

Data Visualization (DSC 530/CIS 602-01)

Tabular Data

Dr. David Koop

D3 Key Features

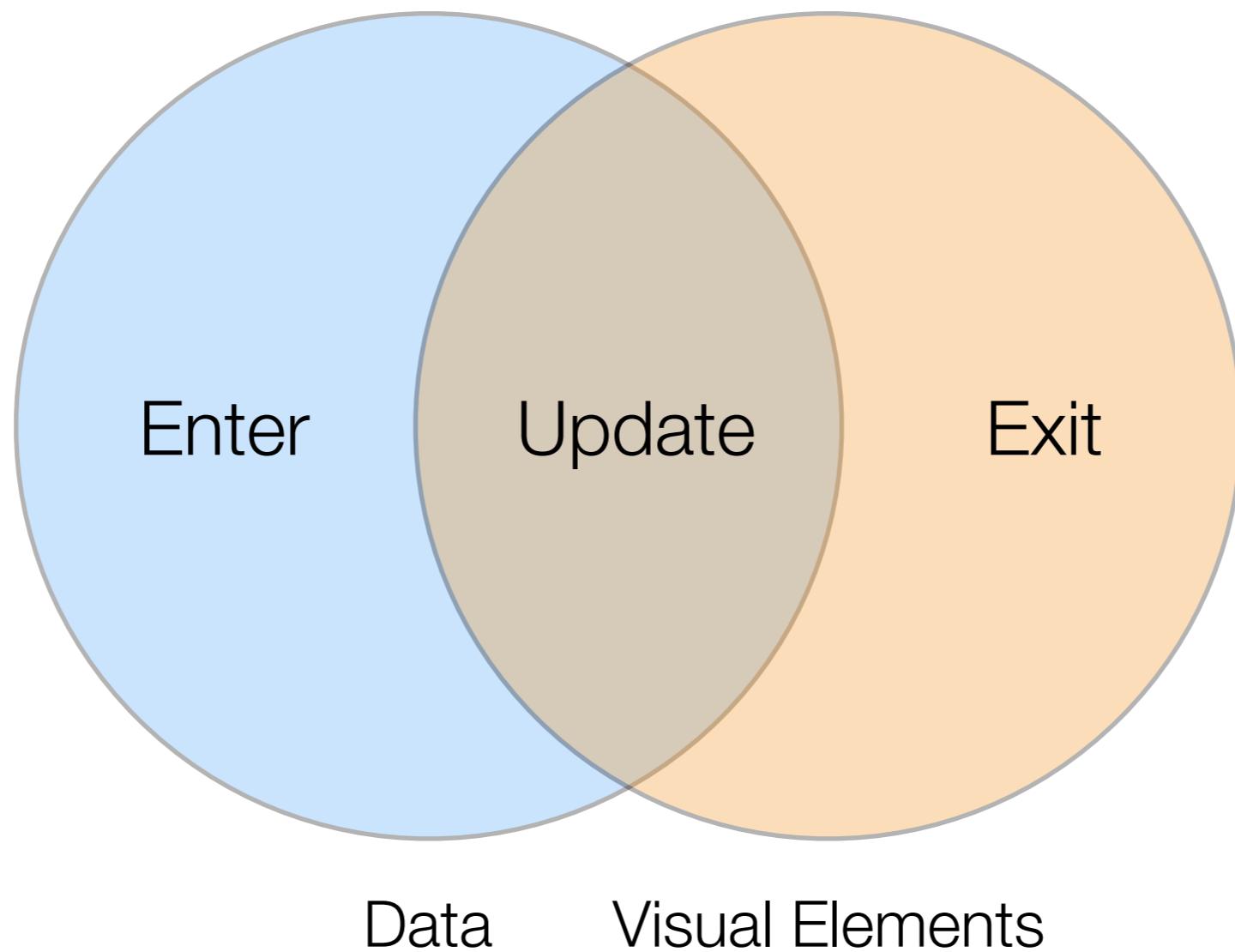
- Supports data as a core piece of Web elements
 - Loading data
 - Dealing with changing data (joins, enter/update/exit)
 - Correspondence between data and DOM elements
- Selections (similar to CSS) that allow greater manipulation
- Method Chaining
- Integrated layout algorithms, axes calculations, etc.
- Focus on interaction support
 - Straightforward support for transitions
 - Event handling support for user-initiated changes

D3 Information and Examples

- Ogievetsky has put together a nice set of interactive examples that show off the major features of D3
- <http://vadim.ogievetsky.com/IntroD3/>
- Other references:
 - Murrary's book on Interactive Data Visualization for the Web
 - The D3 website: d3js.org

D3 Data Joins

- Two groups: data and visual elements
- Three parts of the join between them: enter, update, and exit
- enter: `s.enter()`, update: `s`, exit: `s.exit()`

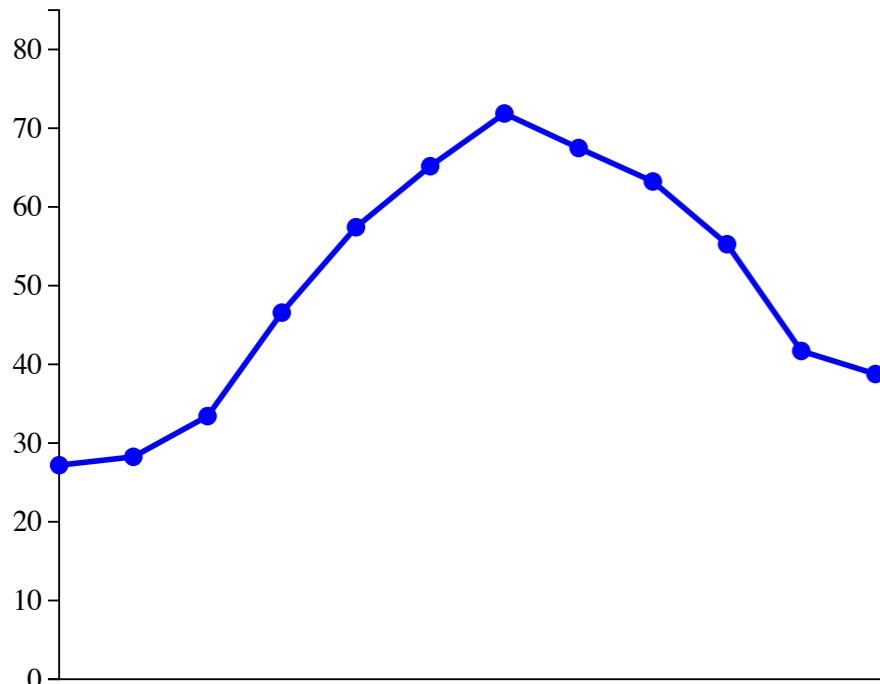


D3 Updates

- Deal with data items with no visual element (enter)
- Deal with visual elements with no data items (exit)
- Update visual elements based on changes in data (update)
- Enter+update:
 - `s.enter().append("rect");
s.attr("height", function(d) { return d; });`
 - Appending to the enter selection adds the new pairs to the update part of the selection!
- Even if the number of data items and visual elements stays the same, a change to the update selection can cause major changes
- Generally end with the same number of data items and visual elements

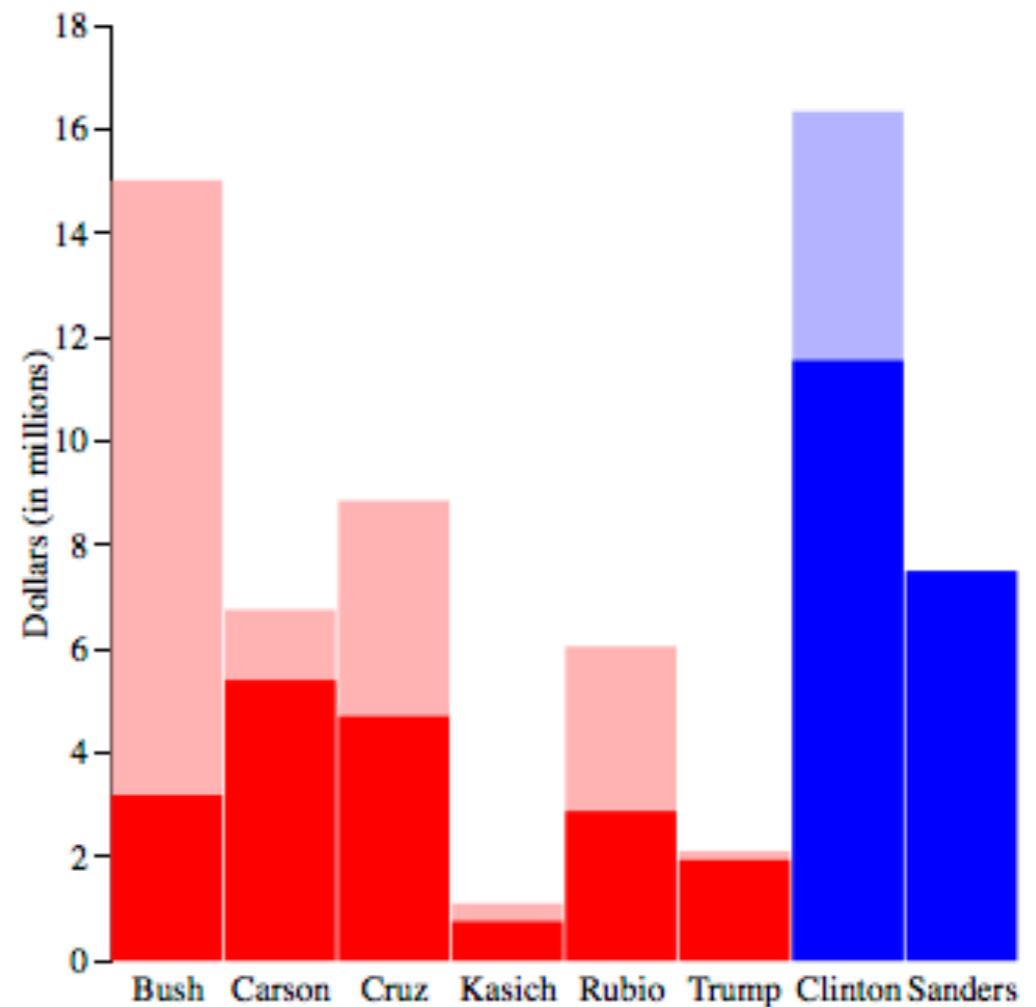
Example: Line Chart in D3

- Part 3b of Assignment 1 in D3
- Start:
<http://codepen.io/dakoop/pen/YwogrP>
- Important Features:
 - Margin Convention
 - Axis Helper
 - Selections and Classes
- Potential Solution:
<http://codepen.io/dakoop/pen/eJwbaW>



Assignment 2

- [www.cis.umassd.edu/~dkoop/
dsc530/assignment2.html](http://www.cis.umassd.edu/~dkoop/dsc530/assignment2.html)
- Use D3!
- 2016 Campaign Finance Data
 - Changes from mid- to end-2015
 - Including SuperPAC receipts
 - Extra Credit: Transitions
- Cheating
- Questions?

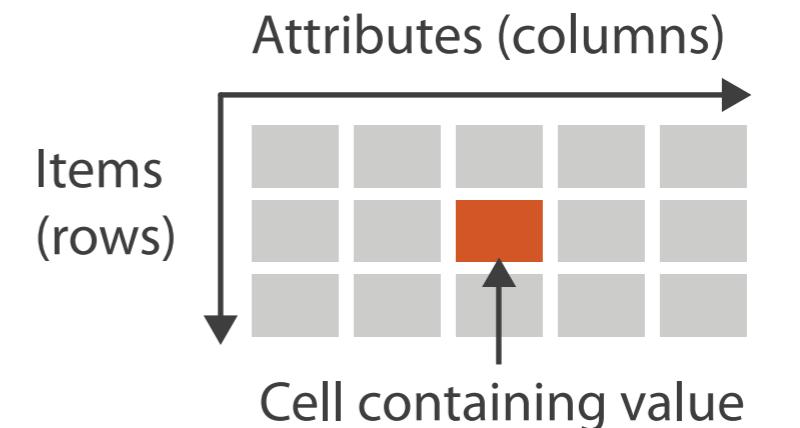


Tables

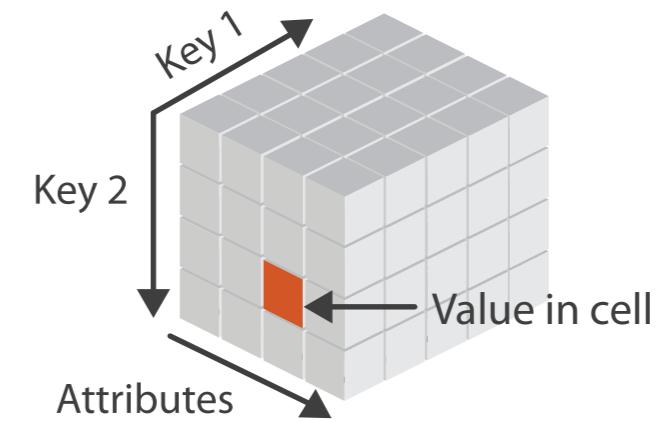
	REMOTE	STATION	FF	SEN/DIS	7-D AFAS UNL	D AFAS/RMF I	JOINT RR TKT	7-D UNL	30-D UNL
1	R011	42ND STREET & 8TH AVENUE	00228985	00008471	00000441	00001455	00000134	00033341	00071255
2	R170	14TH STREET-UNION SQUARE	00224603	00011051	00000827	00003026	00000660	00089367	00199841
3	R046	42ND STREET & GRAND CENTRAL	00207758	00007908	00000323	00001183	00003001	00040759	00096613
4	R012	34TH STREET & 8TH AVENUE	00188311	00006490	00000498	00001279	00003622	00035527	00067483
5	R293	34TH STREET - PENN STATION	00168768	00006155	00000523	00001065	00005031	00030645	00054376
6	R033	42ND STREET/TIMES SQUARE	00159382	00005945	00000378	00001205	00000690	00058931	00078644
7	R022	34TH STREET & 6TH AVENUE	00156008	00006276	00000487	00001543	00000712	00058910	00110466
8	R084	59TH STREET/COLUMBUS CIRCLE	00155262	00009484	00000589	00002071	00000542	00053397	00113966
9	R020	47-50 STREETS/ROCKEFELLER	00143500	00006402	00000384	00001159	00000723	00037978	00090745
10	R179	86TH STREET-LEXINGTON AVE	00142169	00010367	00000470	00001839	00000271	00050328	00125250
11	R023	34TH STREET & 6TH AVENUE	00134052	00005005	00000348	00001112	00000649	00031531	00075040
12	R029	PARK PLACE	00121614	00004311	00000287	00000931	00000792	00025404	00065362
13	R047	42ND STREET & GRAND CENTRAL	00100742	00004273	00000185	00000704	00001241	00022808	00068216
14	R031	34TH STREET & 7TH AVENUE	00095076	00003990	00000232	00000727	00001459	00024284	00038671
15	R017	LEXINGTON AVENUE	00094655	00004688	00000190	00000833	00000754	00020018	00055066
16	R175	8TH AVENUE-14TH STREET	00094313	00003907	00000286	00001144	00000256	00038272	00074661
17	R057	BARCLAYS CENTER	00093804	00004204	00000454	00001386	00001491	00039113	00068119
18	R138	WEST 4TH ST-WASHINGTON SO	00093562	00004677	00000251	00000965	00000127	00031628	00074458

