
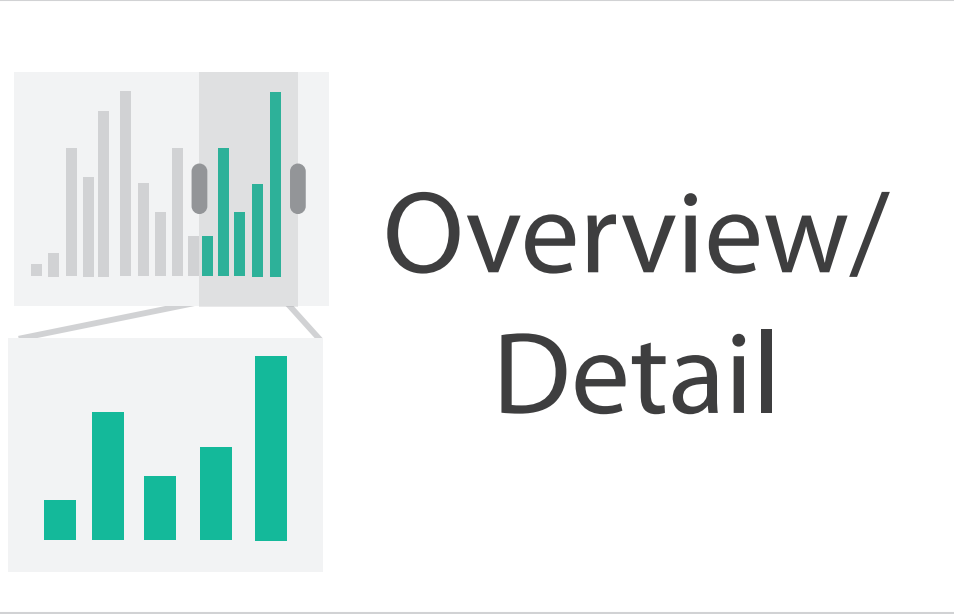
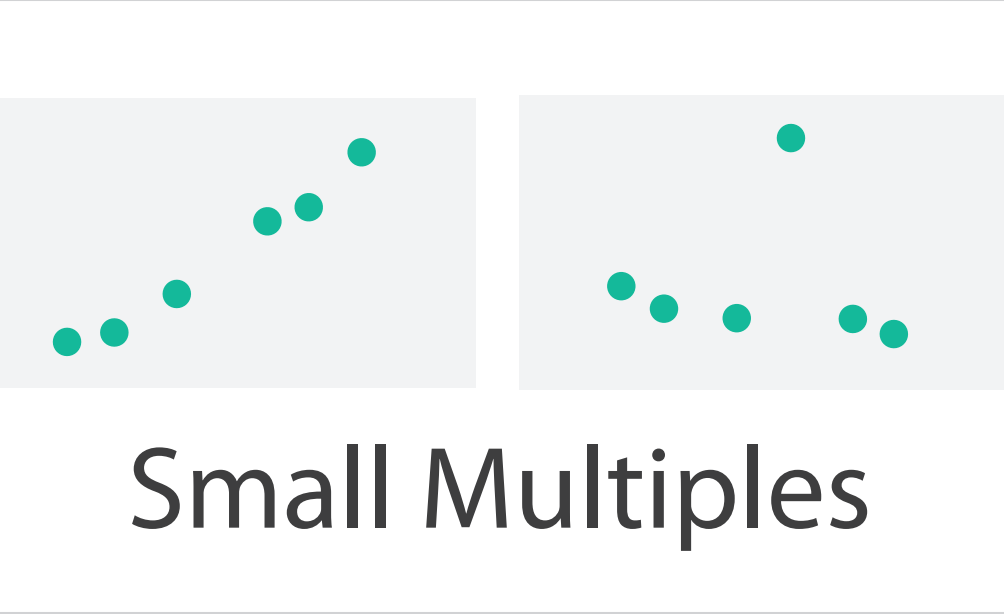

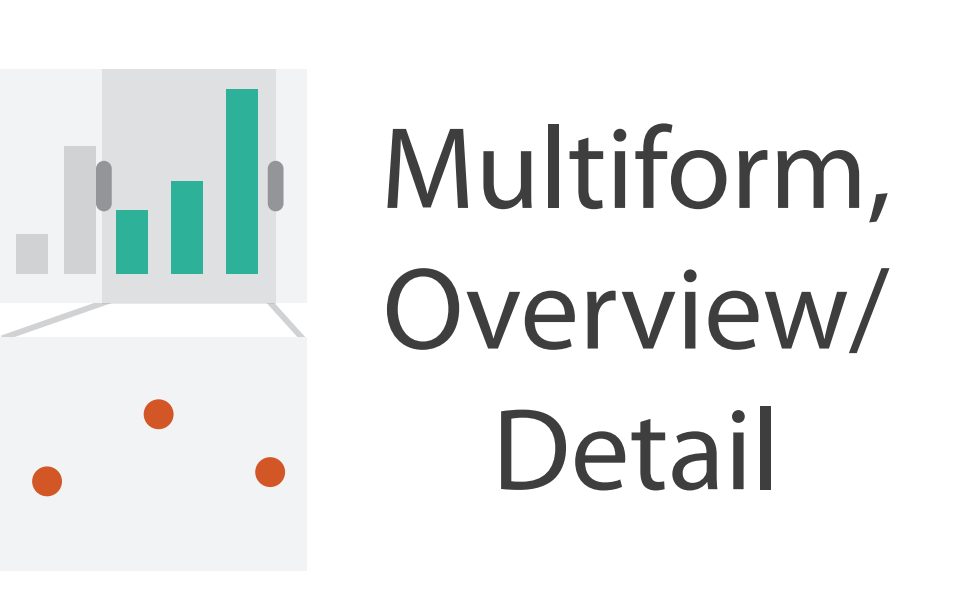



Data Visualization (CSCI 627/490)

Filtering & Aggregation

Dr. David Koop

Multiple Views

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

[Munzner (ill. Maguire), 2014]

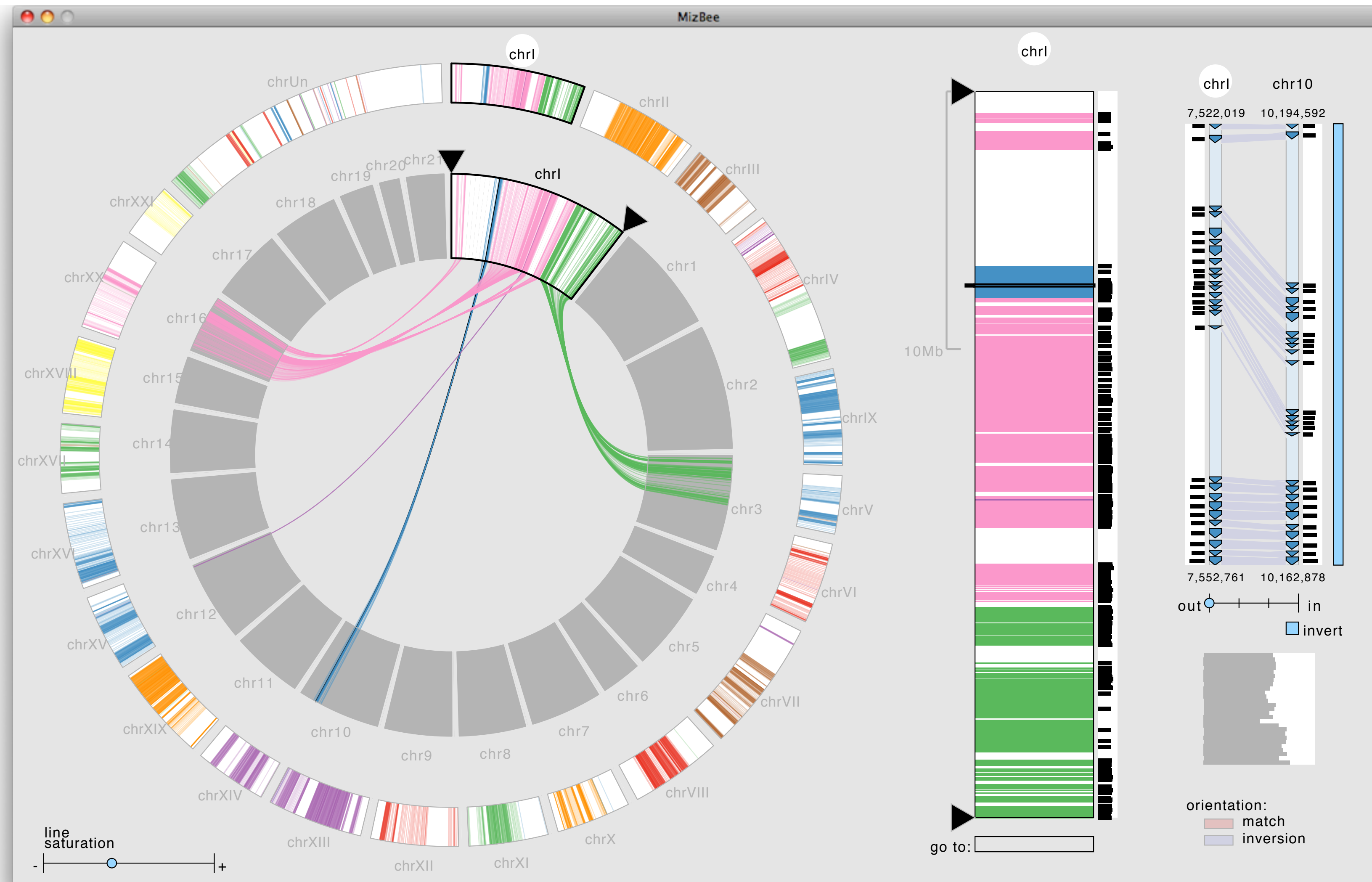
Multiform

The screenshot displays a GIS application window with a menu bar (File, Edit, Misc, Base, Meta, Help) and several panels:

- Map Overview:** A small map of the United States with Michigan highlighted.
- States:** A table listing US states with columns for Name, Area, and Population. Michigan is selected.
- Map:** A detailed map of Michigan showing county boundaries and various data layers.
- Counties:** A table listing Michigan counties with columns for Name, Area, Population, and other metrics. Wayne County is selected.
- Cities:** A table listing cities within Wayne County, including their names, counties, and populations. Detroit is selected.
- Airports & Seaplane Bases:** A table listing airports and seaplane bases with columns for Name, Elevation, and County.
- Census Values (Scatterplots):** A grid of scatterplots comparing various census variables (e.g., 10^4 SQ_MILE, 10^4 POPO00SQMIL, Under18, MEDAGE2000, PEROVER65) across two census values.
- Census Values (Scatterplot):** A larger scatterplot showing the relationship between two census values.
- Color Scheme:** A panel for selecting a color scheme, currently set to 'Sequential Non-Gray'.
- City-City Distances:** A table showing distances between various cities, with columns for city names and distance values.
- Show/Label:** Checkboxes for displaying and labeling different data layers like Counties, Cities, Roads, Railroads, Airports, and Urban Areas.

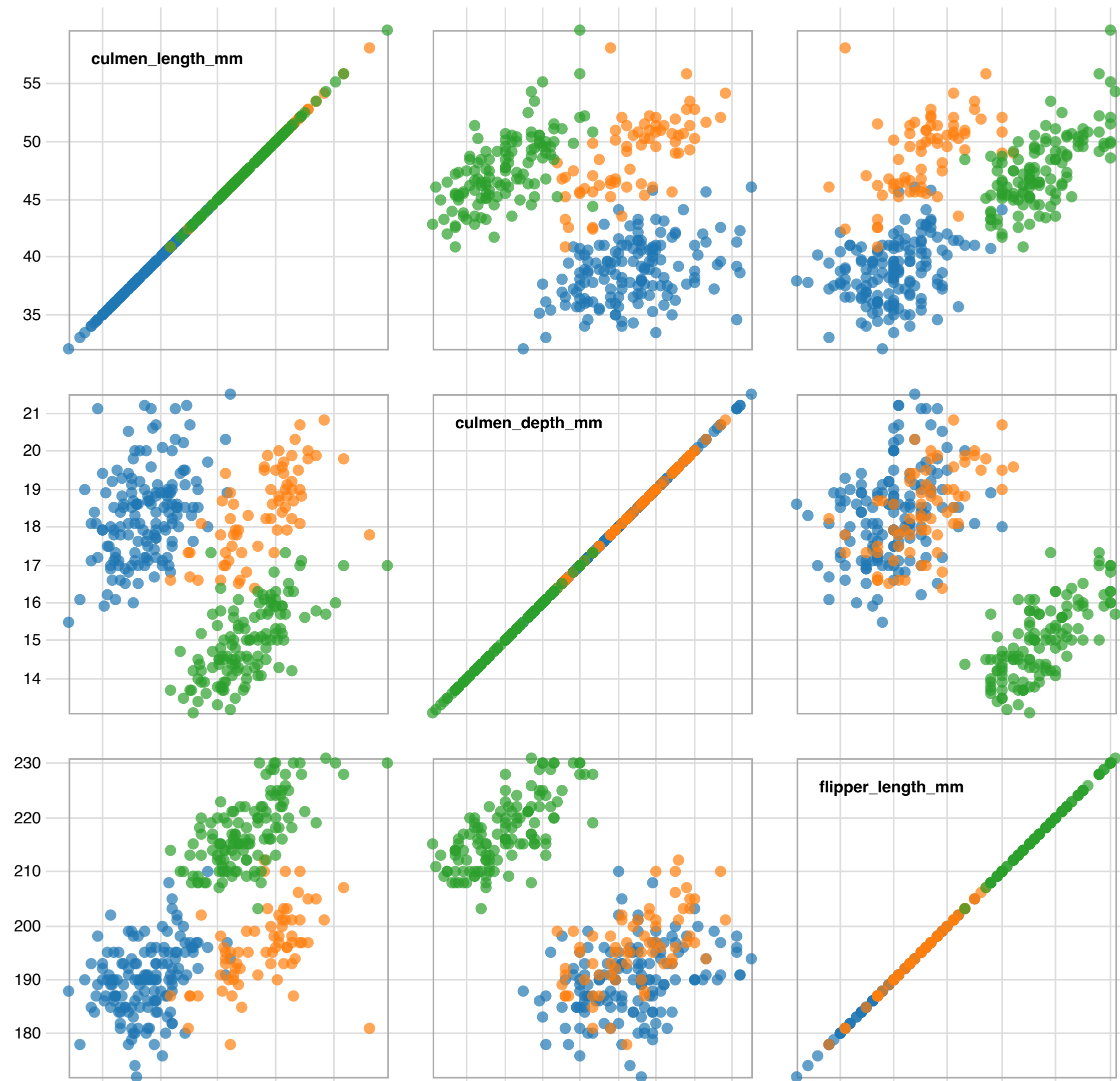
[Improvise, Weaver, 2004]

Overview-Detail (Different Encoding)



[M. Meyer et al.]

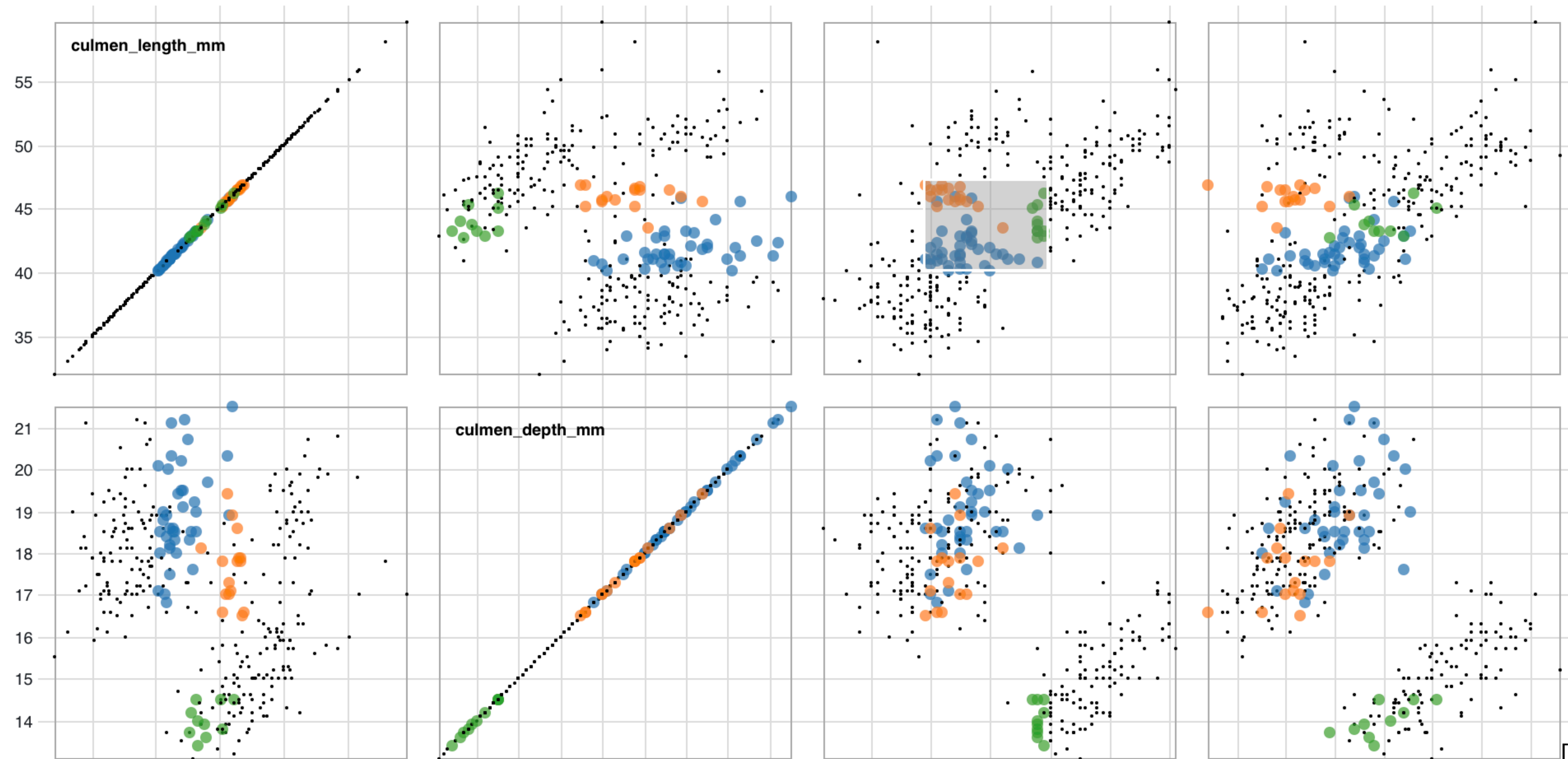
Small Multiples: Scatterplot Matrix



[D3 Example]

