

CSC 544

Data Visualization

Joshua Levine
josh@arizona.edu

Lecture 12

Tabular Arrangements

Feb. 22, 2023

Today's Agenda

- Reminders:
 - P01 due! P02 posted.
 - A03 questions?
- Goals for today:
 - Introduce spatial arrangements of tables

Reminder: Main Themes for CSC544

- **Mechanics:** how do I build a visualization?
 - Javascript, CSS, HTML, d3
- **Principles:** why should I build it in this way?
 - mathematical and perceptual arguments
- **Techniques:** what do I use to turn principles and mechanics in an actual visualization?
 - algorithms, software libraries

Reminder: Main Themes for CSC544

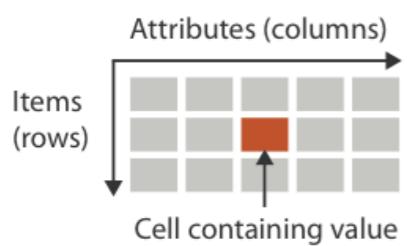
- **Mechanics:** how do I build a visualization?
 - Javascript, CSS, HTML, d3
- **Principles:** why should I build it in this way?
 - mathematical and perceptual arguments
- **Techniques:** what do I use to turn principles and mechanics in an actual visualization?
 - algorithms, software libraries

→ Data and Dataset Types

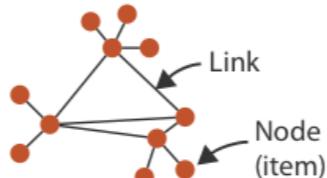
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

→ Dataset Types

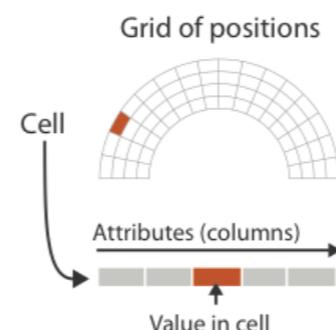
→ Tables



→ Networks



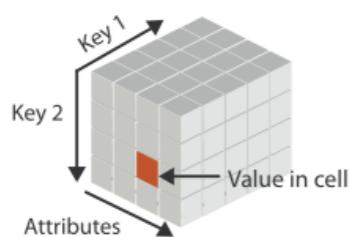
→ Fields (Continuous)



→ Geometry (Spatial)



→ Multidimensional Table



→ Trees



Arranging Tables

Arrange Tables

→ Express Values

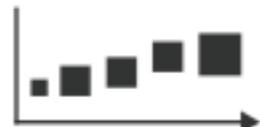


→ Separate, Order, Align Regions

→ Separate



→ Order



→ Align



→ 1 Key
List



→ 2 Keys
Matrix



→ 3 Keys
Volume



→ Many Keys
Recursive Subdivision



→ Axis Orientation

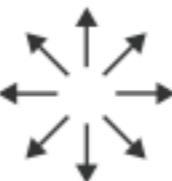
→ Rectilinear



→ Parallel



→ Radial



→ Layout Density

→ Dense



→ Space-Filling



**Arrange is the Focus of
All Four Design Choices
for Tabular Data**

(how much)

④ **Magnitude Channels: Ordered Attributes**

Position on common scale



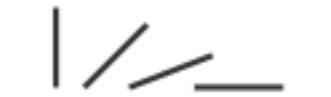
Position on unaligned scale



Length (1D size)



Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



Color saturation



Curvature



Volume (3D size)



(what or where)

④ **Identity Channels: Categorical Attributes**

Spatial region



Color hue



Motion



Shape



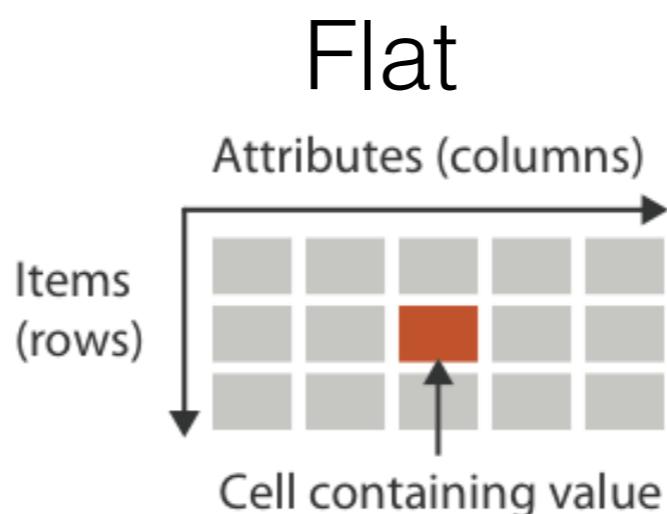
Effectiveness
Least ↑ Same ↓ Most

There are no nonspatial channels that are highly effective for all attribute types!