Visualization of Tables

- Items and attributes
- For now, attributes are not known to be positions
- Keys and values
 - **key** is an independent attribute that is unique and identifies item
 - **value** tells some aspect of an item
- Keys: categorical/ordinal
- Values: +quantitative
- Levels: unique *values* of categorical or ordered attributes



→ *Multidimensional Table*



[Munzner (ill. Maguire), 2014]

Arrange Tables

Arrange Tables

④ Express Values



④ Separate, Order, Align Regions

→ Separate



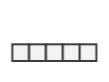
→ Order



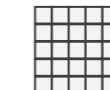
→ Align



→ 1 Key
List



→ 2 Keys
Matrix



→ 3 Keys
Volume

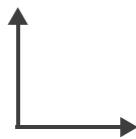


→ Many Keys
Recursive Subdivision



④ Axis Orientation

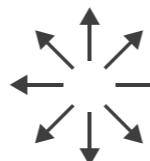
→ Rectilinear



→ Parallel

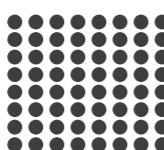


→ Radial



④ Layout Density

→ Dense

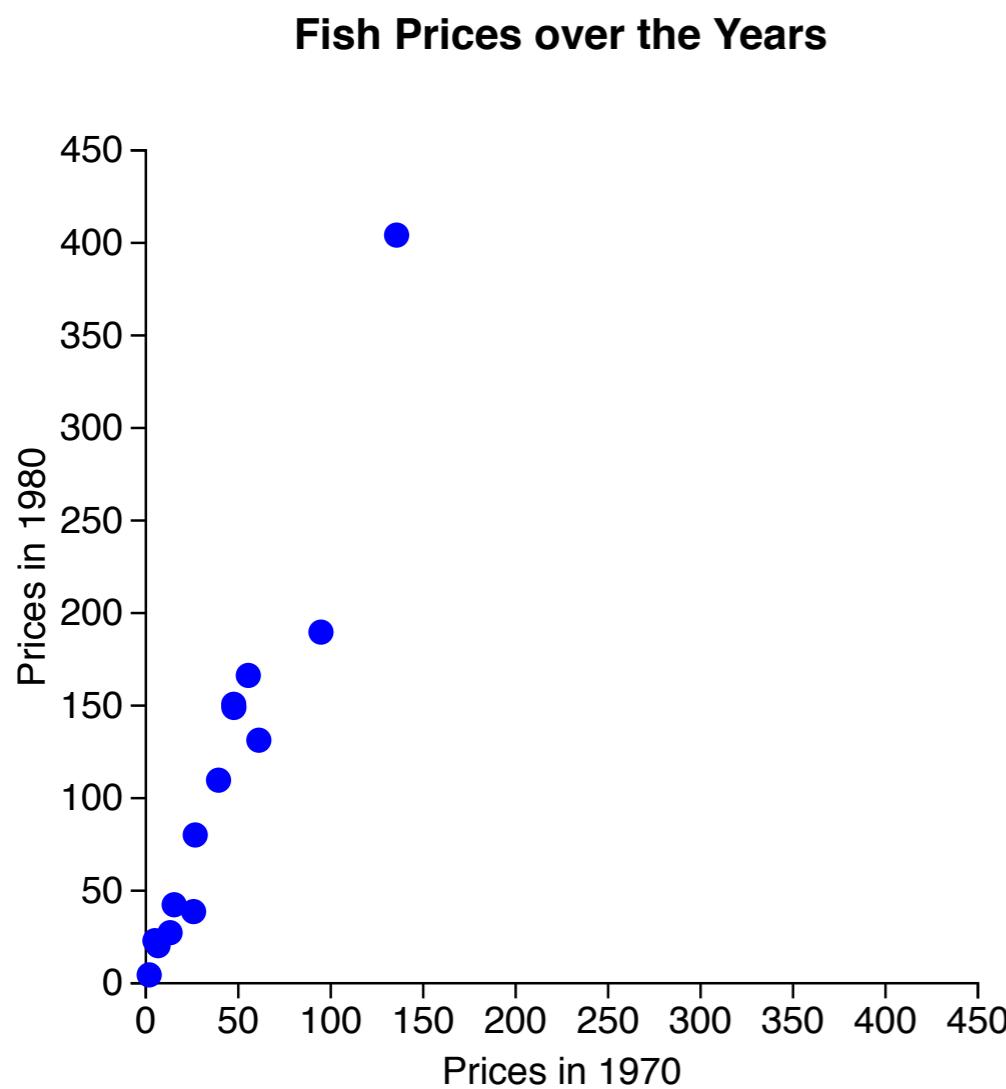


→ Space-Filling



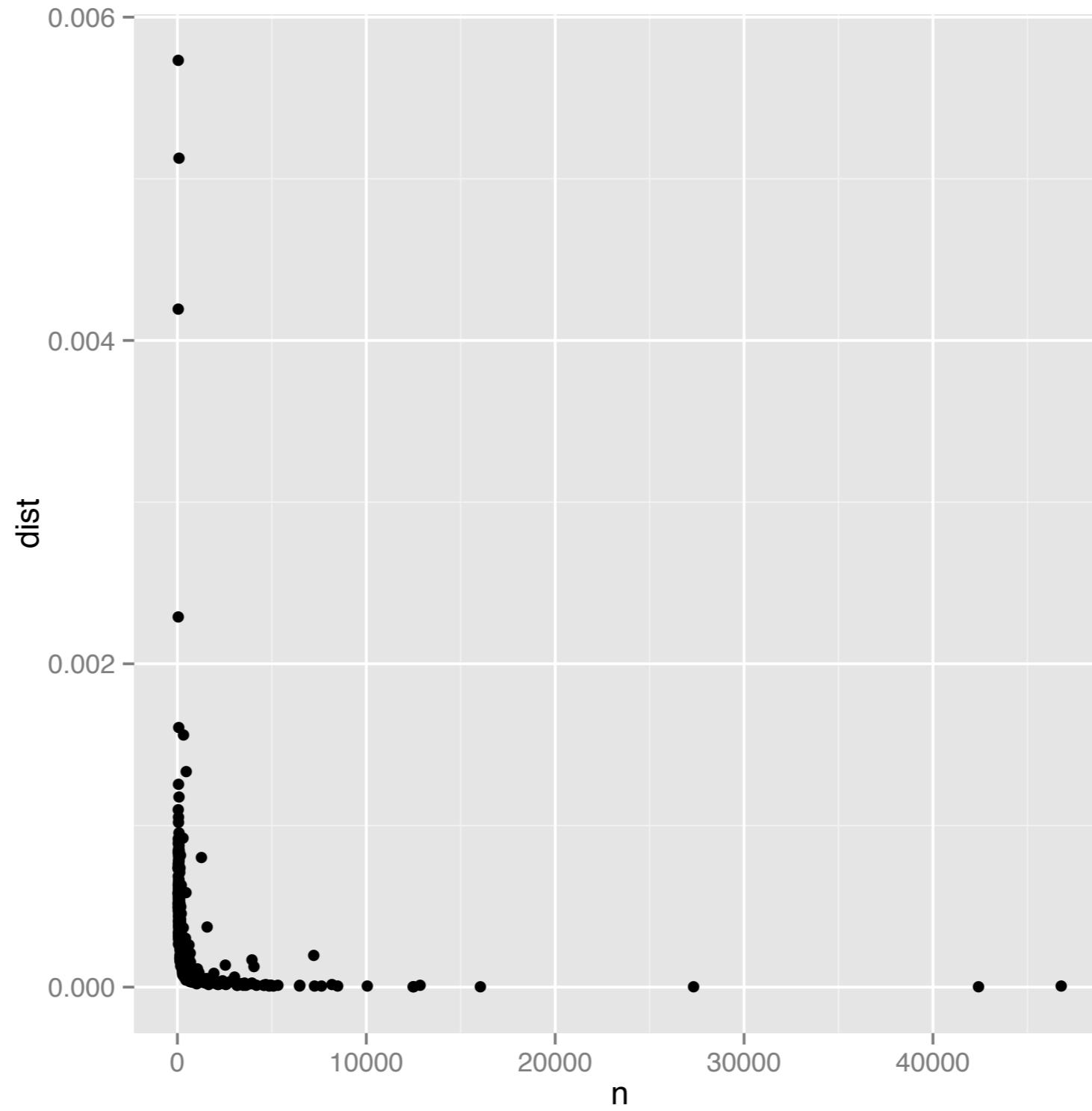
[Munzner (ill. Maguire), 2014]

Express Values: Scatterplots



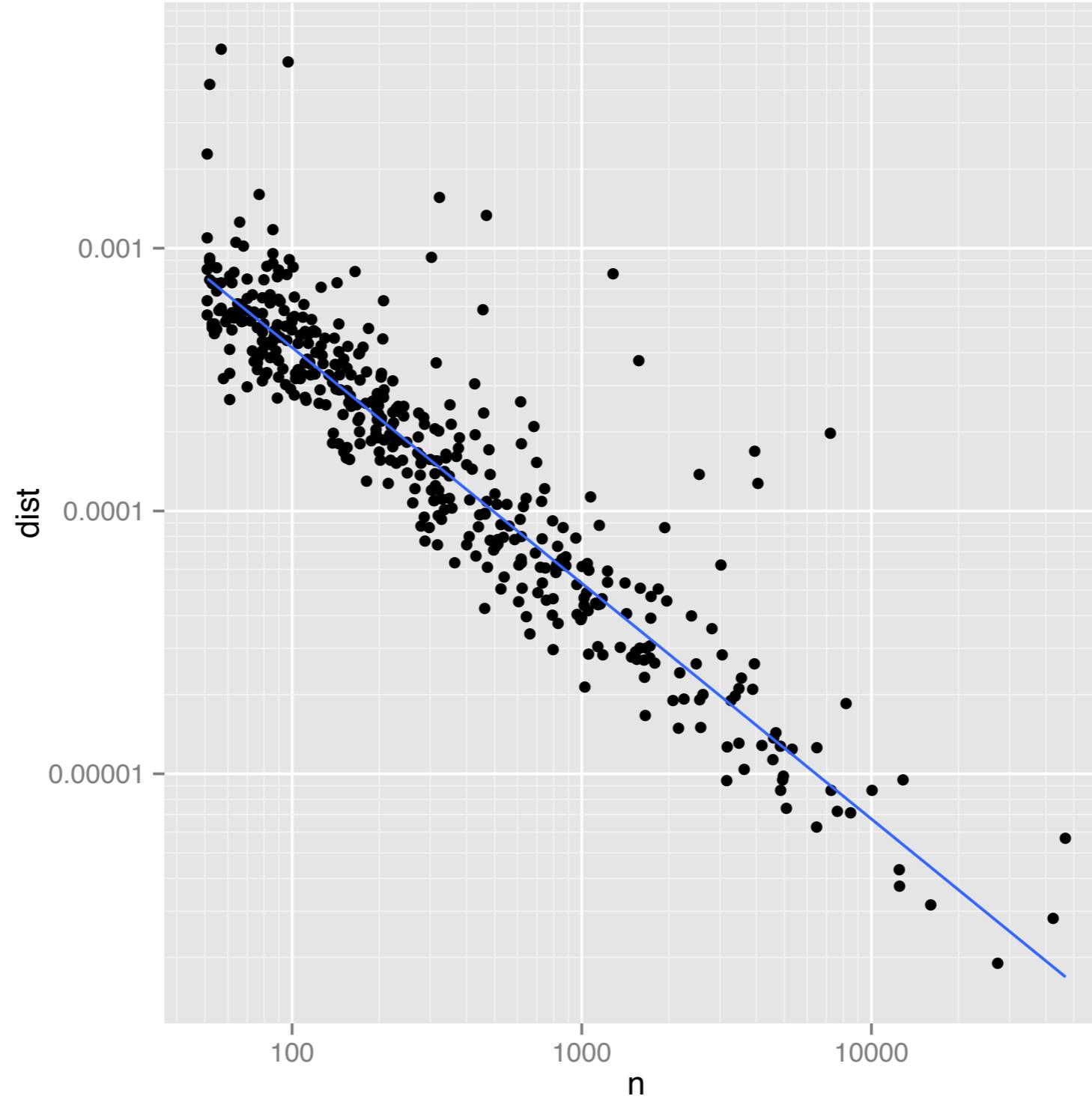
- Data: two quantitative values
- Task: find trends, clusters, outliers
- How: marks at spatial position in horizontal and vertical directions
- Correlation: dependence between two attributes
 - Positive and negative correlation
 - Indicated by lines
- Coordinate system (axes) and labels are important!