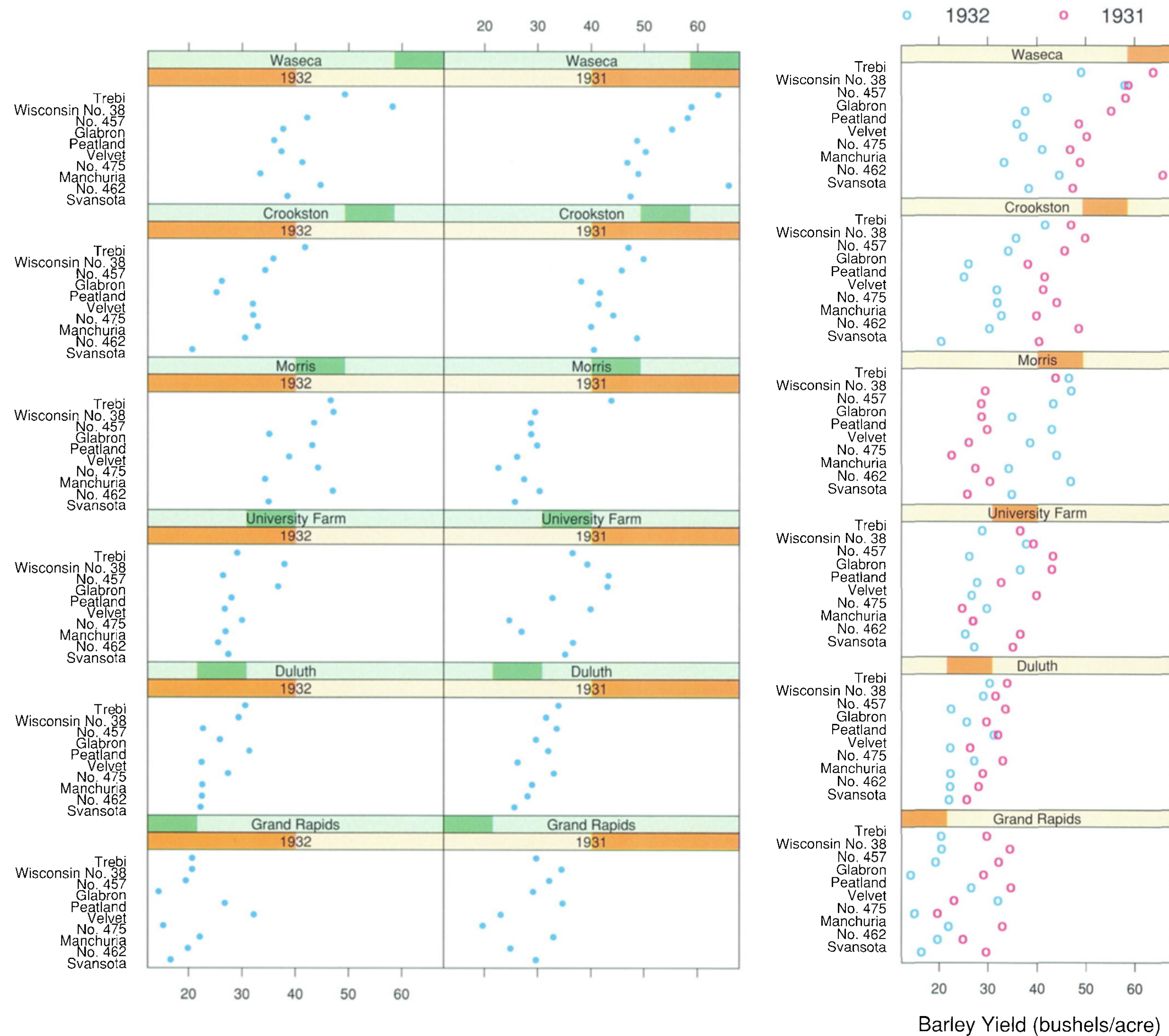
Brushing

■ Adelie ■ Chinstrap ■ Gentoo



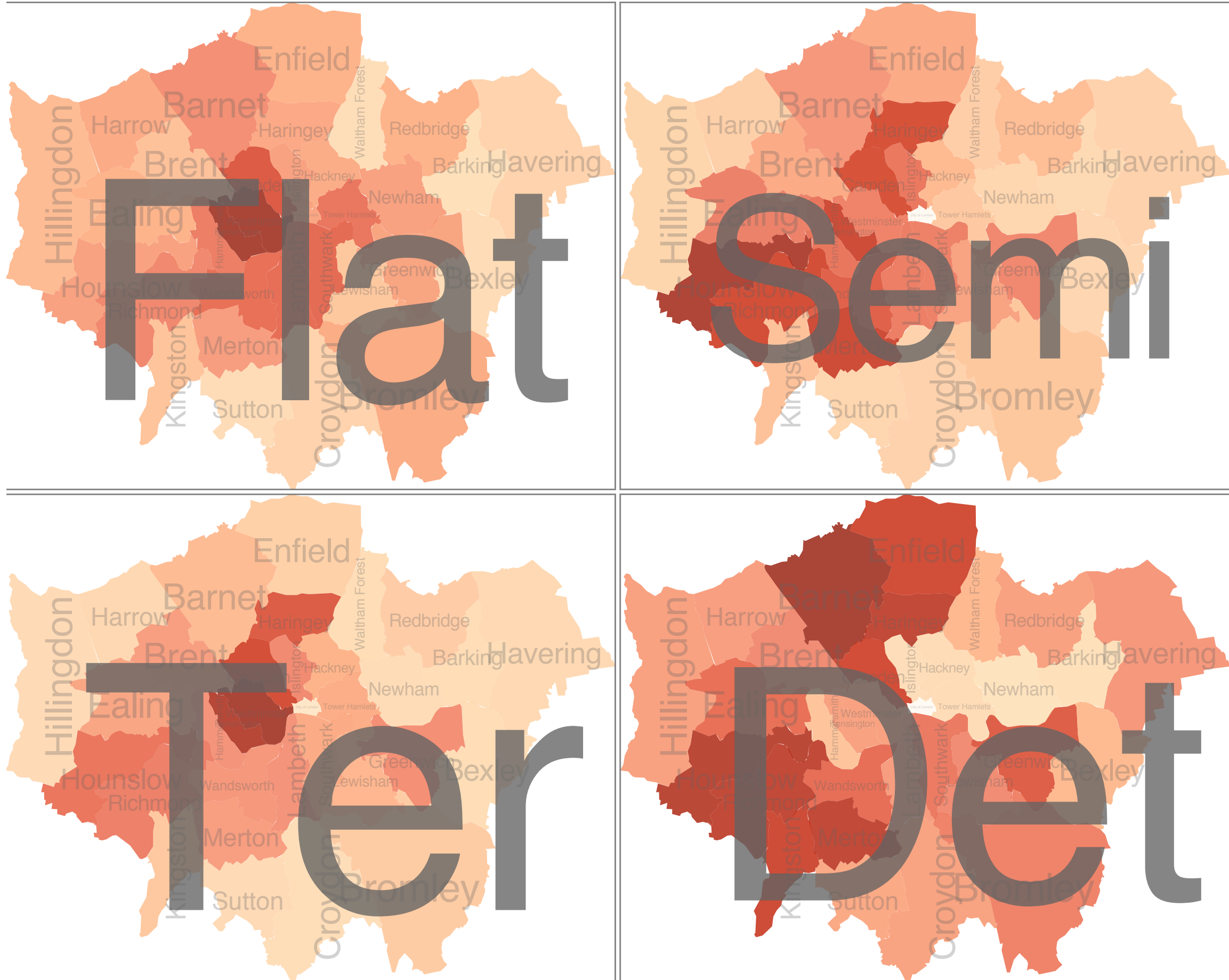
[D3 Example]

Partitioning: Main-Effects Ordering



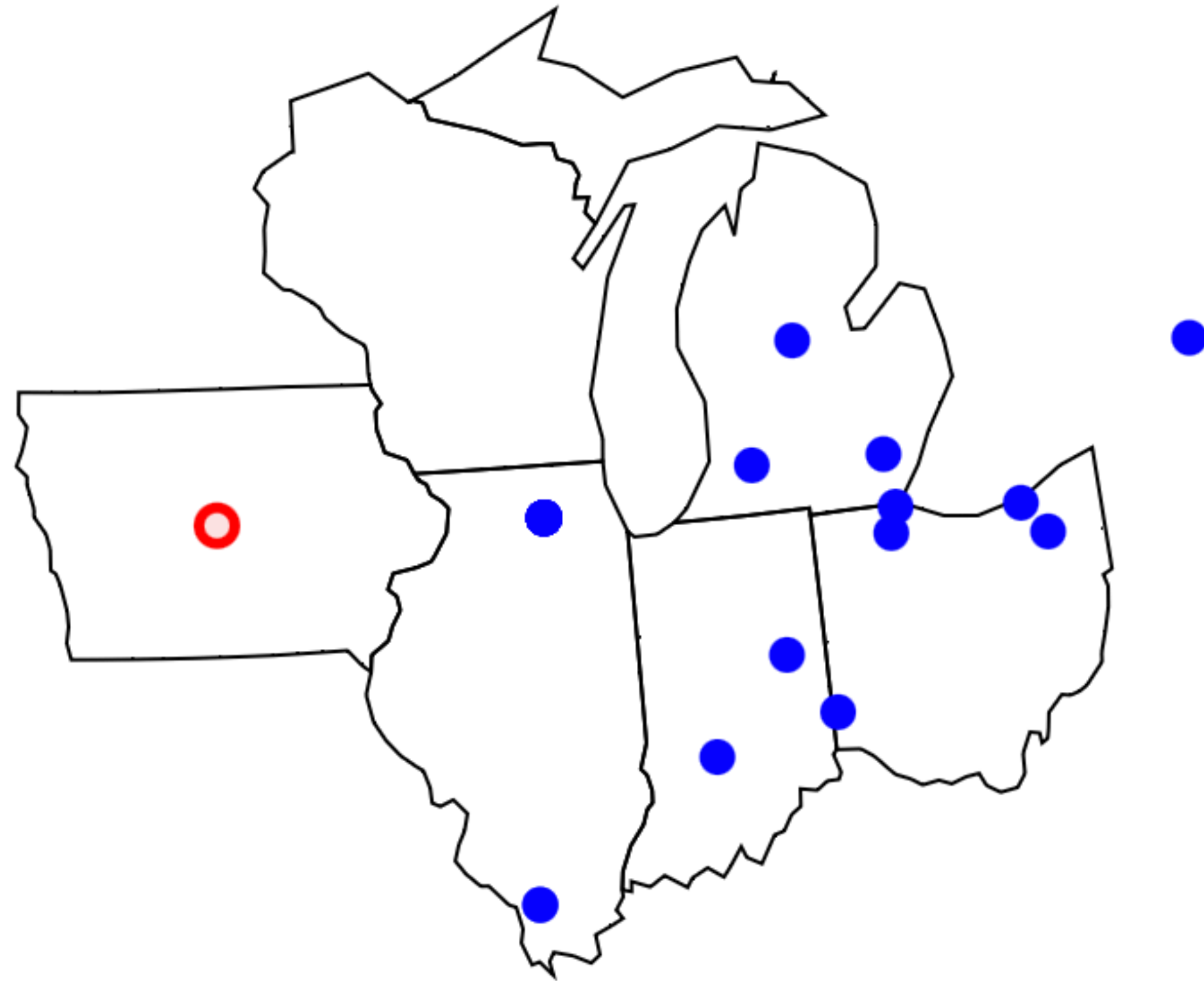
[Becker et al., 1996]

Recursive Subdivision



[Slingsby et al., 2009]

Linked Highlighting Example



Date	Opponent	Result	Site
Nov 8, 2018	Yale	W 89–80	DeKalb, IL
Nov 11, 2018	at Iowa State	L 60–70	Ames, IA
Nov 18, 2018	Northern Iowa	W 70–59	DeKalb, IL
Nov 25, 2018	at Indiana	L 73–91	Bloomington, IN
Nov 28, 2018	North Dakota State	W 81–63	DeKalb, IL
Dec 1, 2018	at Montana	L 70–86	Missoula, MT
Dec 2, 2018	vs. Nevada	W 98–69	Missoula, MT
Dec 7, 2018	at Southern Illinois	L 73–82	Carbondale, IL
Dec 8, 2018	vs. Western Illinois	W 86–61	Carbondale, IL
Dec 17, 2018	Eastern Illinois	W 78–59	DeKalb, IL
Dec 21, 2018	Chicago State	W 114–52	DeKalb, IL
Dec 31, 2018	Brown	W 109–102	DeKalb, IL
Jan 5, 2019	Miami (OH)	W 82–71	DeKalb, IL
Jan 9, 2019	at Kent State	L 78–87	Kent, OH

Reducing Complexity

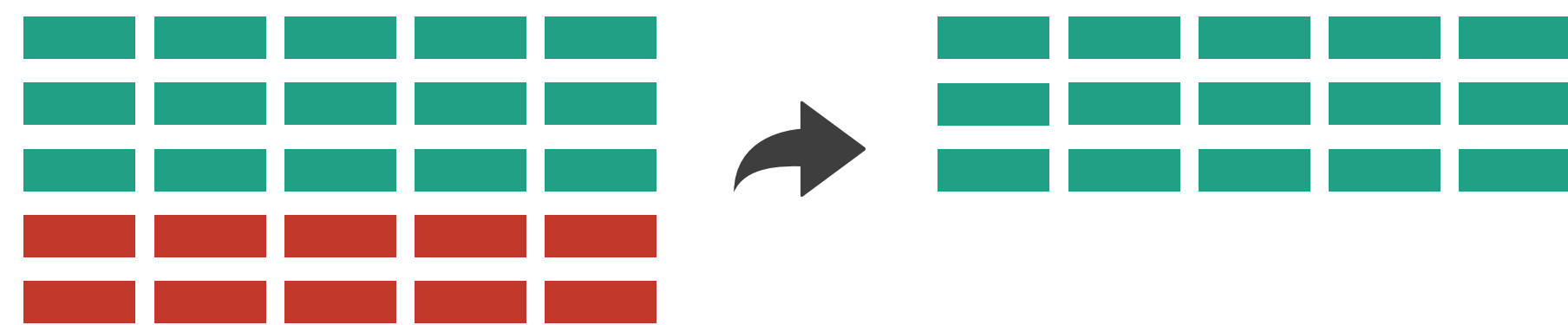
Reducing Complexity

- Too many items or attributes lead to visual clutter
- Interaction and Multiple Views can help, but often lose the ability to start understanding an entire dataset at first glance
- **Reduction** techniques show less data to reduce complexity
- Can reduce items or attributes (both are **elements**)
- **Filtering**: eliminate elements from the current view
 - "out of sight, out of mind"
- **Aggregation**: replace elements with a new element that represents the replaced elements
 - summarization is often challenging to design
- Another method is **focus+context**: show details in the context of an overview

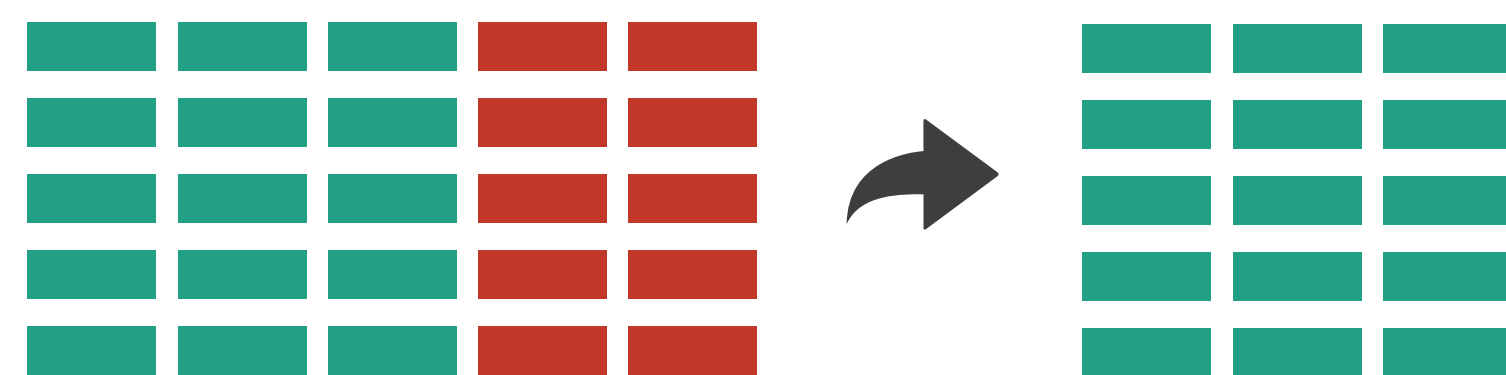
Overview: Reducing Items & Attributes

➔ Filter

➔ Items

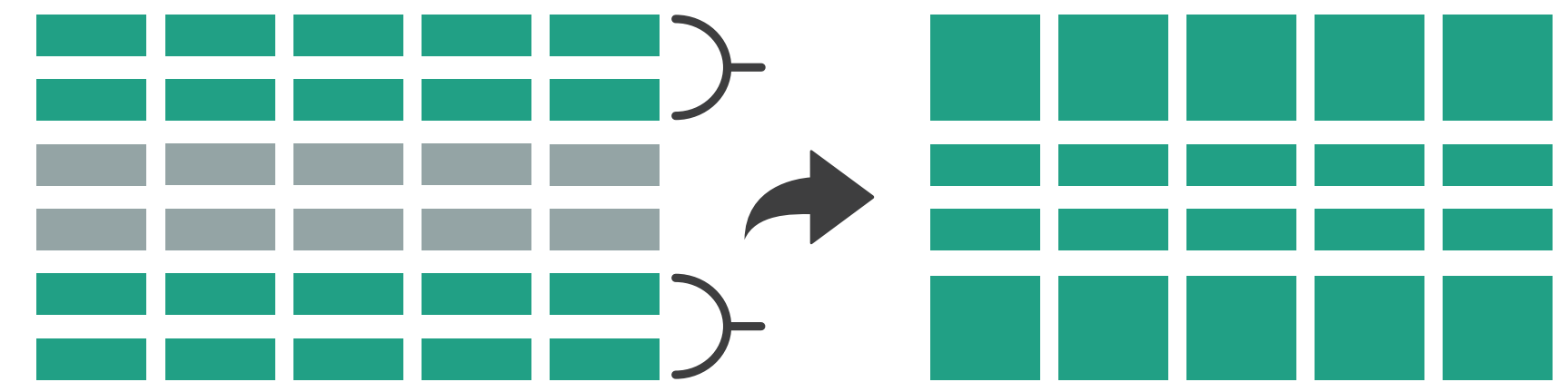


➔ Attributes

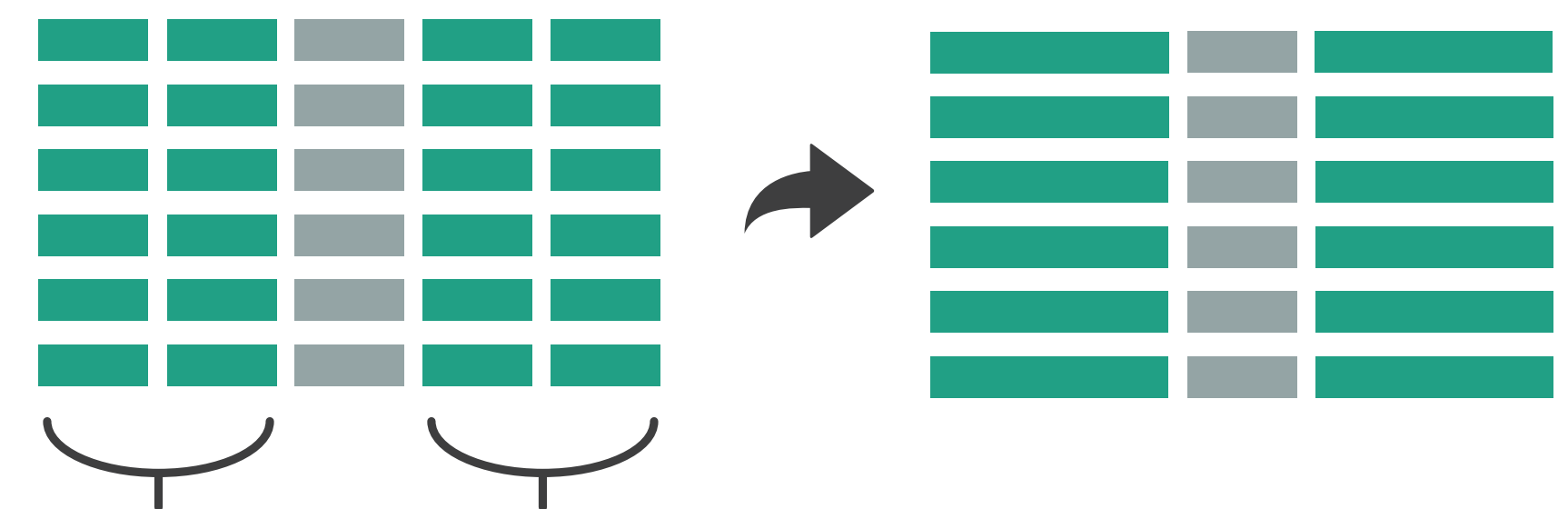


➔ Aggregate

➔ Items



➔ Attributes



[Munzner (ill. Maguire), 2014]