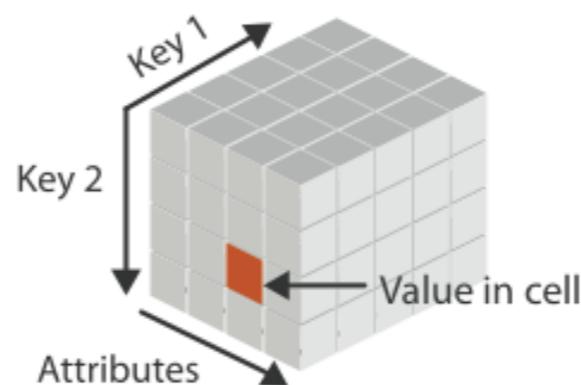
Recall: Attribute Semantics

When we arrange tabular data, attributes are chosen to be
keys and **values**

Tables

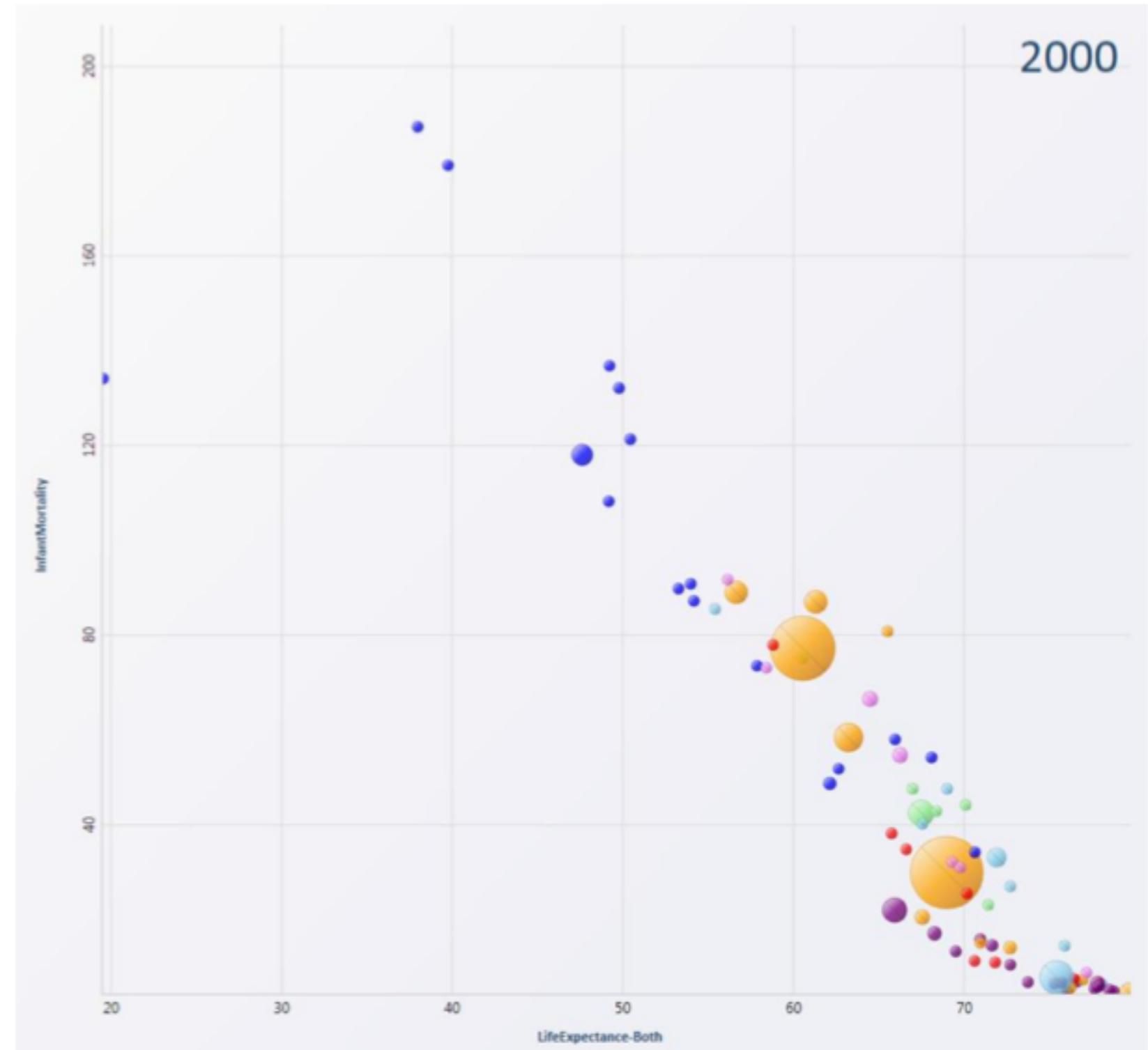


Multidimensional



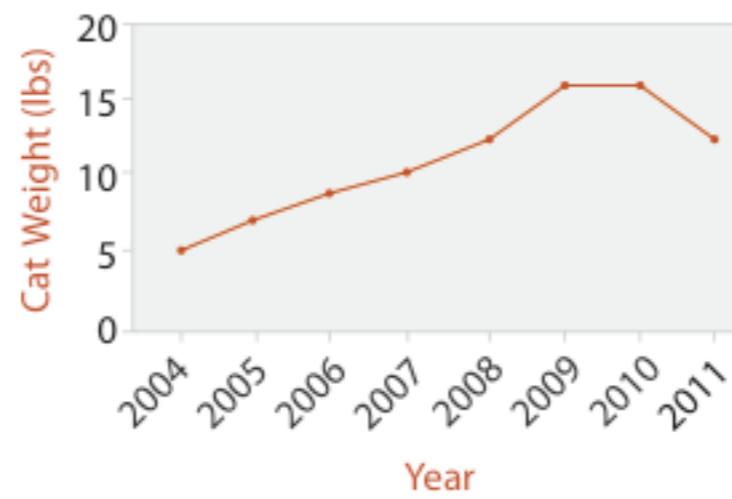
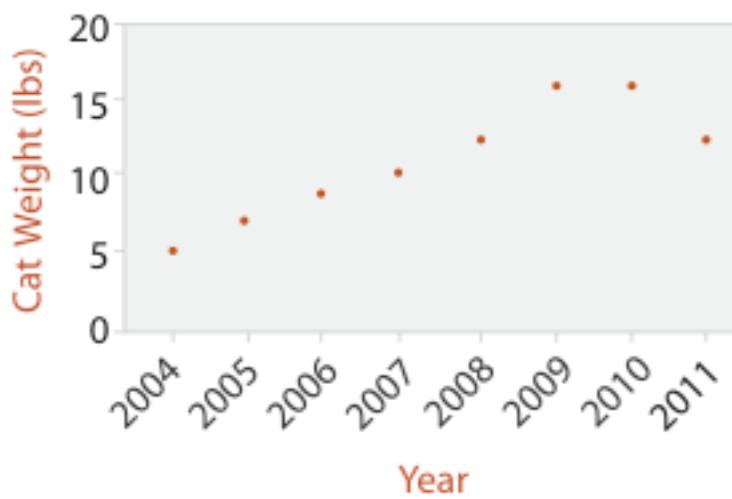
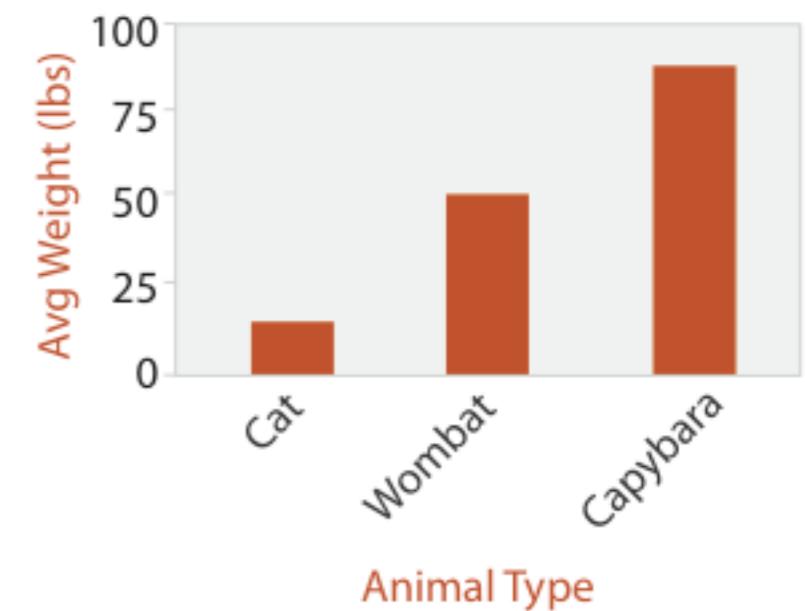
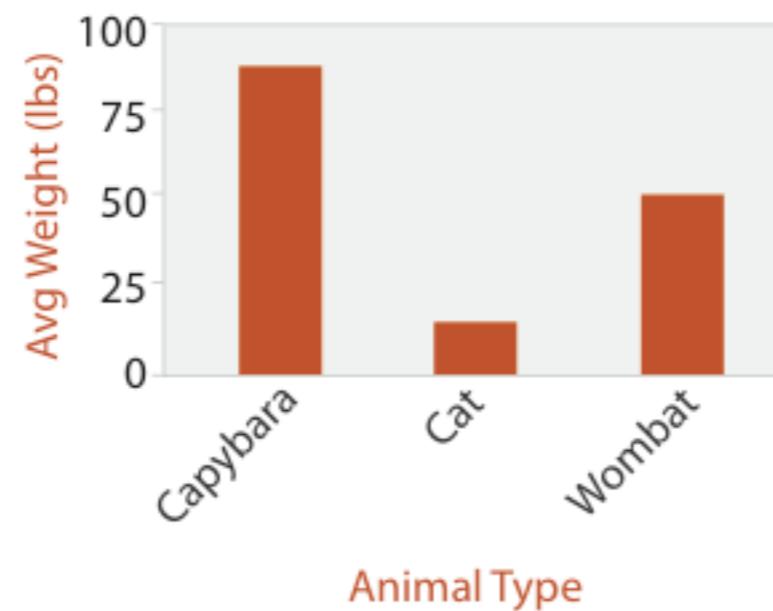
Encode Using Zero Keys: Scatterplots

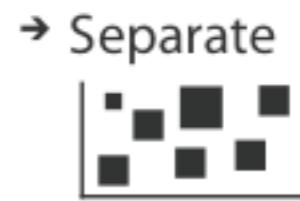
④ Express Values



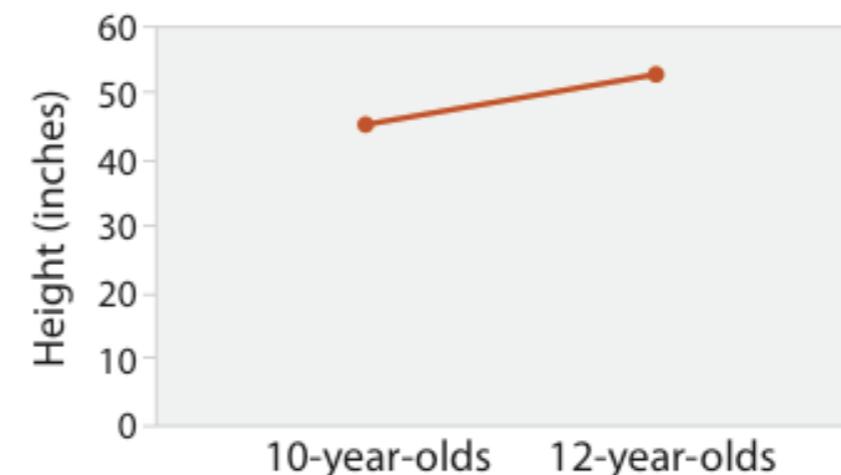
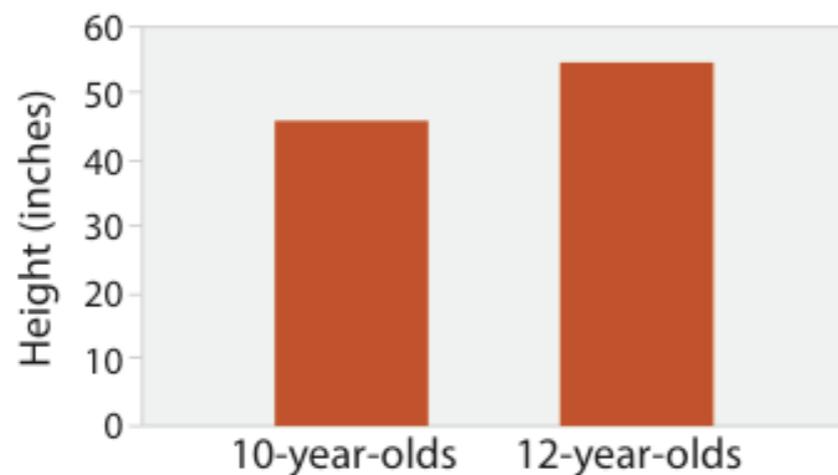
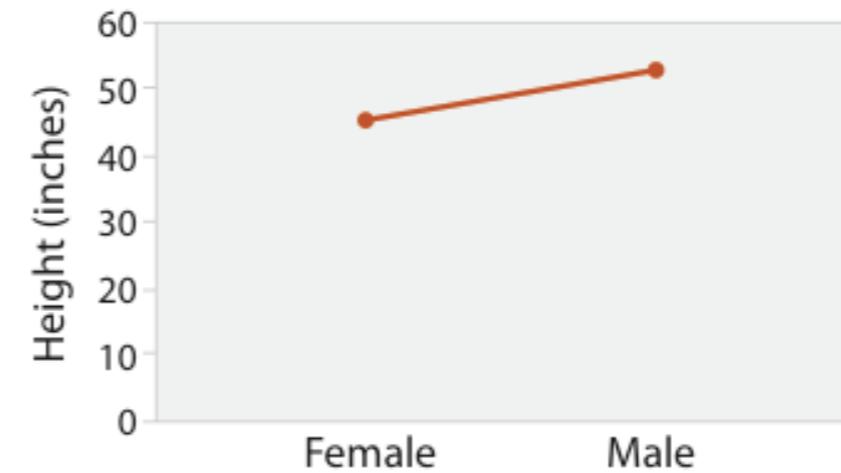
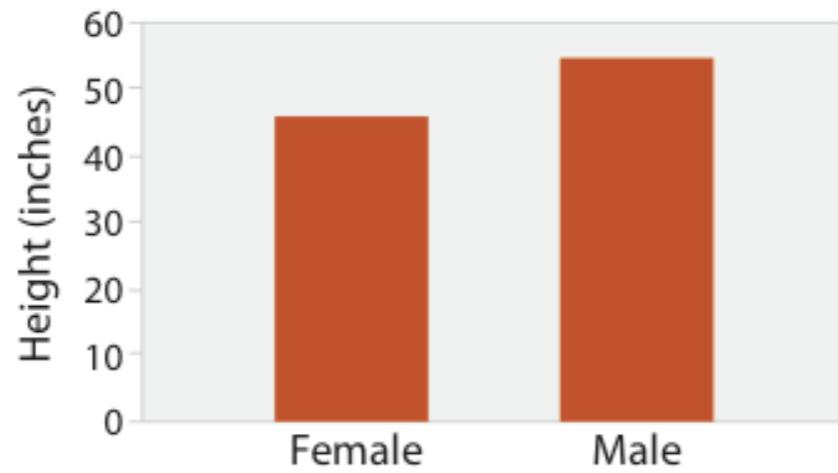
Encode One Attribute as a Key: Bar, Dot, Line Charts

