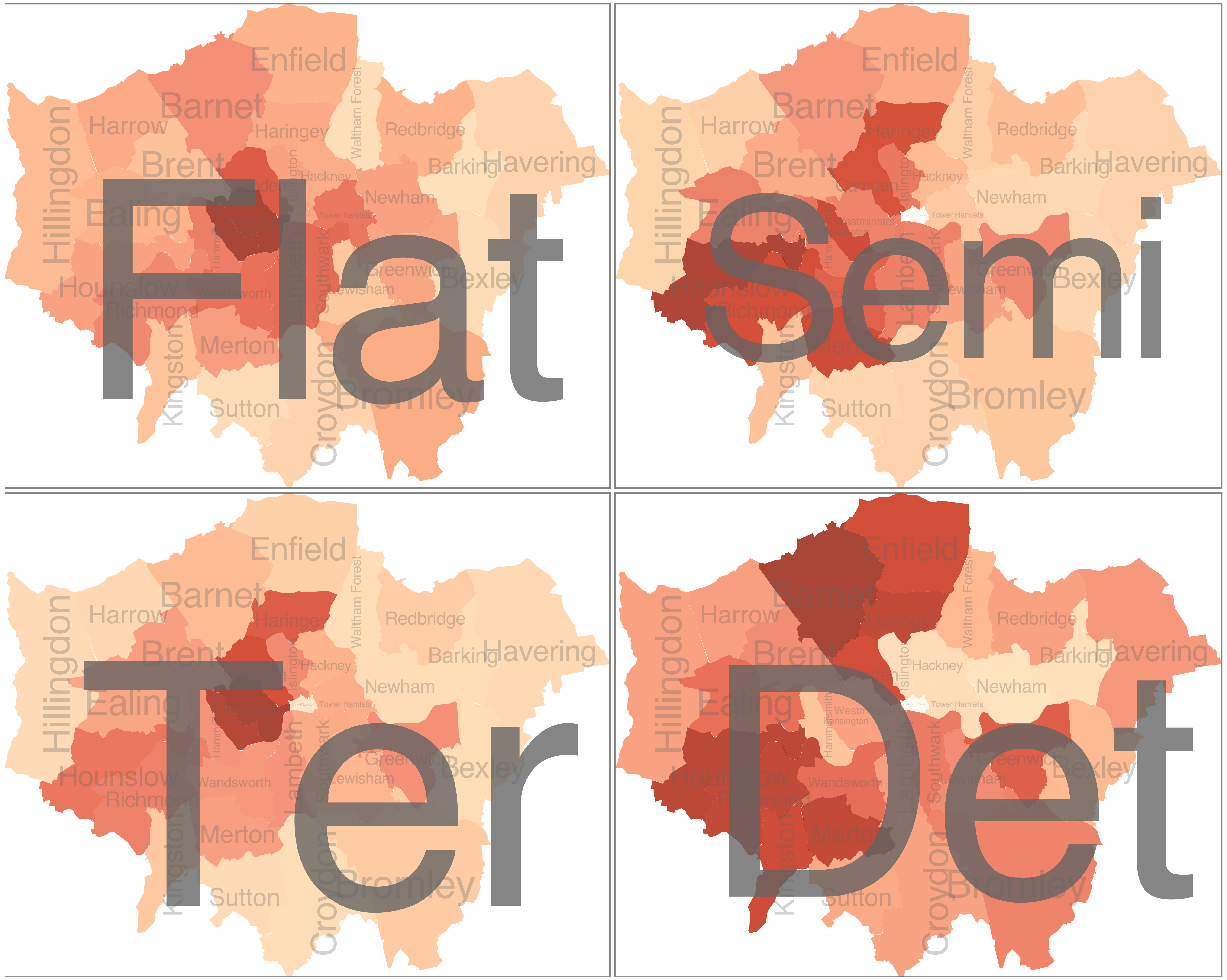
- Split dataset into groups and visualize each group
- Extremes: one item per group, one group for all items
- Can be a hierarchy
  - Order: which splits are more "related"?
  - Which attributes are used to split? usually categorical

Partitioned Views: Trellis Matrix Alignment



[Becker et al., 1996]

Recursive Subdivision: HiVE System



[Slingsby et al., 2009]

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 Friday, Nov. 13

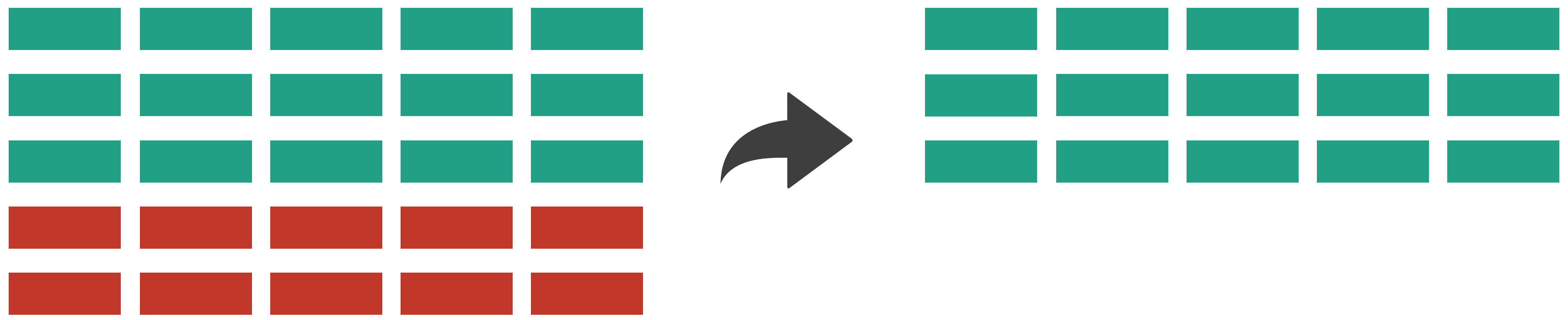
Assignment 5

- To be released soon
- Citi Bike NYC Data
  - Trips between neighborhoods
- Covers
  - Multiple Views
  - Filtering
  - Aggregation
  - Brushing

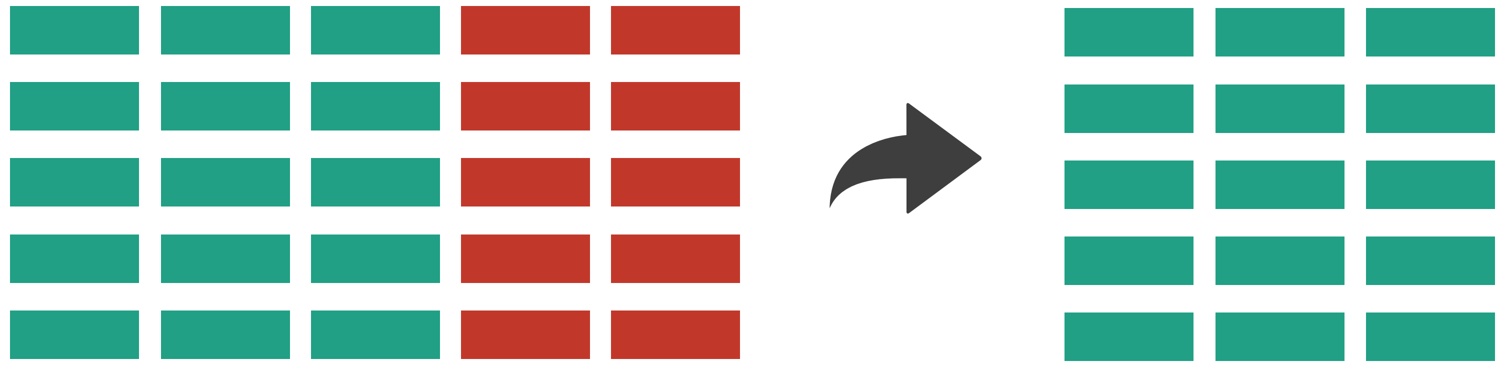
Overview: Reducing Items & Attributes

Filter

- Items

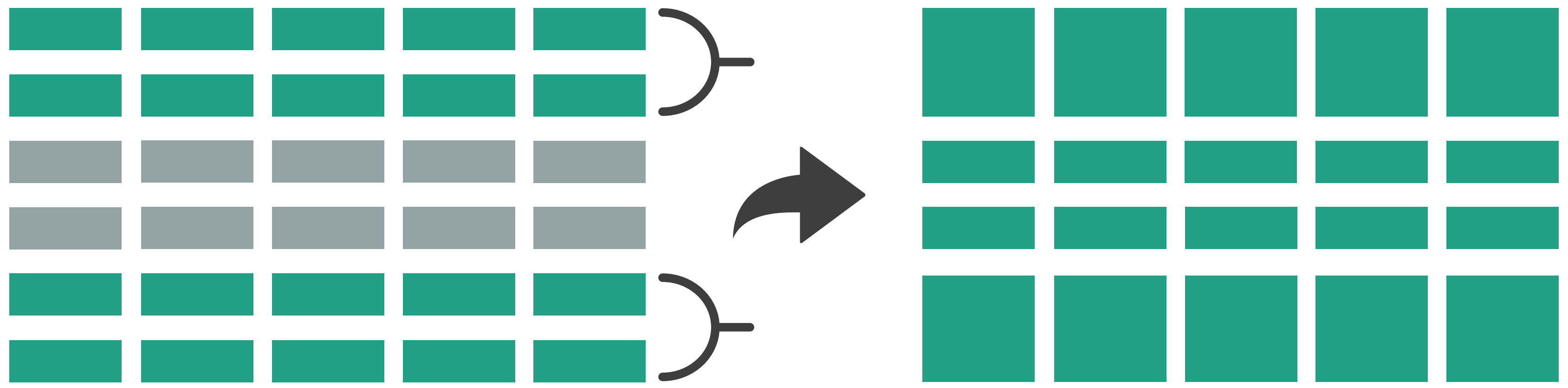


- Attributes

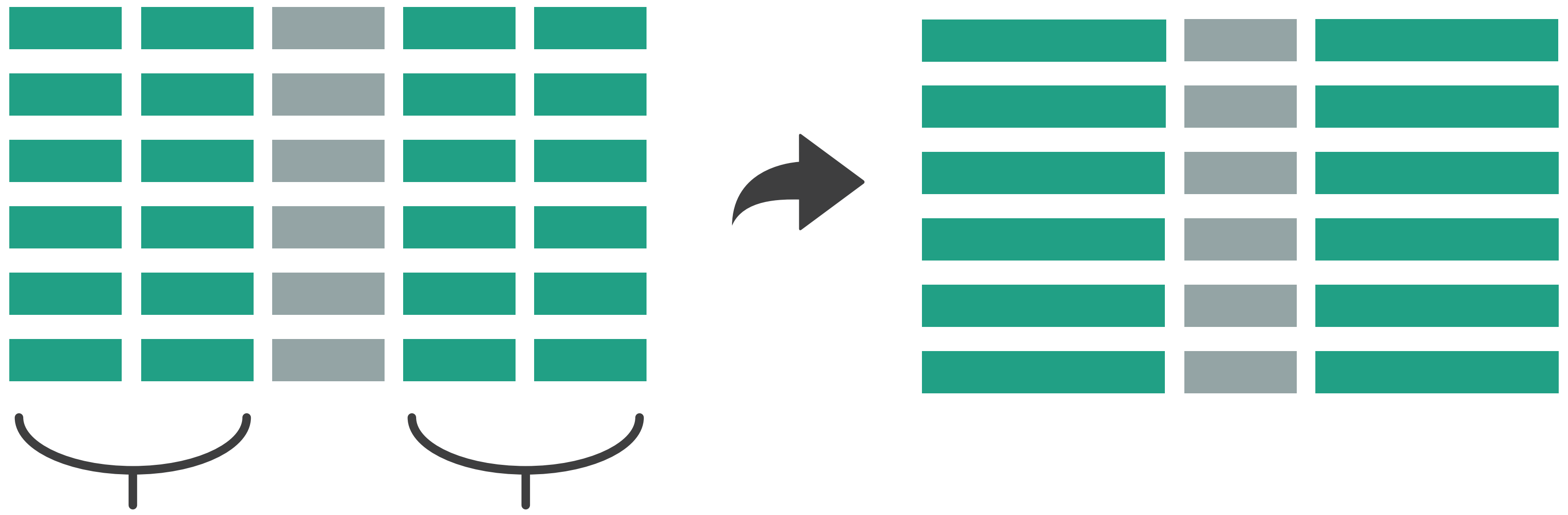


Aggregate

- Items



- Attributes



[Munzner (ill. Maguire), 2014]