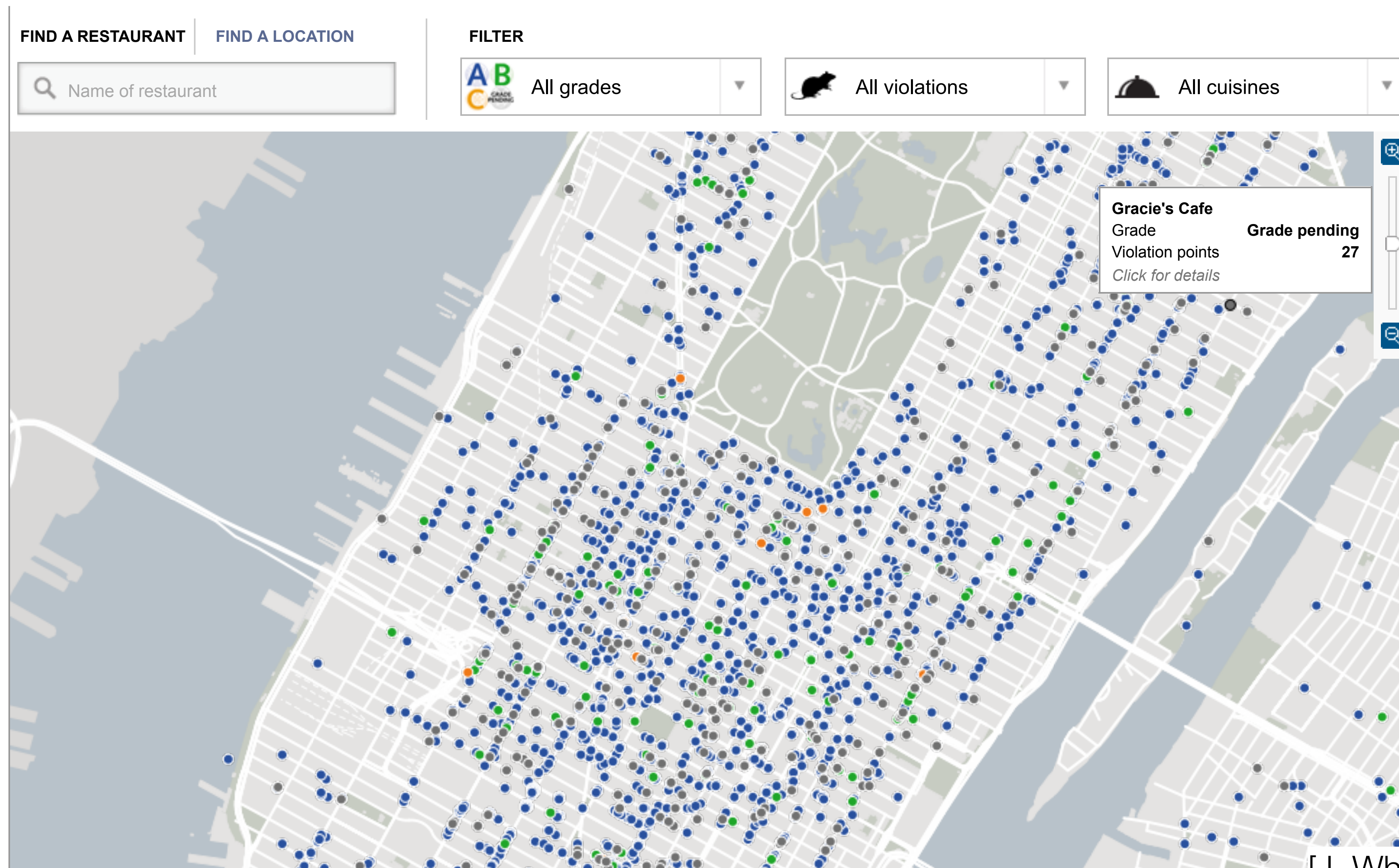
Filtering

- 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

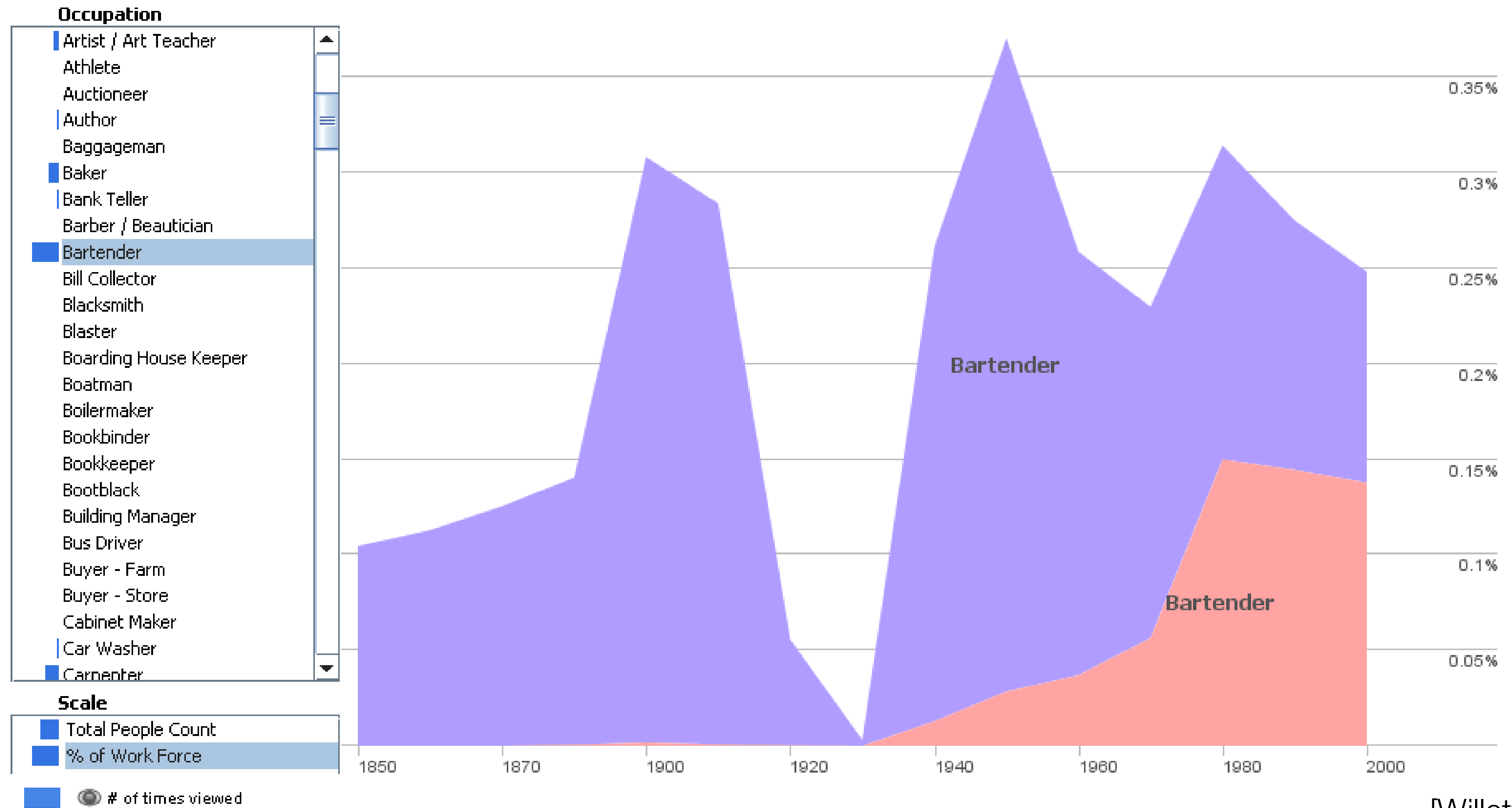


[J. White, New York Times]

Dynamic Filters

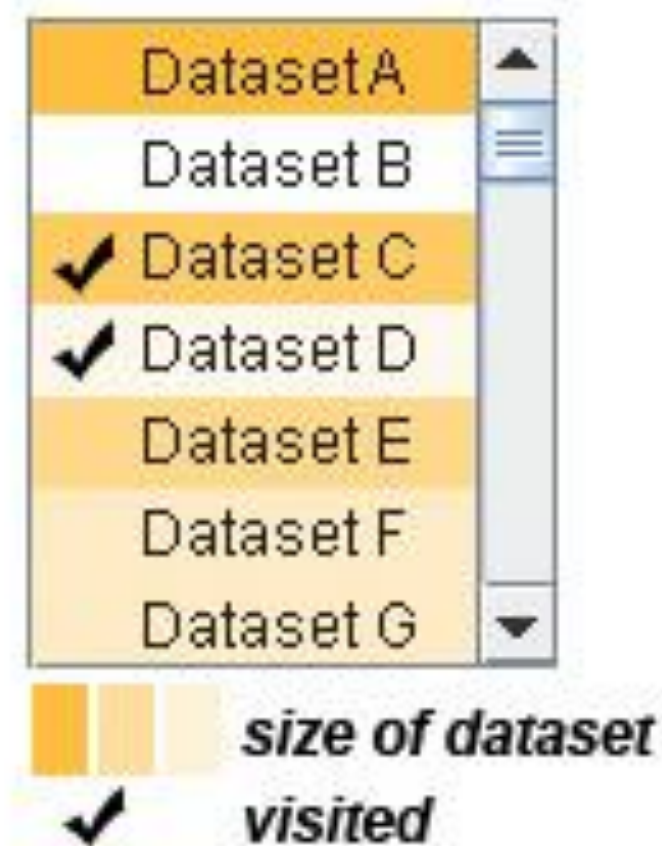
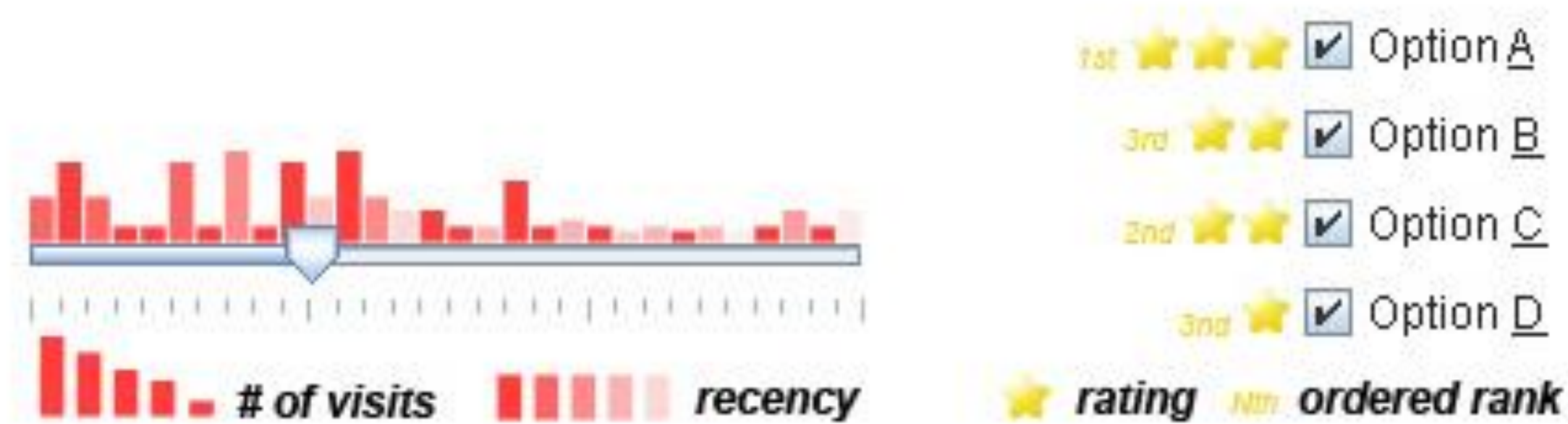
- Interaction need not be with the visualization itself
- Users interact with **widgets** that control which items are shown
 - Sliders, Combo boxes, Text Fields
- Often tied to attribute values
- Examples:
 - All restaurants with an "A" Grade
 - All pizza places
 - All pizza places with an "A" Grade

Scented Widgets



[Willett et al., 2007]