→ Order

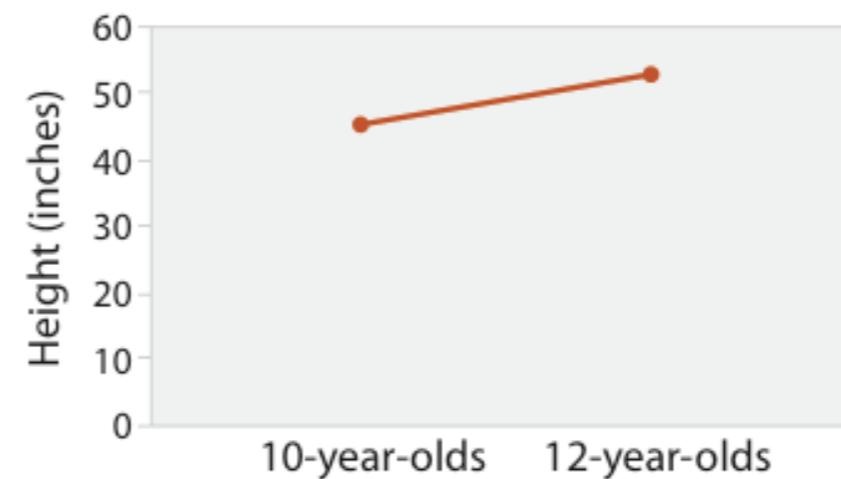
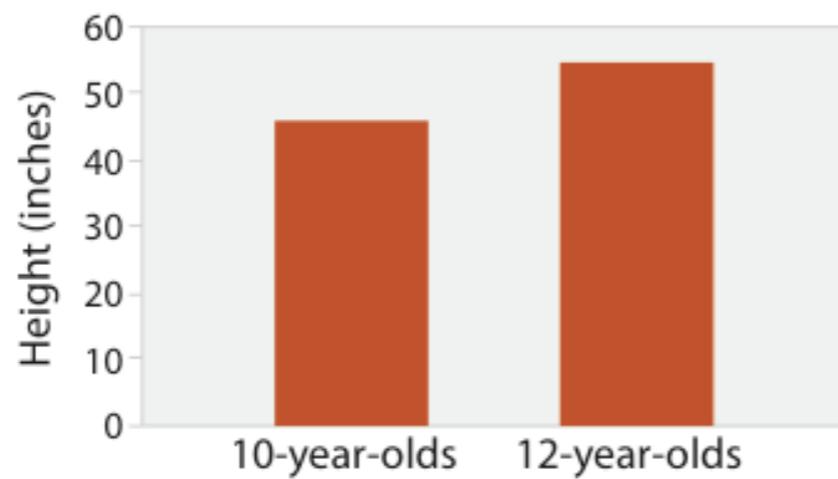
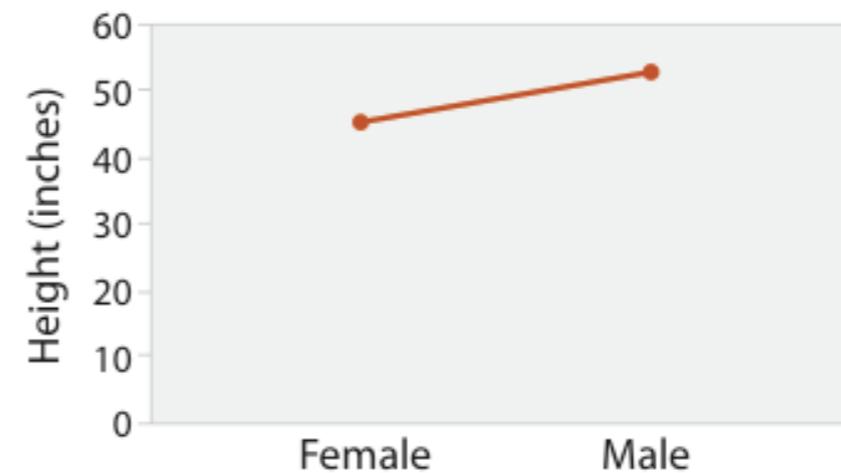
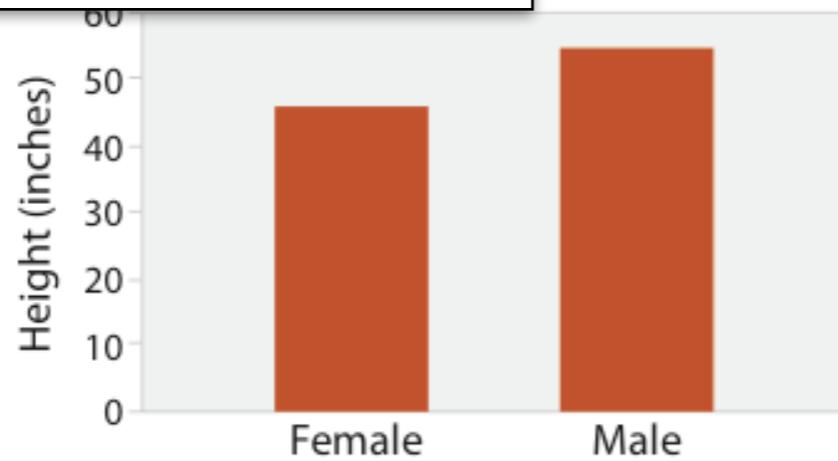
→ Separate


Line Charts Should Not Be Used for Categorical Keys



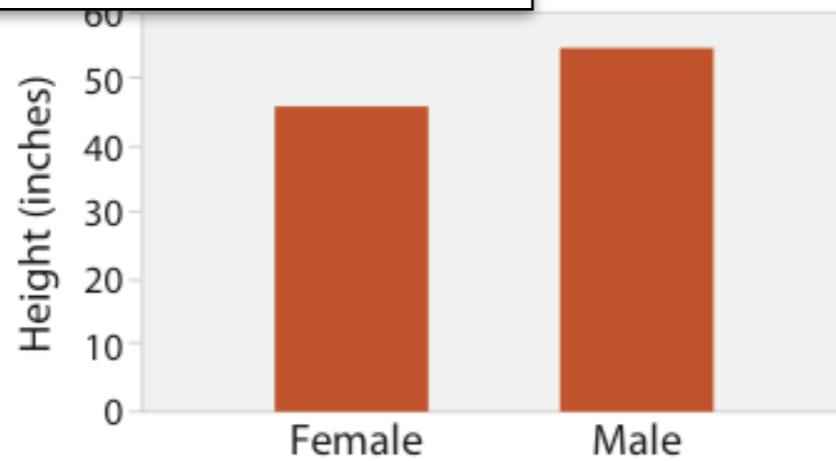
Line Charts Should Not Be Used for Categorical Keys

OK: “Men are taller than women (on average)”

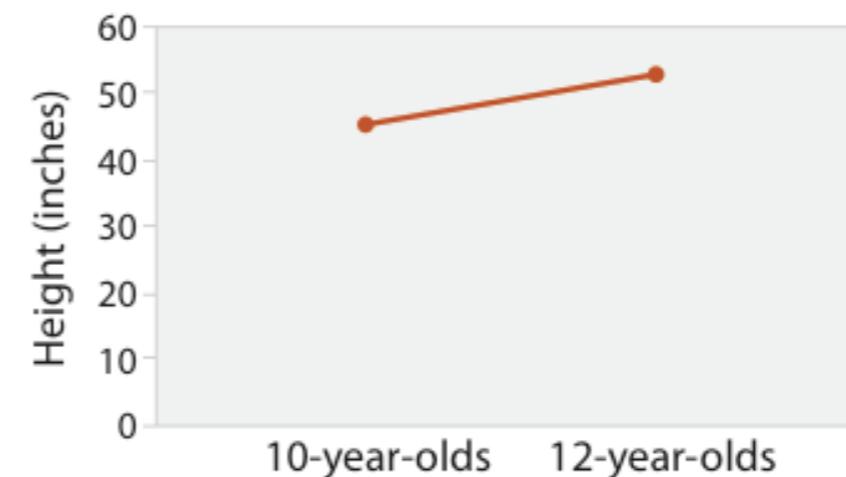
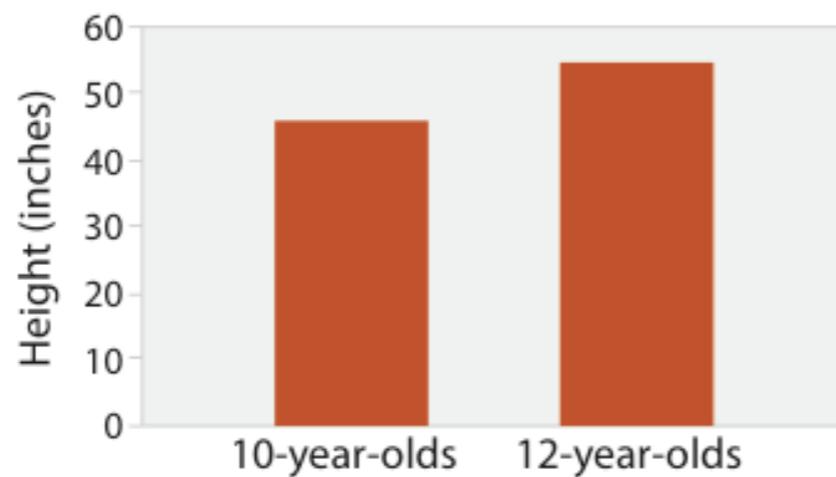
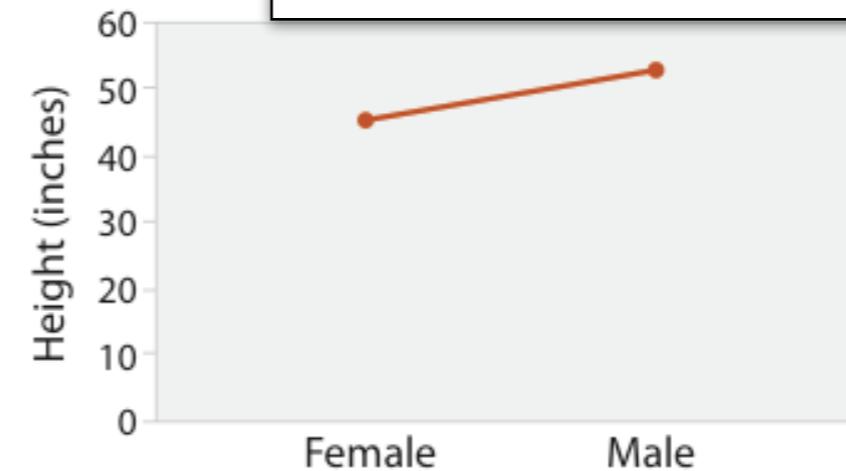


Line Charts Should Not Be Used for Categorical Keys

OK: “Men are taller than women (on average)”