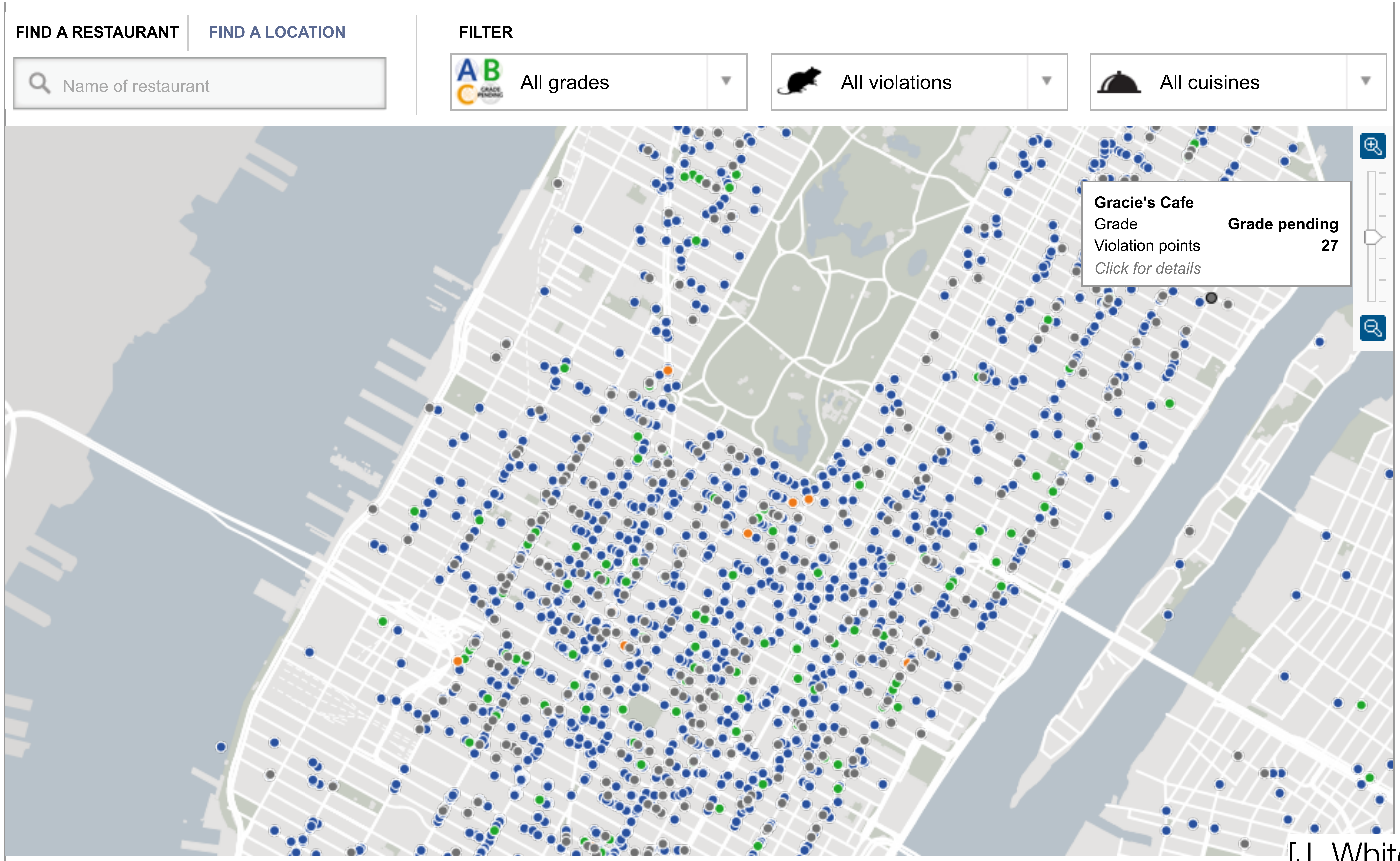
Filtering

- Just don't show certain elements
- Item filtering: most common, eliminate marks for filtered items
- Attribute filtering:
  - attributes often mapped to different channels
  - if mapped to same channel, allows many attributes (e.g. parallel coordinates, star plots), can filter
- How to specify which elements?
  - Pre-defined rules
  - User selection

Filter vs. Query

- Queries start with an empty set of items and **add** items
- Filters start with all items and **remove** items

Example: NYC Health Dept. Restaurant Ratings

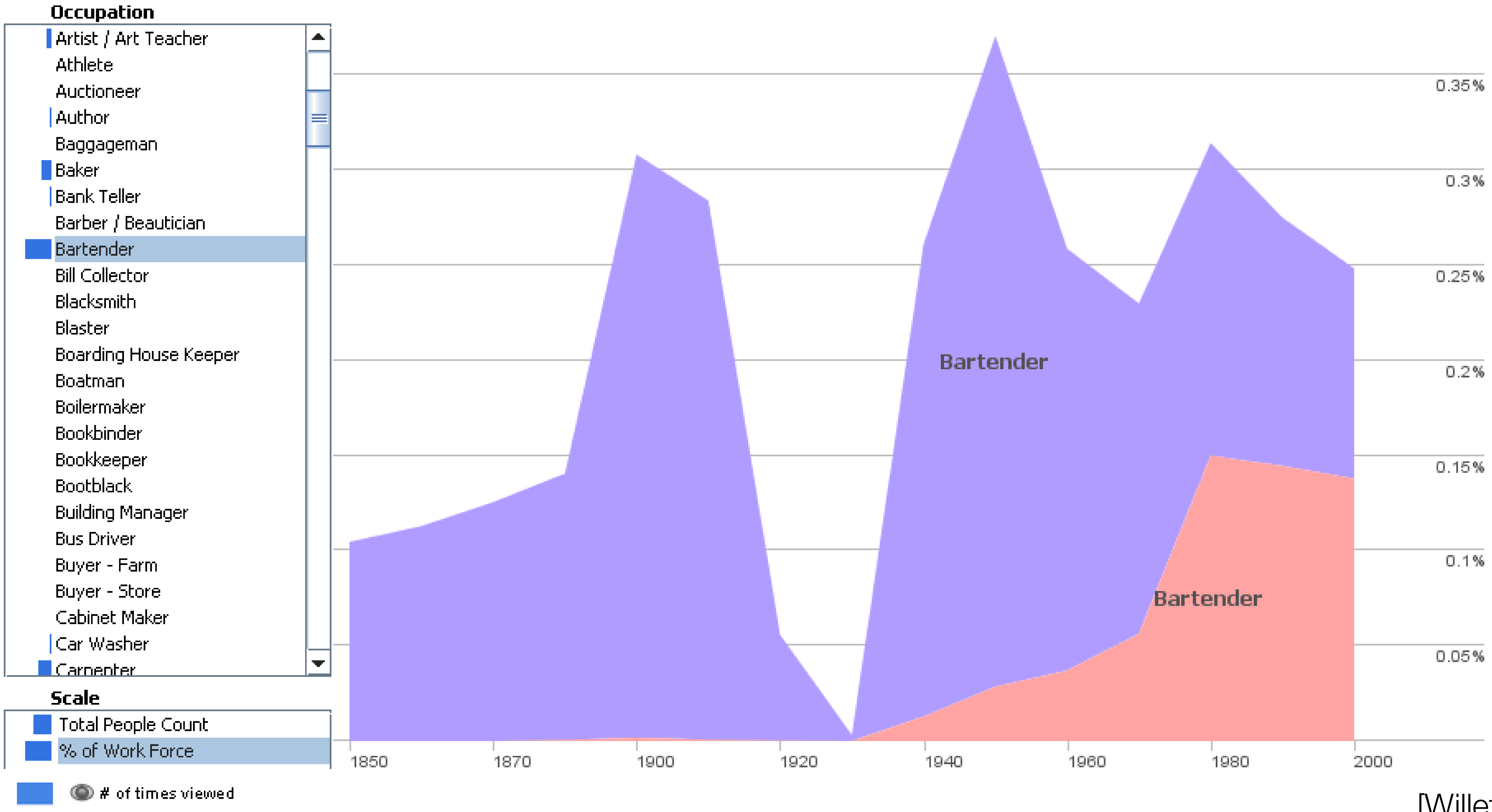


[J. White, New York Times]

Dynamic Filters

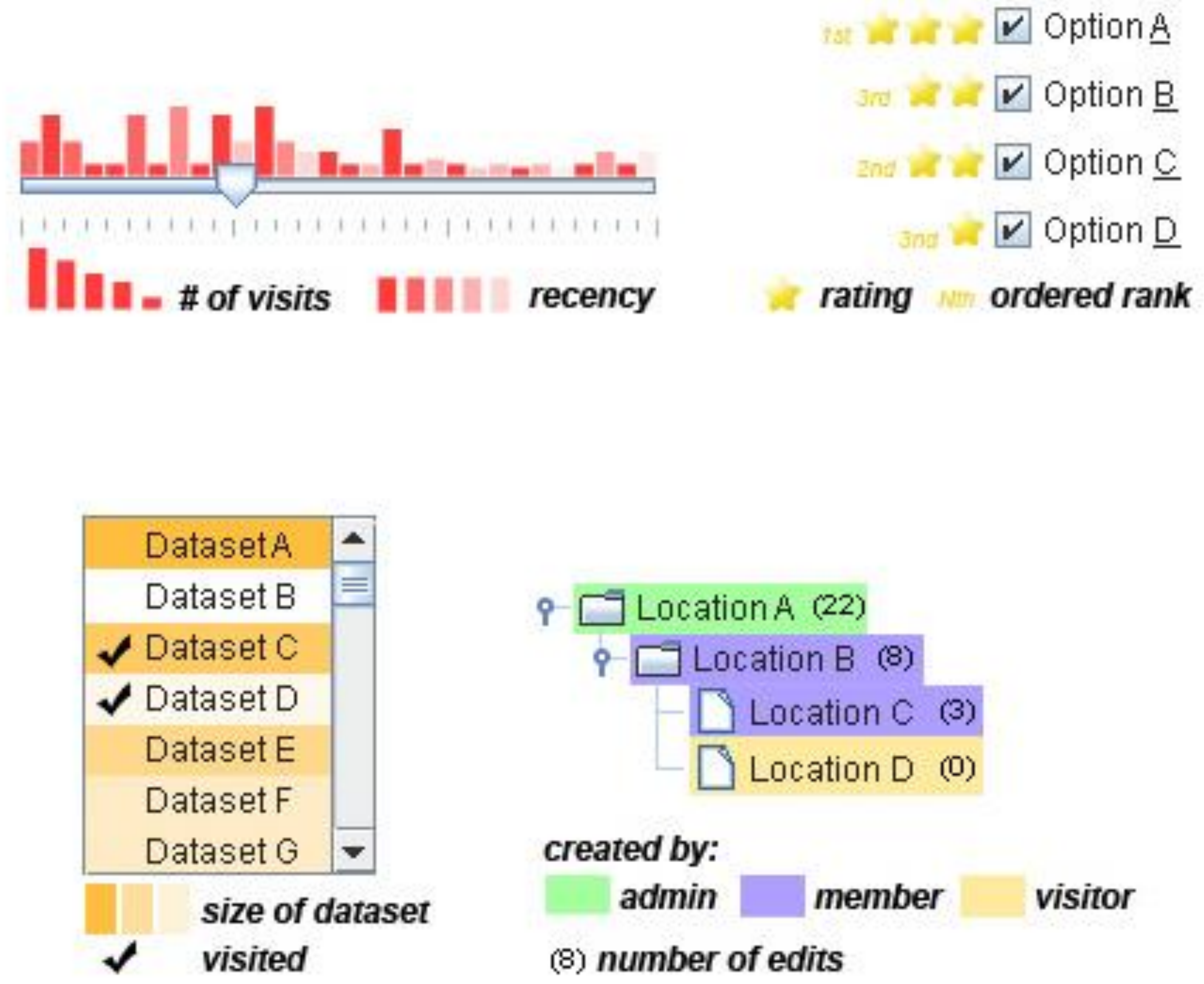
- Interaction need not be with the visualization itself
- Users interact with **widgets** that control which items are shown
  - Sliders, Combo boxes, Text Fields
- Often tied to attribute values
- Examples:
  - All restaurants with an "A" Grade
  - All pizza places
  - All pizza places with an "A" Grade

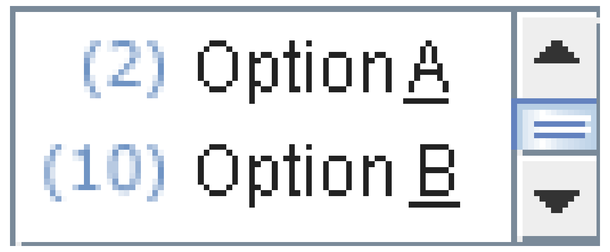
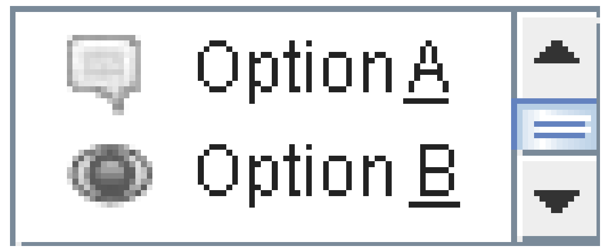
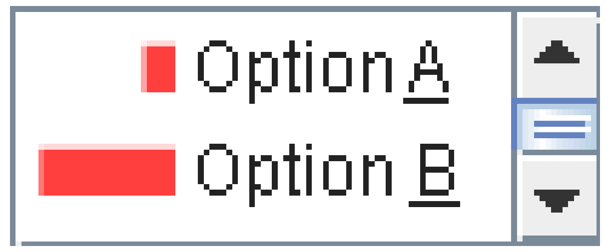
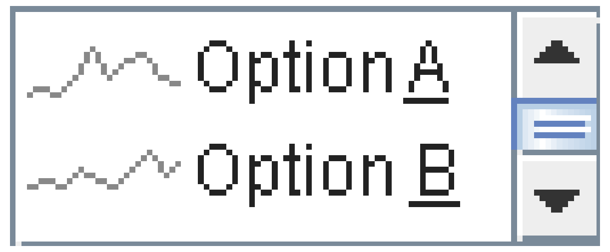
Scented Widgets



[Willett et al., 2007]