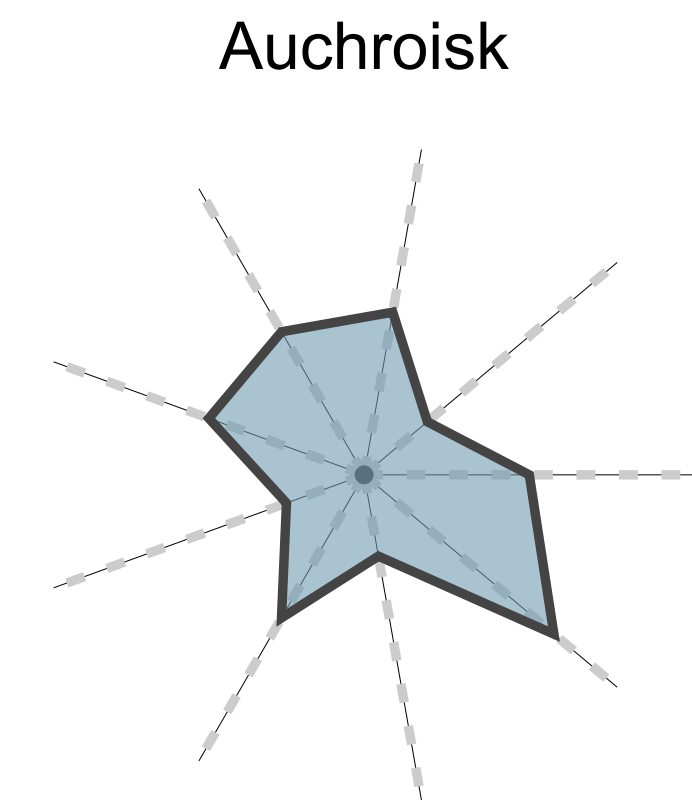
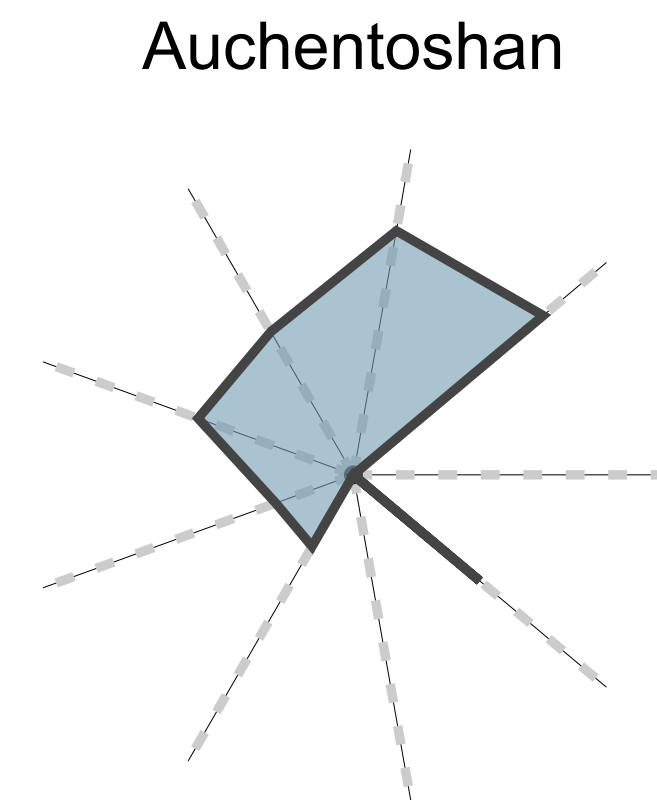
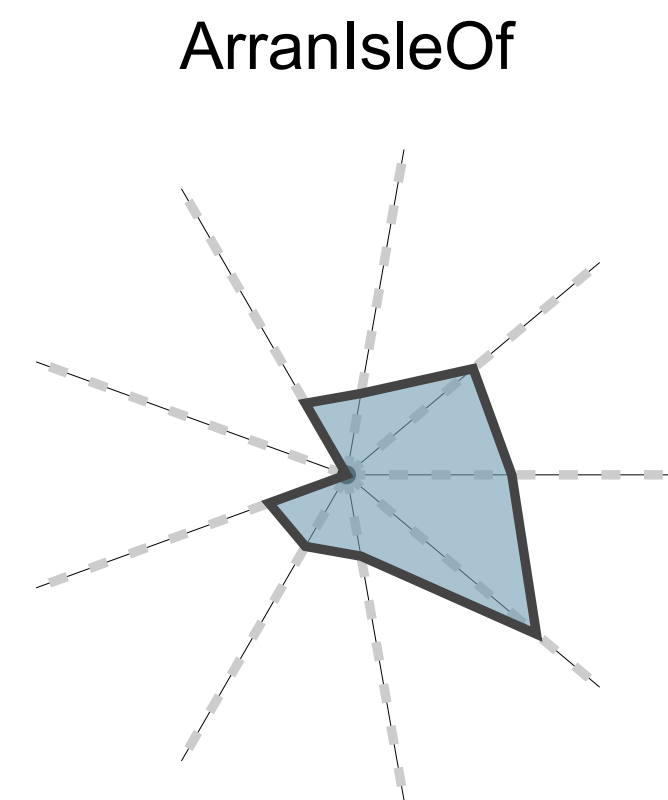
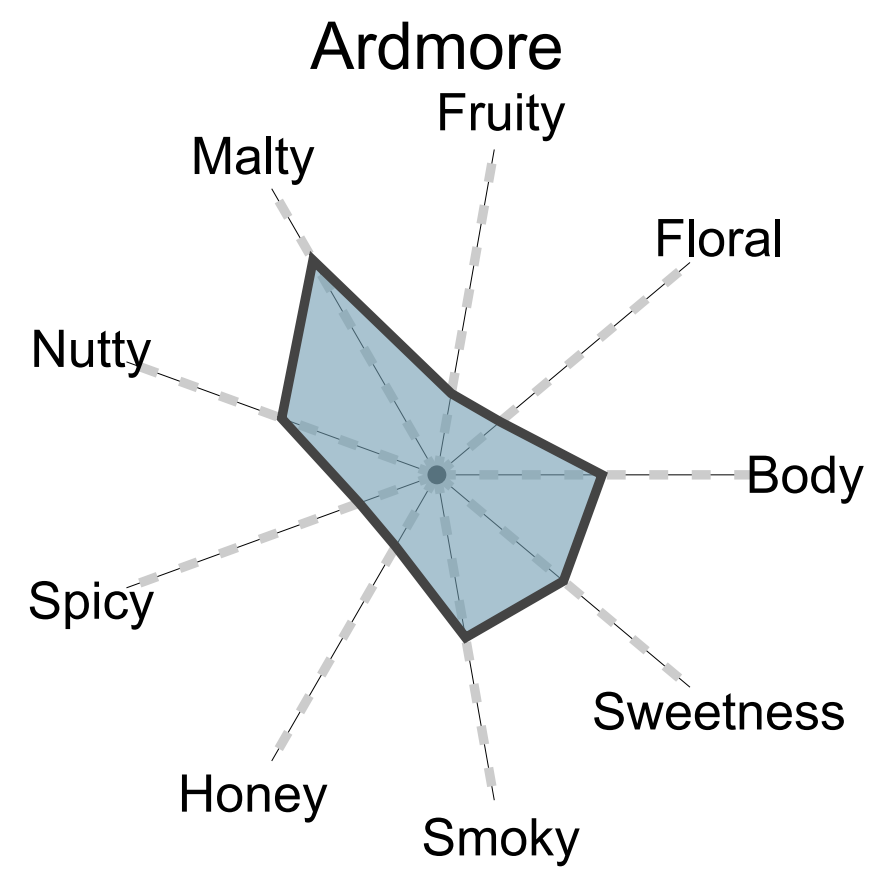
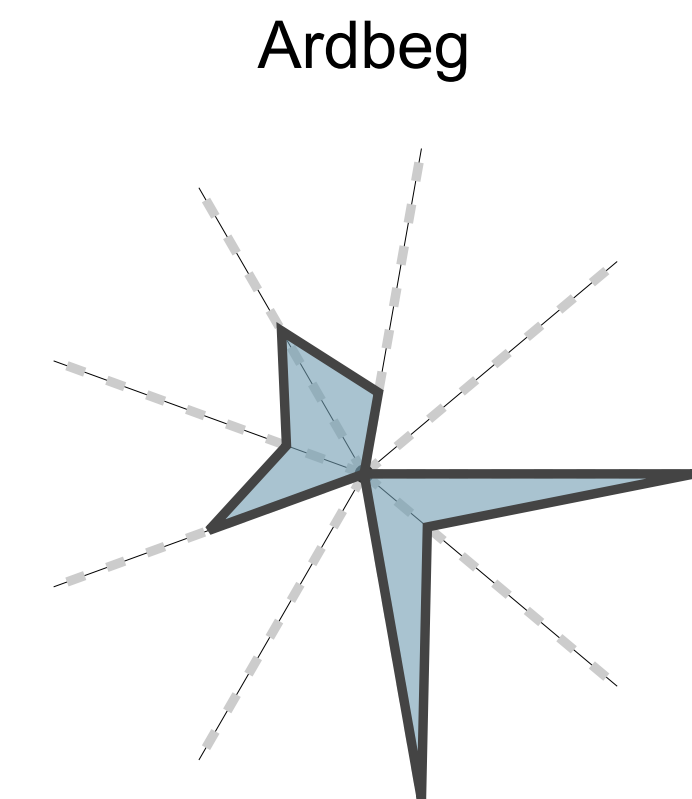
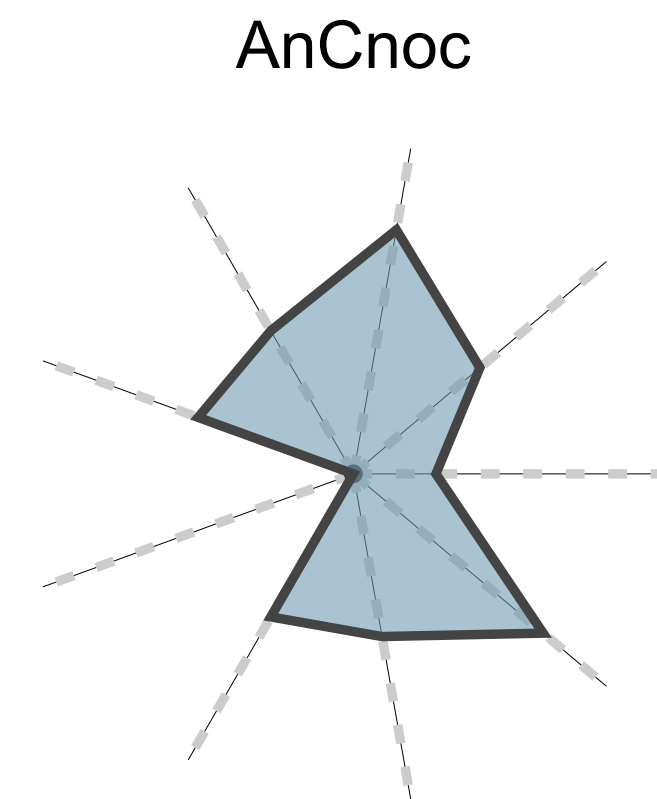
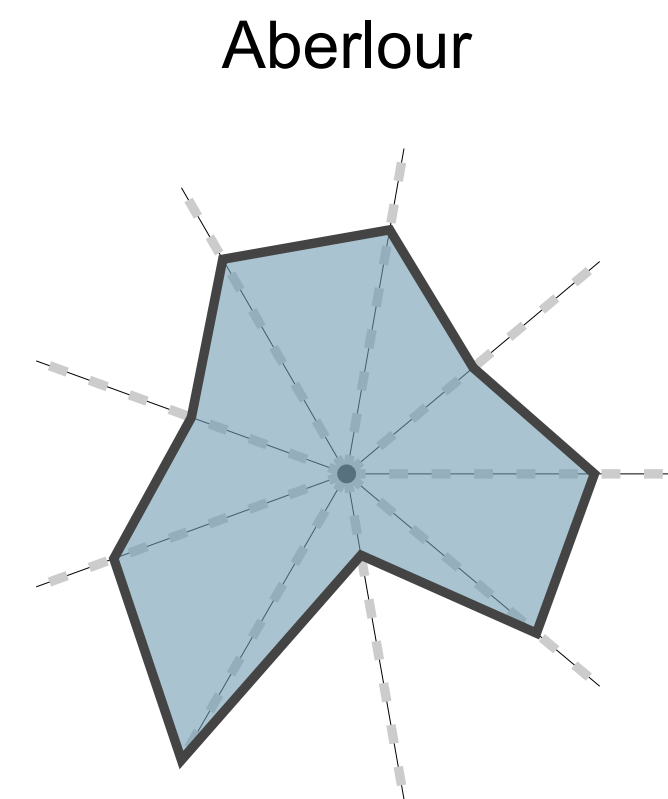
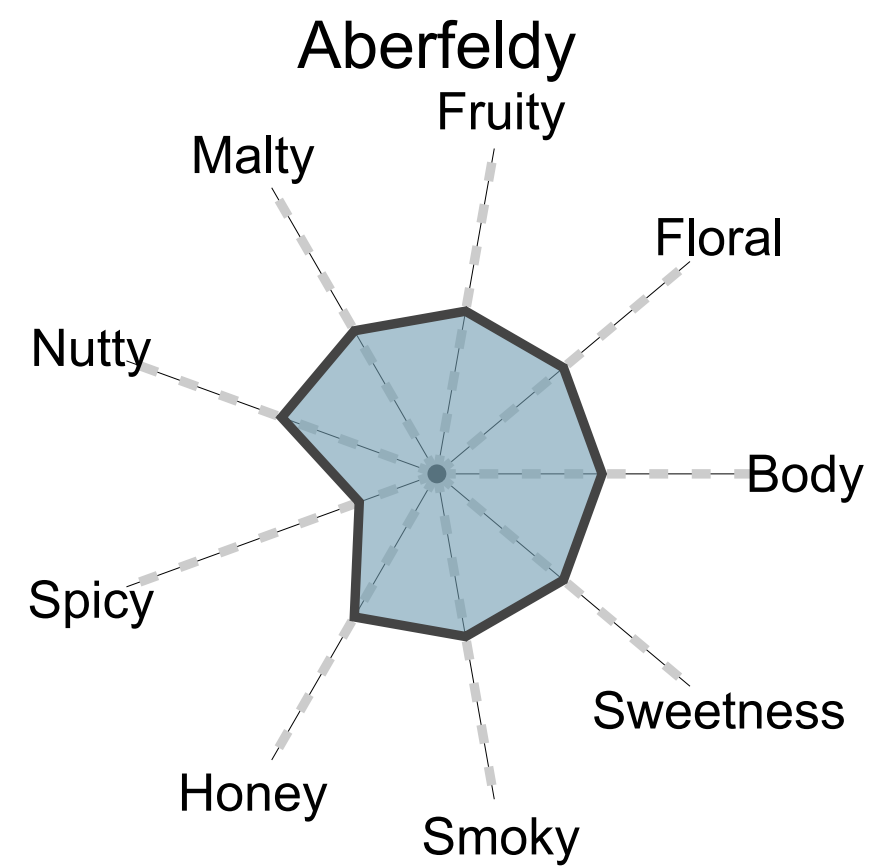
Scented Widgets



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	

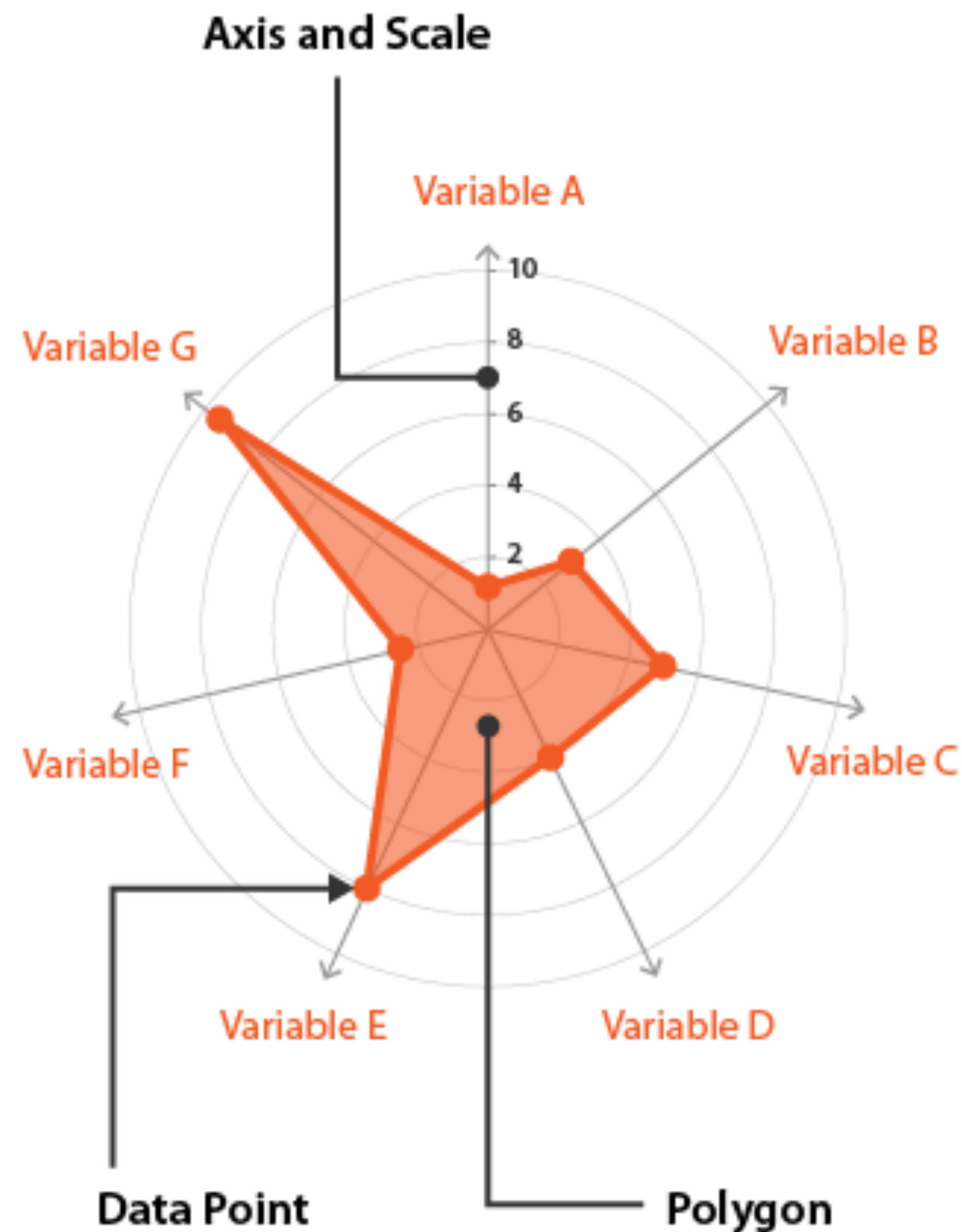
[Willett et al., 2007]

Star Plots (aka Radar Charts)



[K. Schau]