Coordinate Systems



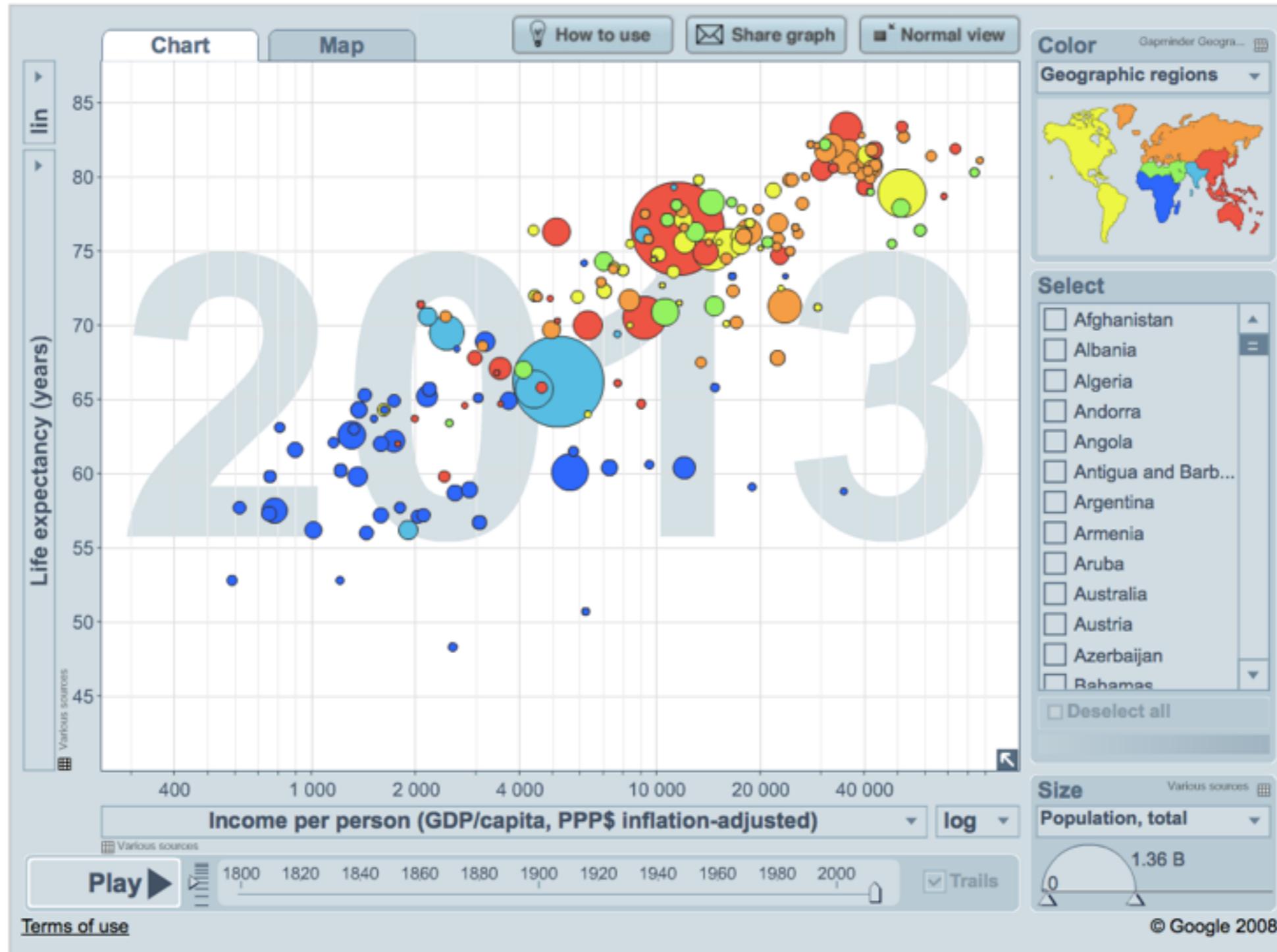
[Wickham, 2014]

Log-Log Plot



[Wickham, 2014]

Bubble Plot



[Gapminder, Wealth & Health of Nations]

Scatterplot

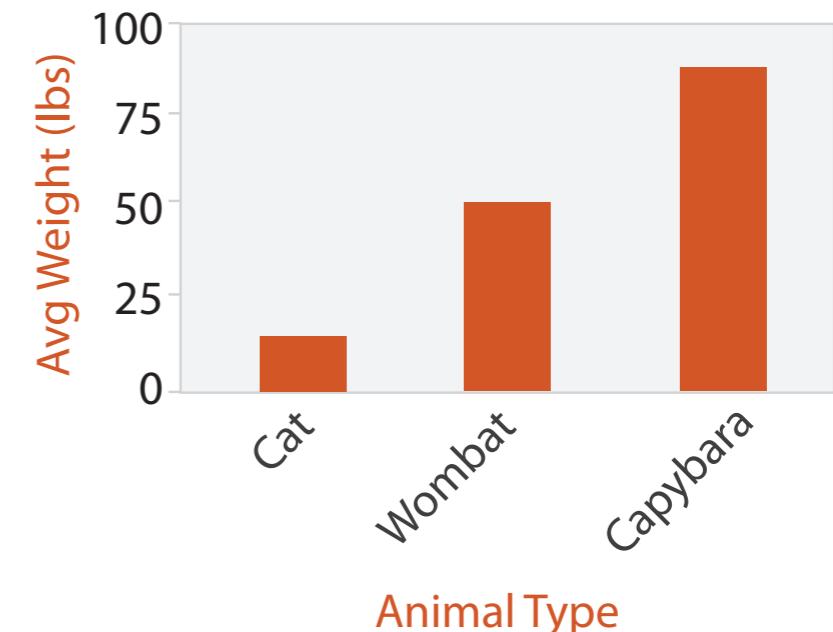
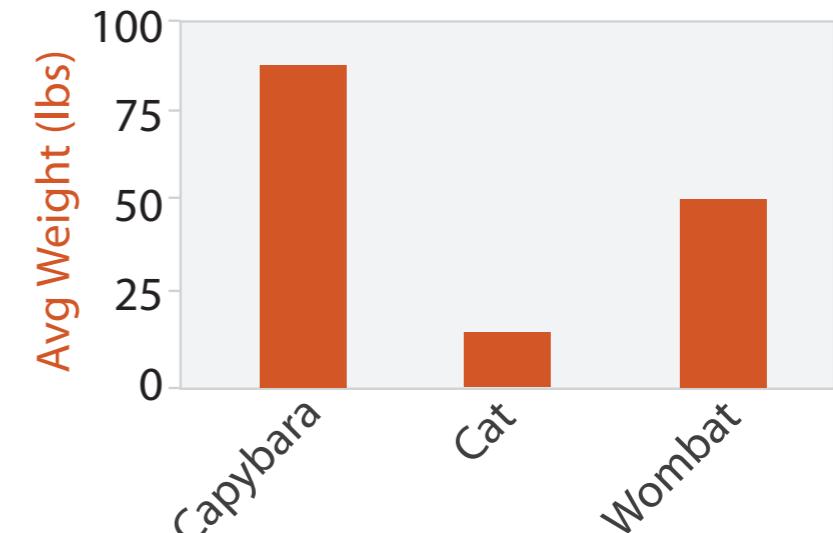
- Data: two quantitative values
 - Task: find trends, clusters, outliers
 - How: marks at spatial position in horizontal and vertical directions
 - **Scalability:** hundreds of items
-
- Cool recent result from Harrison et al., "Ranking Visualizations of Correlation Using Weber's Law", 2014:
 - Correlation perception can be modeled via Weber's Law
 - Scatterplots are one of the best visualizations for both positive and negative correlation
 - Further analysis: M. Kay and J. Heer, "Beyond Weber's Law", 2015

Separate, Order, and Align: Categorical Regions

- Categorical: `=`, `!=`
- Spatial position can be used for categorical attributes
- Use **regions**, distinct contiguous bounded areas, to encode categorical attributes
- Three operations on the regions:
 - Separate (use categorical attribute)
 - Align (use some other ordered attribute)
 - Order
- Alignment and order can use same or different attribute

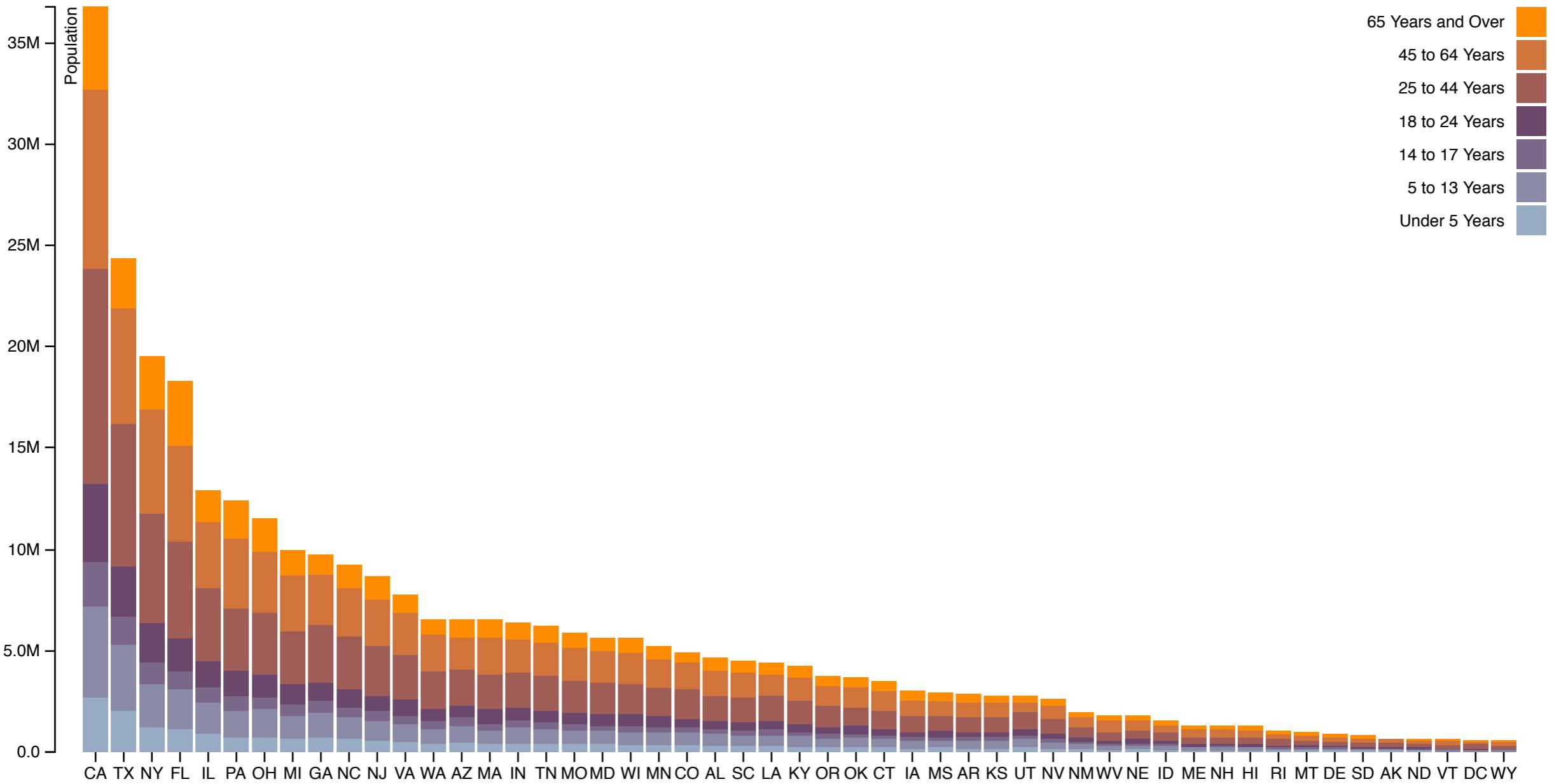
List Alignment: Bar Charts

- Data: one quantitative attribute, one categorical attribute
- Task: lookup & compare values
- How: line marks, vertical position (quantitative), horizontal position (categorical)
- What about **length**?
- Ordering criteria: alphabetical or using quantitative attribute
- Scalability: distinguishability
 - bars at least one pixel wide
 - hundreds



[Munzner (ill. Maguire), 2014]

Stacked Bar Charts



[Bostock, 2012]

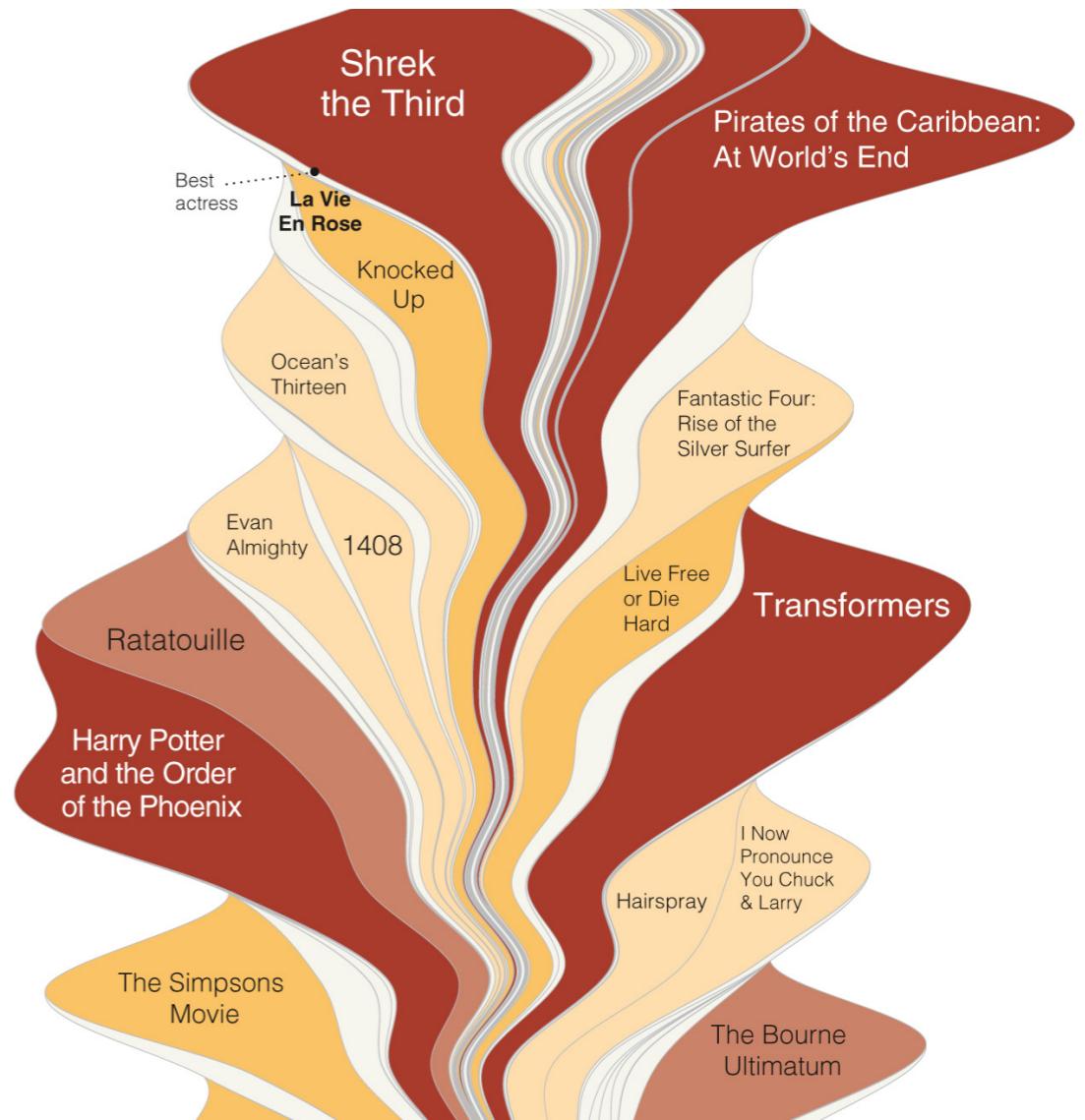
Stacked Bar Charts

- Data: multidimensional table: one quantitative, **two** categorical
- Task: lookup values, part-to-whole relationship, trends
- How: line marks: position (both horizontal & vertical),
subcomponent line marks: length, color
- Scalability: main axis (hundreds like bar chart), bar classes (<12)

- Orientation: vertical or horizontal (swap how horizontal and vertical
position are used.)