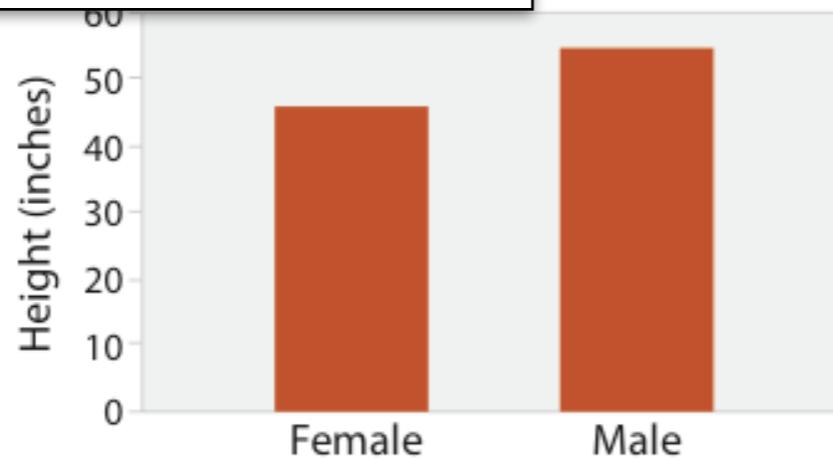


Not OK: “The more male a person is, the taller he/she is”

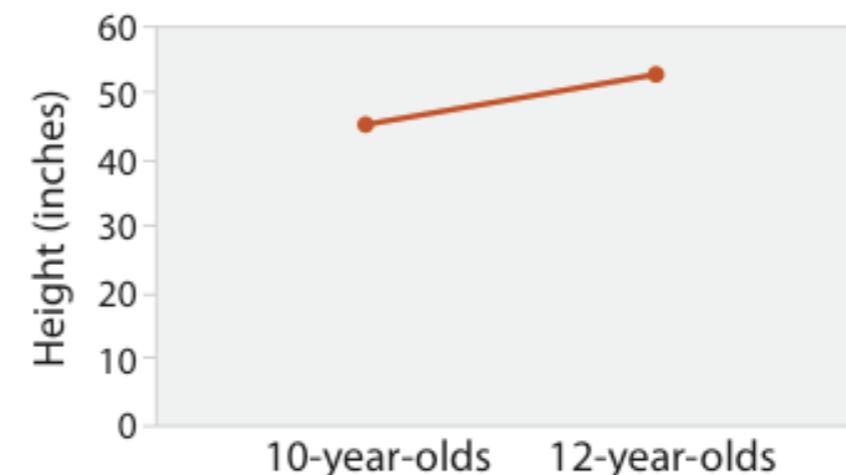
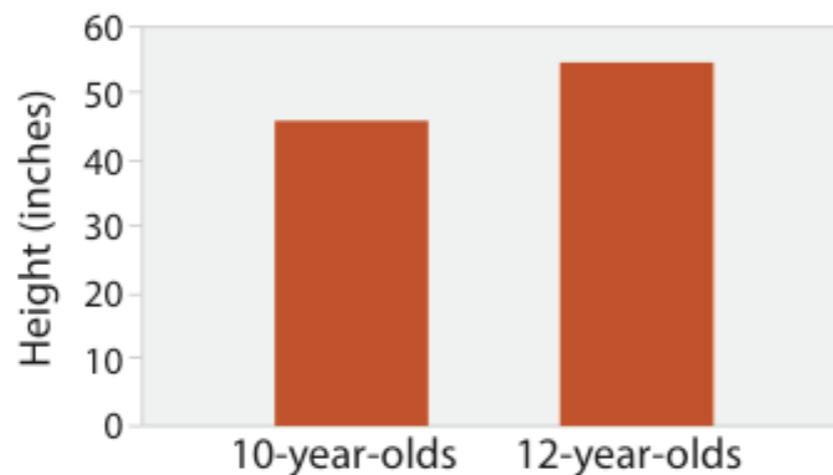
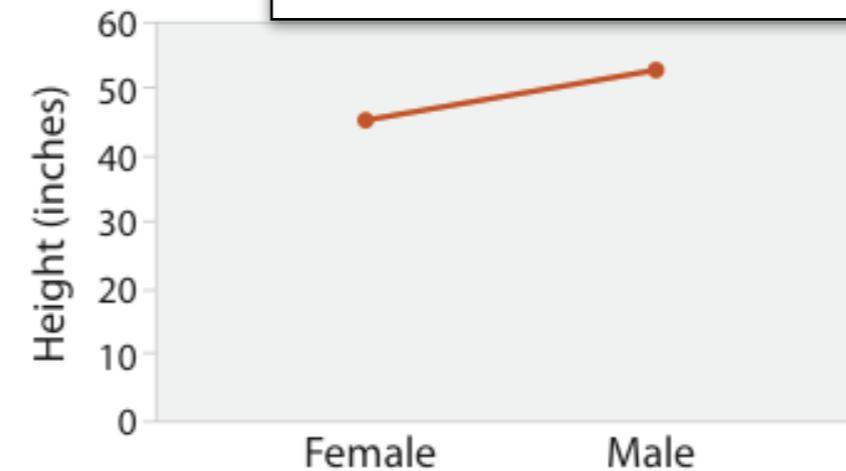


Line Charts Should Not Be Used for Categorical Keys

OK: “Men are taller than women (on average)”



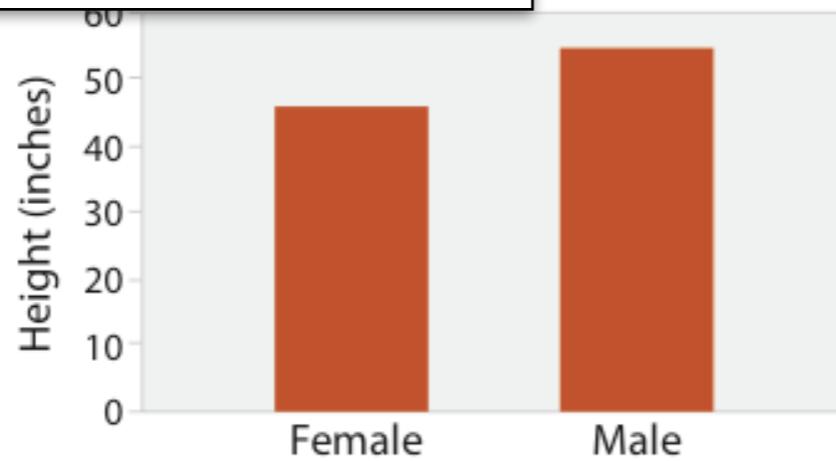
Not OK: “The more male a person is, the taller he/she is”



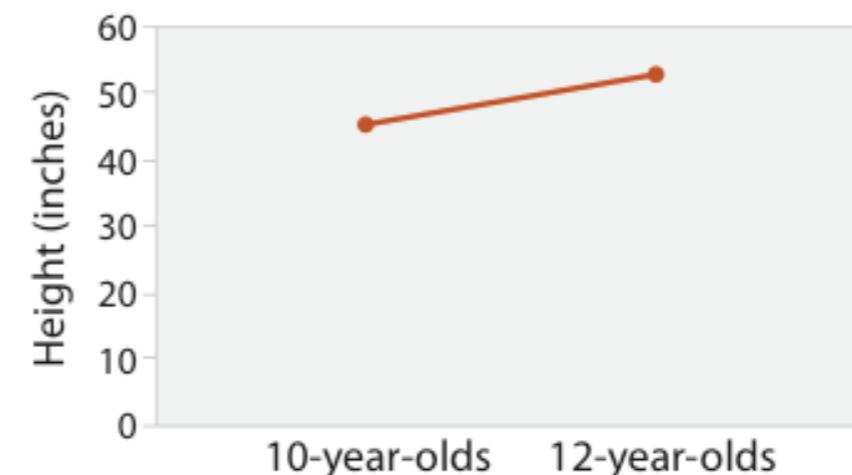
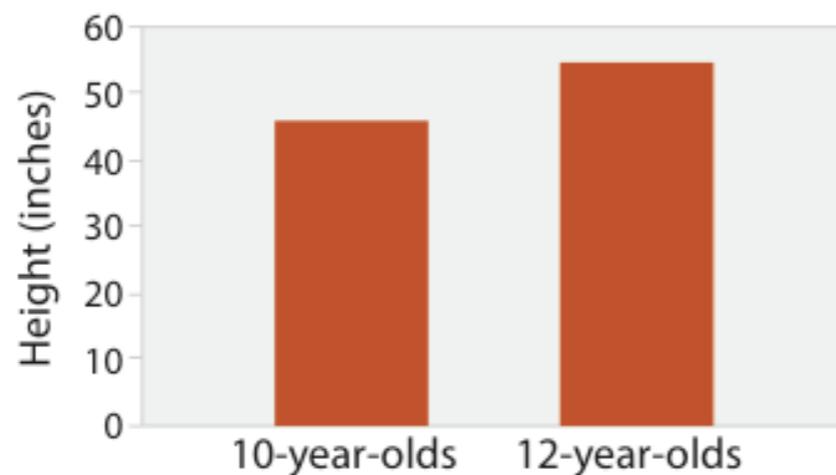
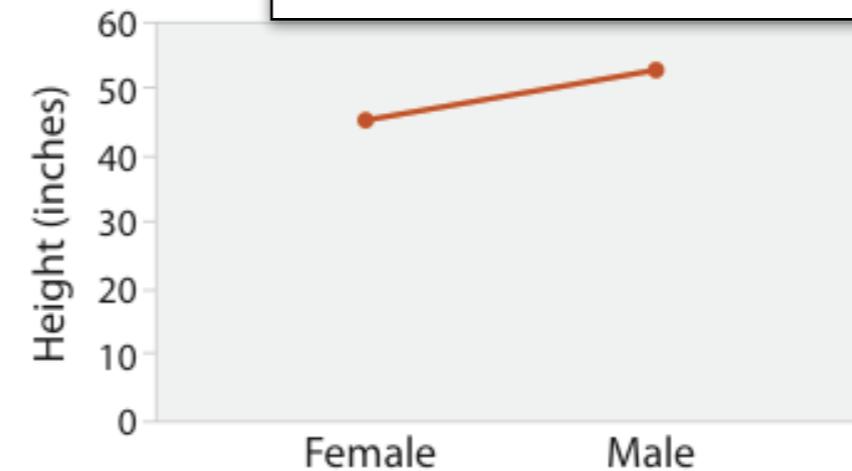
OK: “Twelve year olds are taller than ten year olds”

Line Charts Should Not Be Used for Categorical Keys

OK: “Men are taller than women (on average)”