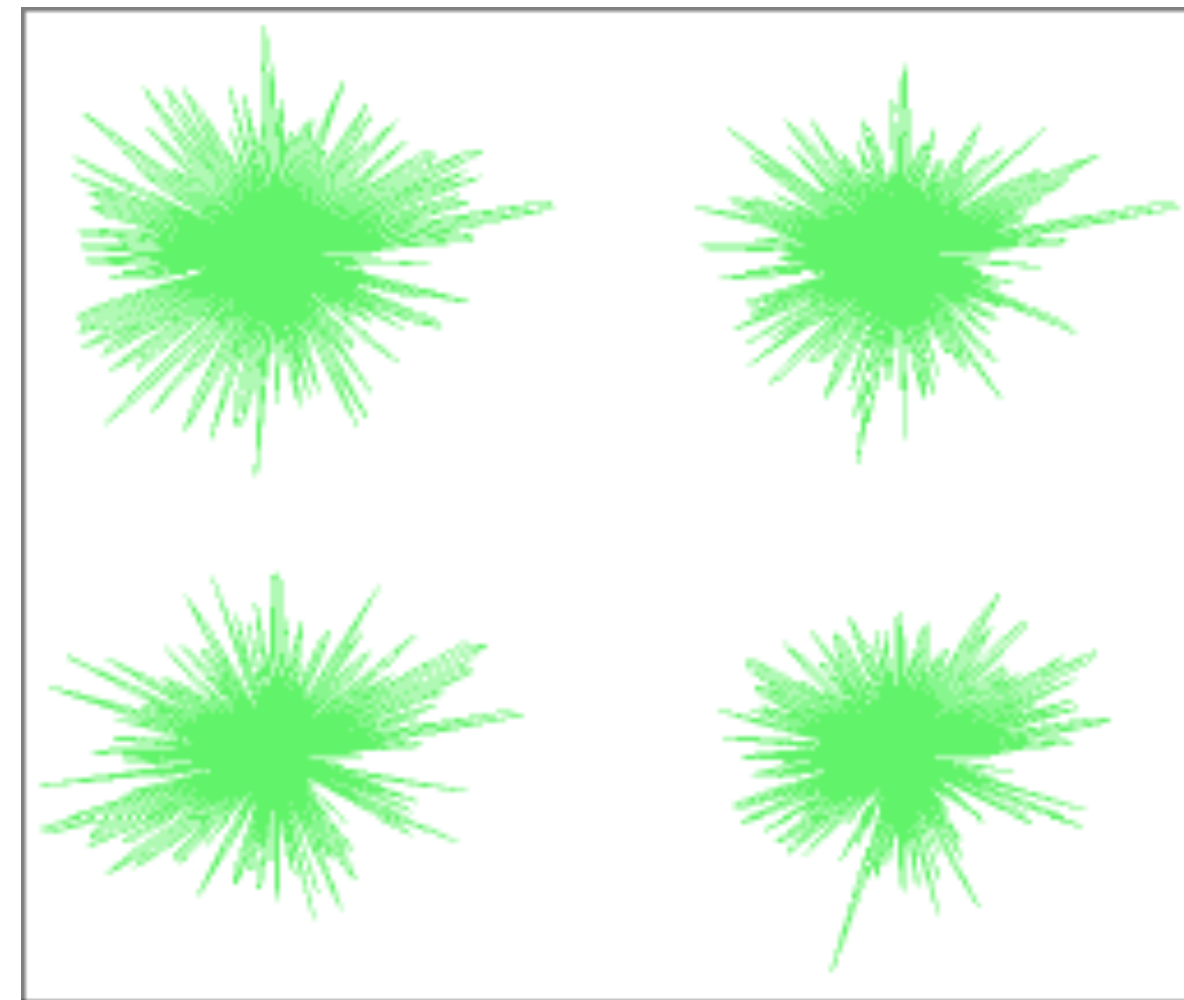
Star Plot / Radar Chart



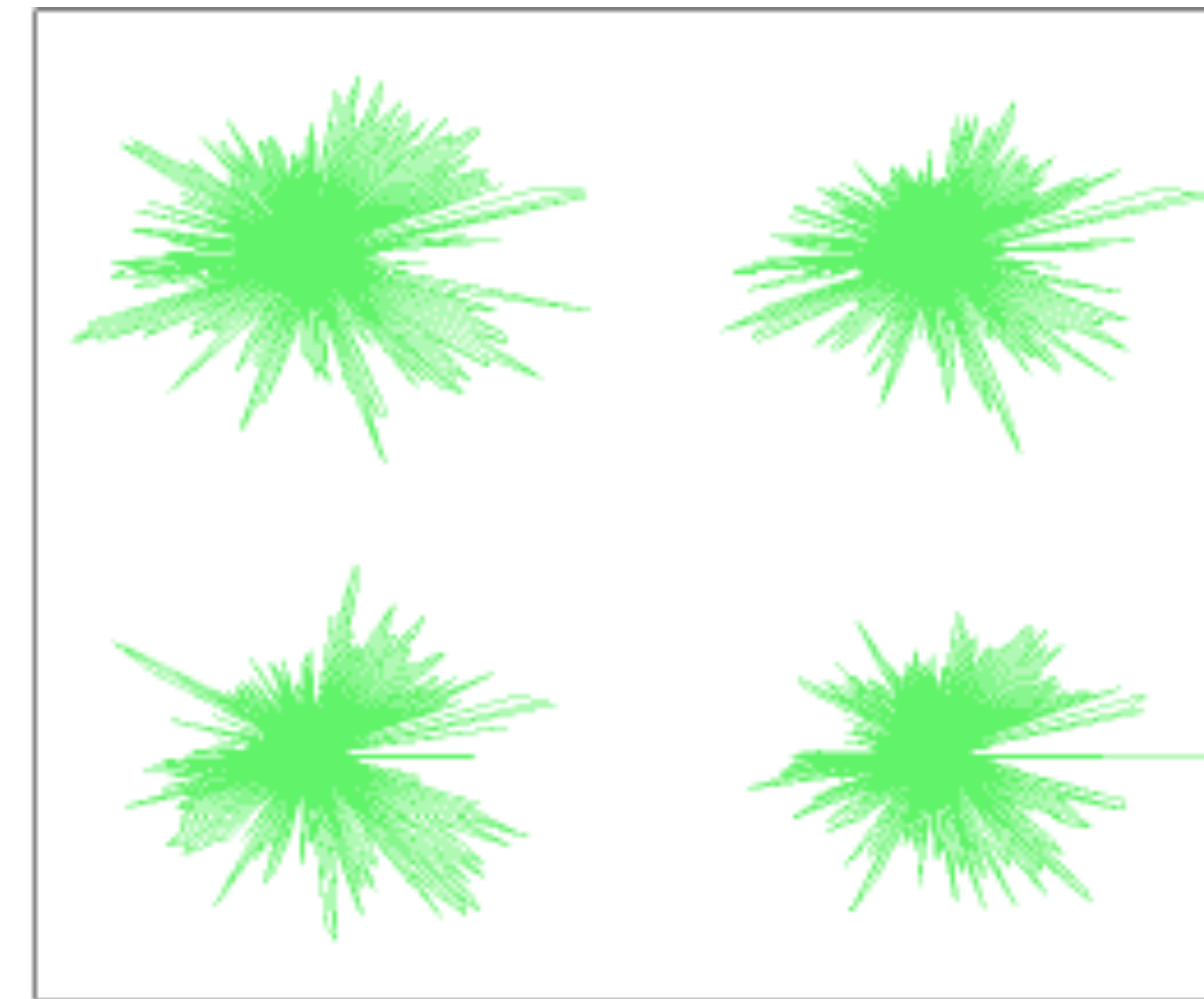
- Use:
 - Compare variables
 - Similarities/differences of items
 - Locate outliers
- Considerations:
 - Order of axes
 - Too many axes cause problems

[S. Ribecca]

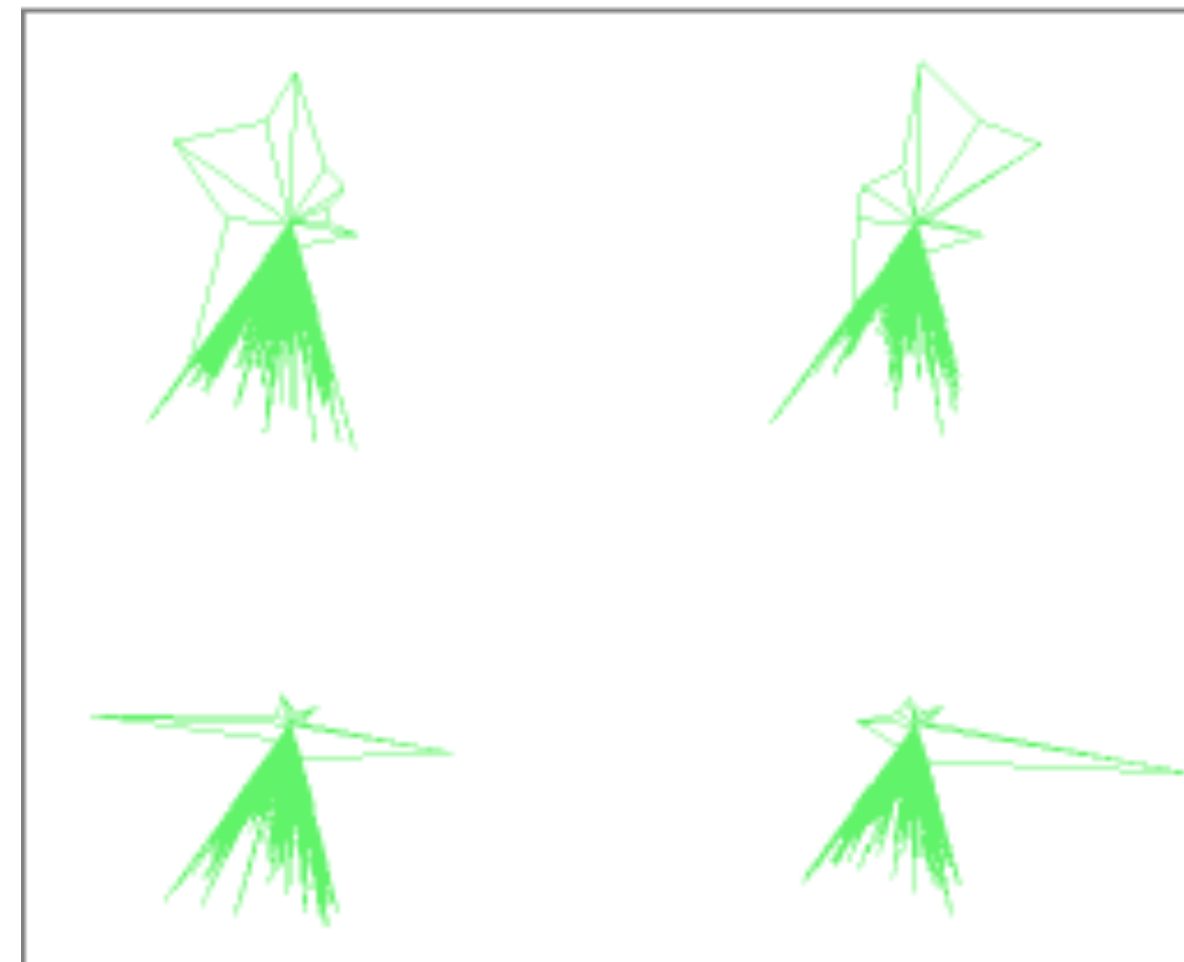
Attribute Filtering on Star Plots



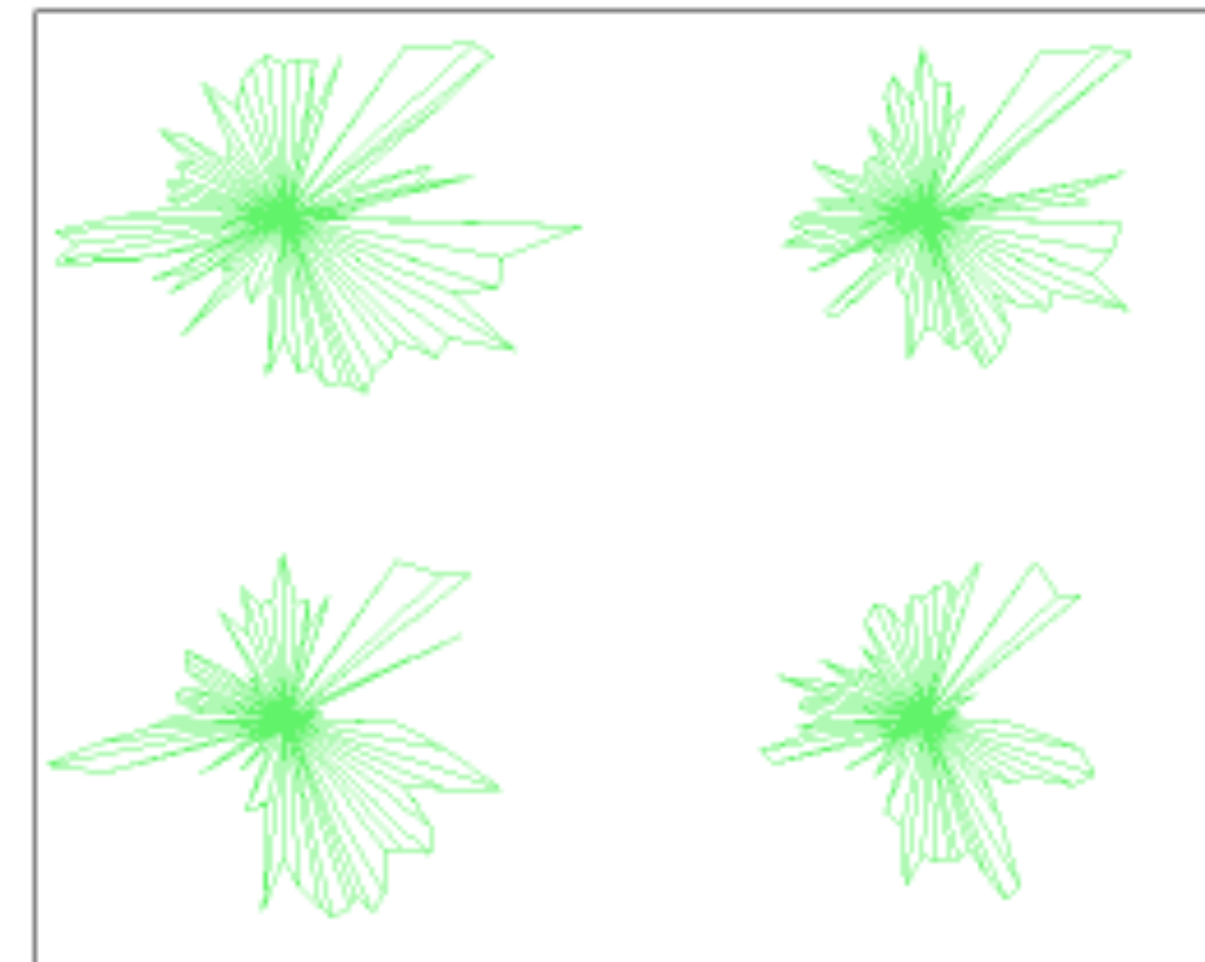
(a)



(b)



(c)



(d)

[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

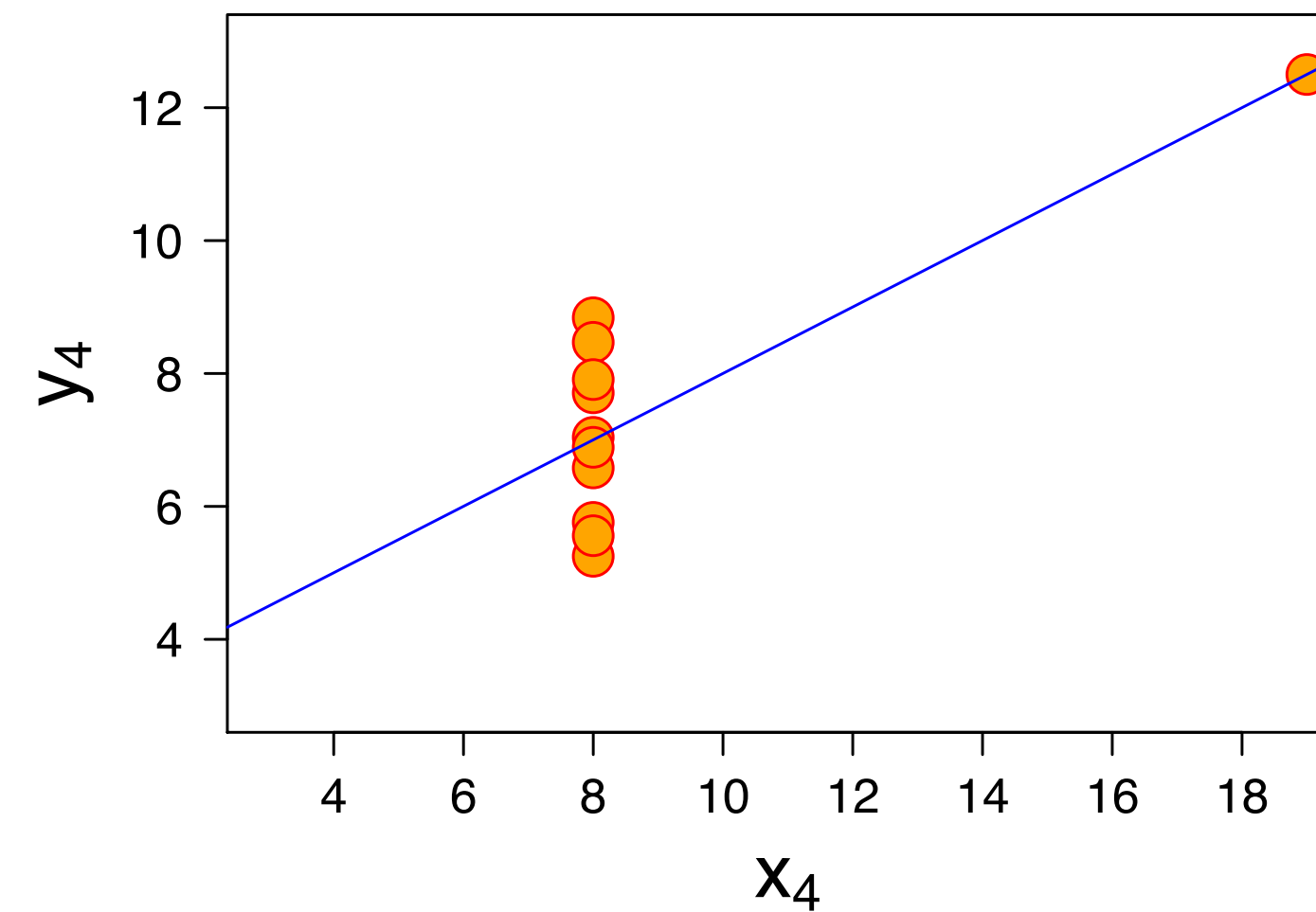
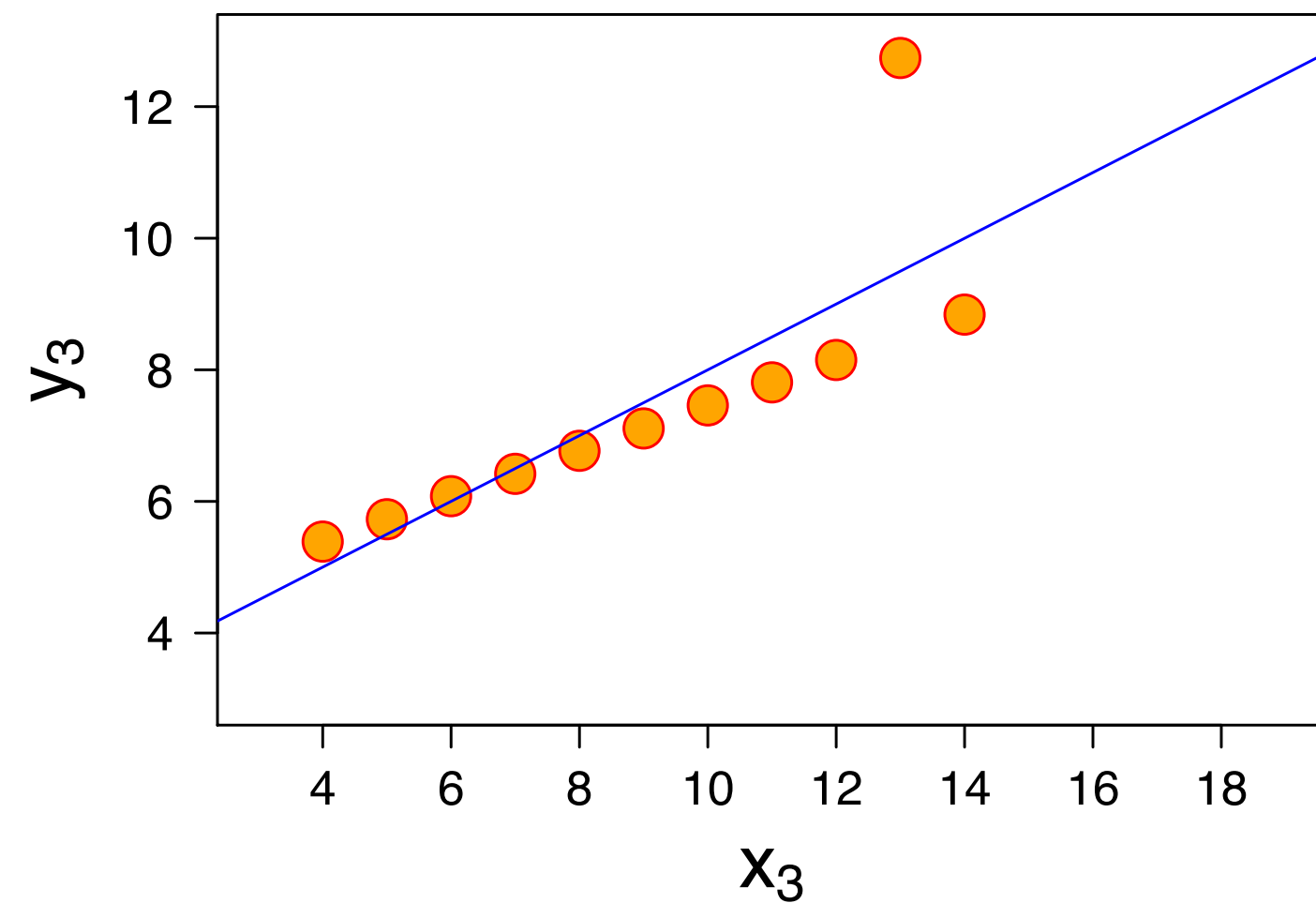
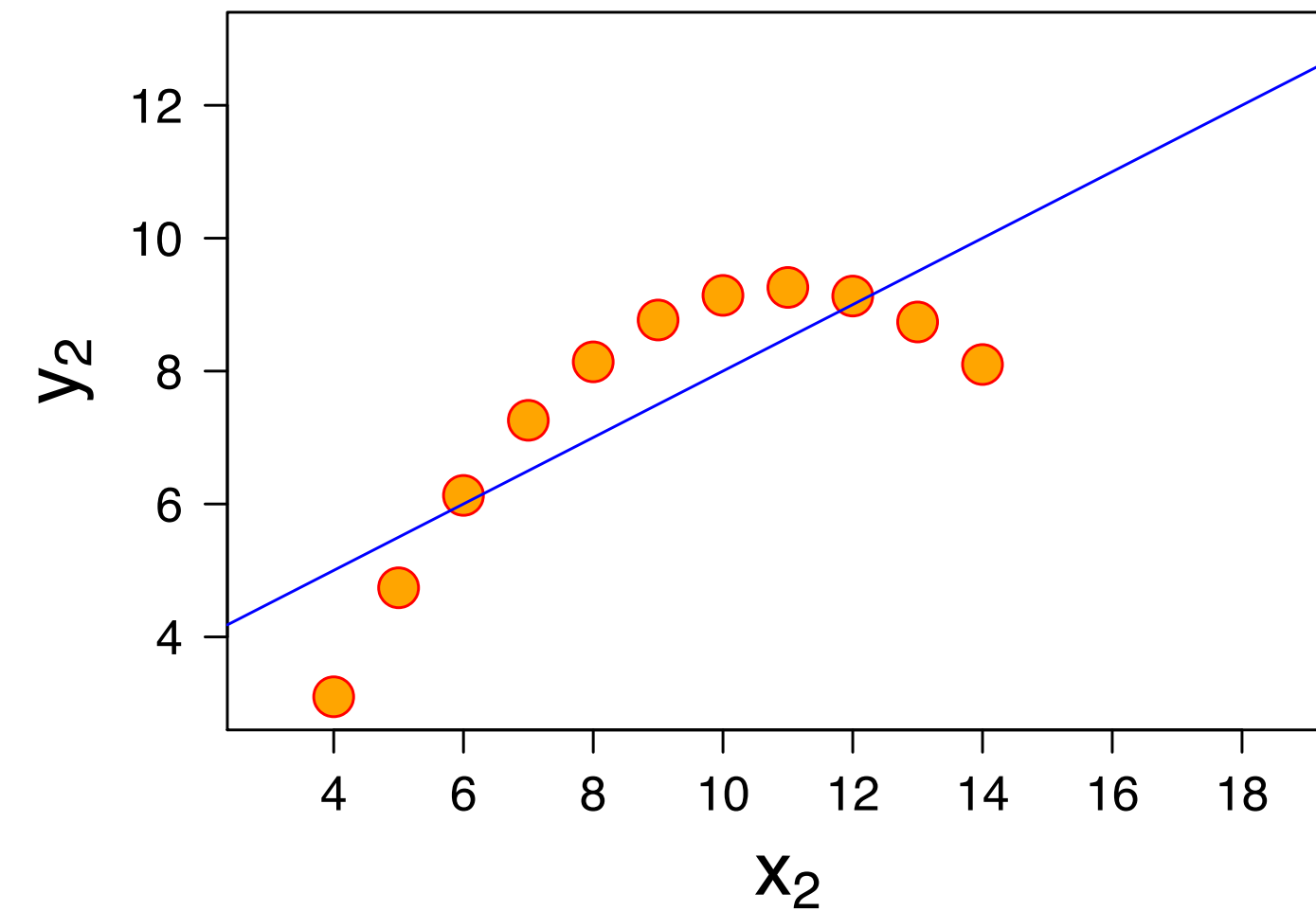
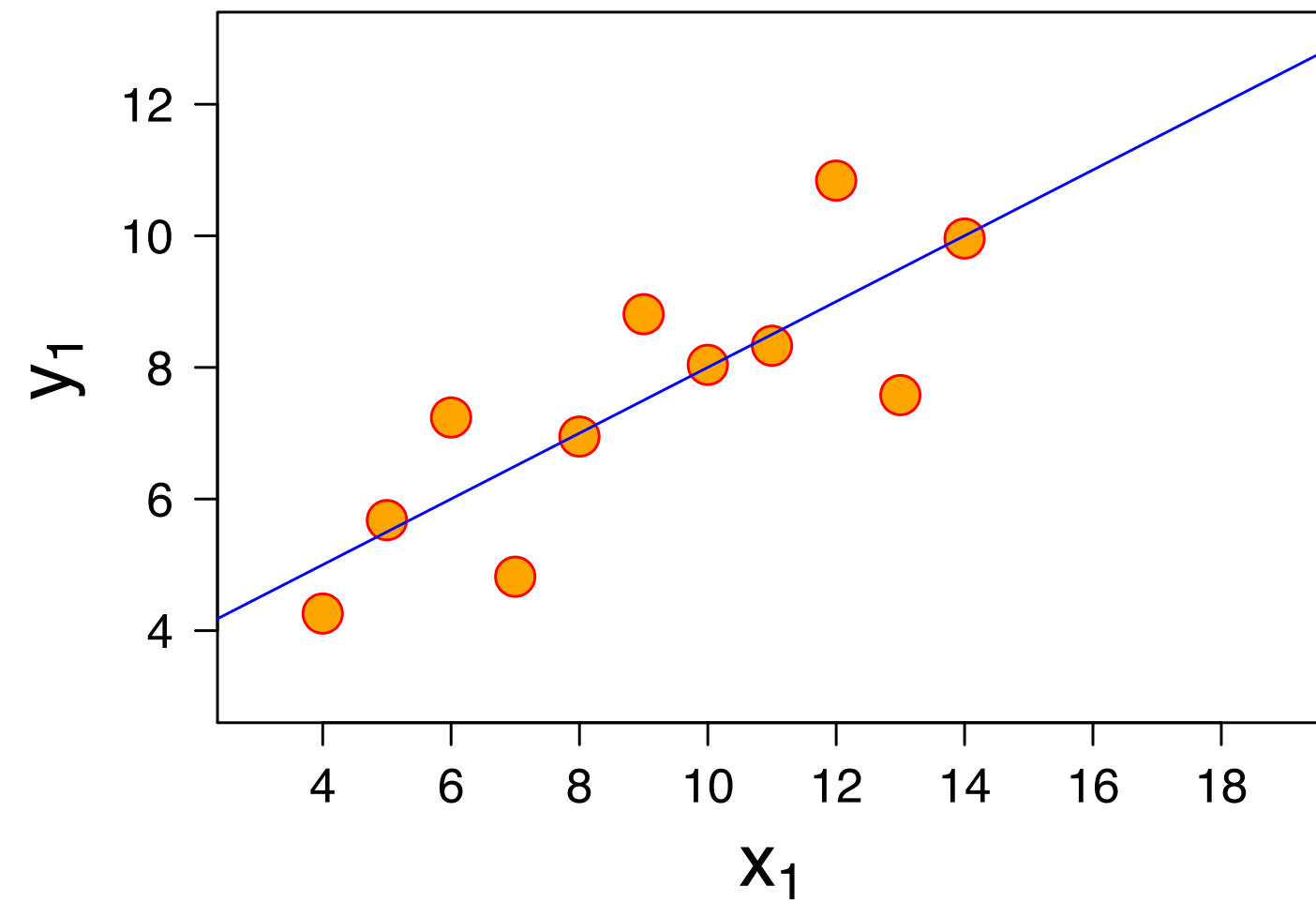
Aggregation