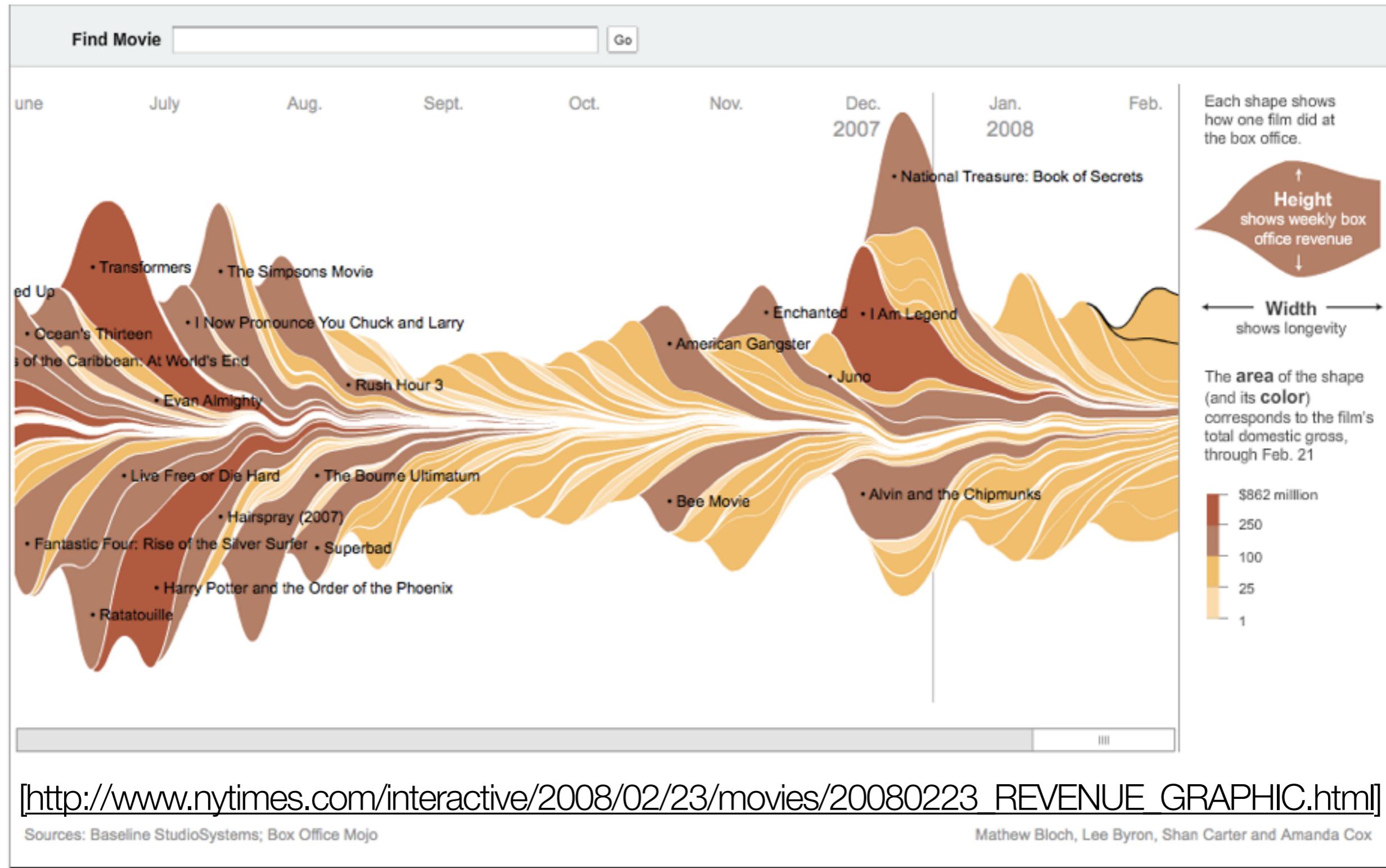
Streamgraphs

- Include a time attribute
- Data: multidimensional table, one quantitative attribute (count), one ordered key attribute (time), one categorical key attribute
- + derived attribute: layer ordering (quantitative)
- Task: analyze trends in time, find (maxmial) outliers
- How: derived position+geometry, length, color
- Scalability: more categories than stacked bar charts

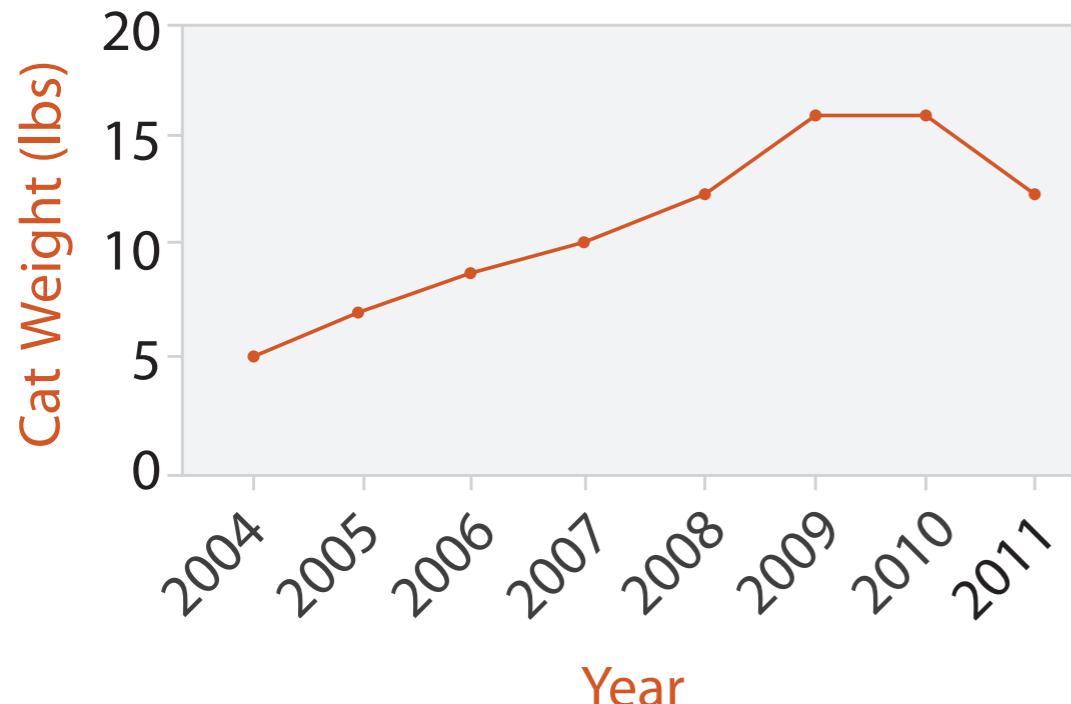
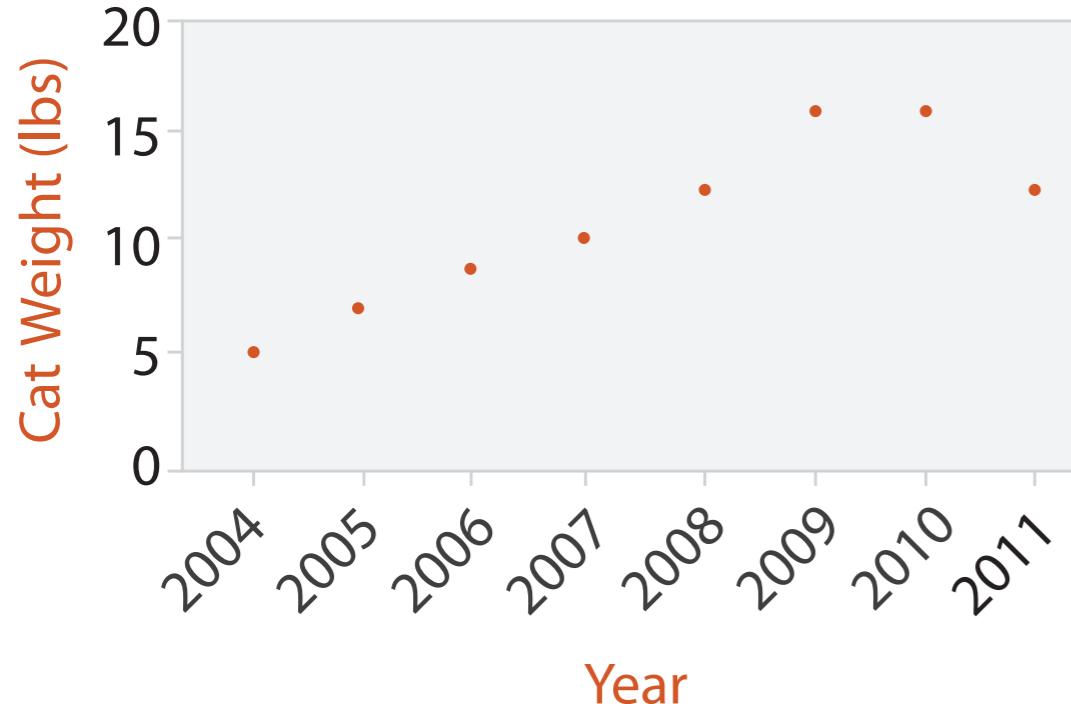


[Byron and Wattenberg, 2012]

NYTimes “Ebb and Flow of Movies”



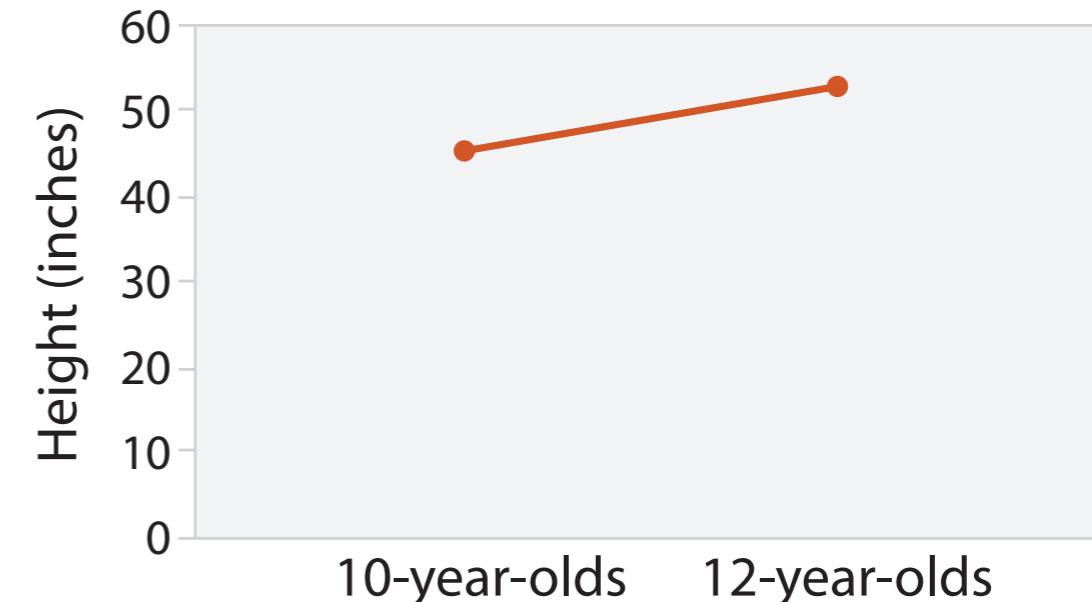
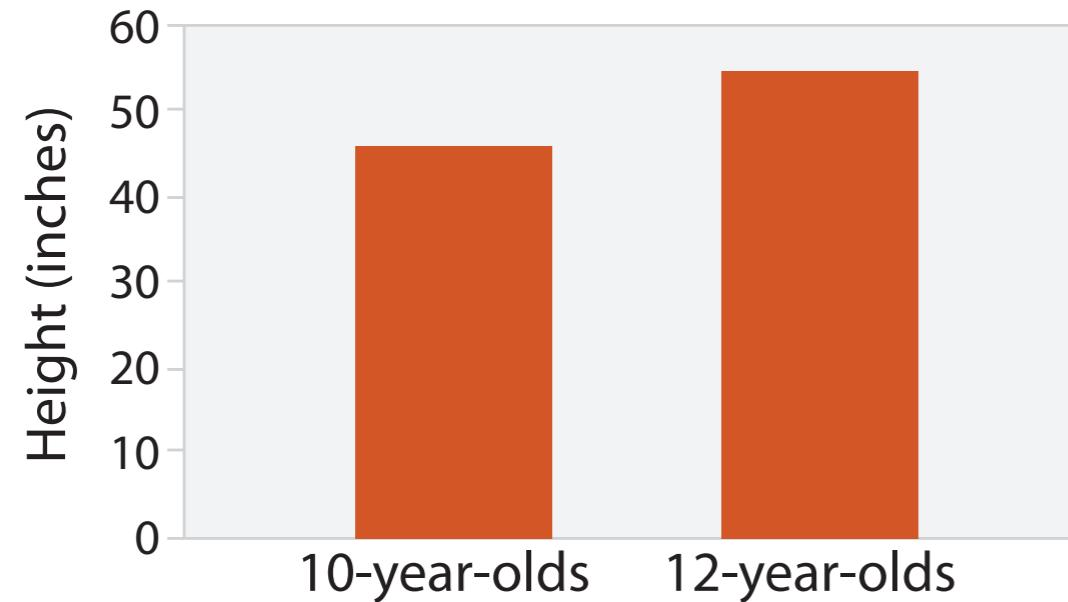
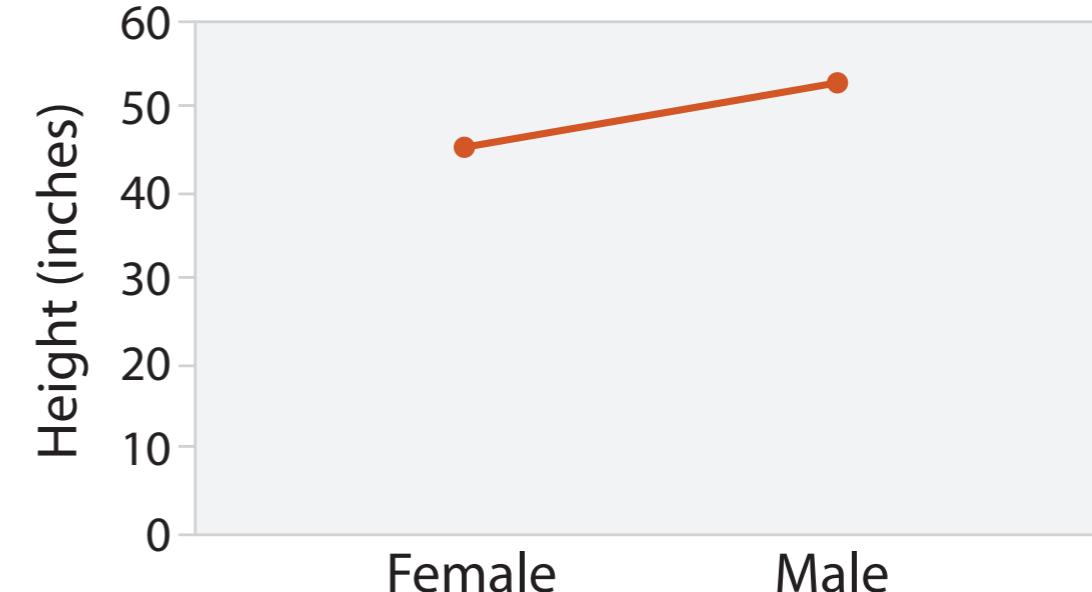
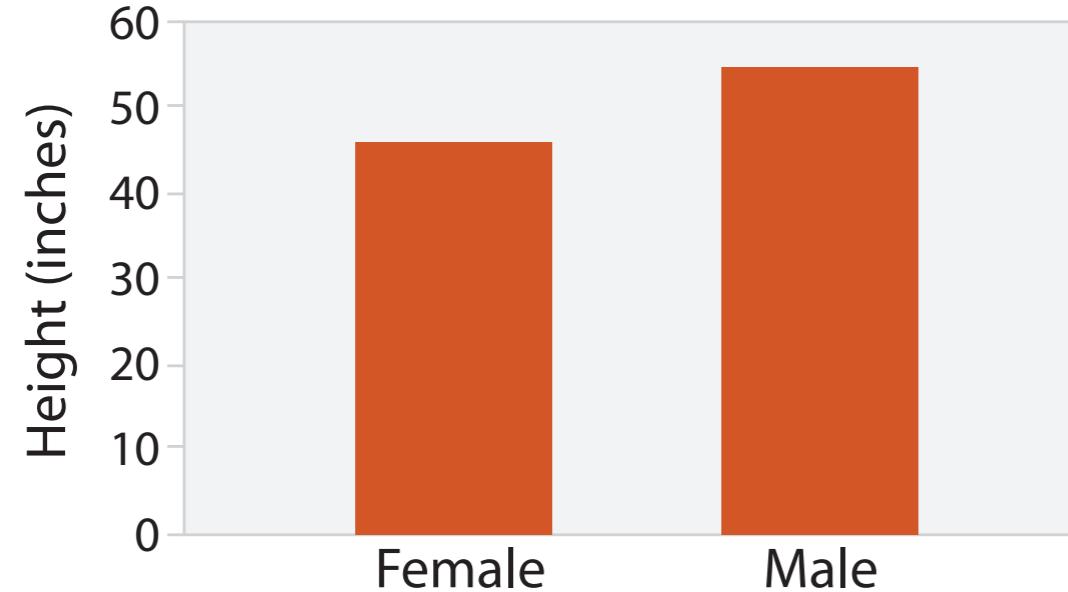
Dot and Line Charts