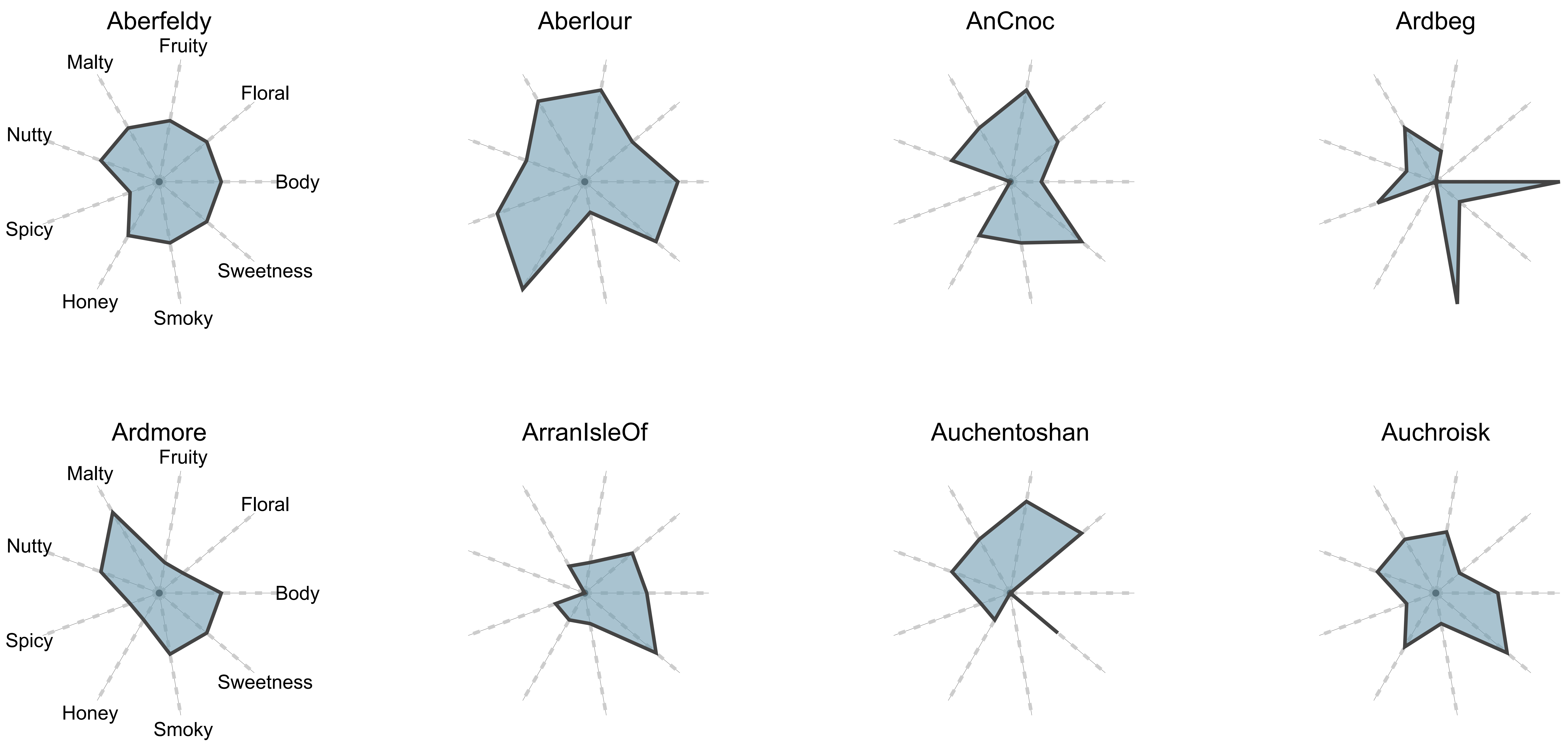
Scented Widgets



Name	Description	Example
Hue	Varies the hue of the widget (or of a visualization embedded in it)	
Saturation	Varies the saturation of the widget (or of a visualization embedded in it)	
Opacity	Varies the saturation of the widget (or of a visualization embedded in it)	
Text	Inserts one or more small text figures into the widget	
Icon	Inserts one or more small icons into the widget.	
Bar Chart	Inserts one or more small bar chart visualizations into the widget	
Line Chart	Inserts one or more small line chart visualizations into the widget	

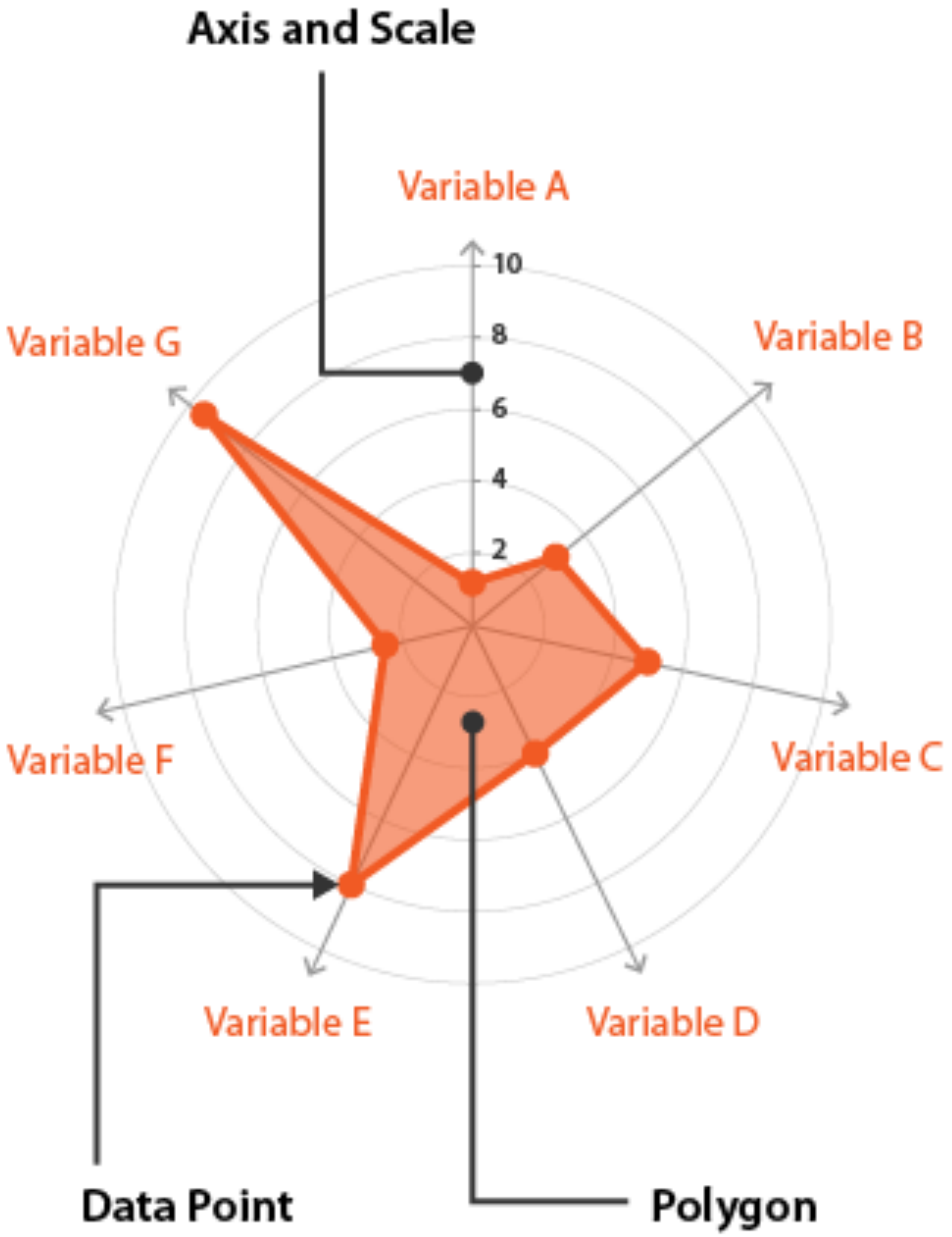
[Willett et al., 2007]

Star Plots (aka Radar Charts)



[K. Schaul]

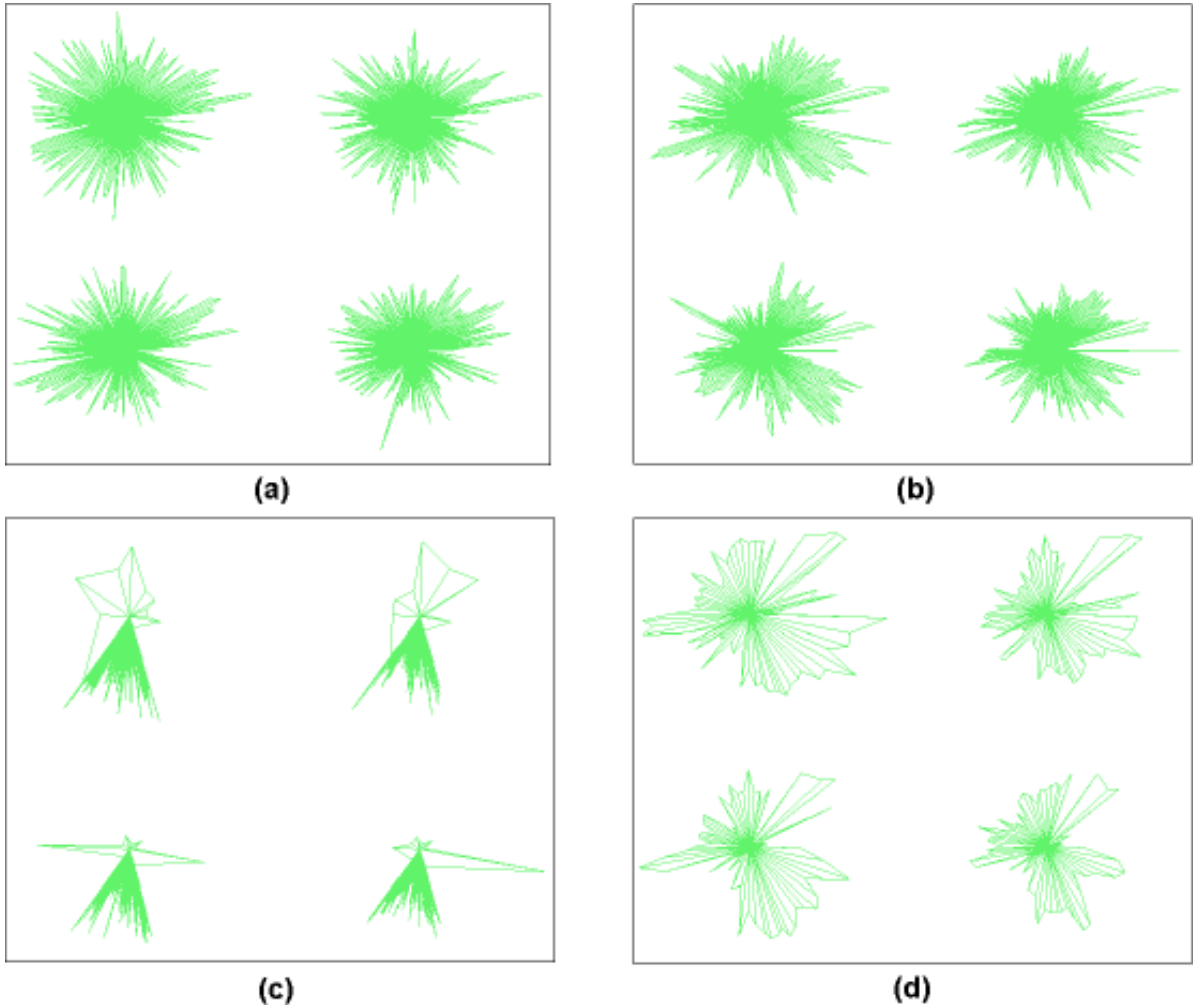
Star Plot / Radar Chart



- Use:
  - Compare variables
  - Similarities/differences of items
  - Locate outliers
- Considerations:
  - Order of axes
  - Too many axes cause problems

[S. Ribecca]

Attribute Filtering on Star Plots



[Yang et al., 2003]

Attribute Filtering

- How to choose which attributes should be filtered?
  - User selection?
  - Statistics: similarity measures, attributes with low variance are not as interesting when comparing items