- 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

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

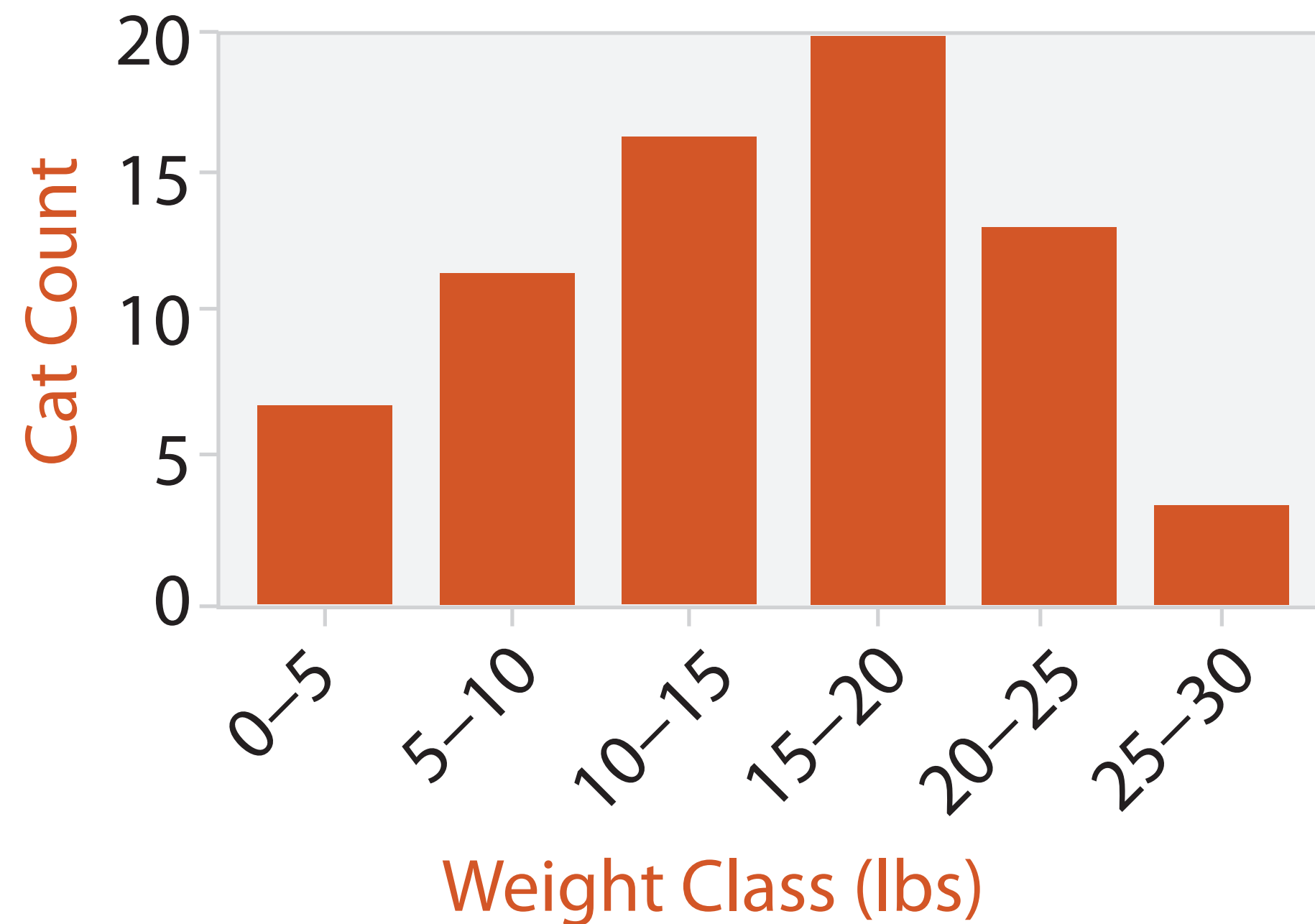
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	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

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

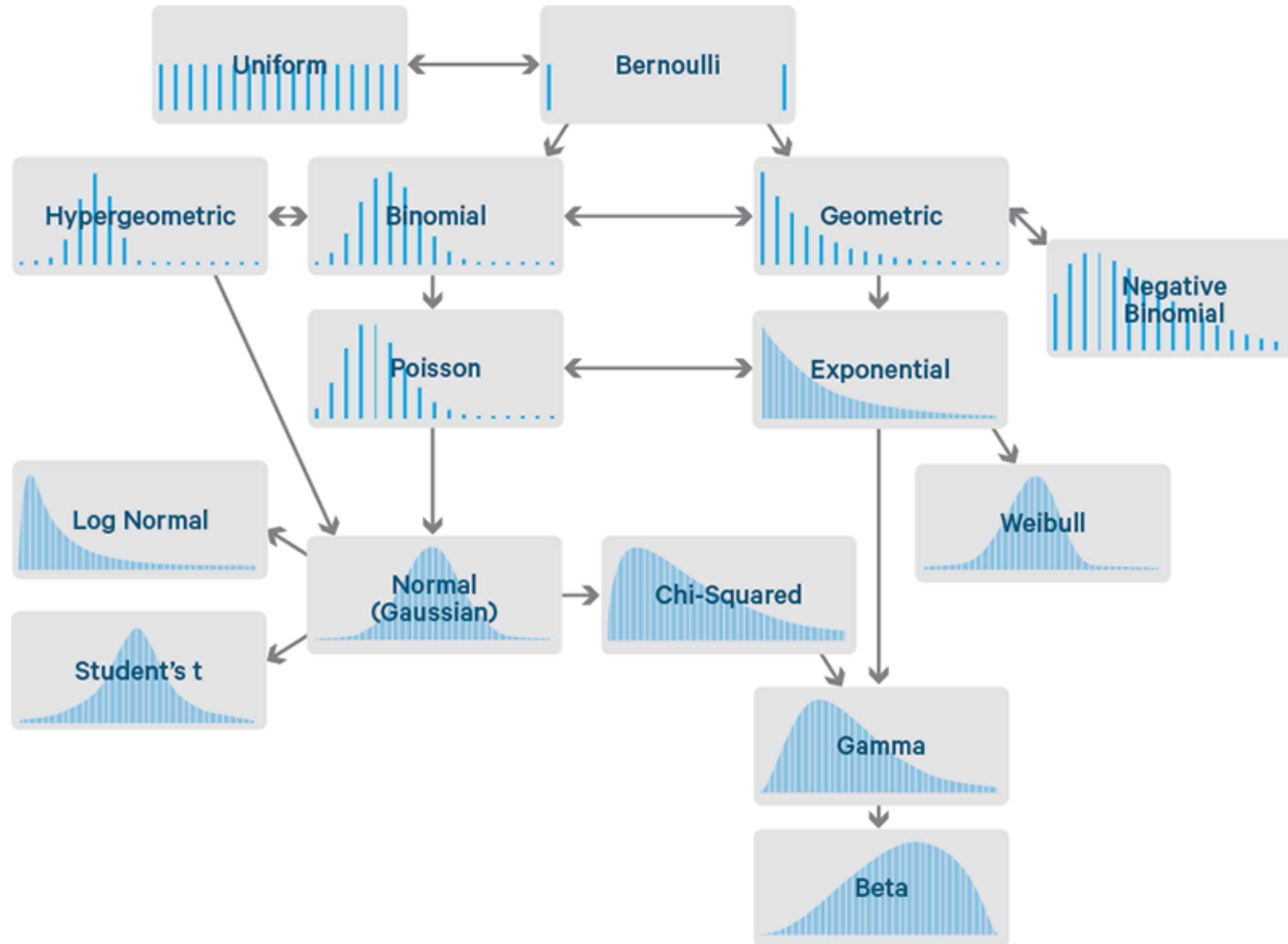
Observation Frequency



Observed ranks of posts by subreddit

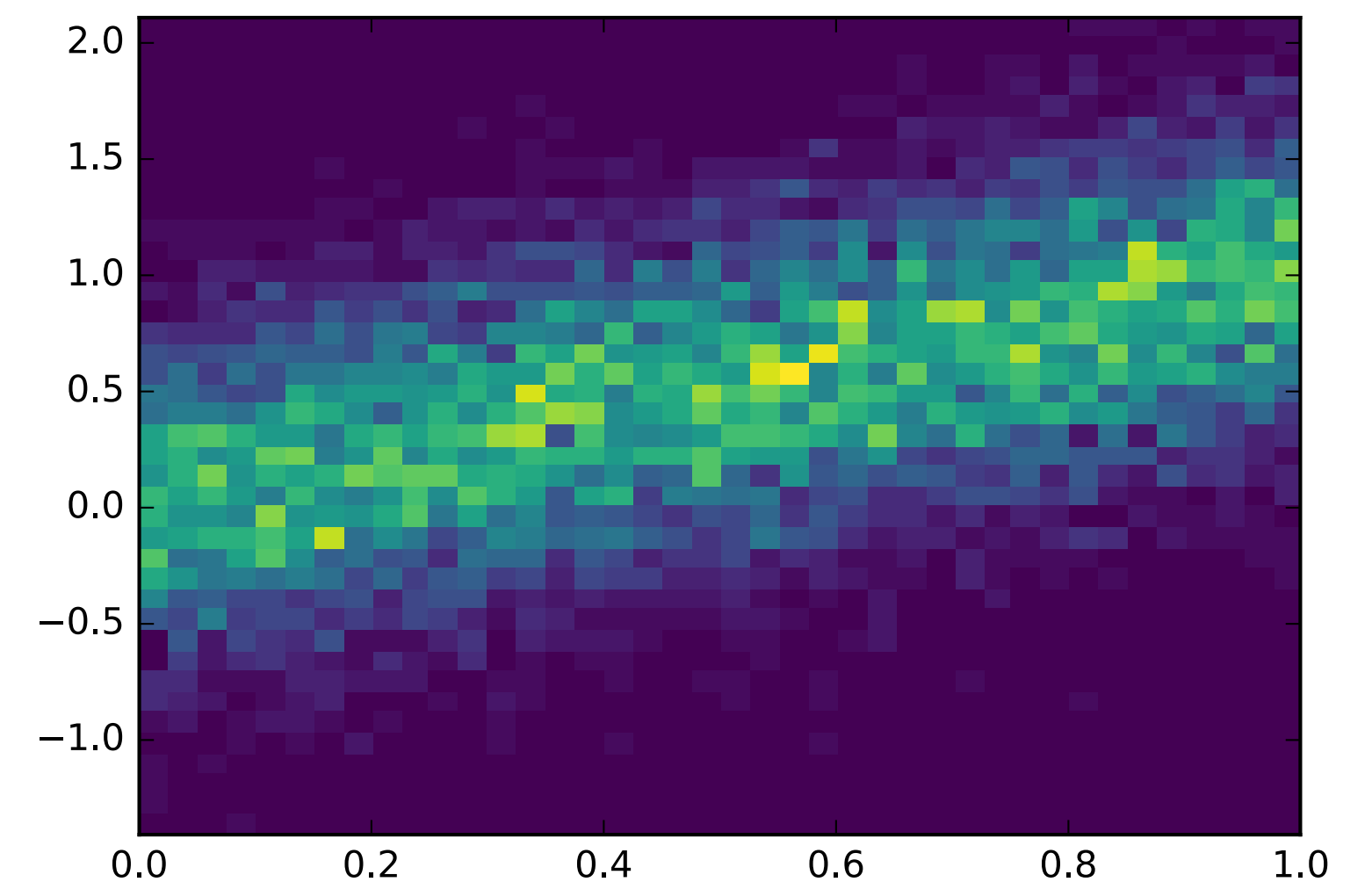
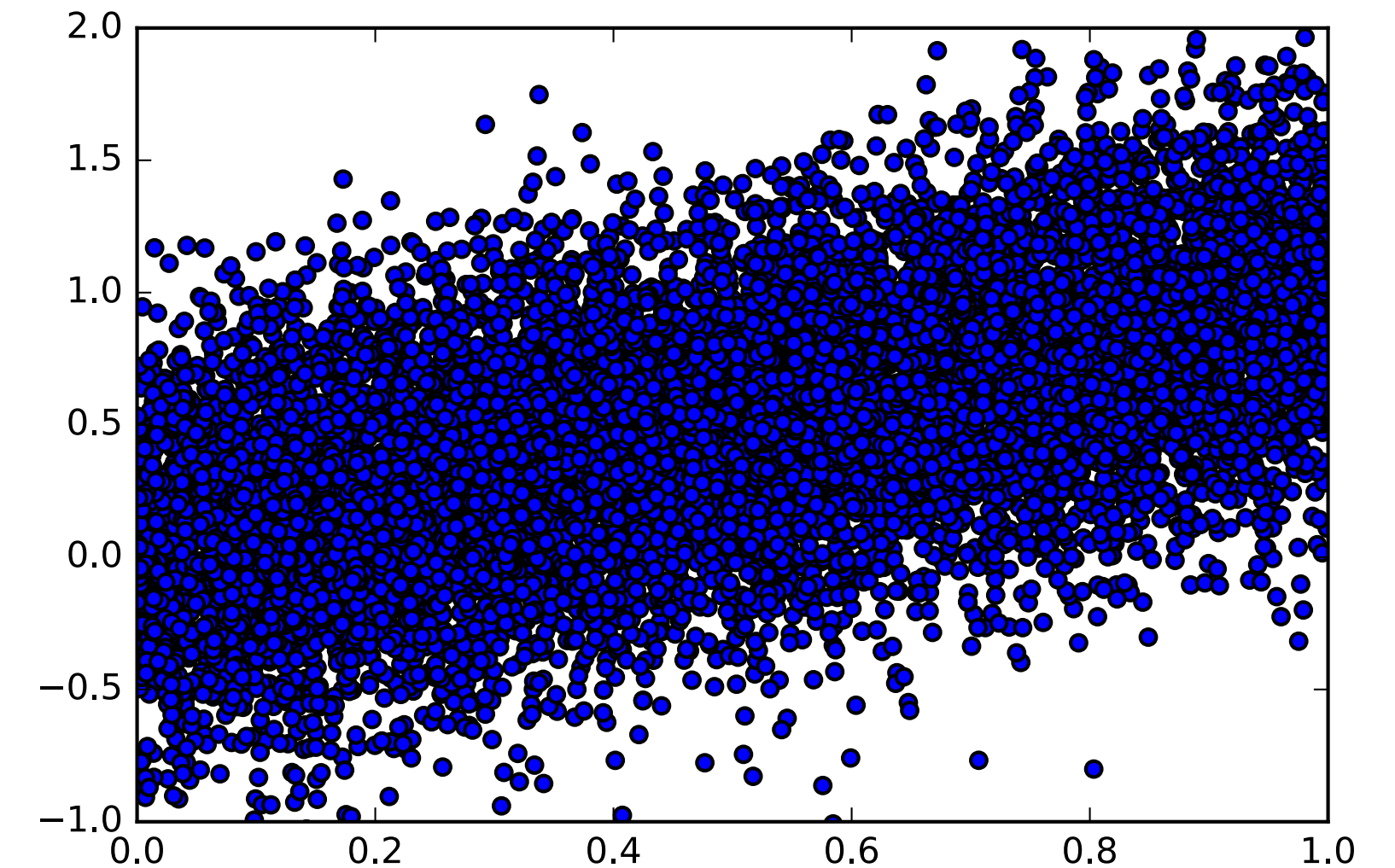
[["The reddit Front Page is Not a Meritocracy"](#), T. W. Schneider]