- Data: one quantitative attribute, one **ordered** attribute
- Task: lookup values, find outliers and trends
- How: point mark and positions
- Line Charts: add **connection mark** (line)
- Similar to scatterplots but allow ordered attribute

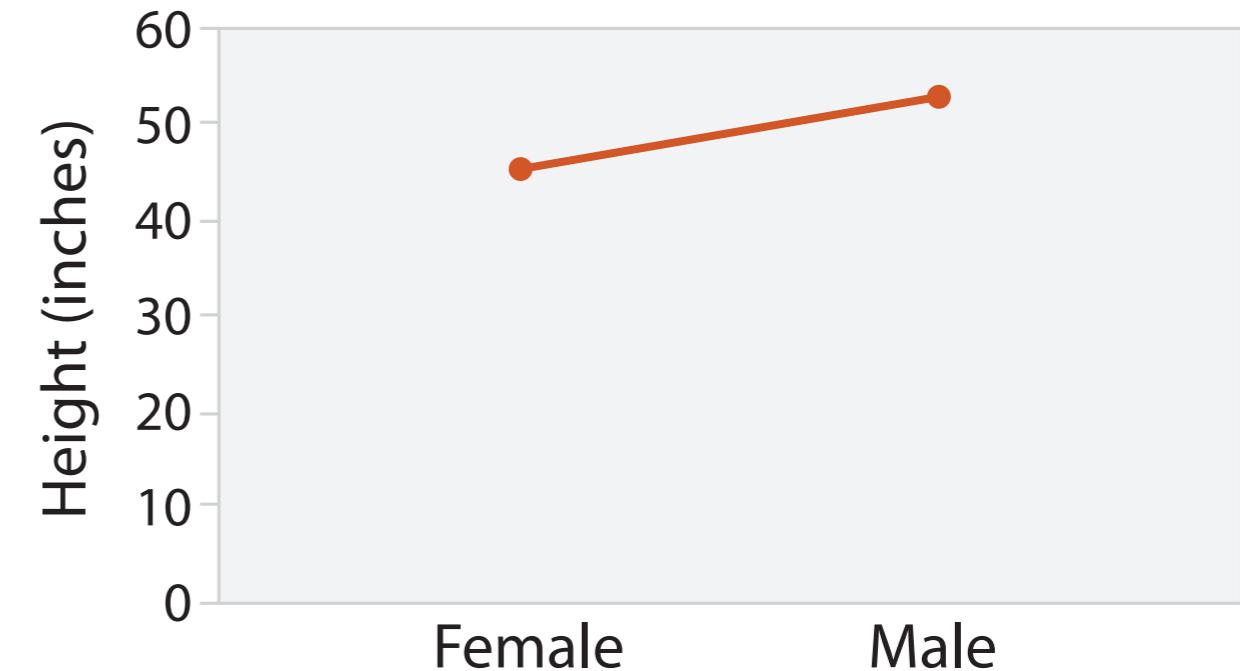
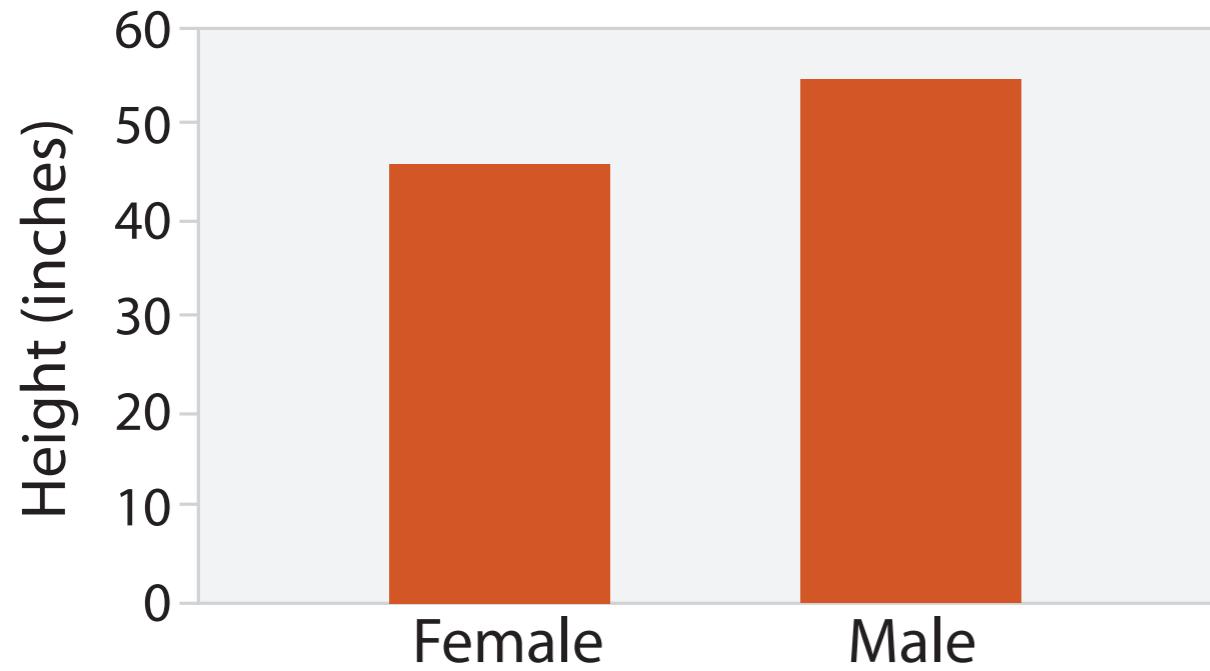
[Munzner (ill. Maguire), 2014]

Proper Use of Line and Bar Charts



[Zacks and Tversky, 1999, Munzner (ill. Maguire), 2014]

Proper Use of Line and Bar Charts



[Zacks and Tversky, 1999, Munzner (ill. Maguire), 2014]

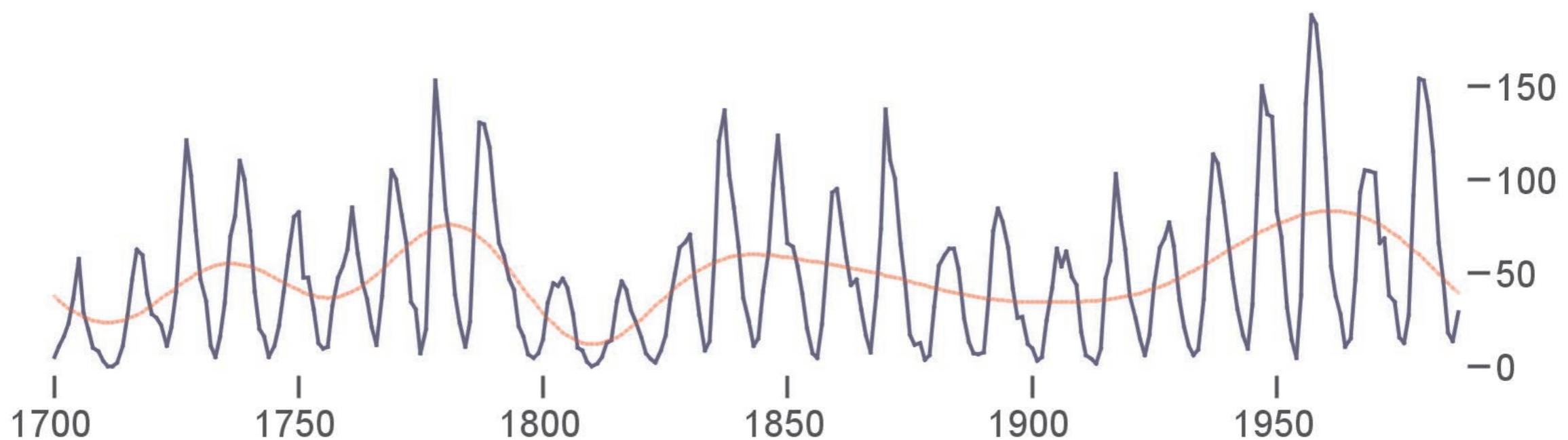
Aspect Ratio

- Trends in line charts are more apparent because we are using angle as a channel
- Perception of angle (and the **relative difference** between angles) is important
- Initial experiments found people best judge differences in **slope** when angles are around 45 degrees (Cleveland et al., 1988, 1993)

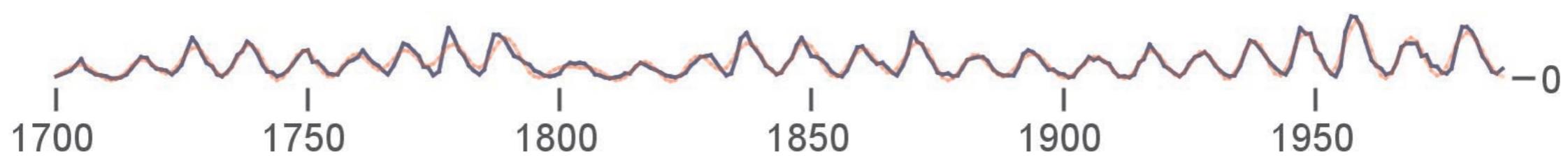
Multiscale Banking

Sunspot Cycles

Aspect Ratio = 3.96



Aspect Ratio = 22.35

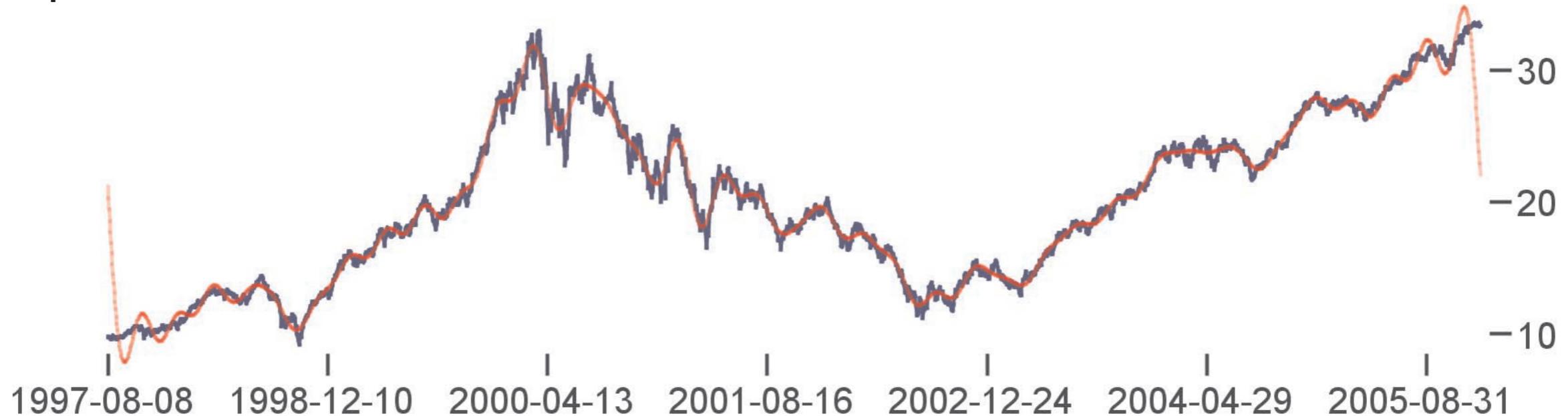


[Heer and Agrawala, 2006]