- Can be combined with item filtering

Aggregation

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

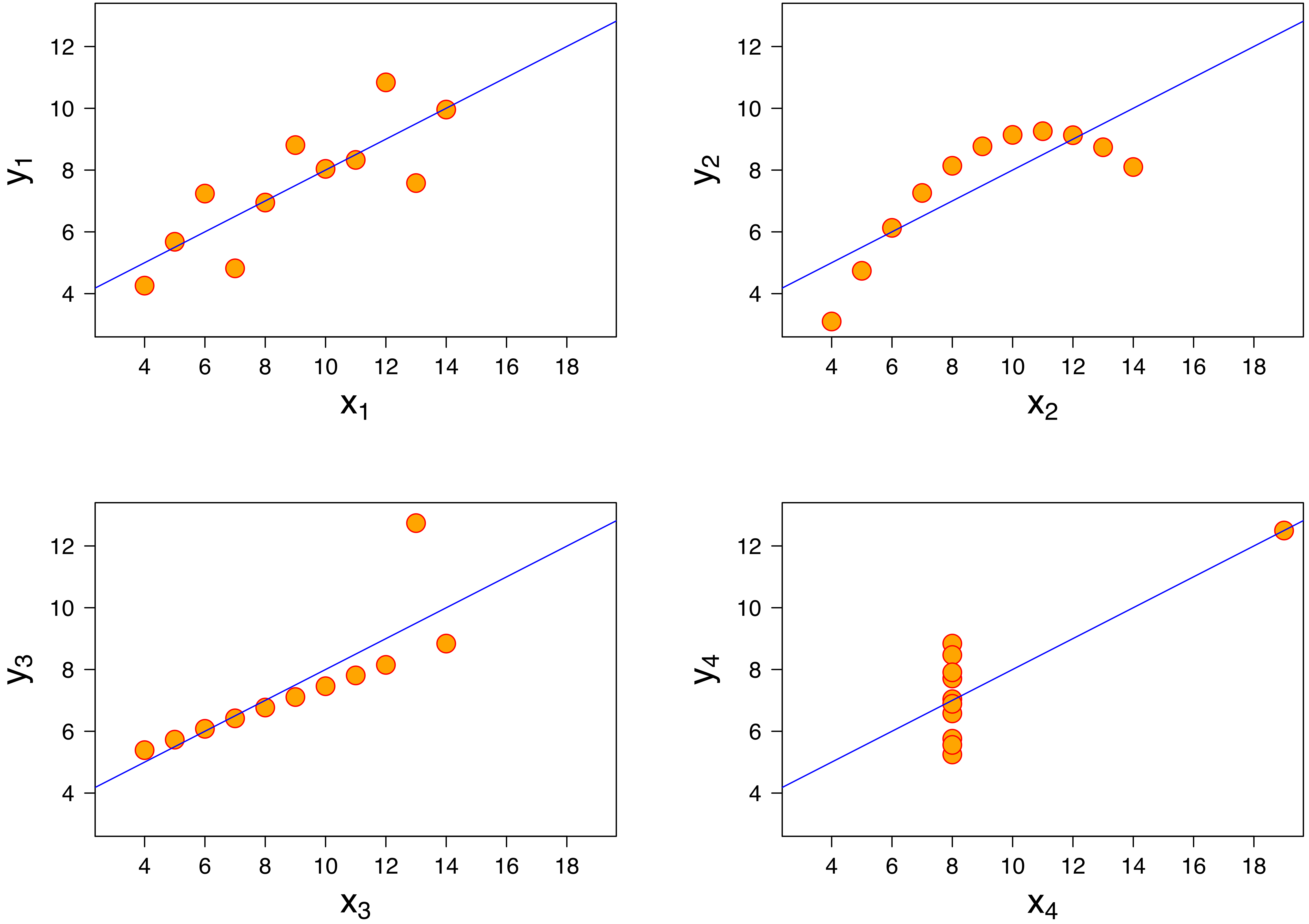
Aggregation

- 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

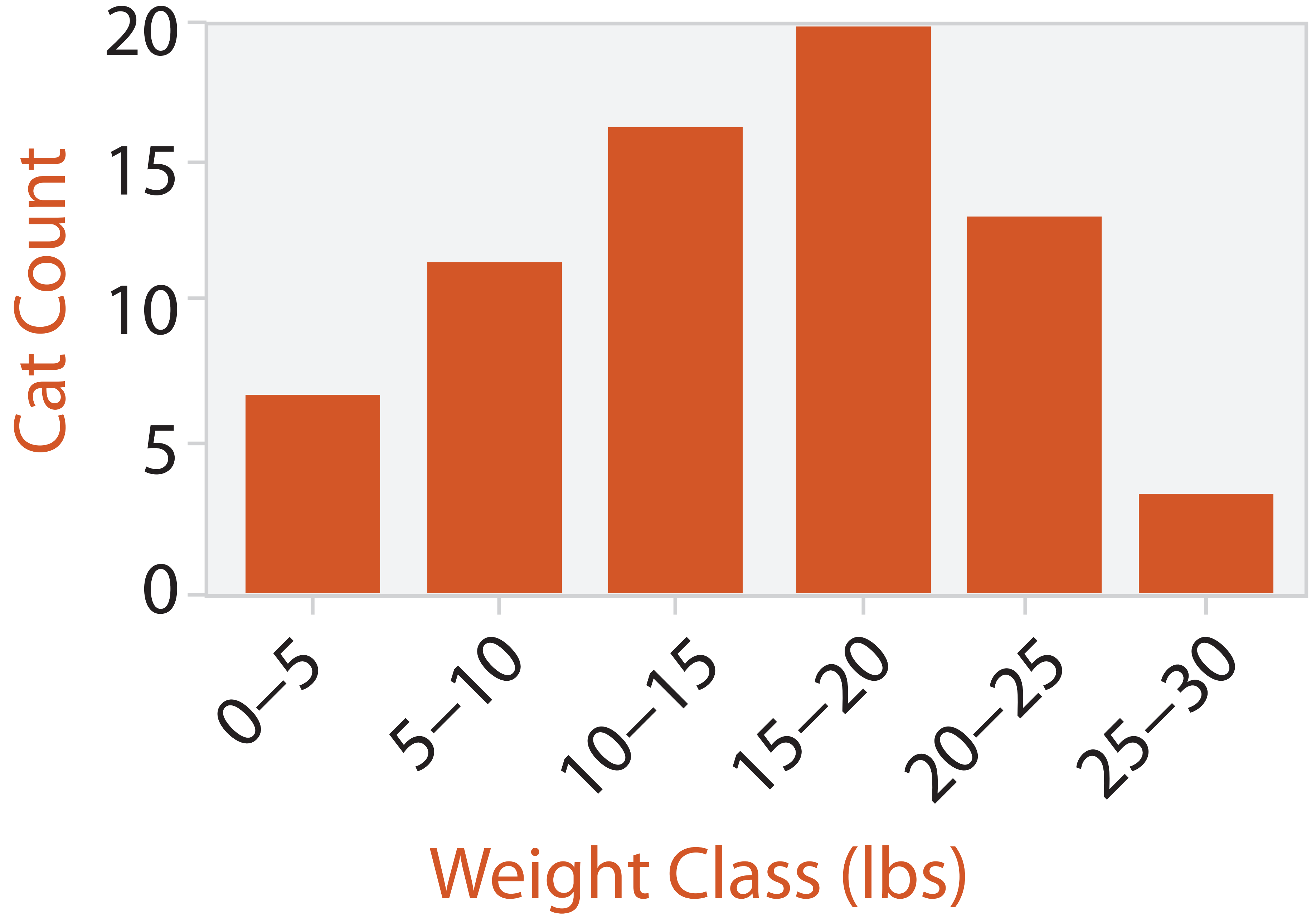
Mean of x	9
Variance of x	11
Mean of y	7.50
Variance of y	4.122
Correlation	0.816

Anscombe's Quartet



[F. J. Anscombe]

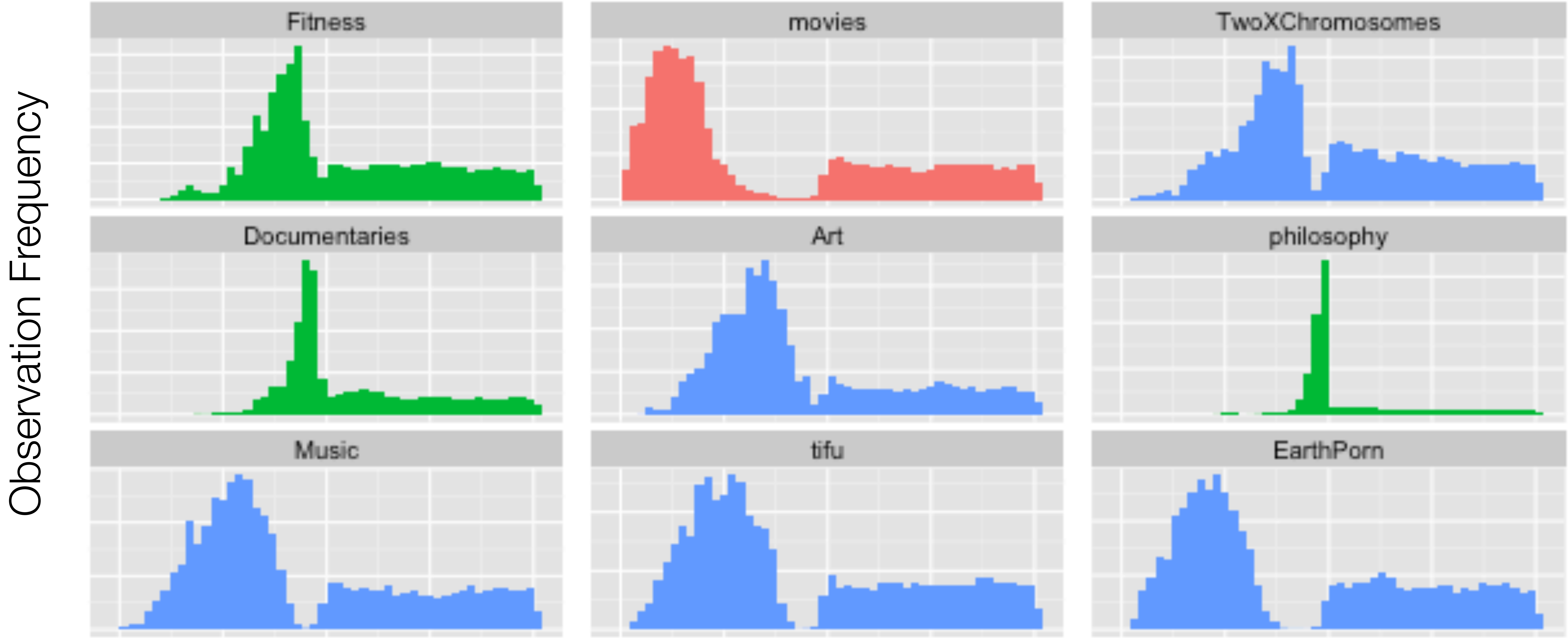
Aggregation: Histograms



- 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]