Common Distributions



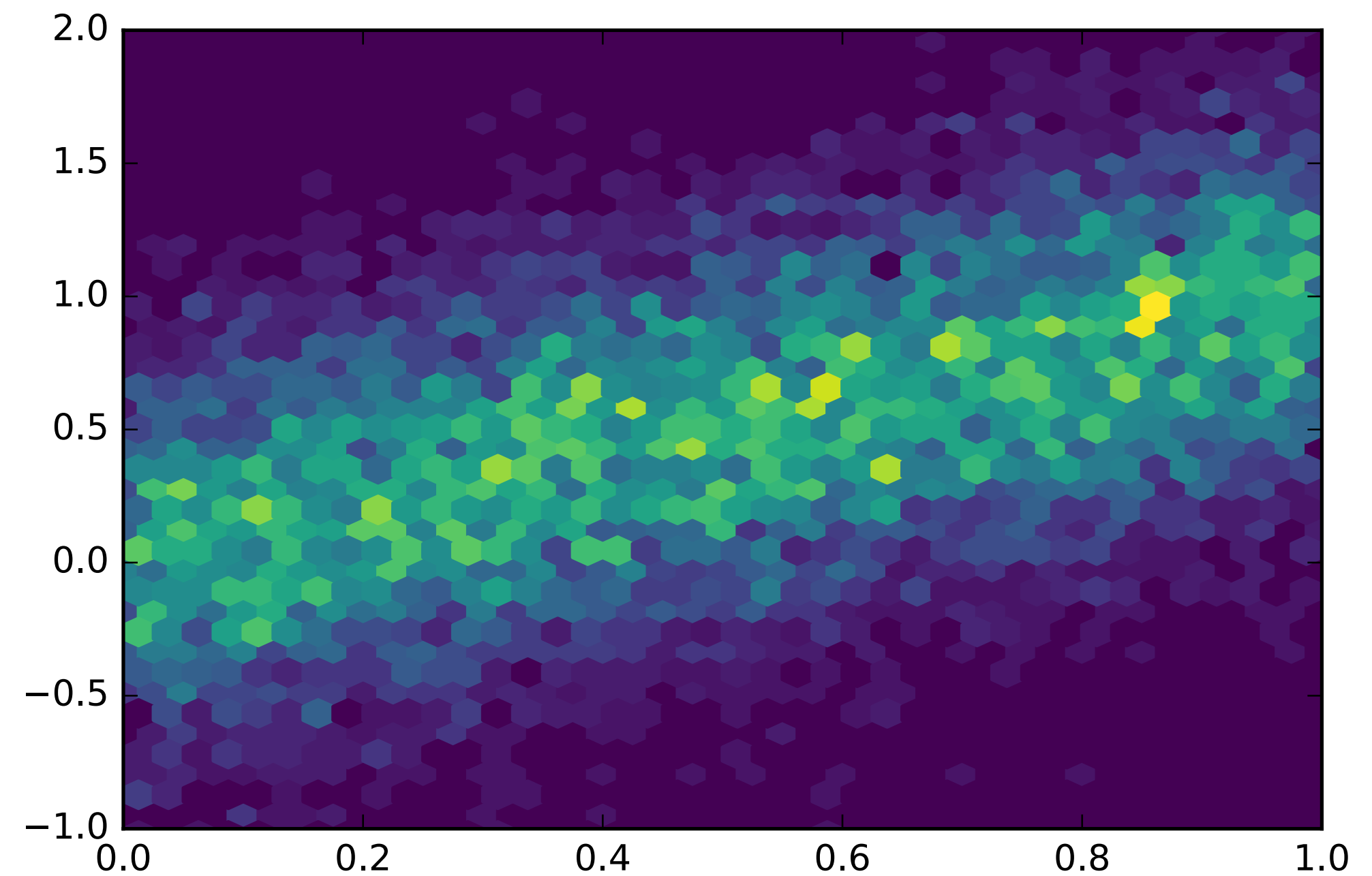
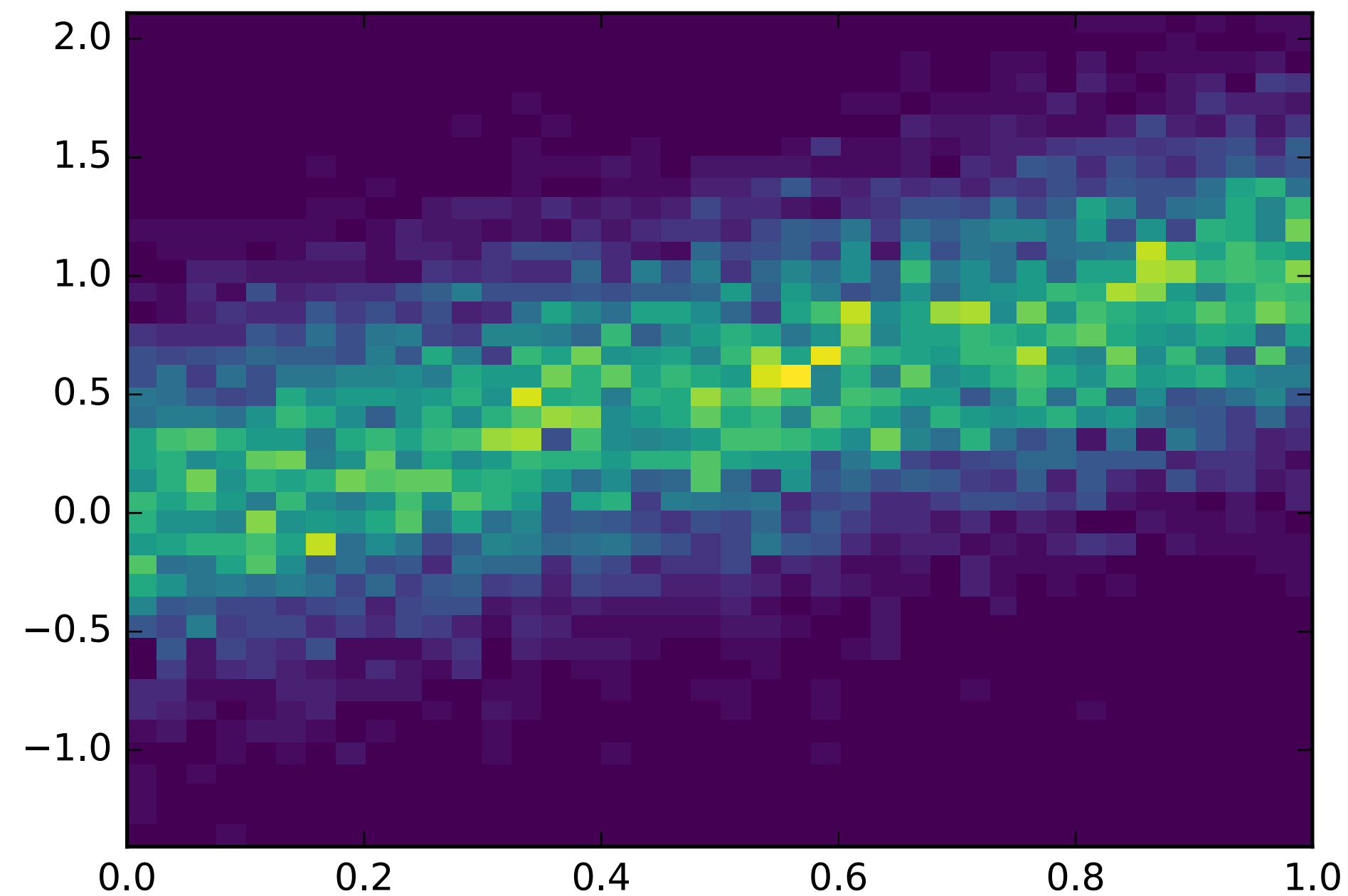
Binning Scatterplots

- 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



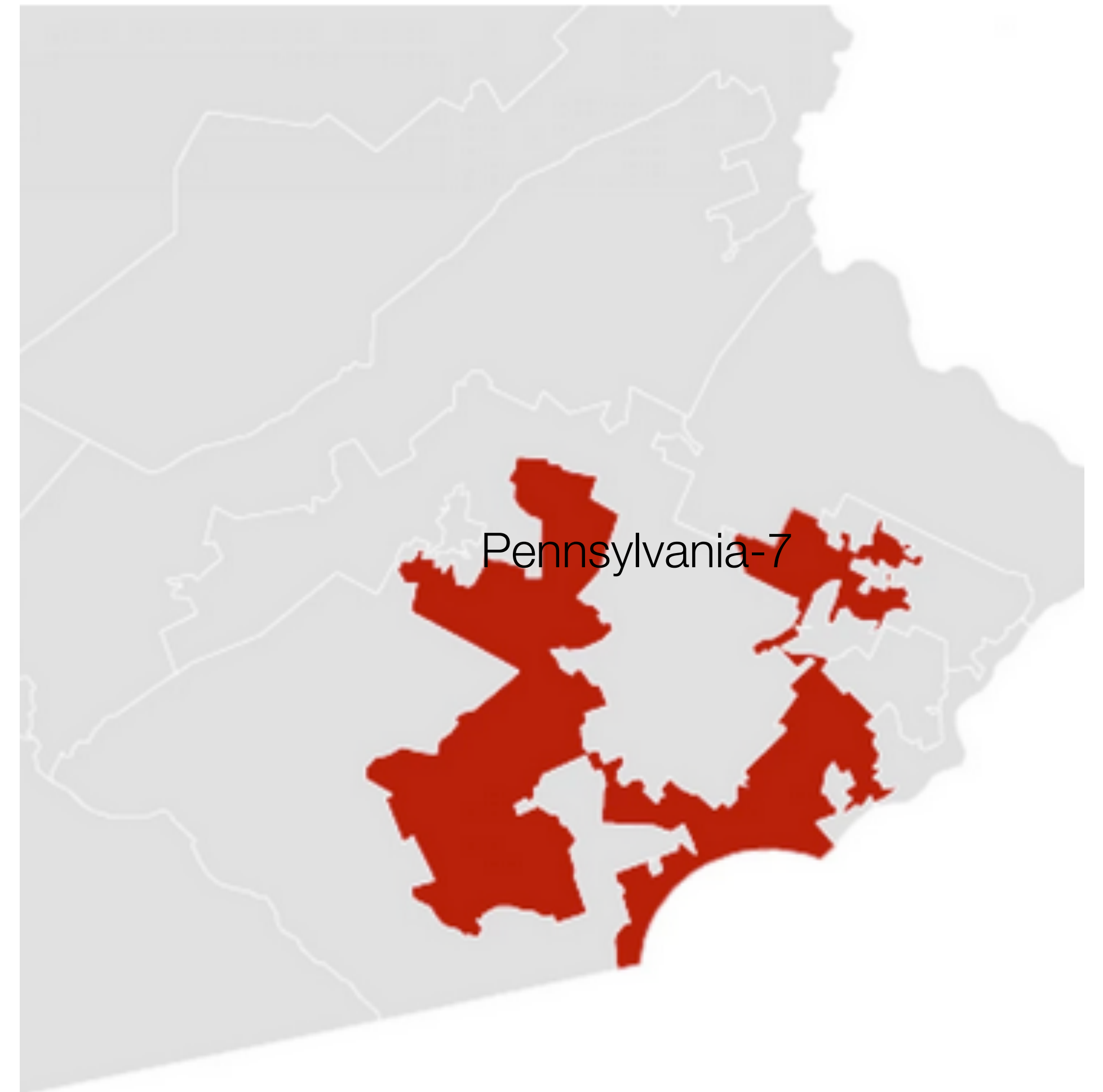
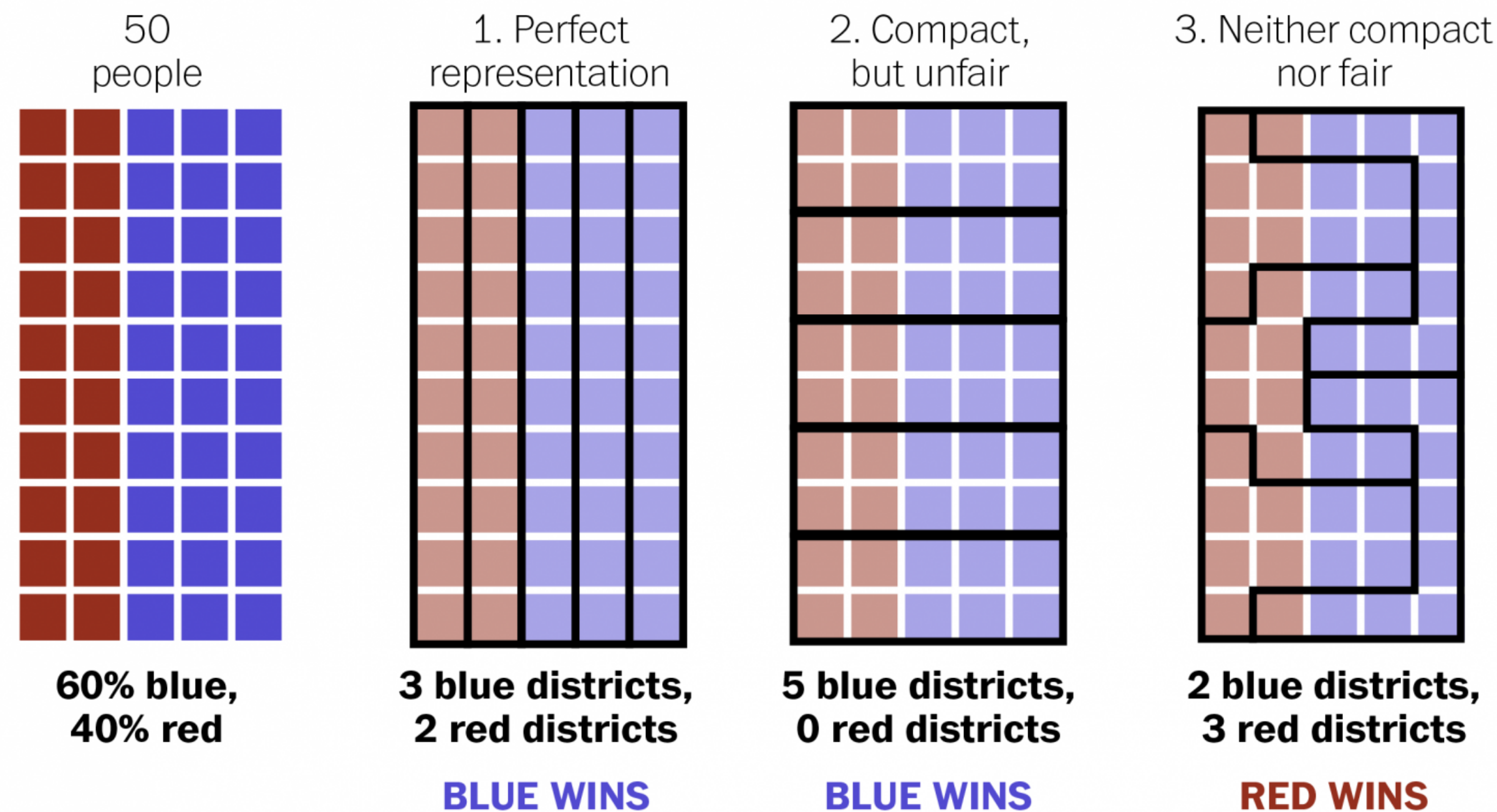
Spatial Aggregation



[Penn State, GEOG 486]