Not OK: “The more male a person is, the taller he/she is”



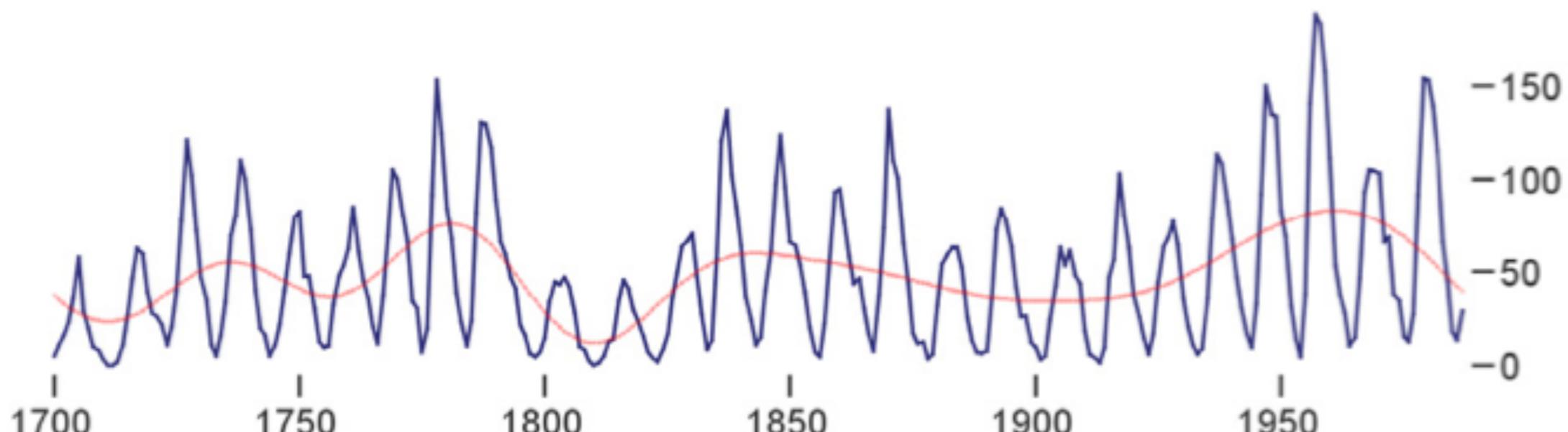
OK: “Twelve year olds are taller than ten year olds”

Sort of OK?: “Height increases with age”

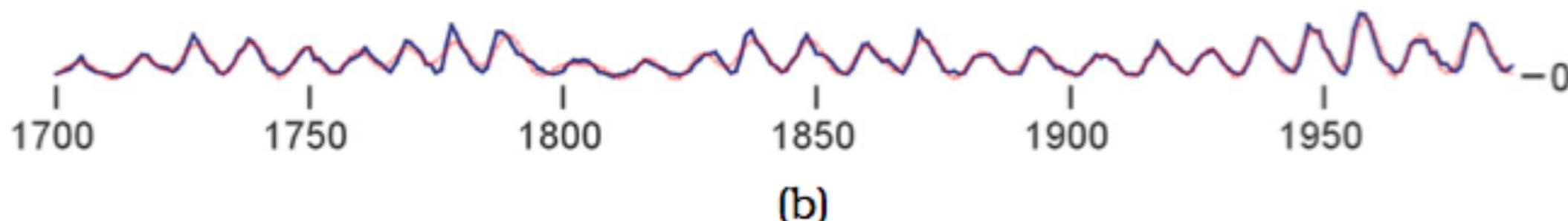
→ Align



Aspect Ratio and Our Ability To Judge Angles



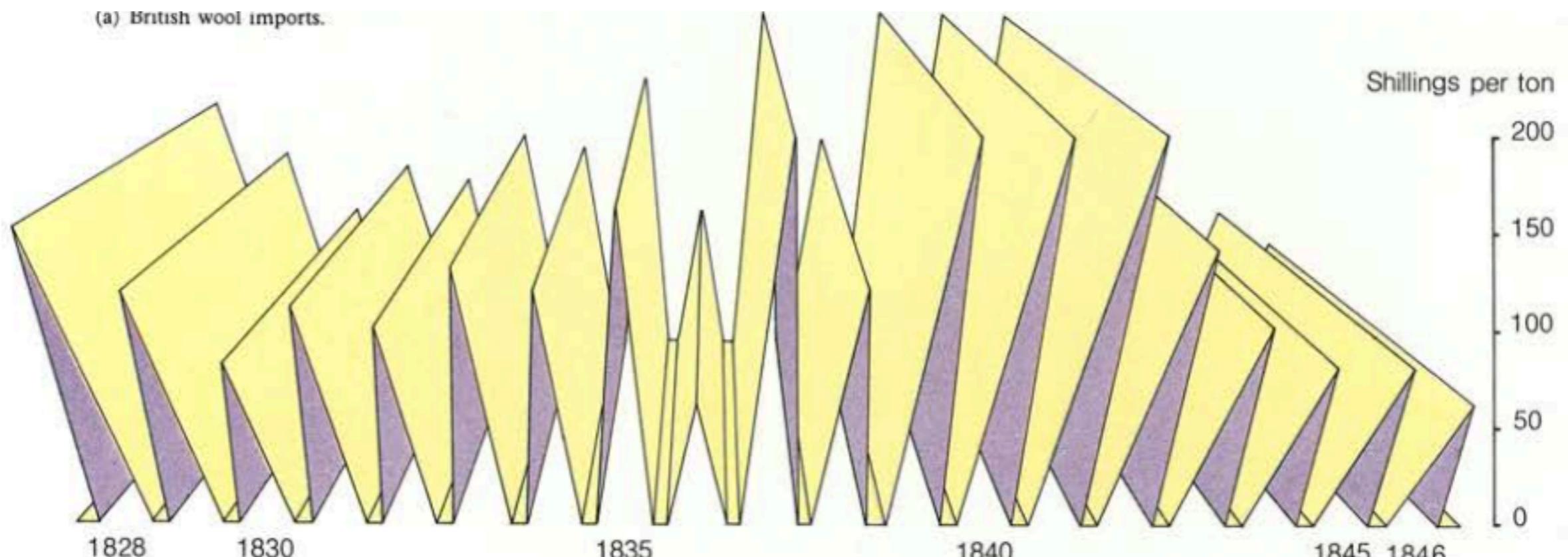
(a)



(b)