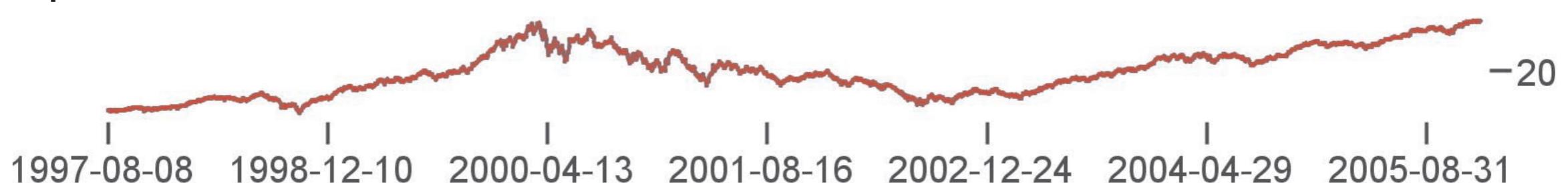
Multiscale Banking

PRMTX Mutual Fund

Aspect Ratio = 4.23



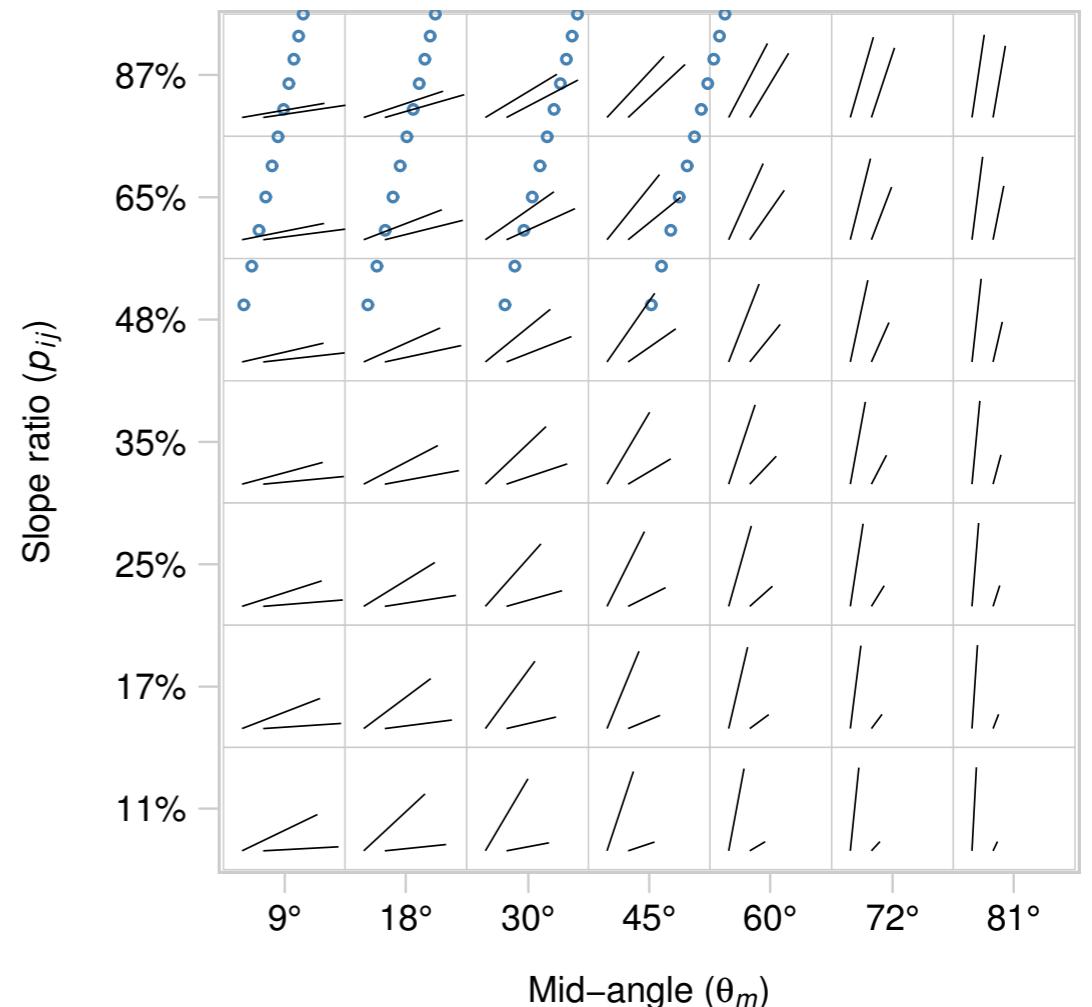
Aspect Ratio = 14.55



[Heer and Agrawala, 2006]

Expanding the Study

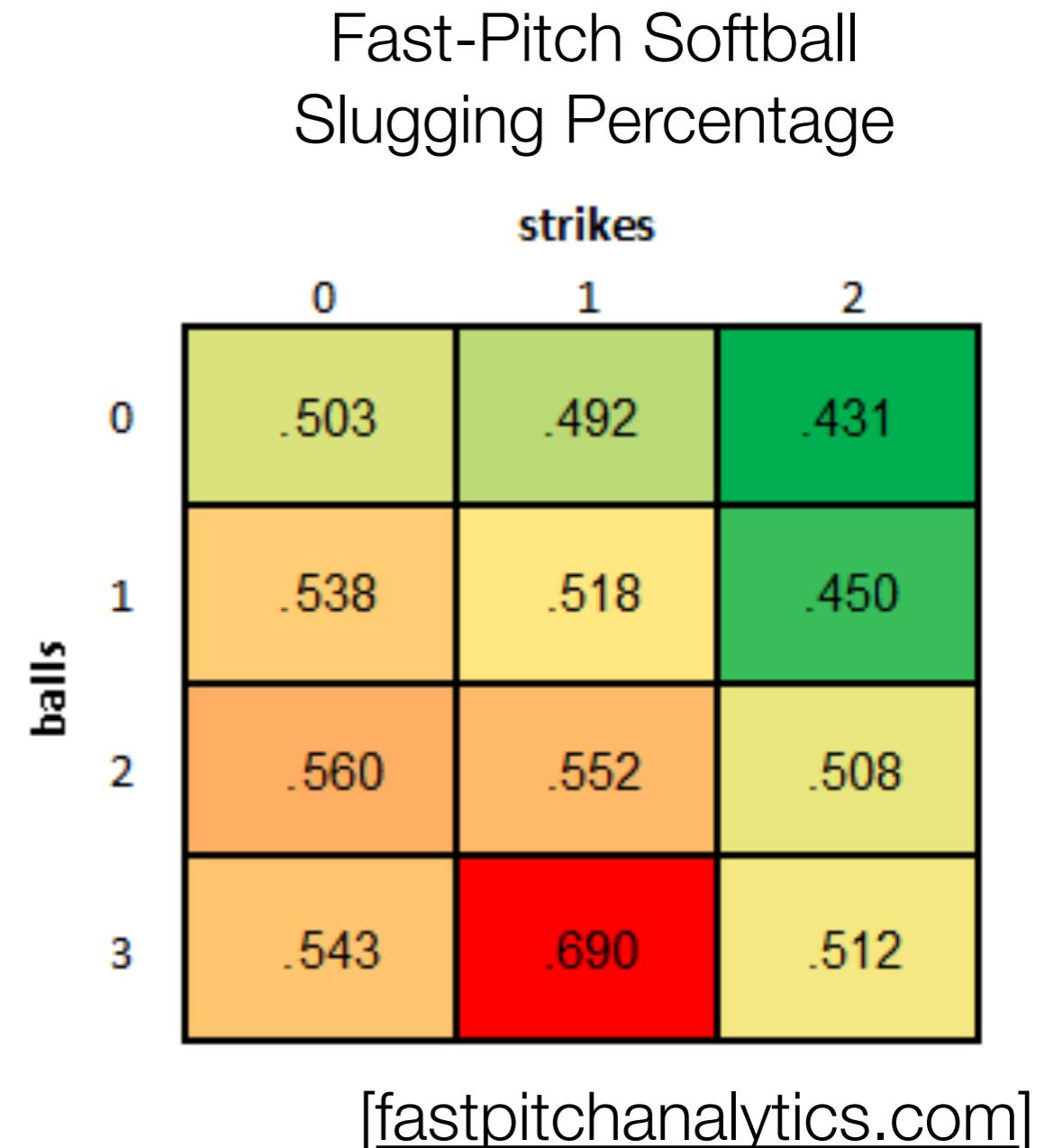
- Cleveland et al. did not study the entire space of slope comparisons and 45 degrees was at the low end of their study (blue marks on right)
- Talbot et al. compared more slopes and found that people do better with smaller slopes
- Baselines may aid with this



[Talbot et al., 2013]

Heatmaps

- Data: Two keys, one quantitative attribute
- Task: Find clusters, outliers, summarize
- How: area marks in grid, color encoding of quantitative attribute
- Scalability: number of pixels for area marks (millions), <12 colors
- Red-green color scales often used
 - Be aware of colorblindness!



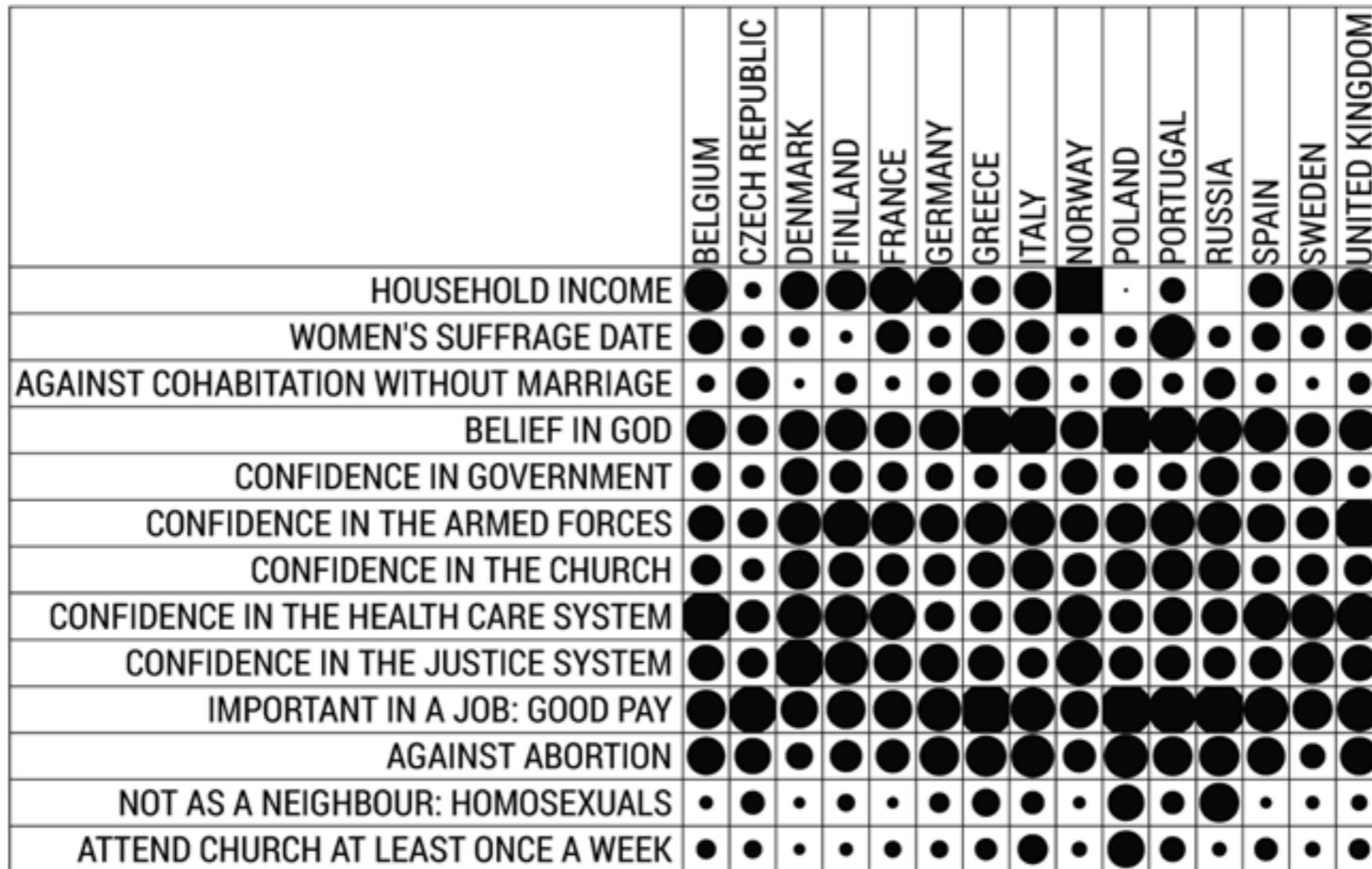
Bertin Matrices

- Must we only use color?
 - What other marks might be appropriate?

[C.Perrin et al., 2014]

Bertin Matrices

- Must we only use color?
 - What other marks might be appropriate?