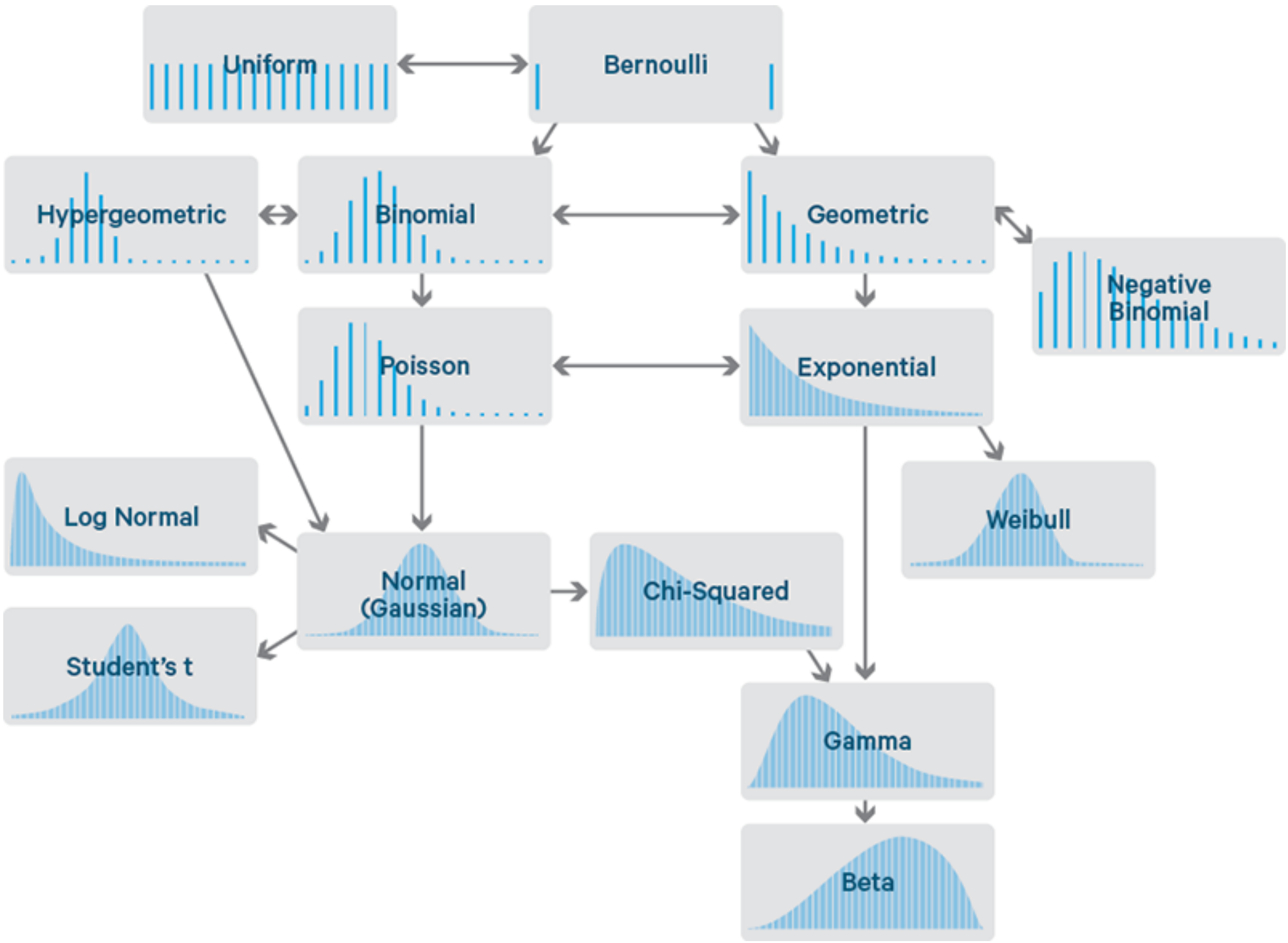
Aggregation: Histograms



Observed ranks of posts by subreddit

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

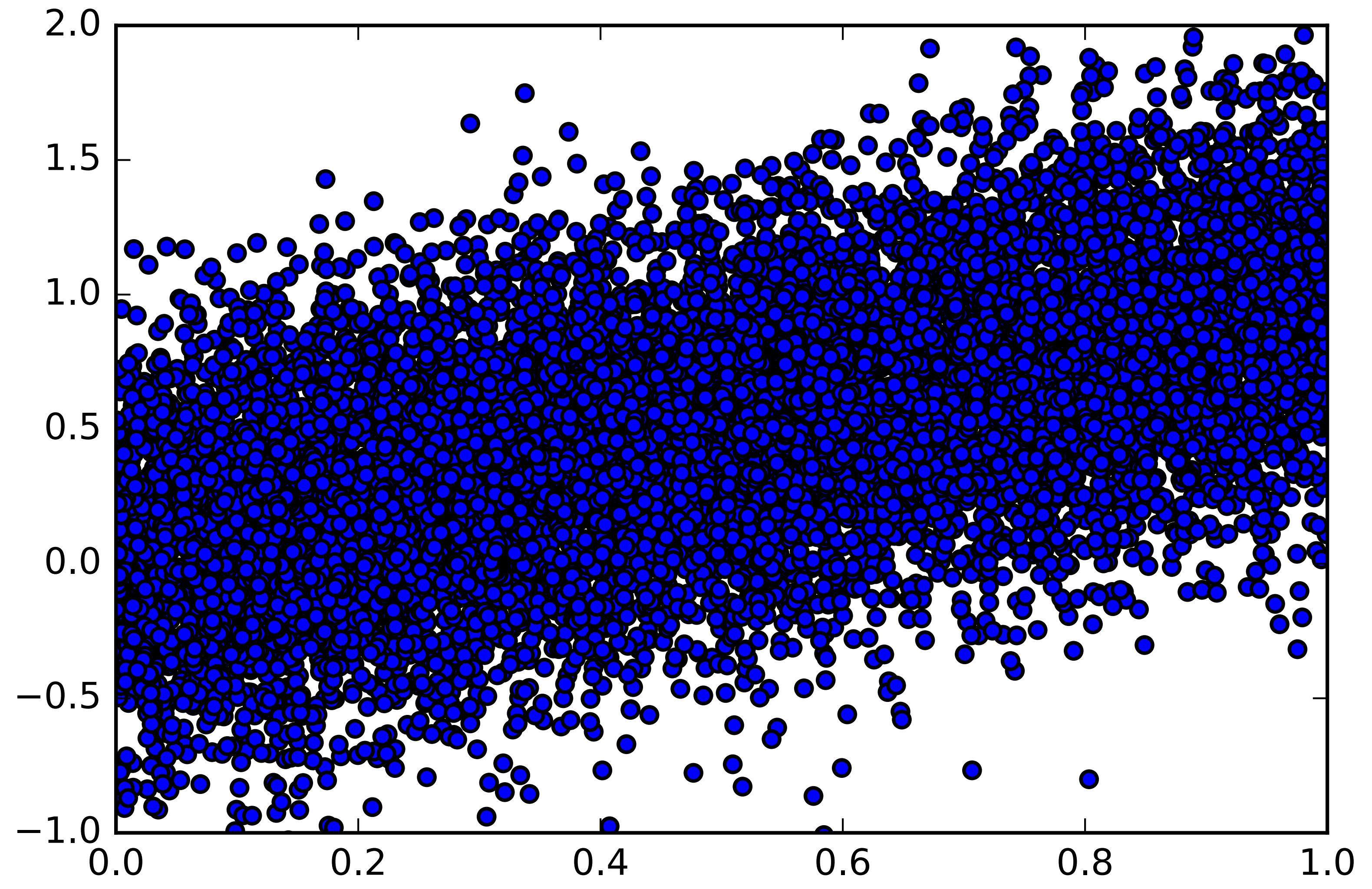
Common Distributions

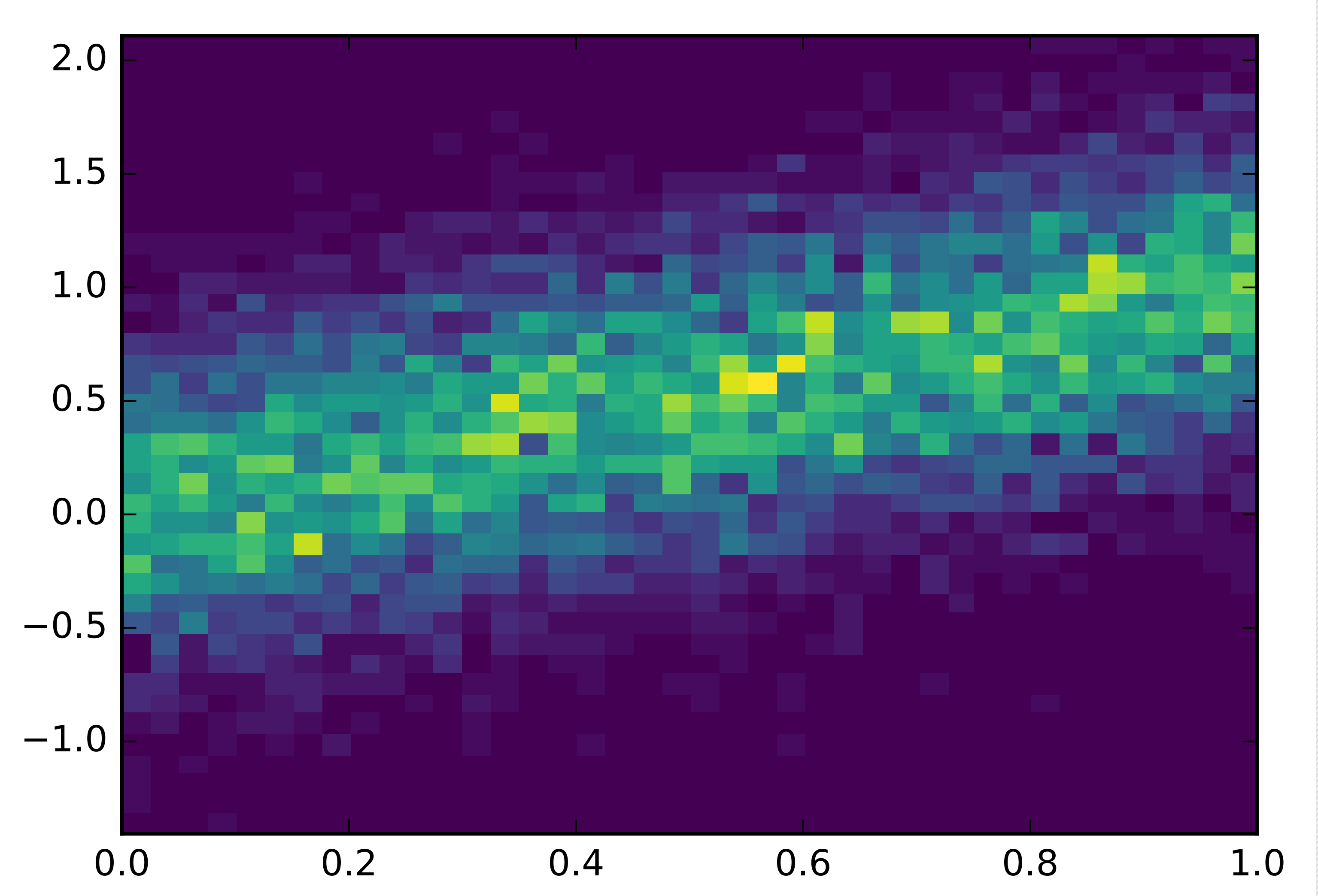


[Cloudera]

Binning Scatterplots

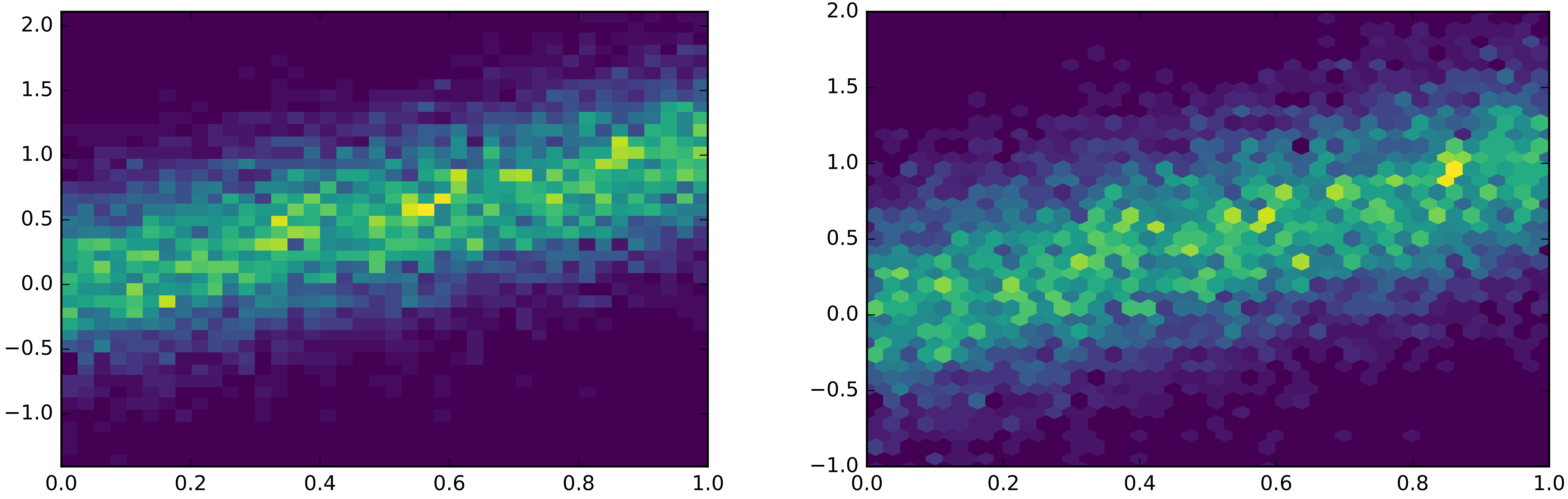
- At some point, cannot see density
- Blobs on top of blobs
- 2D Histogram is a histogram in 2D encoded using color instead of height
- Each region is aggregated



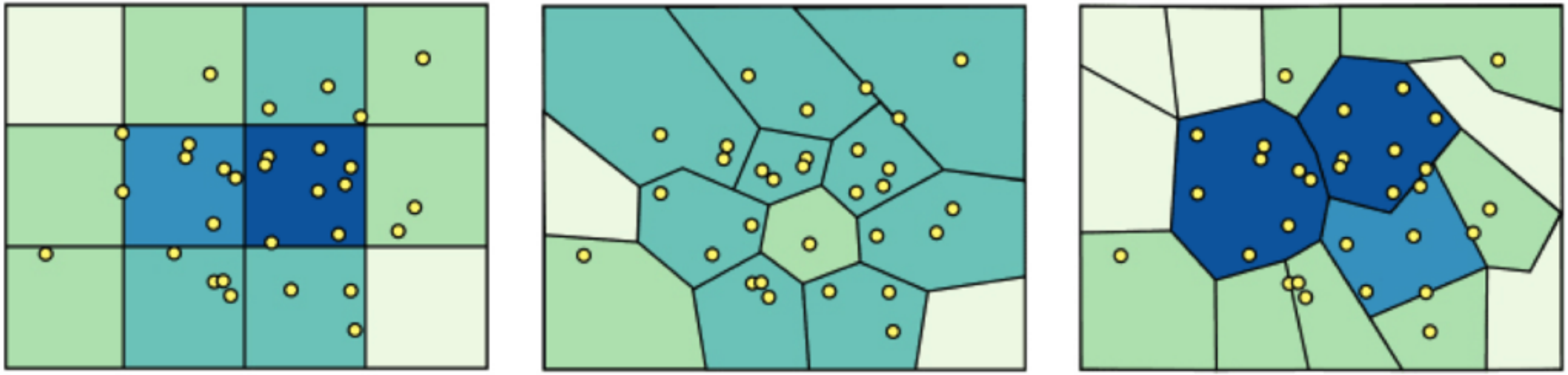


Binning

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



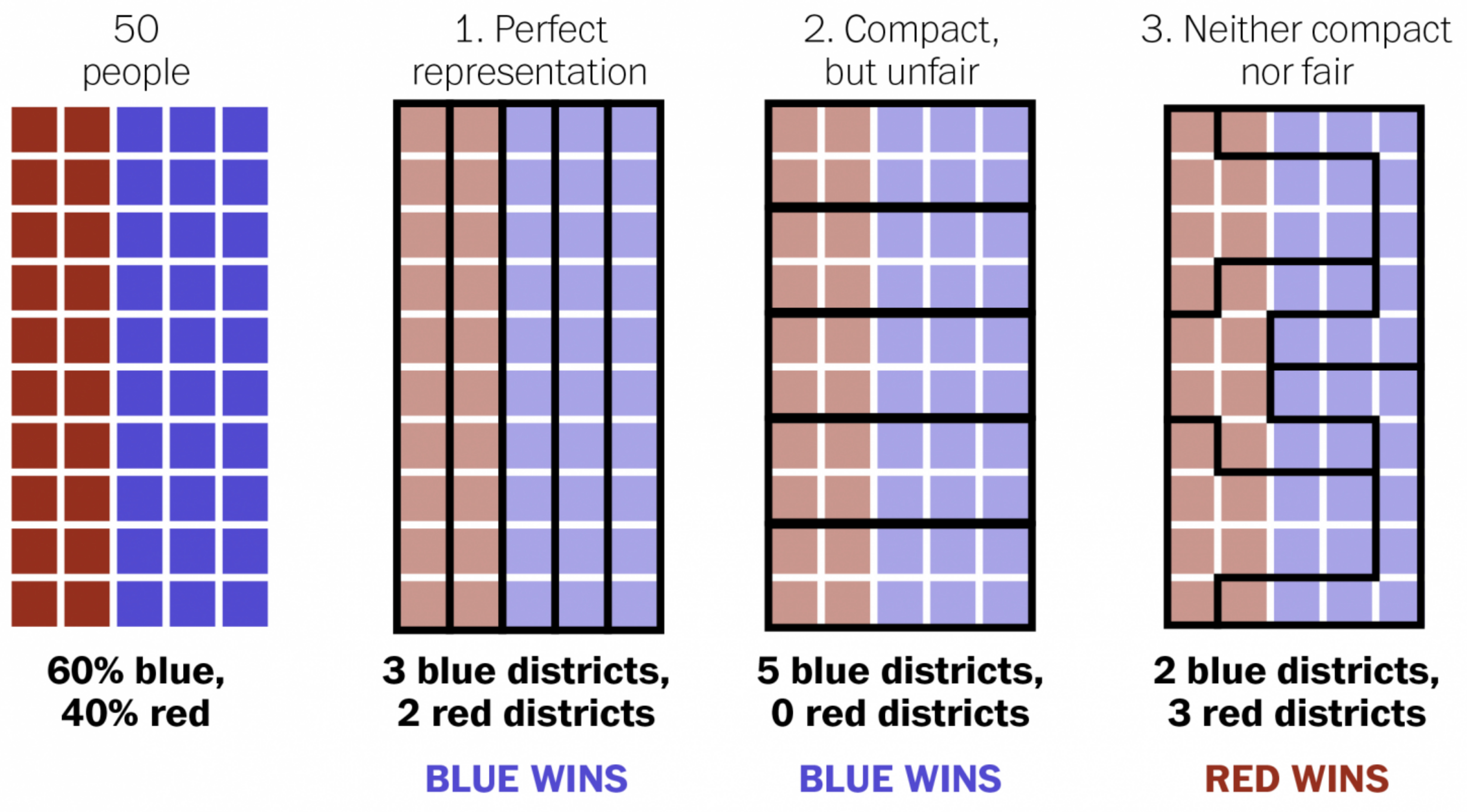
Spatial Aggregation

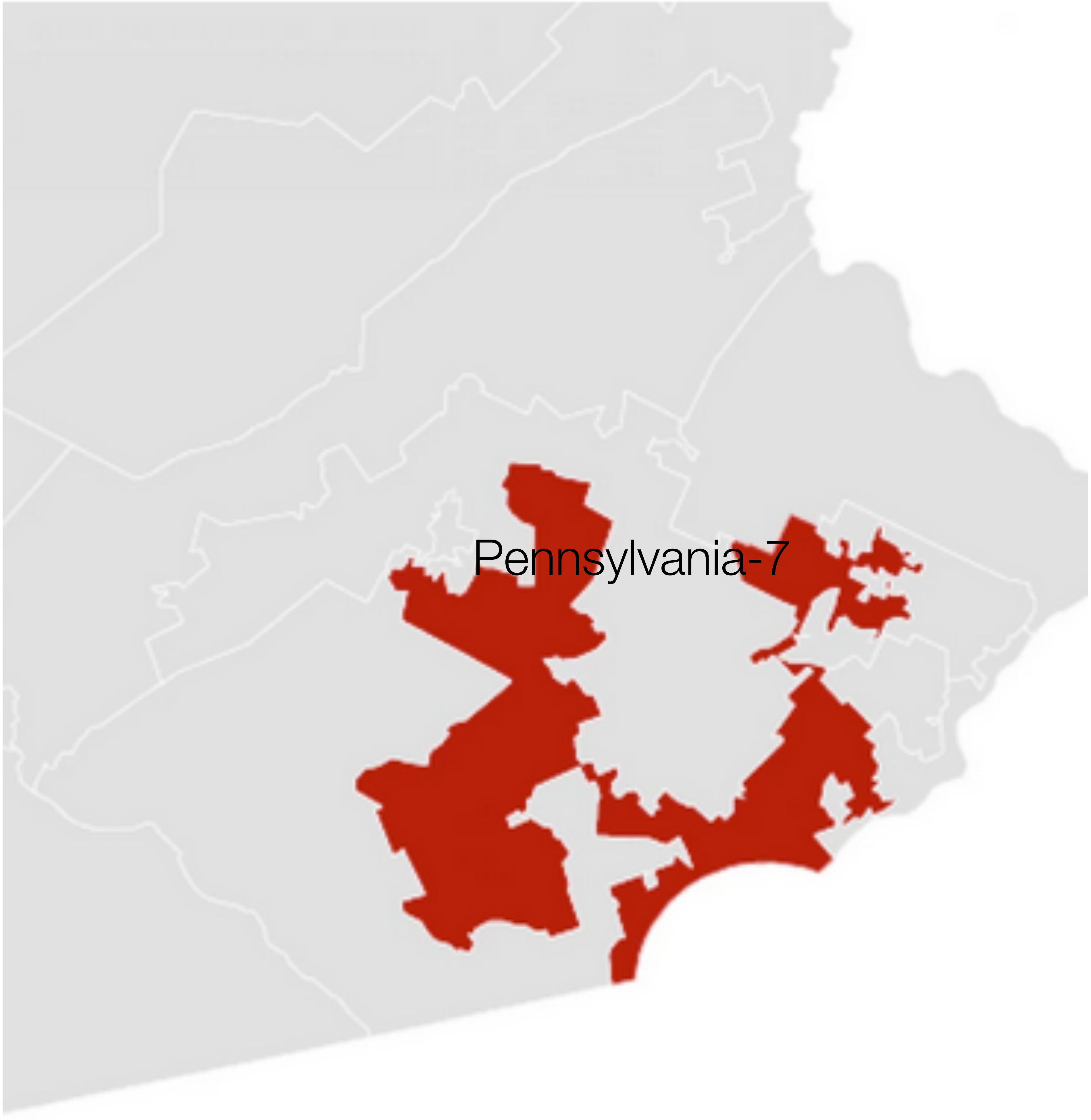


[Penn State, GEOG 486]