→ Align

Aspect Ratio and Our Ability To Judge Angles



Example from this week's design critiques

<https://socialsciences.org.au/library/historical-statistics-chapter-7/>

Alignment with Multiple Keys

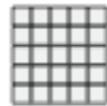
→ Align



→ 1 Key
List



→ 2 Keys
Matrix



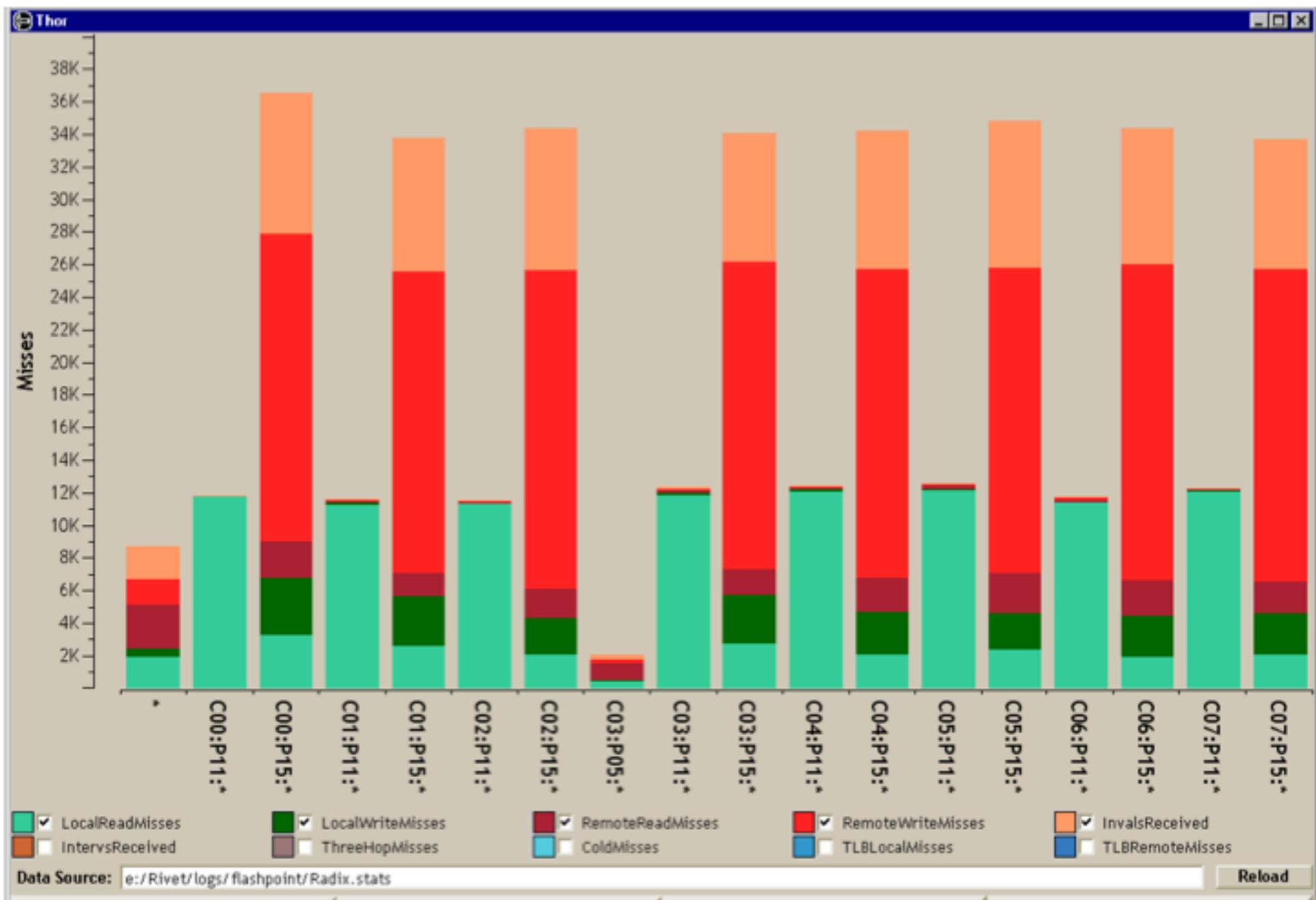
→ 3 Keys
Volume



→ Many Keys
Recursive Subdivision

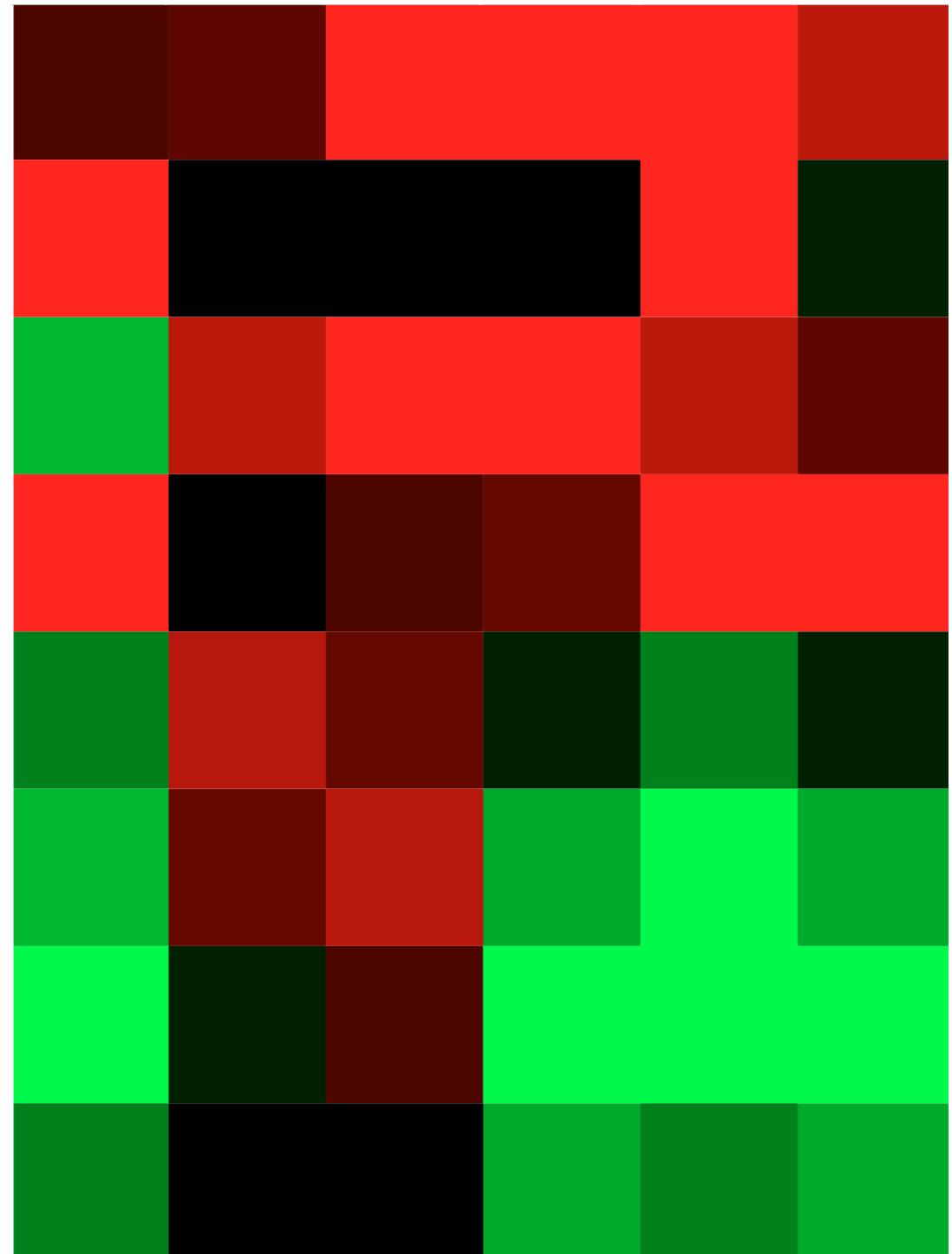


Stacked Bar Charts: Two Categorical Keys

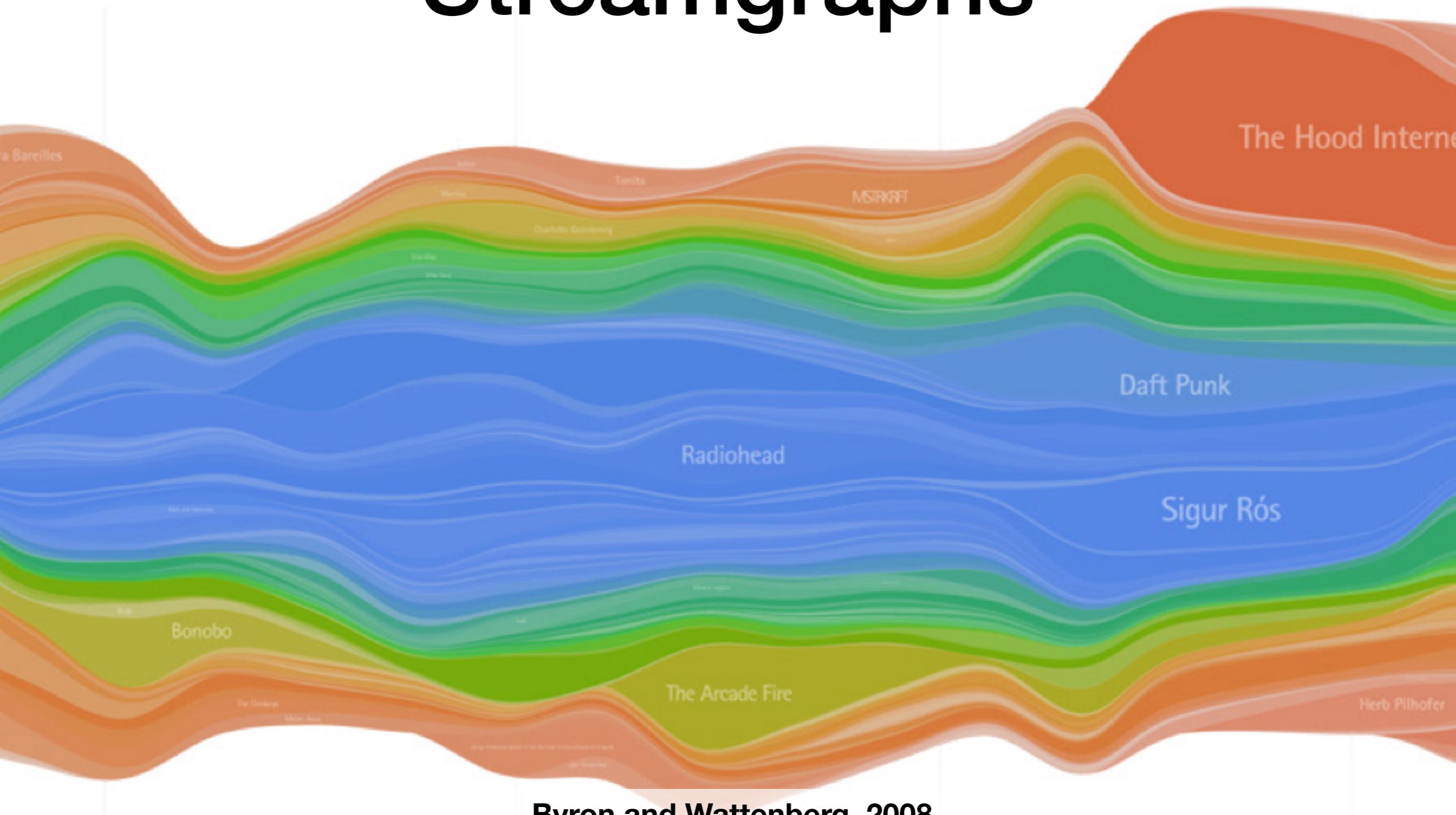


Heatmaps: Two Keys Encoded Spatially And One Value Encoding as Color

- Uses heatmap representation
 - Matrix layout
 - Encode values with color: hotspots. Red Green Used



Multiple Keys, One Value: Streamgraphs



- Original data
 - Set of time series
- Derived data
 - Layer silhouette
 - Consider baseline
 - Consider deviation
 - Consider “wiggliness”