[C.Perrin et al., 2014]

Bertin's Encodings

	0	1	2	3	4	5	6	7	8	9	10	11	text
Text													
Grayscale	●											●	X
Circle	—		●	●	●	●	●	●	●	●	●	●	X
Dual bar chart	●											●	X
Bar chart	●											●	X
Line													X
Black and white bar chart	●											●	X
Average bar chart	●											●	X

QUANTITY
OF INK
ENCODINGS

POSITIONAL
ENCODING

MEAN-BASED
ENCODINGS

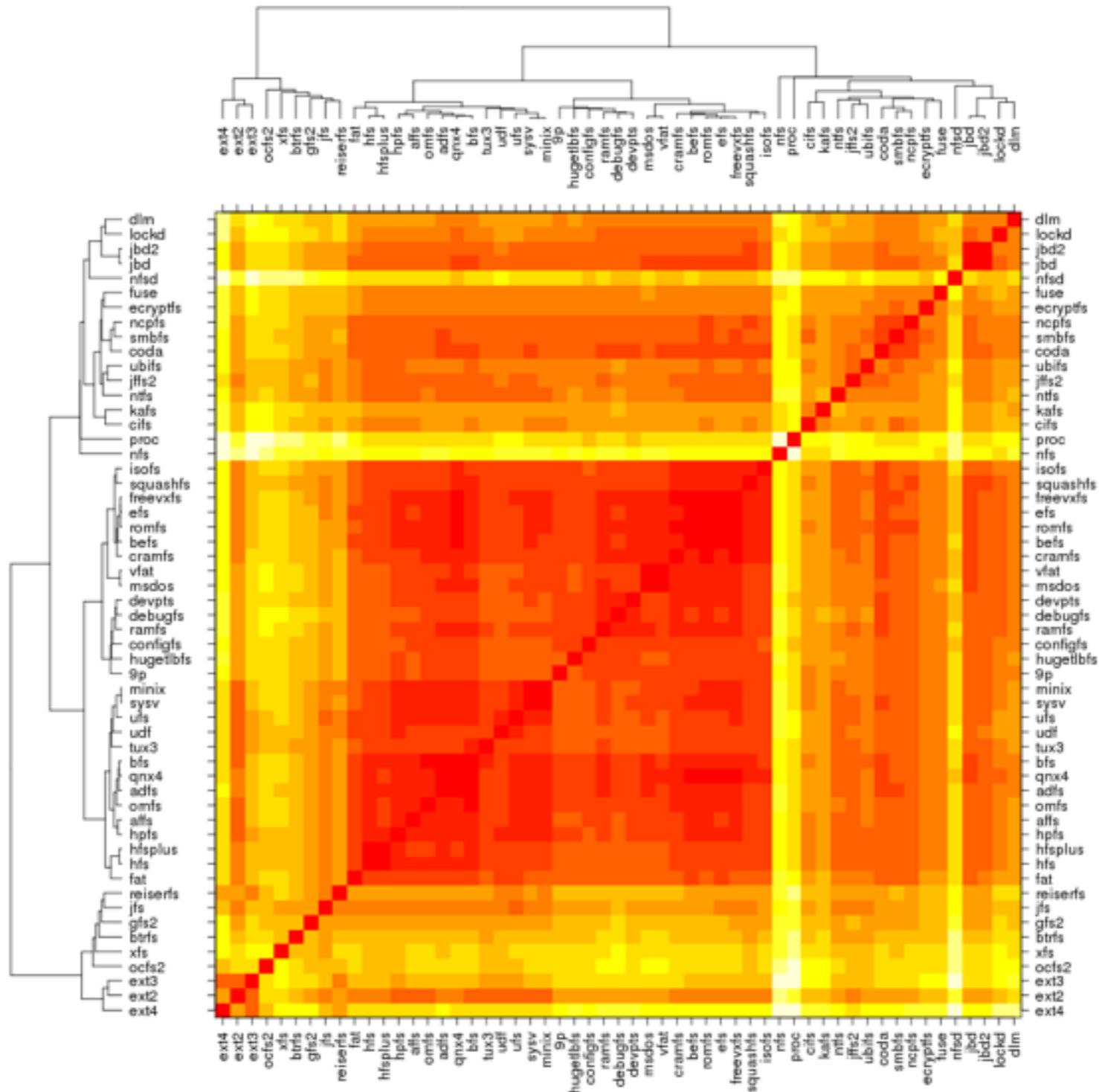
[C.Perrin et al., 2014]

Matrix Reordering



[Bertin Exhibit (INRIA, Vis 2014), Photo by [Robert Kosara](#)]

Cluster Heatmap



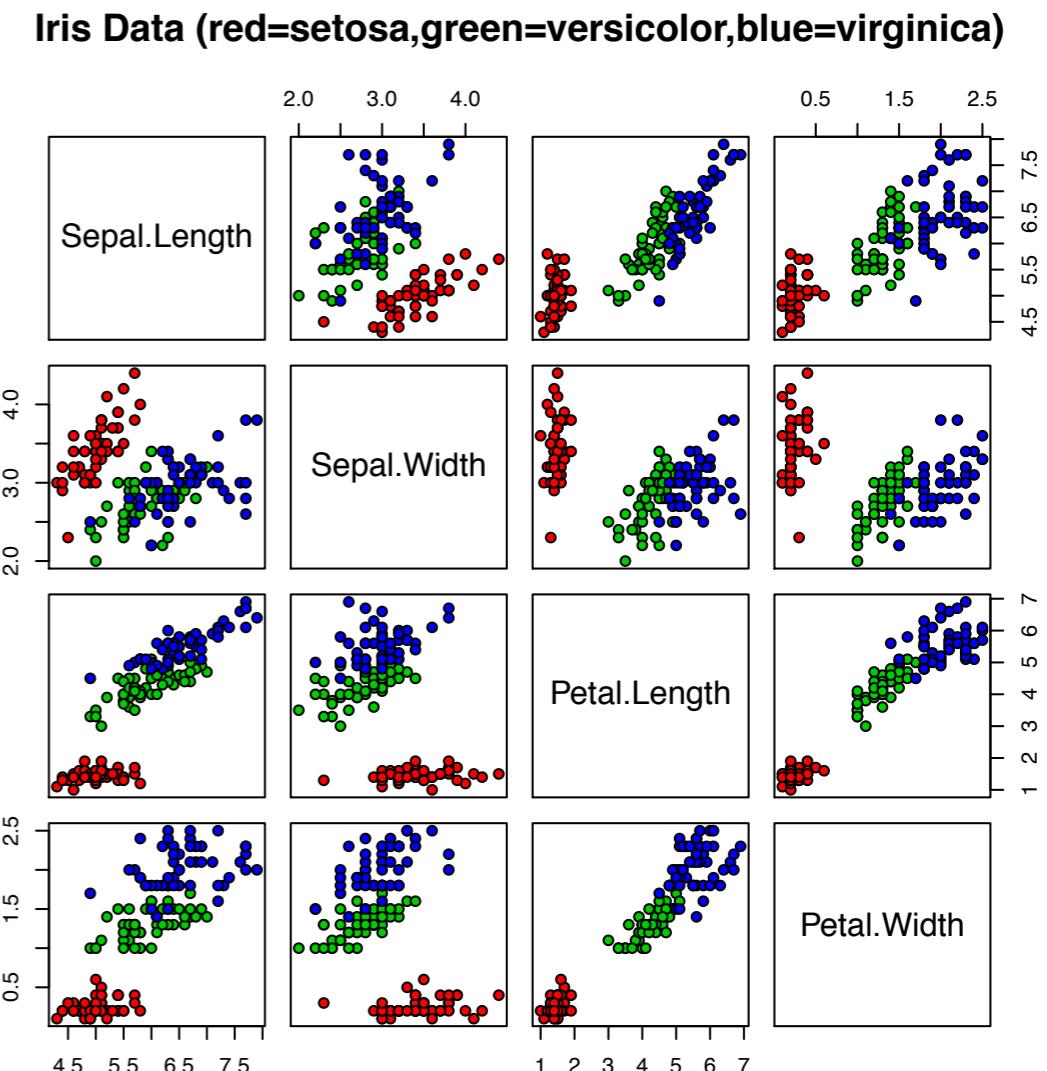
[File System Similarity, R. Musăloiu-E., 2009]

Cluster Heatmap

- Data & Task: Same as Heatmap
 - How: Area marks but matrix is ordered by cluster hierarchies
 - Scalability: limited by the cluster dendrogram
-
- Dendrogram: a visual encoding of tree data with leaves aligned

Scatterplot Matrix (SPLOM)

- Data: Many quantitative attributes
- Derived Data: names of attributes
- Task: Find correlations, trends, outliers
- How: Scatterplots in matrix alignment
- Scale: attributes: ~12, items: hundreds?
- Visualizations in a visualization: at high level, marks are themselves visualizations...



[R command from Wikipedia]

Spatial Axis Orientation

- So far, we have seen the vertical and horizontal axes (a **rectilinear** layout) used to encode almost everything
- What other possibilities are there for axes?

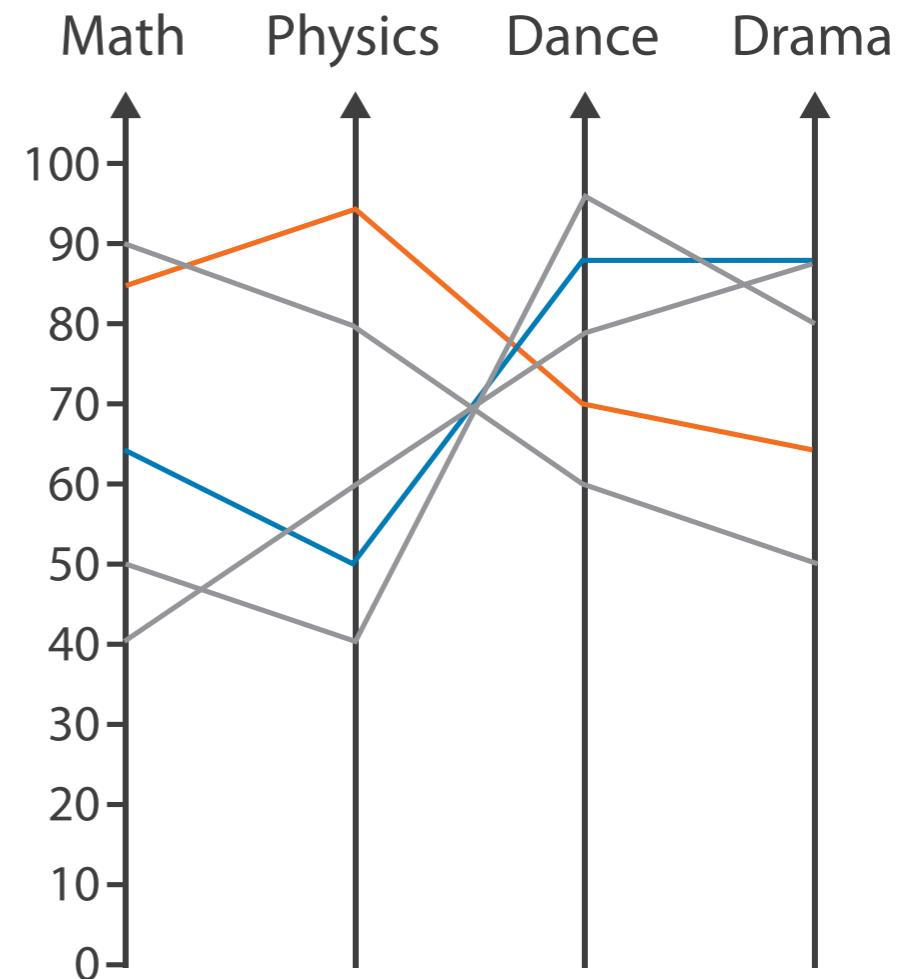
[Munzner (ill. Maguire), 2014]



Spatial Axis Orientation

- So far, we have seen the vertical and horizontal axes (a **rectilinear** layout) used to encode almost everything
- What other possibilities are there for axes?
 - Parallel axes

Parallel Coordinates

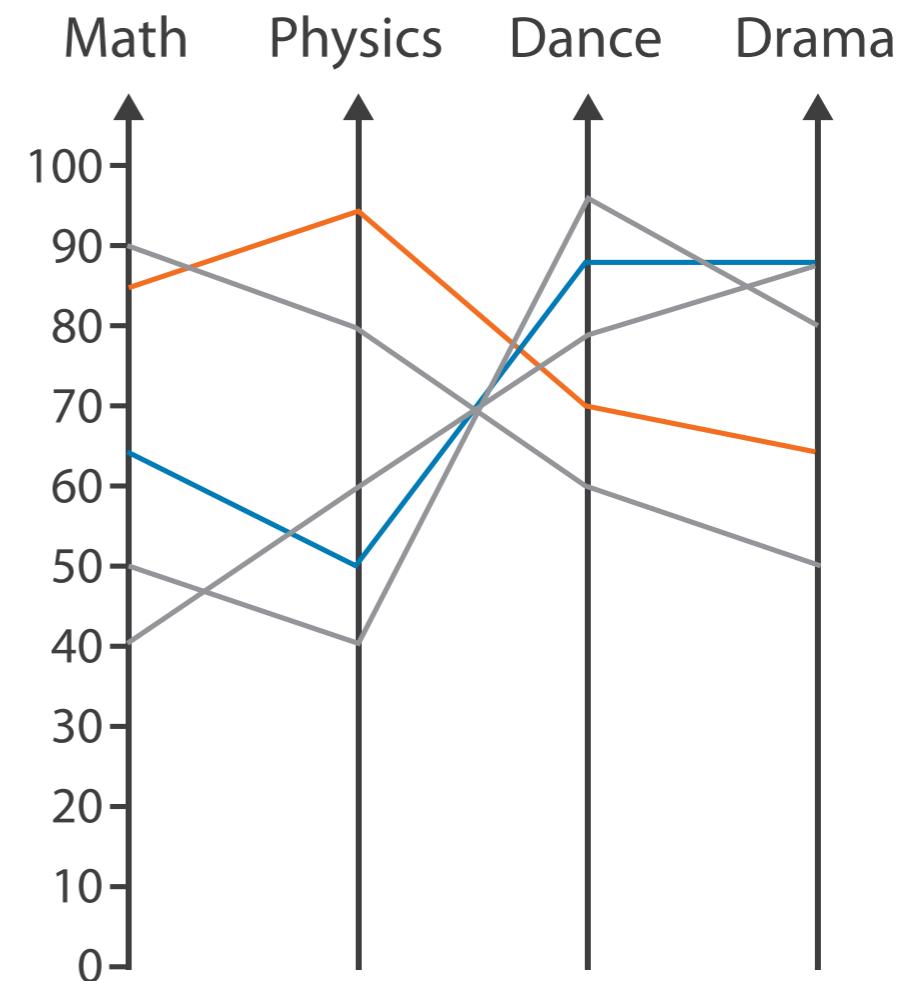


[Munzner (ill. Maguire), 2014]

Spatial Axis Orientation

- So far, we have seen the vertical and horizontal axes (a **rectilinear** layout) used to encode almost everything
- What other possibilities are there for axes?
 - Parallel axes
 - Radial axes