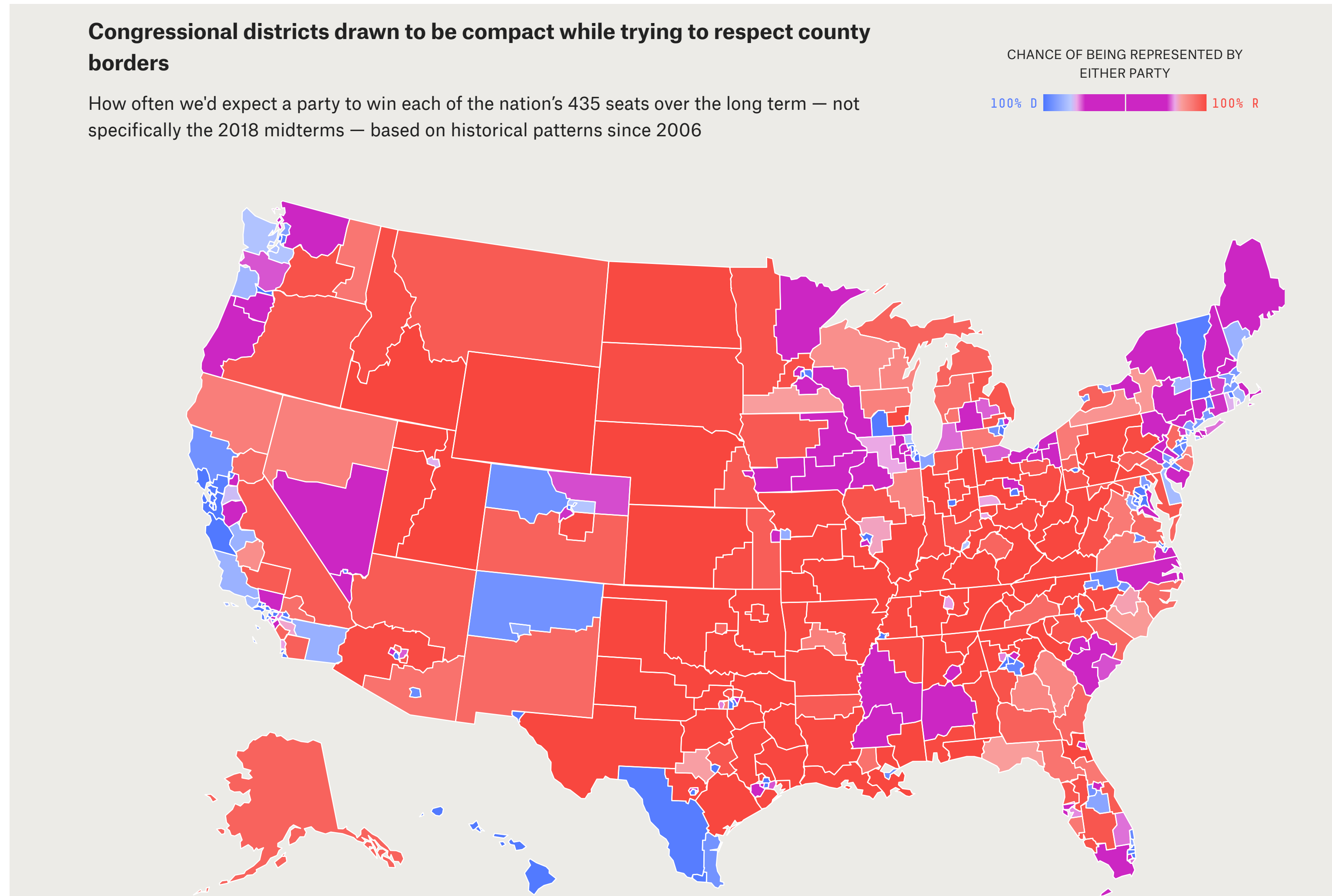
Modifiable Areal Unit Problem

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



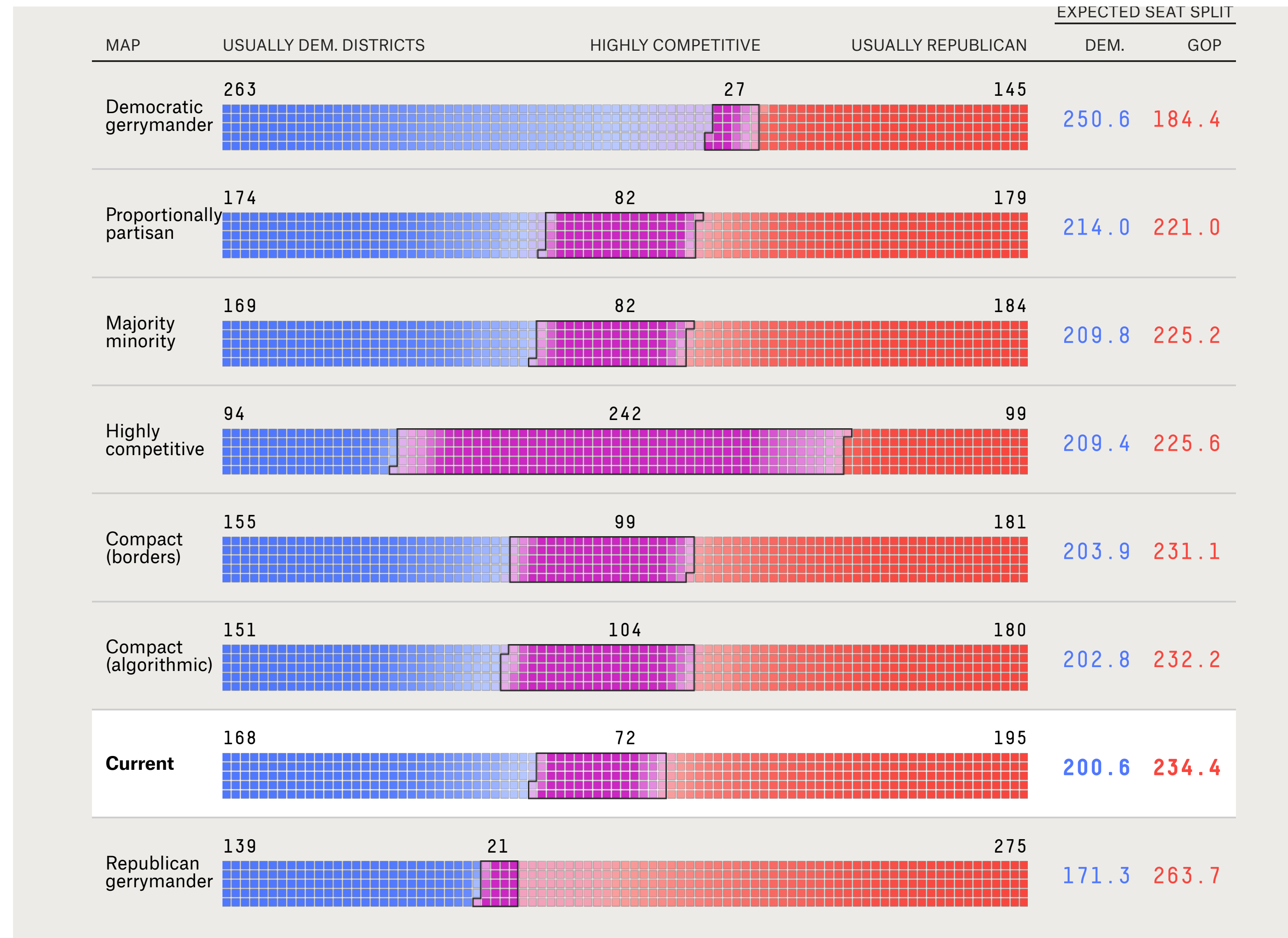
[Wonkblog, Washington Post, Adapted from S. Nass]

Drawing Different Maps: Compactness



[A. Bycoffe et al., 538]

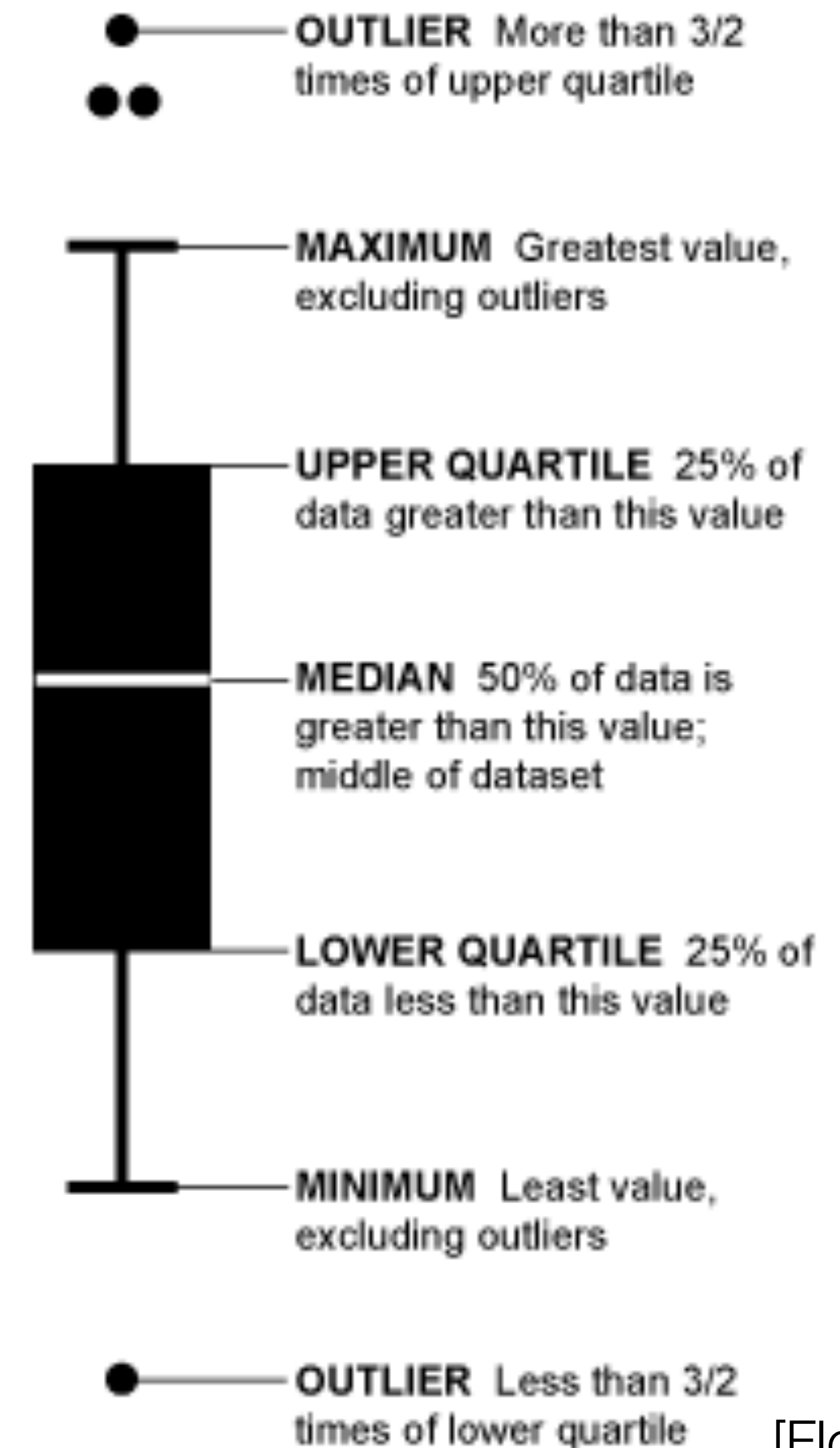
Drawing Different Maps



[A. Bycoffe et al., 538]

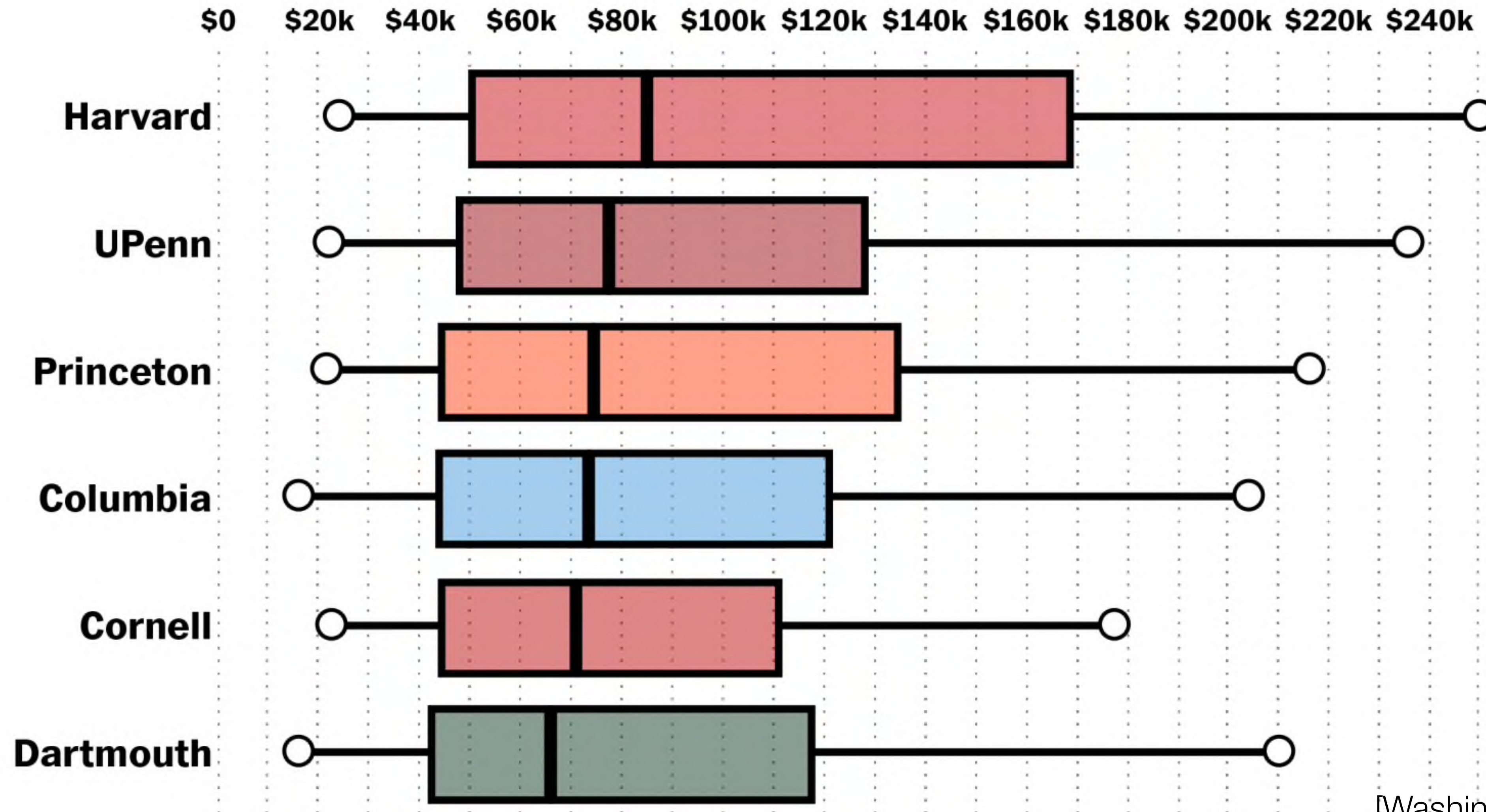
Boxplots

- 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



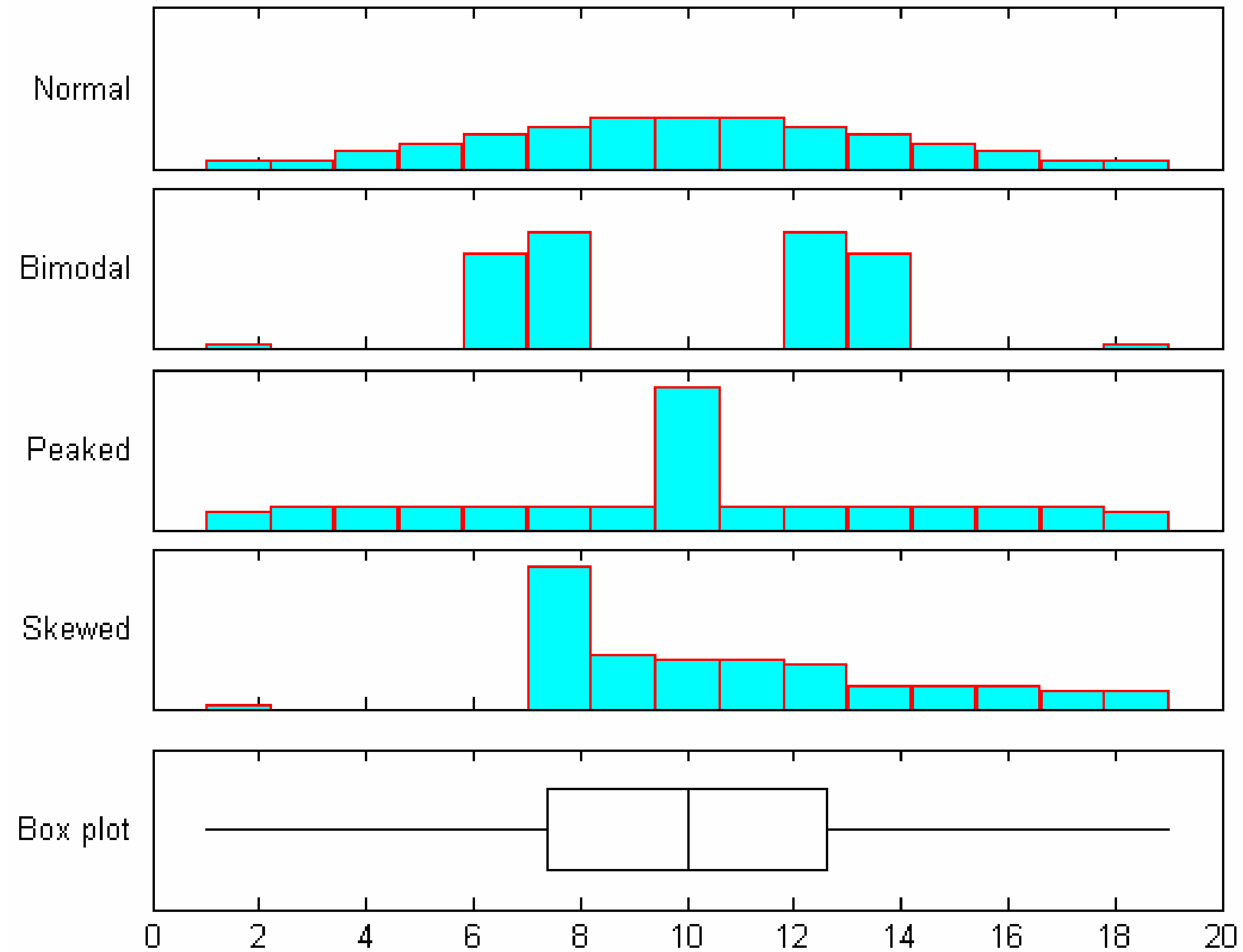
[Flowing Data]

Aggregation: Boxplots



[Washington Post, 2015]

Four Distributions, Same Boxplot...



[C. Choonpradub and D. McNeil, 2005]