Modifiable Areal Unit Problem

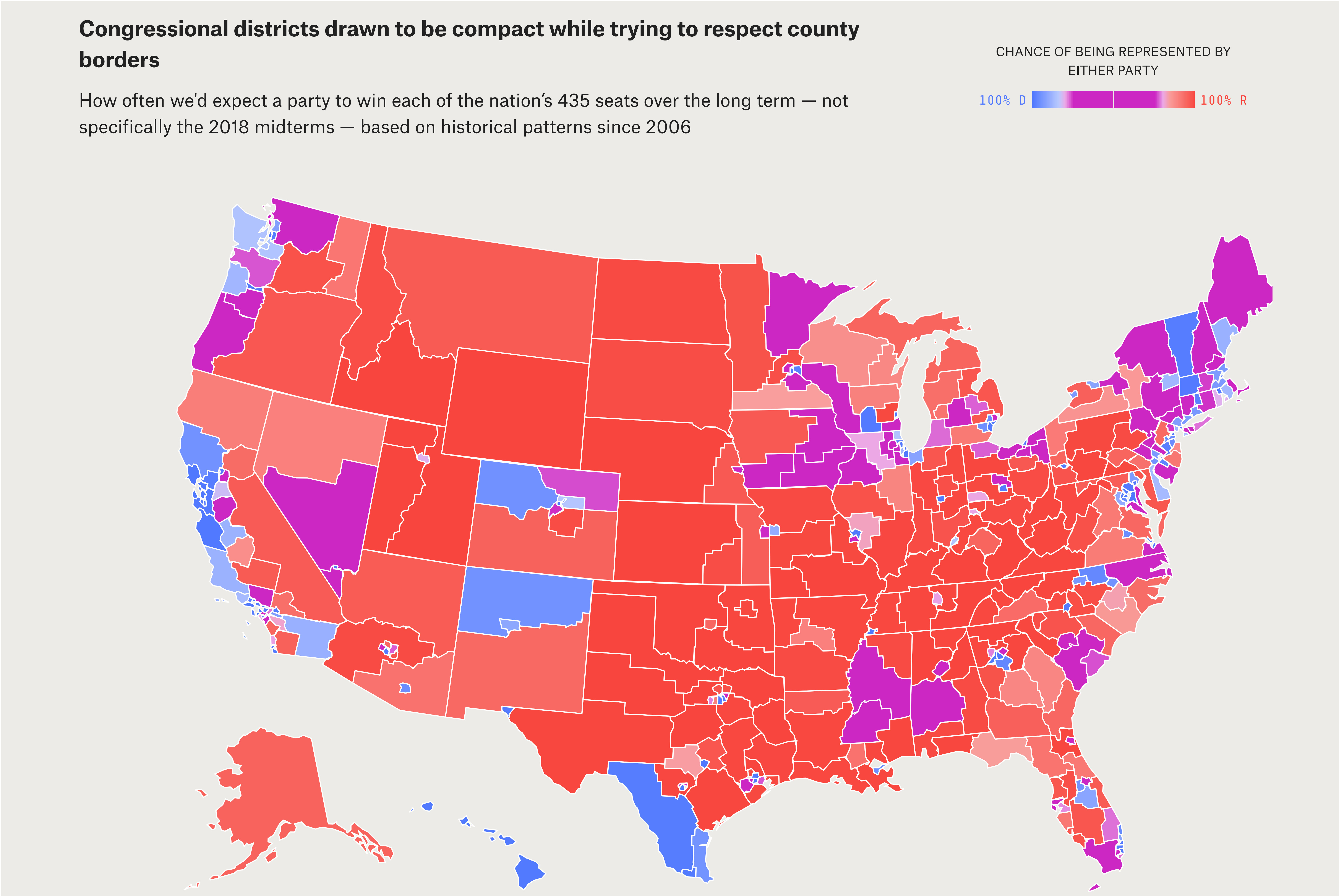
- How you draw boundaries impacts the type of aggregation you get
- Similar to bins in histograms
- Gerrymandering





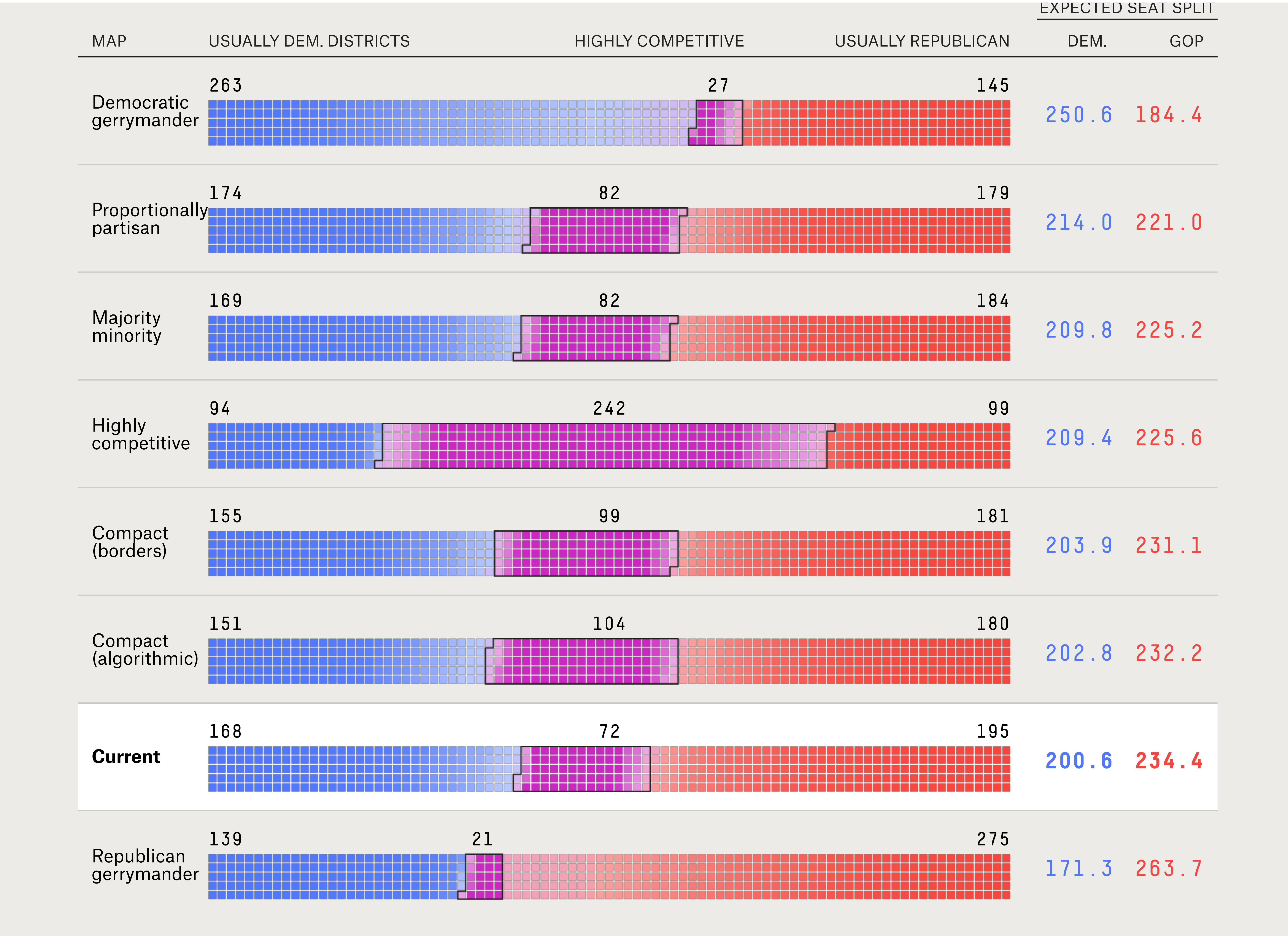
[Wonkblog, Washington Post, Adapted from S. Nass]

Drawing Different Maps: Compactness



[A. Bycoffe et al., 538]

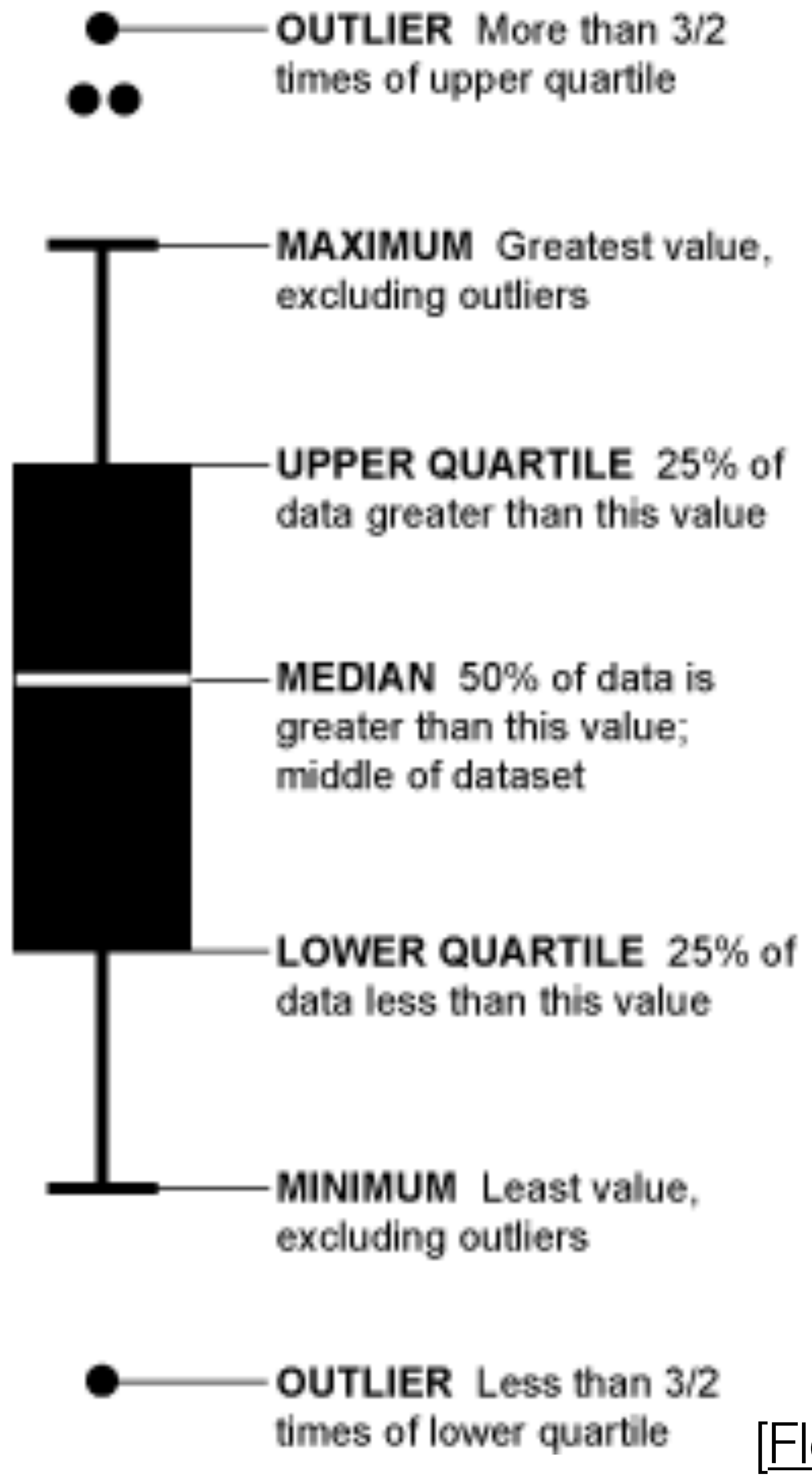
Drawing Different Maps



[A. Bycoffe et al., 538]