Parallel Coordinates

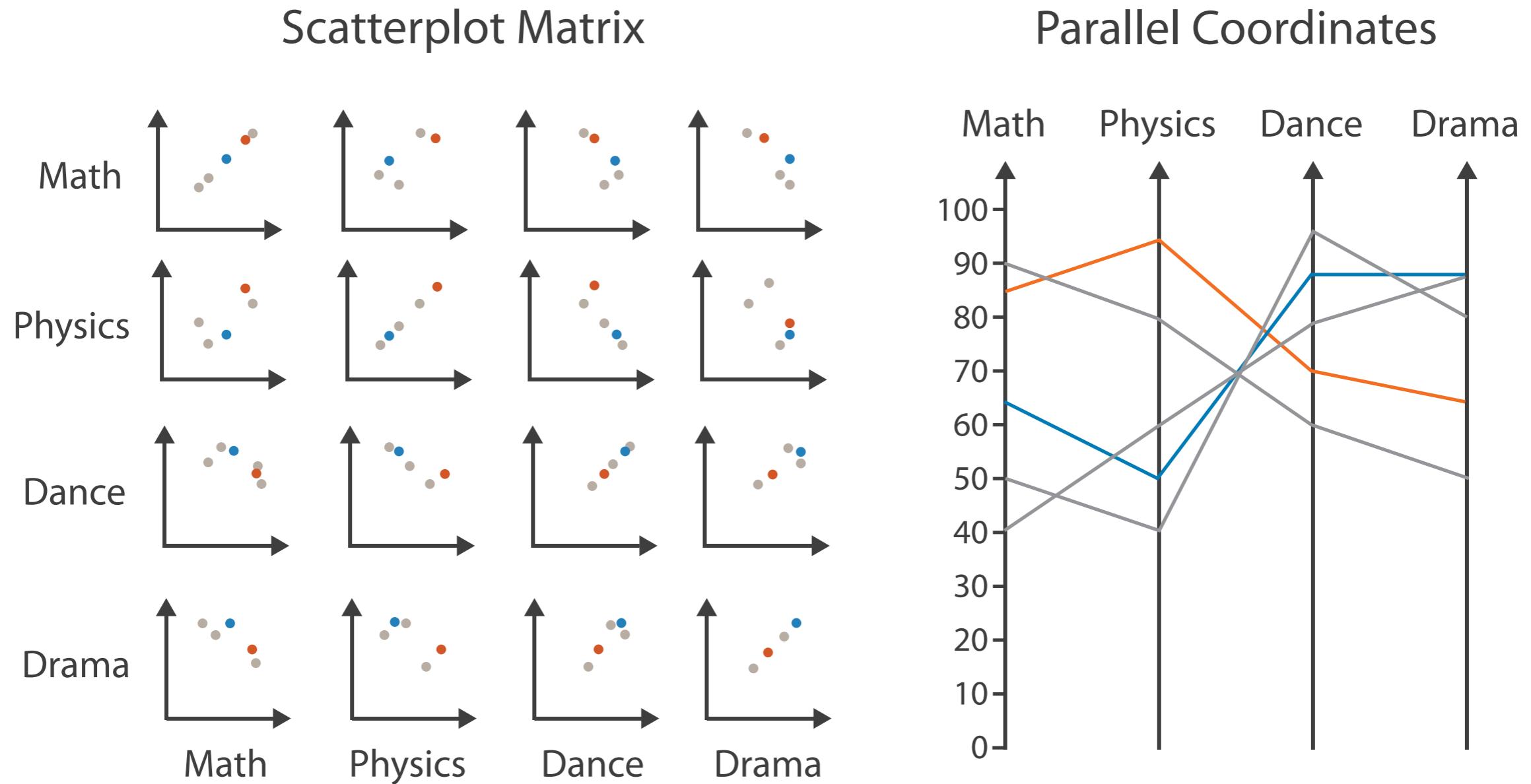


[Munzner (ill. Maguire), 2014]

Parallel Coordinates

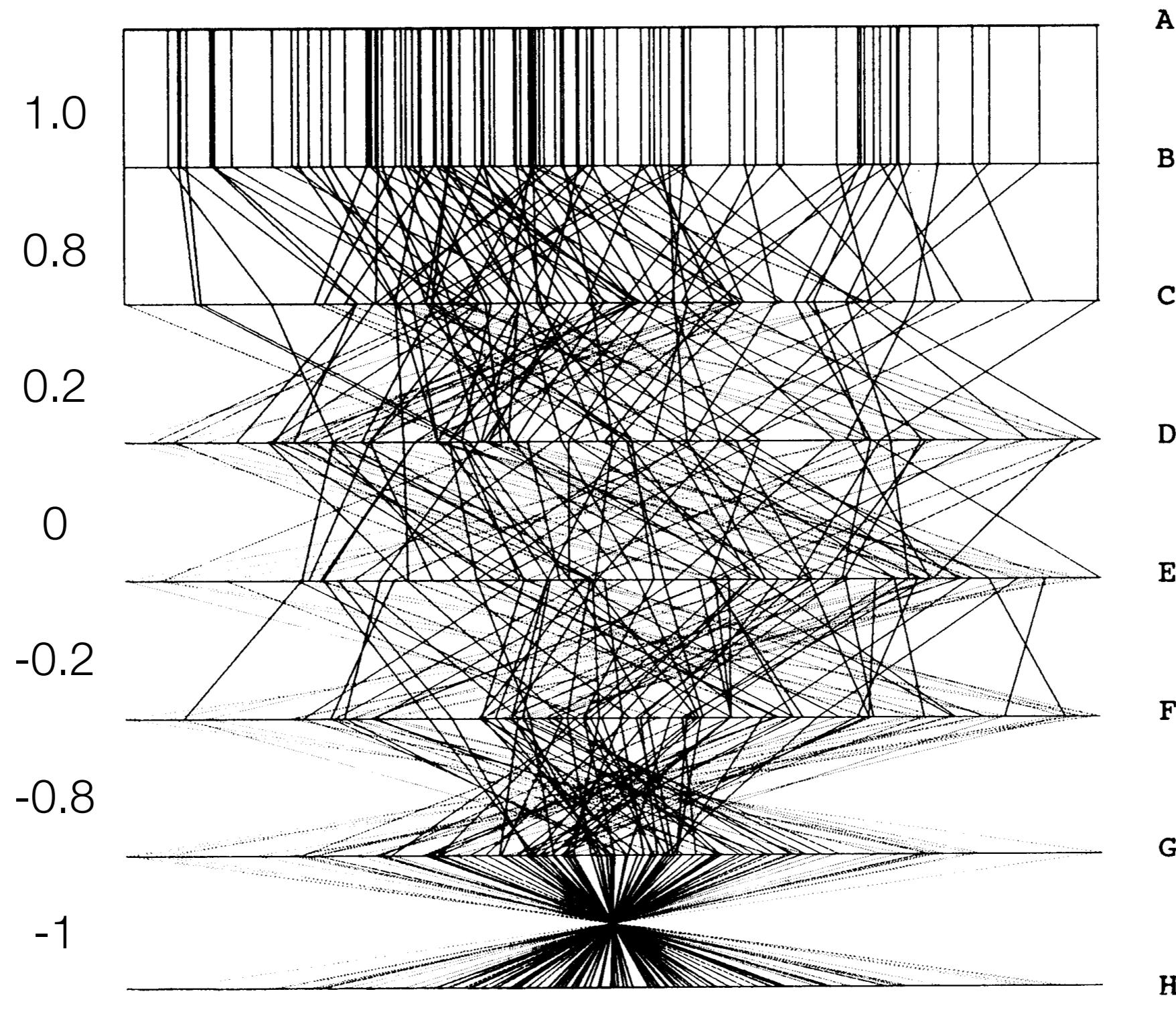
- Data: **many** quantitative attributes
 - Task: Find trends, extremes, correlation
 - How: vertical spatial position for each attribute, connection marks for identity, axes horizontally spaced
 - Scalability: <40 attributes, hundreds of values
-
- Connection marks help visualize trends between particular values
 - **Ordering** the horizontal axes is important
 - Not as well-known, often requires learning

Comparing SPLOMs and Parallel Coordinates

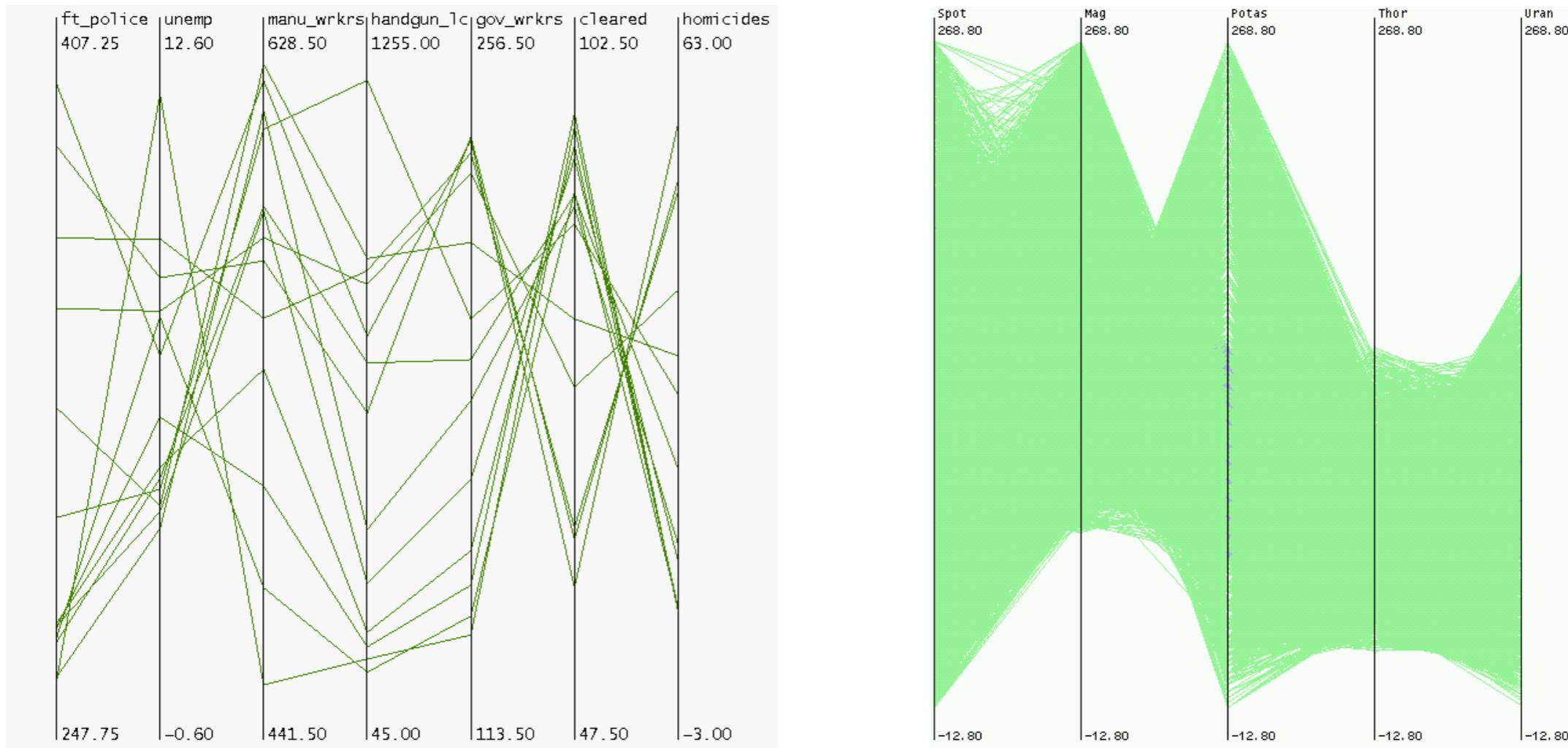


[Munzner (ill. Maguire), 2014]

Correlation in Parallel Coordinates

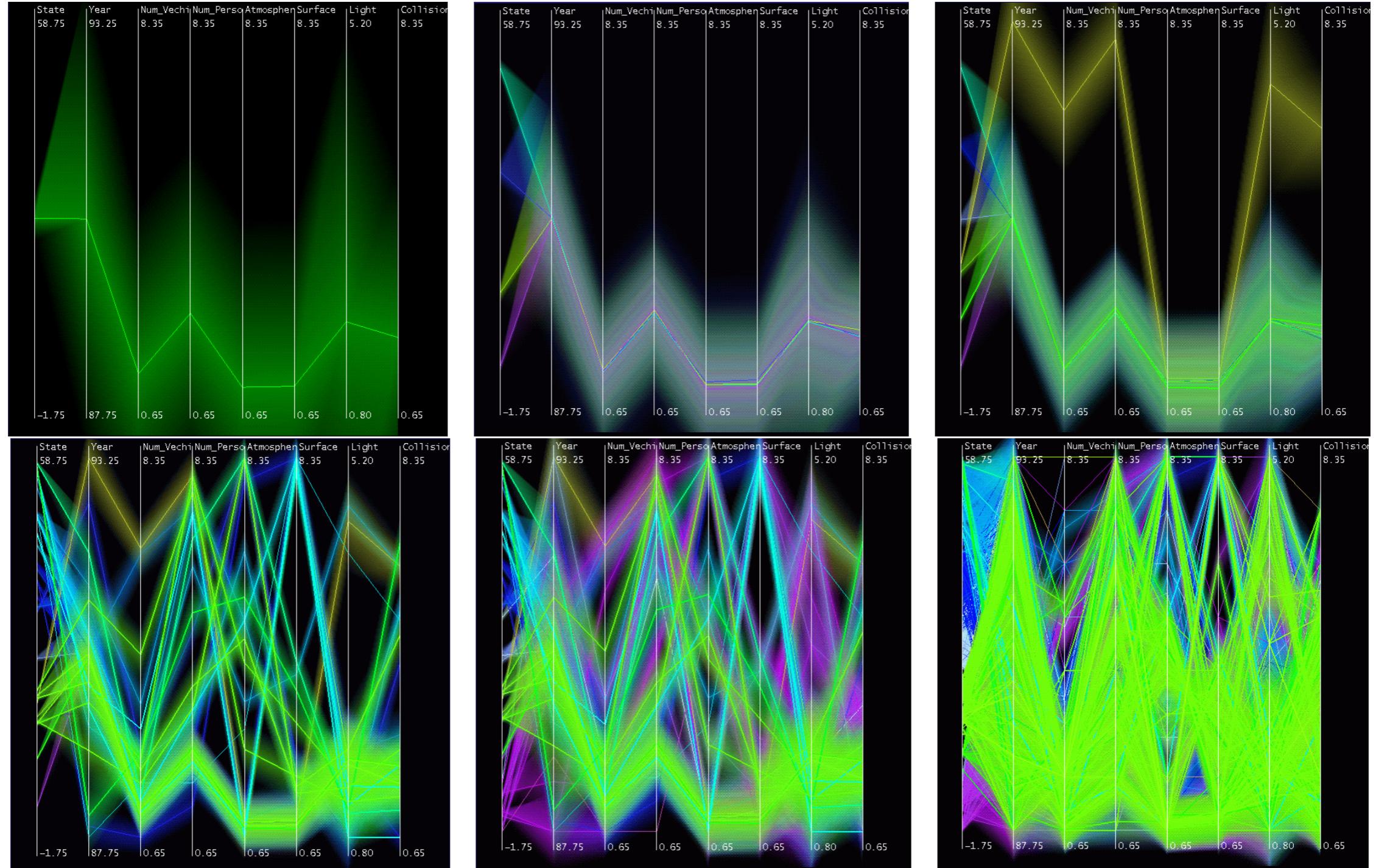


Overdraw in Parallel Coordinates



[Fua et al., 1999]

Hierarchical Parallel Coordinates



[Fua et al., 1999]