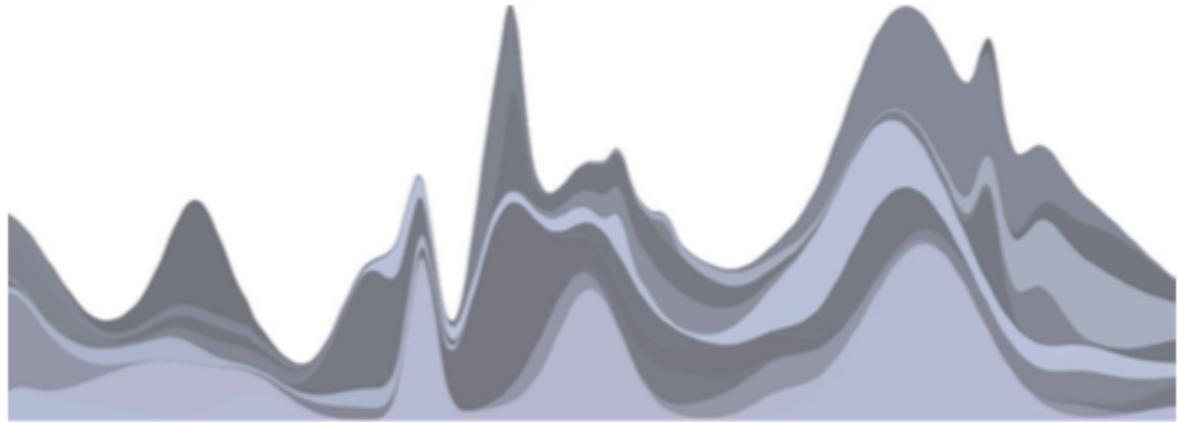


fig 5 – A traditional stacked graph with a baseline $g_0 = 0$

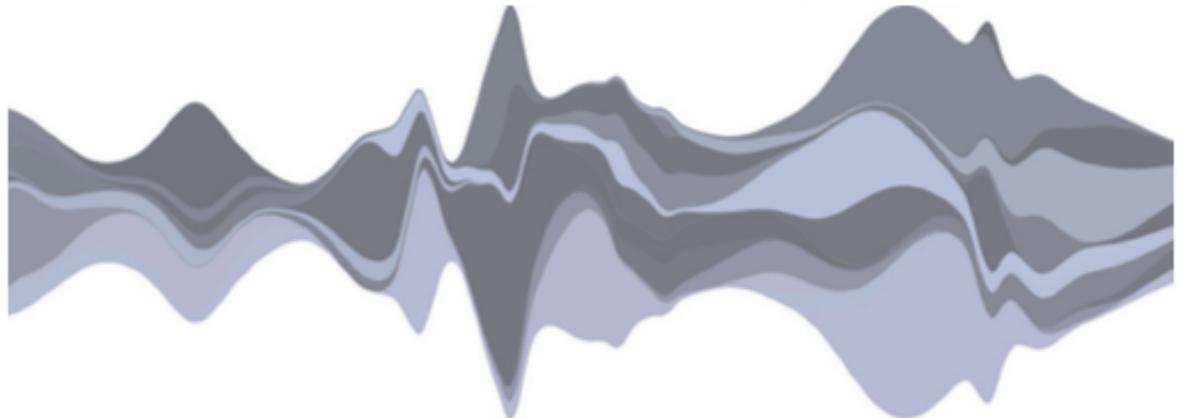


fig 6 – the same data set using the ThemeRiver layout algorithm

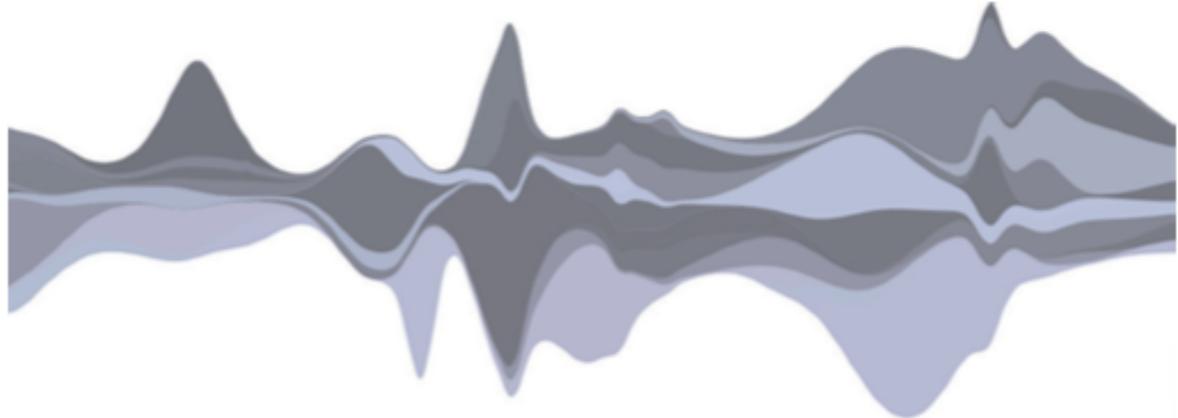


fig 7 – the same data set optimized to reduce the "wiggle" function, or overall variation in slope

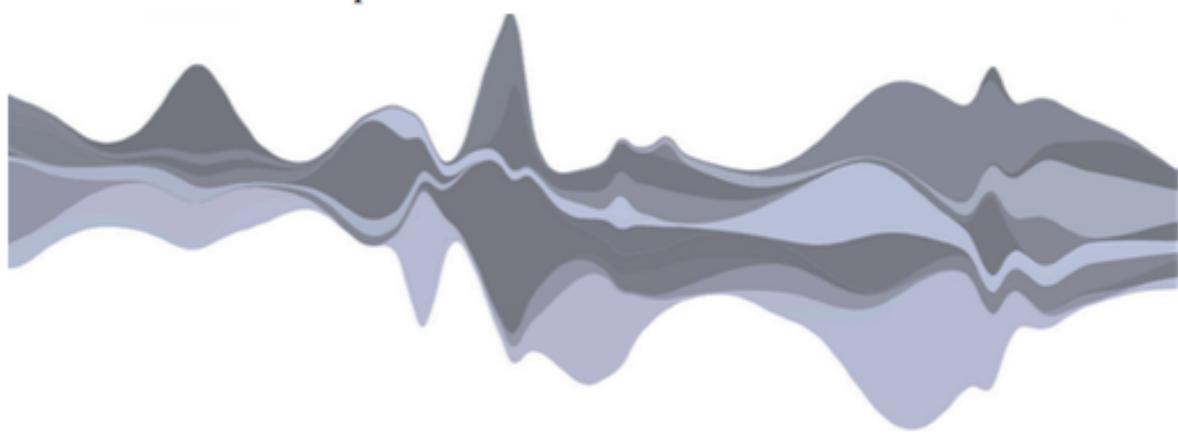
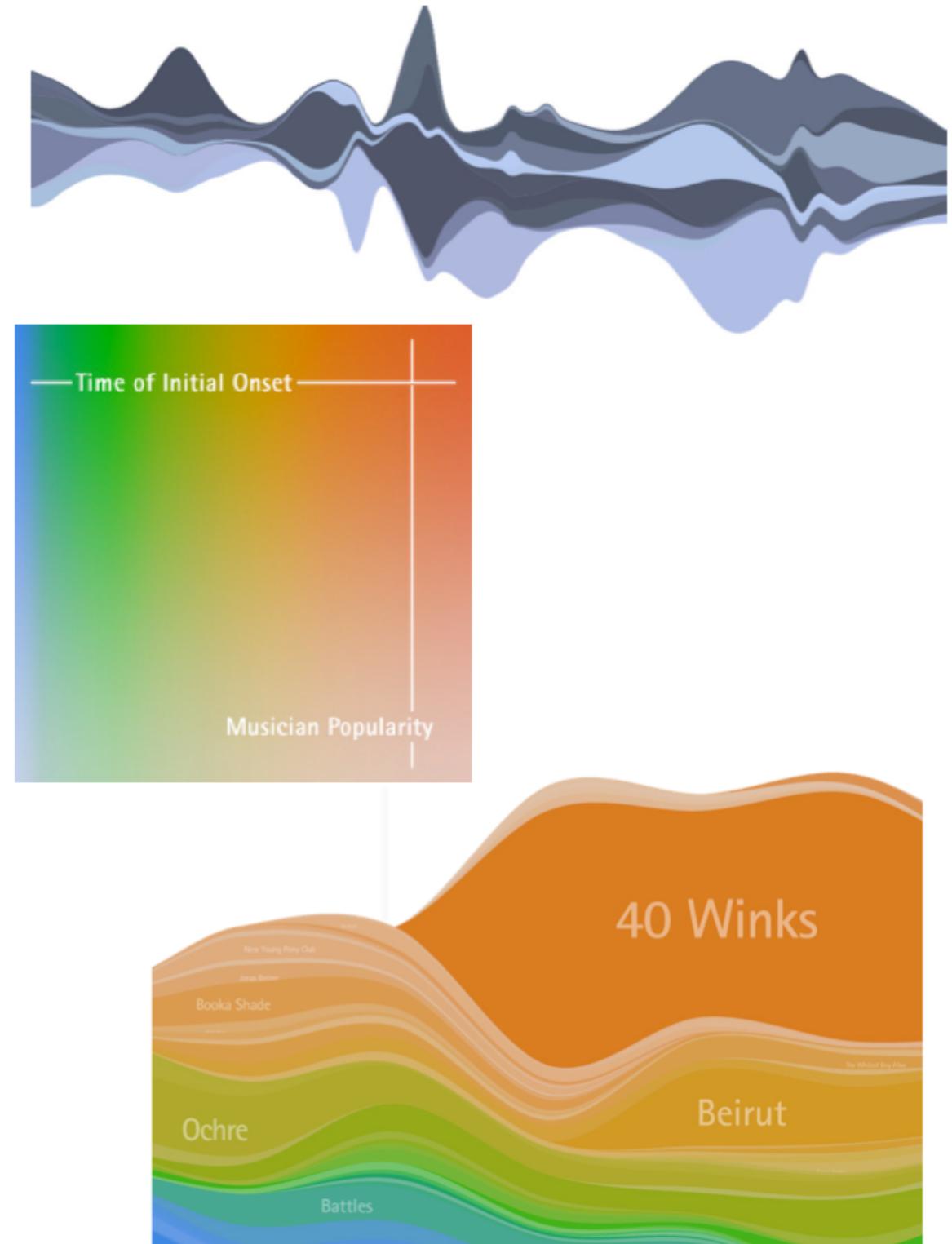


fig 8 – the same data set optimized to reduce the “weighted_wiggle,” the algorithm used in Streamgraph

- Visual representation: stacked graph
 - New technique for minimizing “wiggle of layers”
- Color: 2D colormap
 - Hue: time of onset
 - Saturation: popularity
- Labels
 - placed where embedded labels can be largest



- Layer ordering is inside-out
 - Avoids diagonal striping effect
 - Bursts are on outside which minimizes effect on other layers
 - Prevents drift away from x-axis



fig 12 – an unsorted data set, exhibiting the type of “burstiness” apparent in last fm and box office data sets

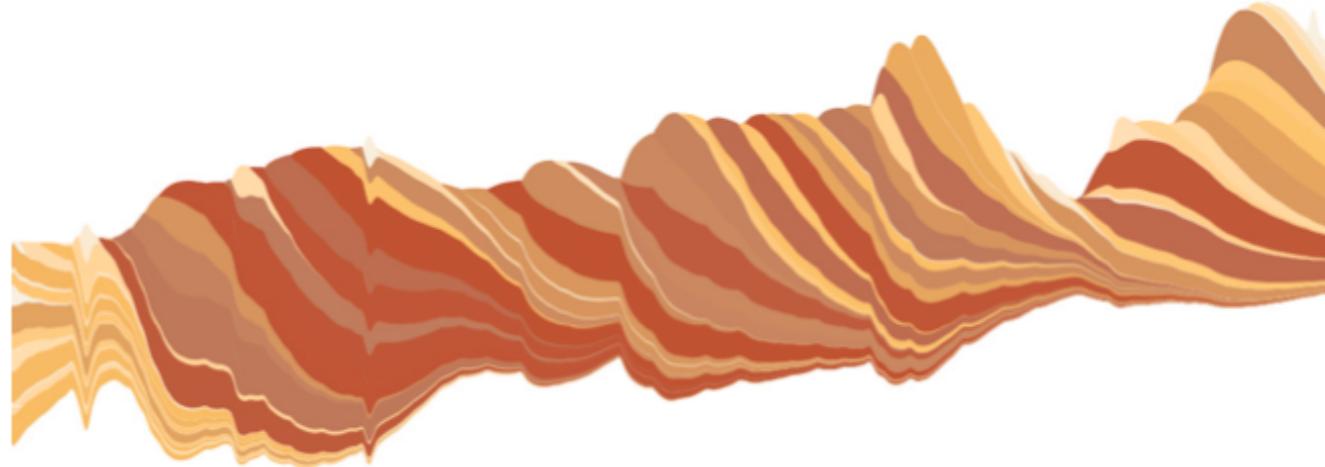


fig 13 – the same data set, naively sorted in order of “onset time” exhibiting the distracting diagonal striping effect