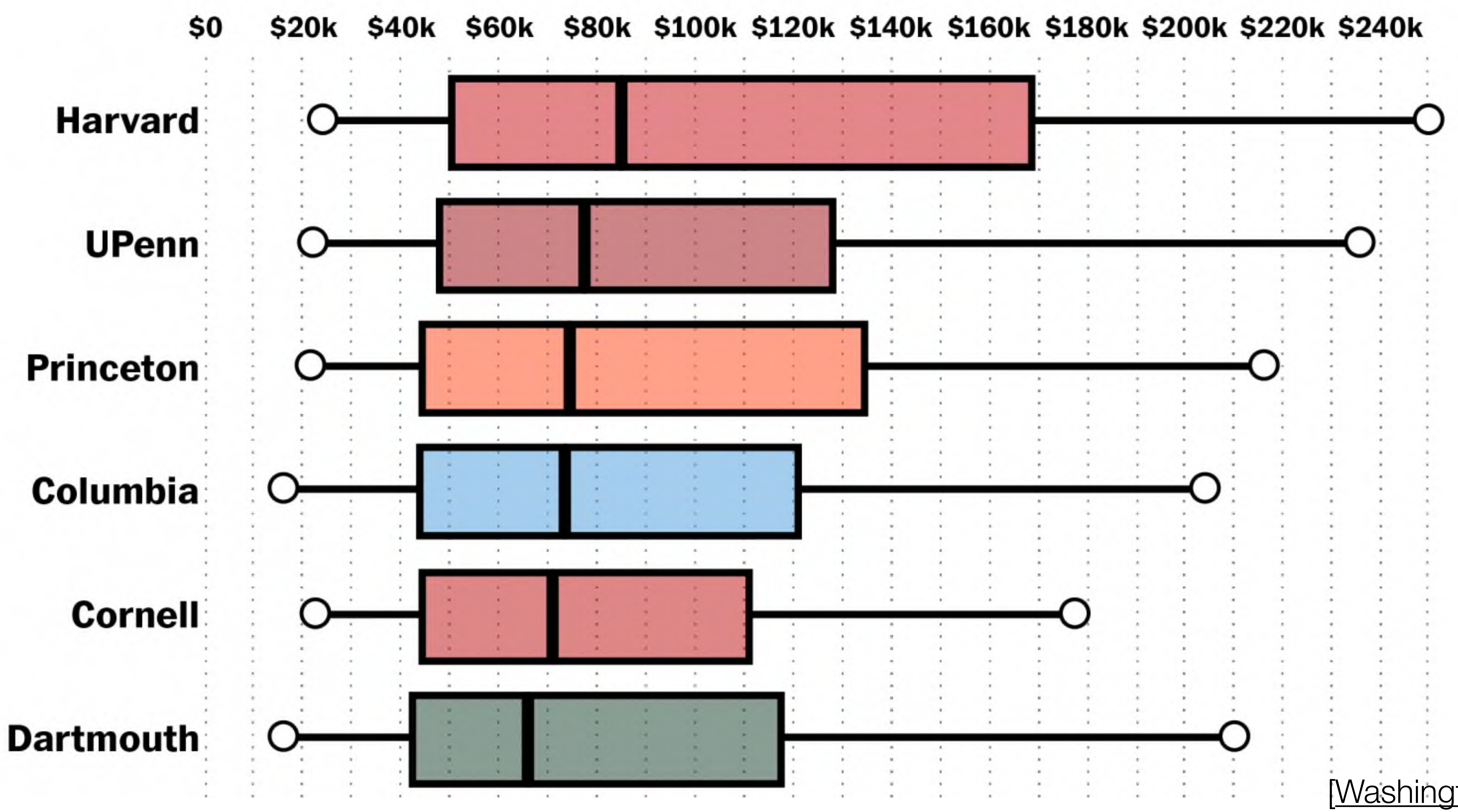
Boxplots

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



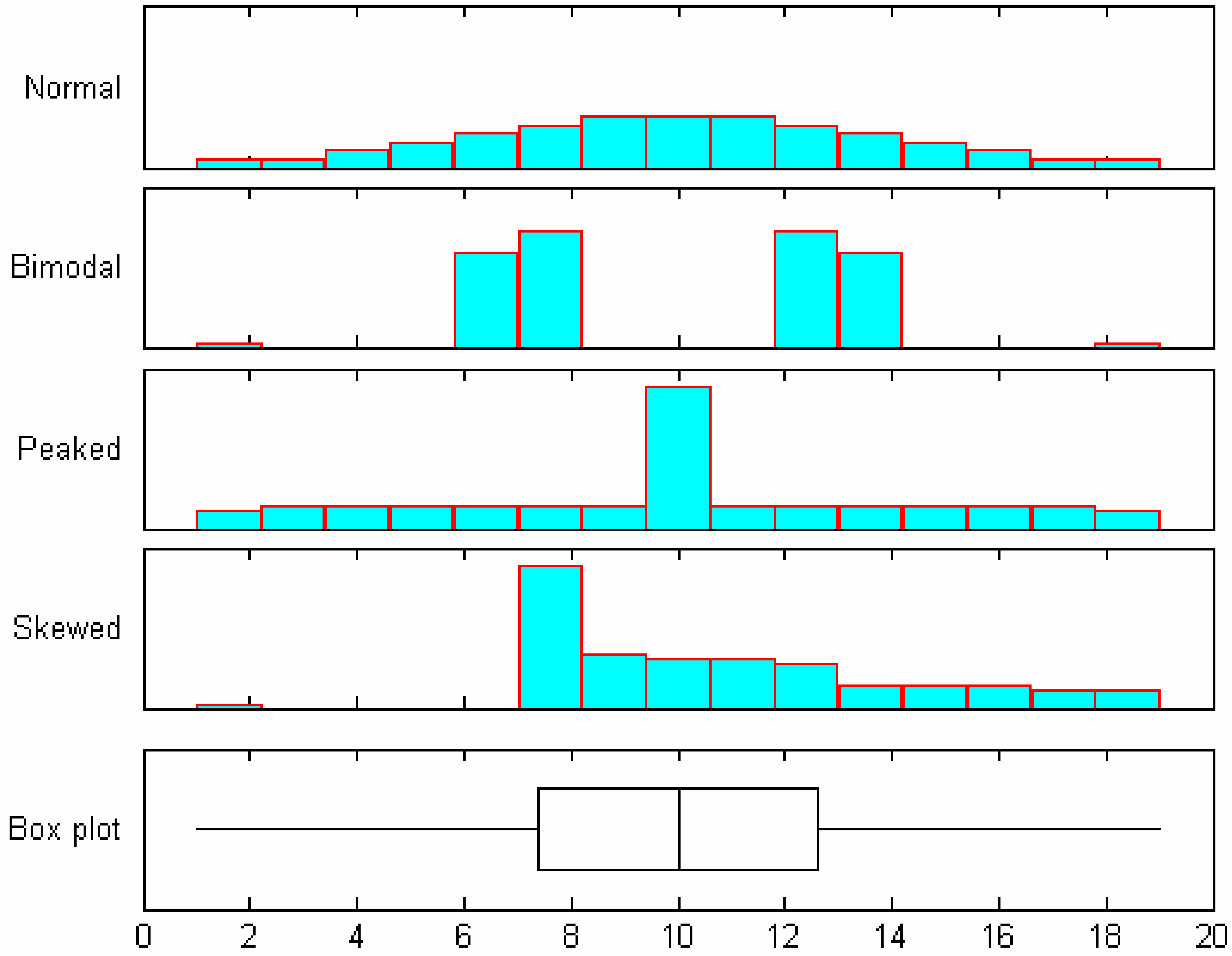
[Flowing Data]

Aggregation: Boxplots



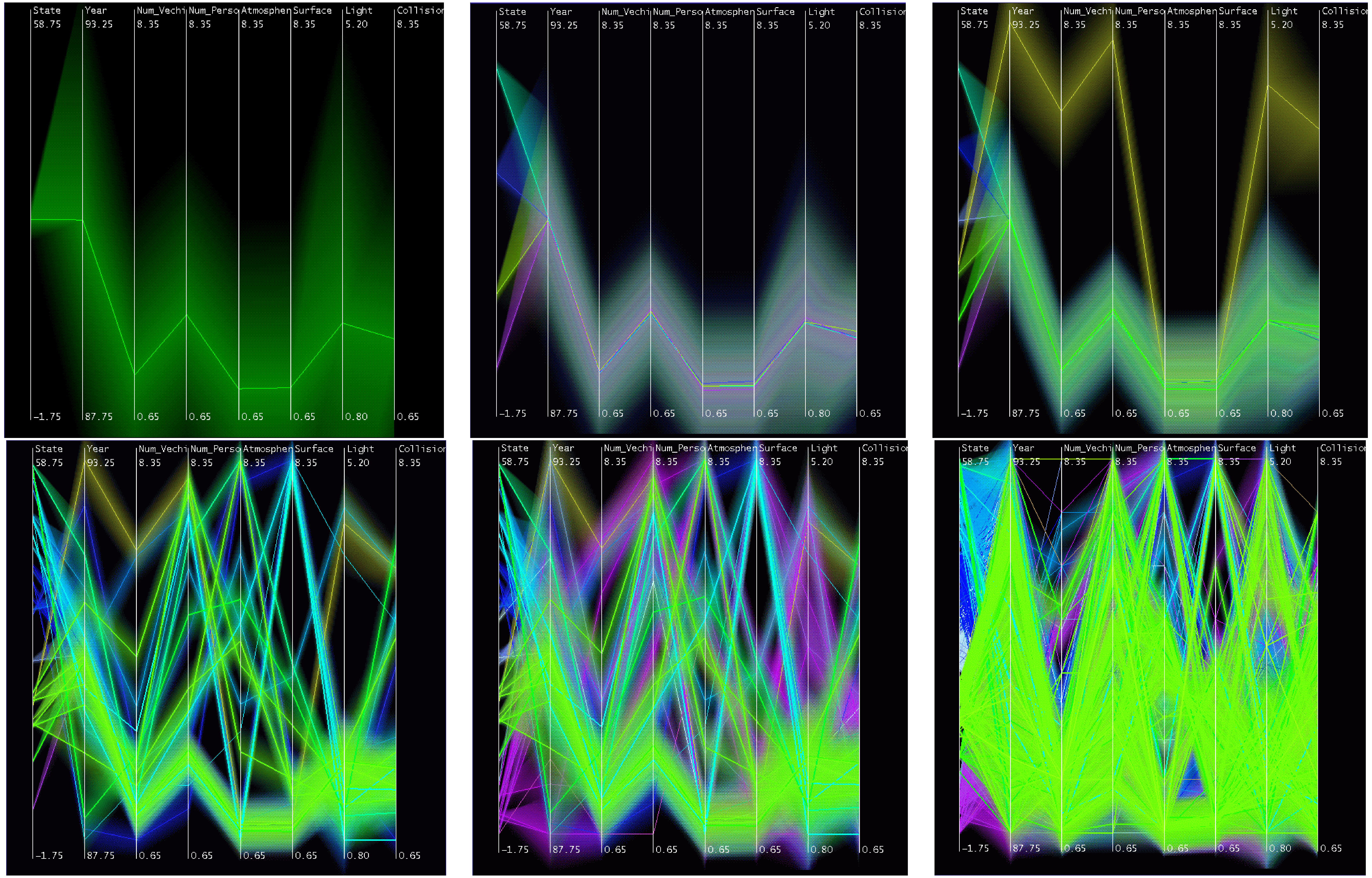
[Washington Post, 2015]

Four Distributions, Same Boxplot…



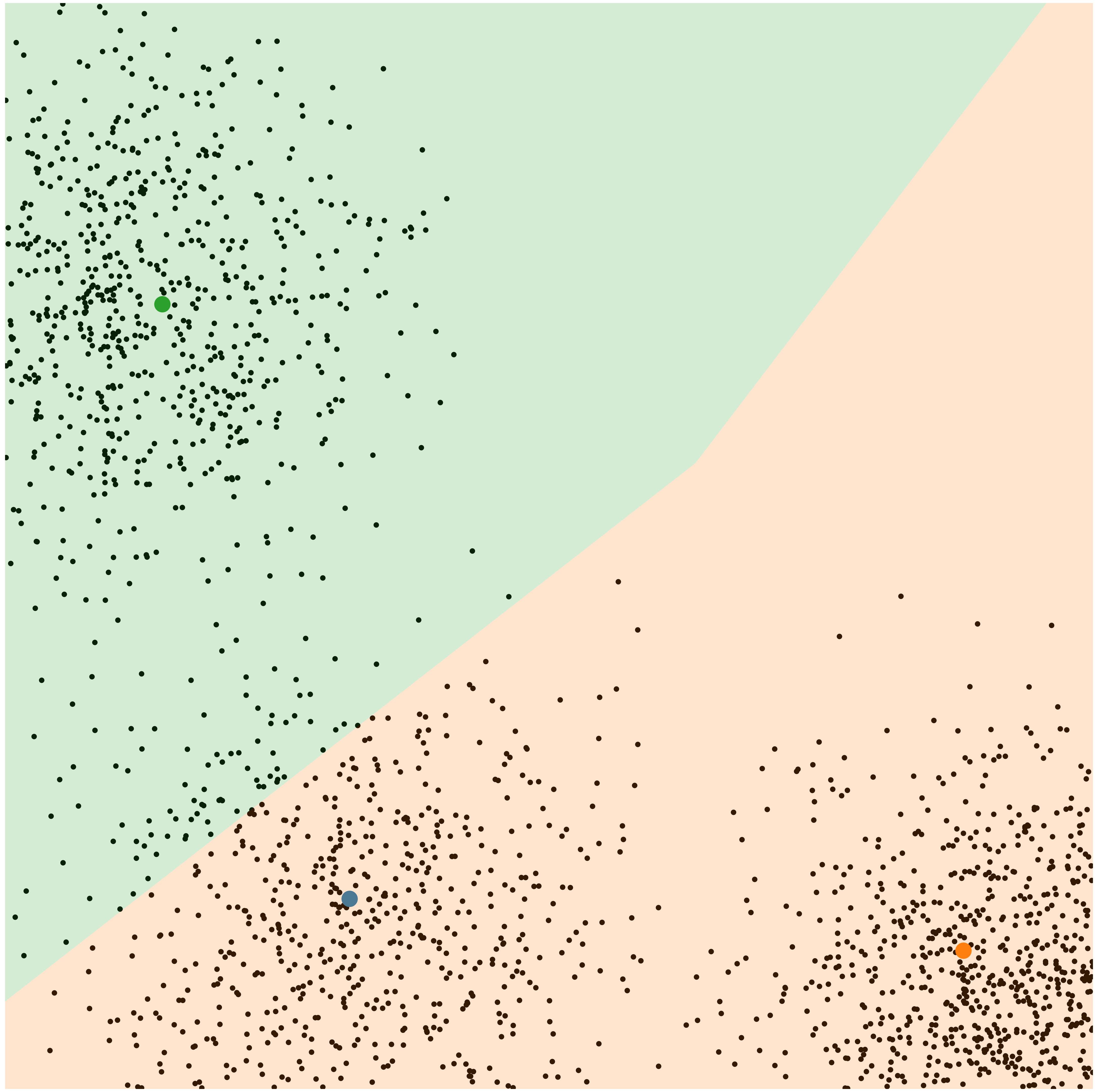
[C. Choonpradub and D. McNeil, 2005]

Hierarchical Parallel Coordinates



[Fua et al., 1999]

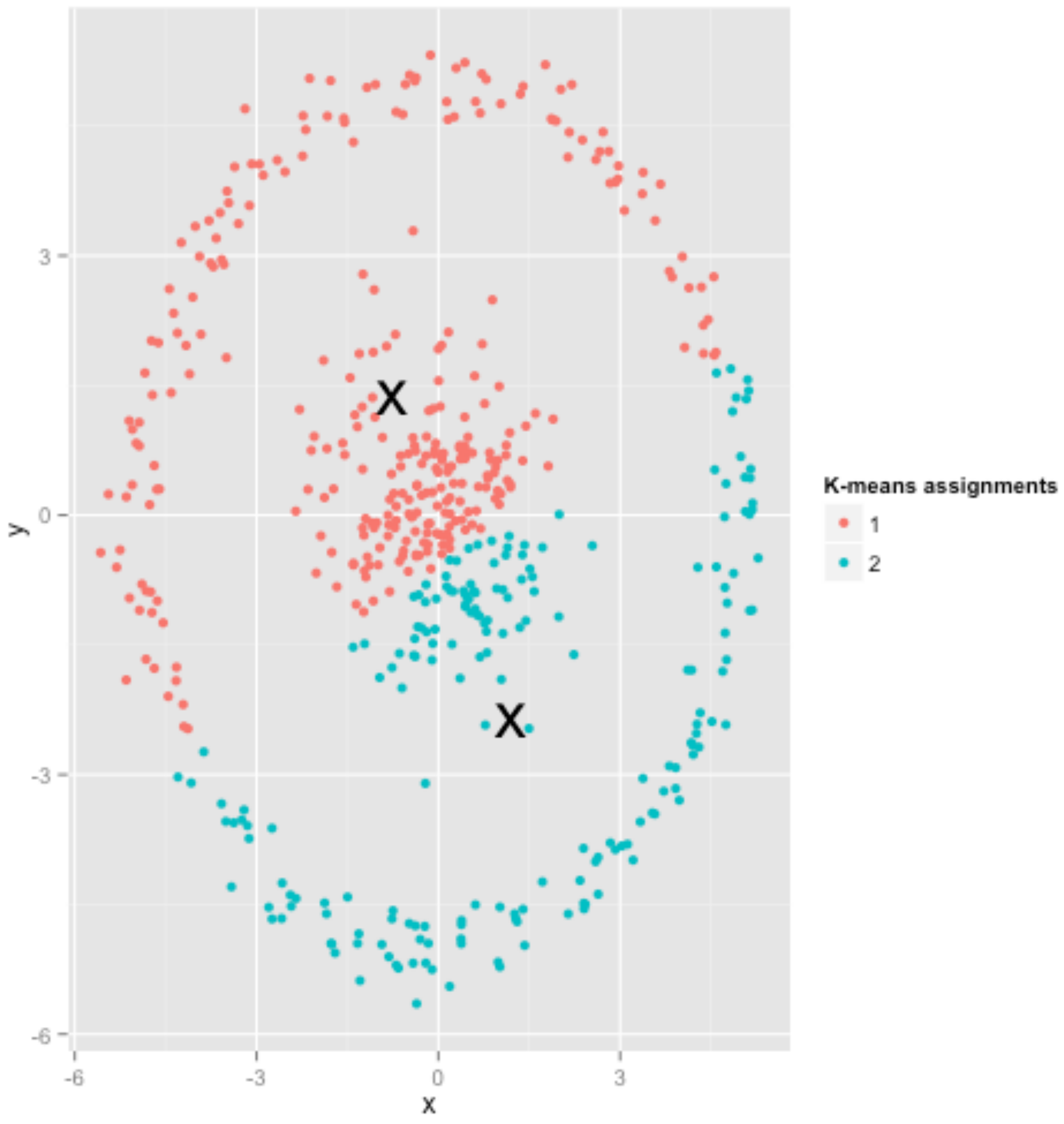
K-Means



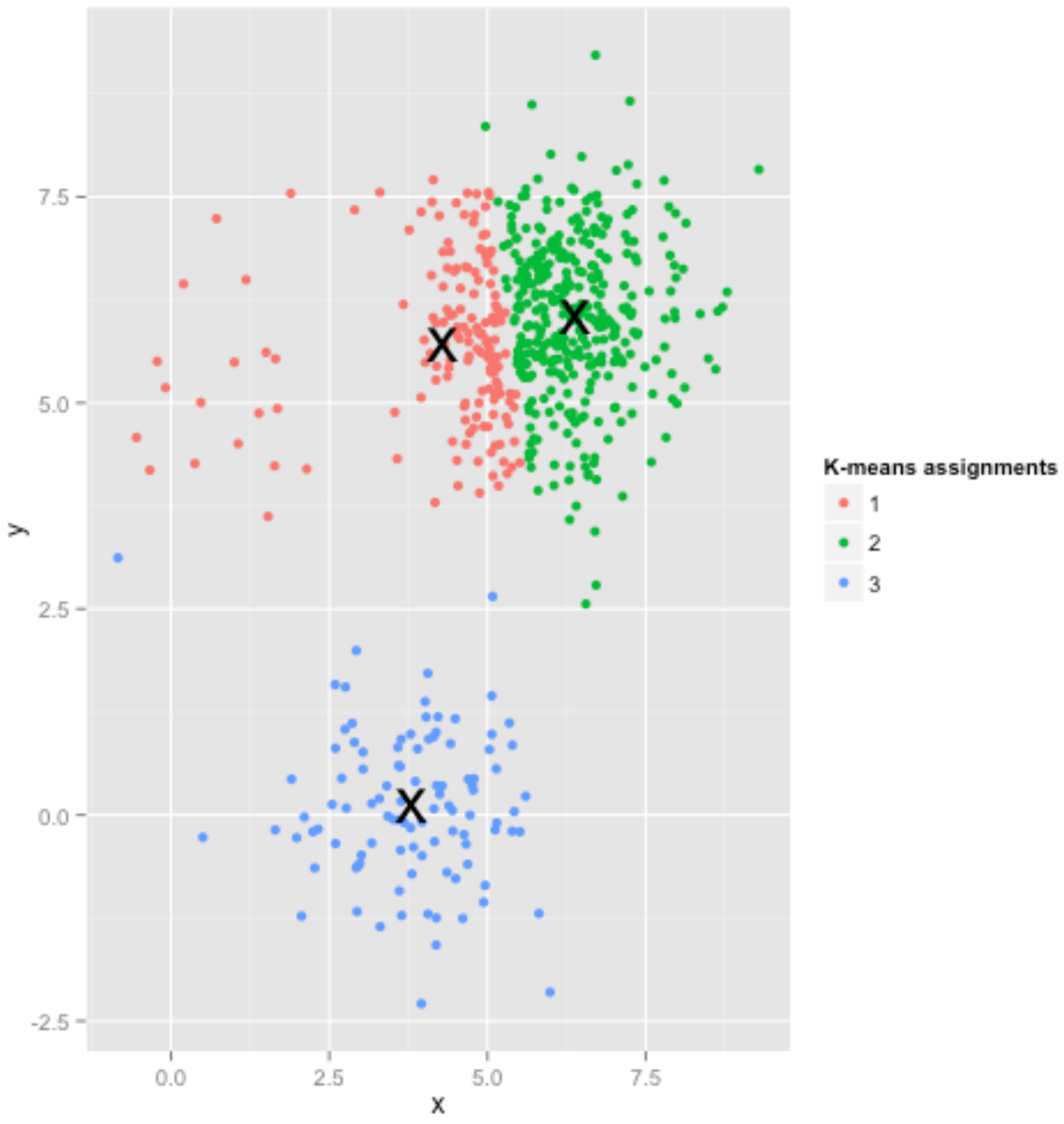
Run

[C. Polis, 2014]

K-Means Issues



Shape



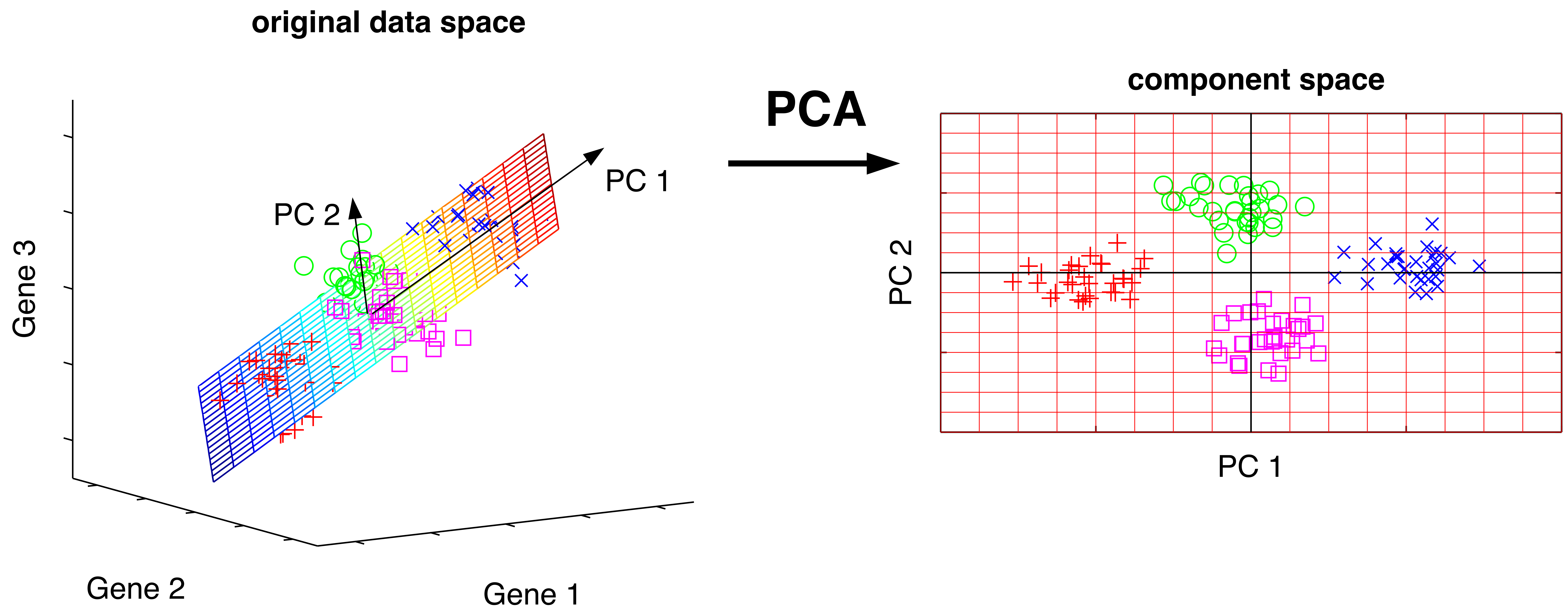
Number of Clusters

[D. Robinson, 2015]

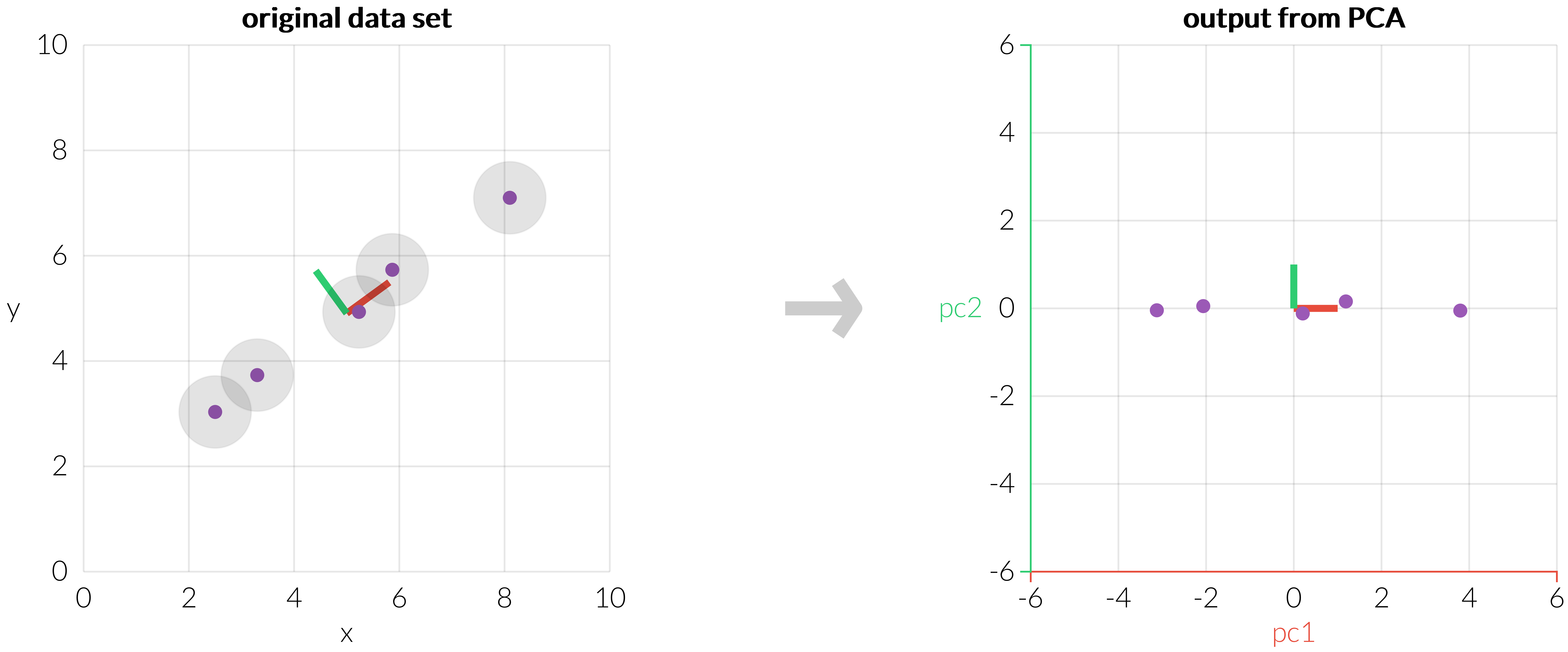
Dimensionality Reduction

- 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]