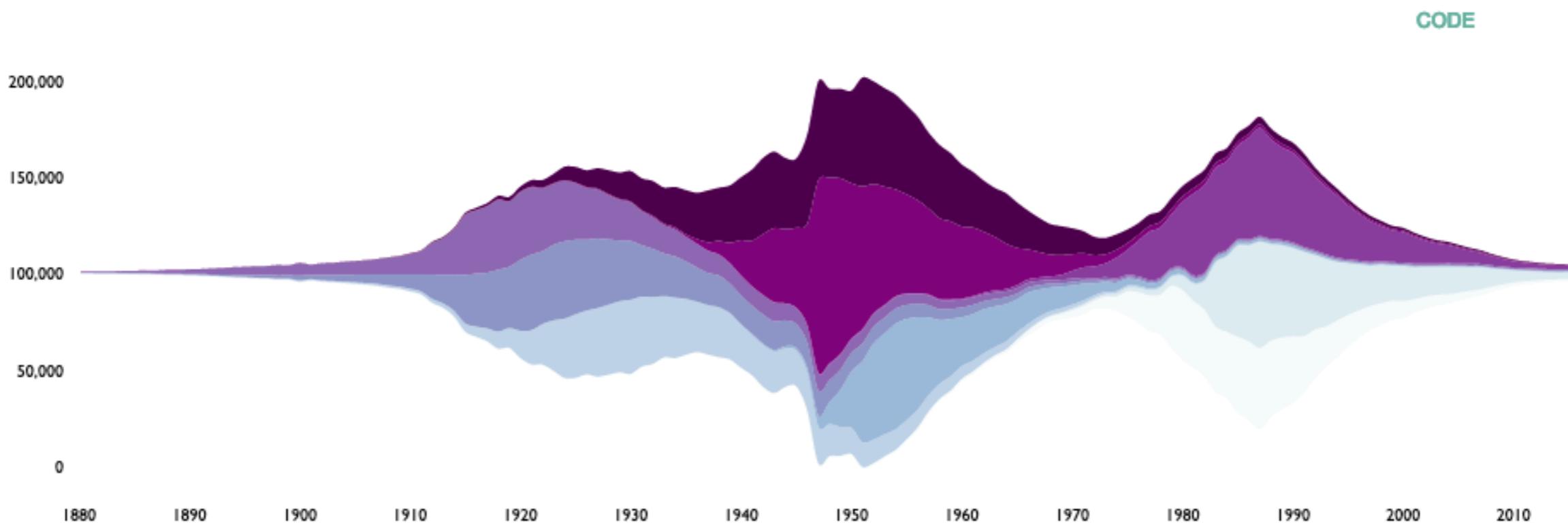
fig 14 – the same data set sorted using the weighted “inside out” strategy to highlight the initial onset of each time series

Definition

A Stream graph is a type of [stacked area chart](#). It displays the evolution of a numeric value (Y axis) following another numeric value (X axis). This evolution is represented for several groups, all with a distinct color.

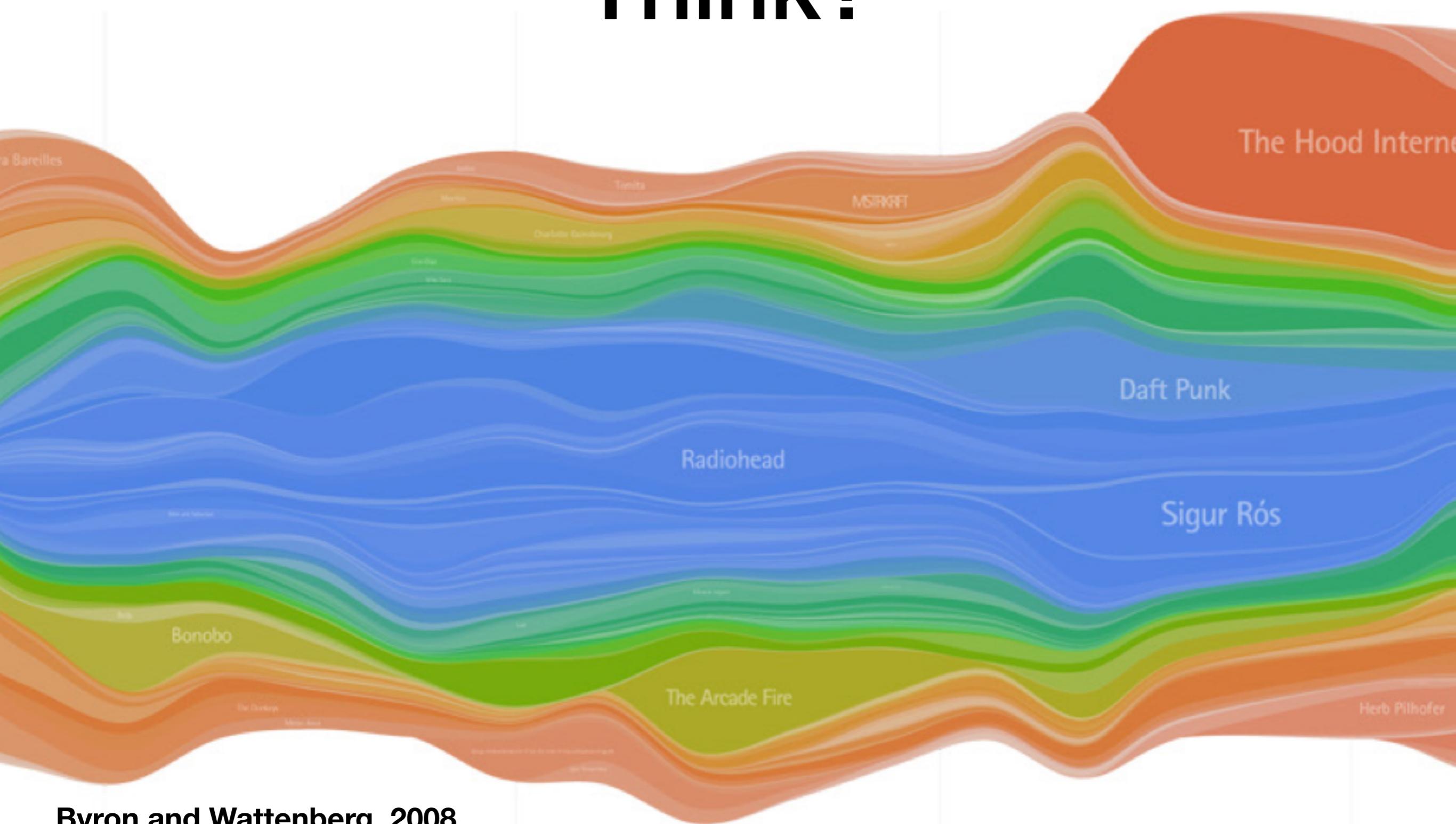
Contrary to a stacked area, there is no corner: edges are rounded what gives this nice impression of flow. Moreover, areas are usually displaced around a central axis, resulting in a flowing and organic shape.

The following example shows the evolution of baby name frequencies in the US between 1880 and 2015.



Note: The dataset is available through the [babynames](#) R library and a [.csv](#) version is available on [github](#).

Critique: What Do You Think?



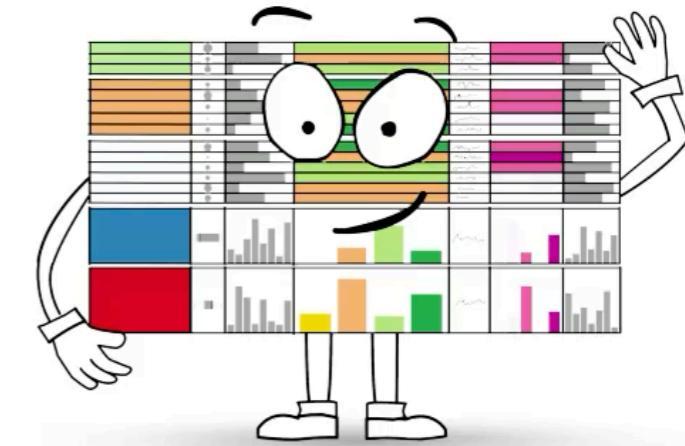
Byron and Wattenberg. 2008

Tabular Representation of Tables



Taggle

Scalable Visualization of Tabular
Data through Aggregation



Katarina Furmanova, Samuel Gratzl, Holger Stitz, Thomas Zichner,
Miroslava Jaresova, Martin Ennemoser, Alexander Lex, and Marc Streit



datavisyn

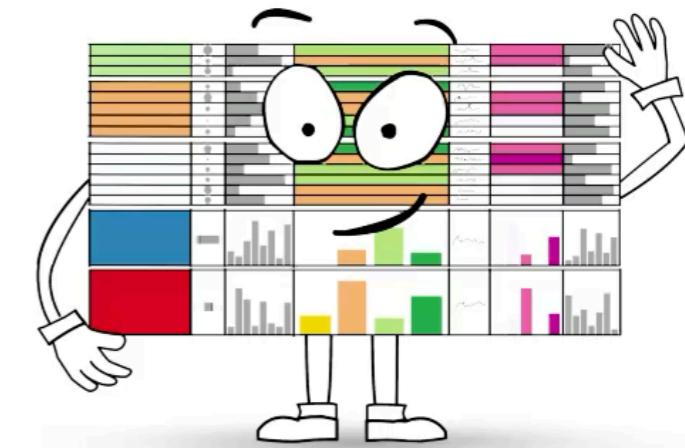


Tabular Representation of Tables



Taggle

Scalable Visualization of Tabular
Data through Aggregation



Katarina Furmanova, Samuel Gratzl, Holger Stitz, Thomas Zichner,
Miroslava Jaresova, Martin Ennemoser, Alexander Lex, and Marc Streit



datavisyn



Taggle

- Problem: scalable visual encoding of multi-attribute tables
- Abstraction: create a tabular view that supports column manipulation
- Design considerations
 - Allow sorting, grouping, and aggregation
 - Multiform visualization : Selection of variety of different plots per column
 - Overview + detail layout.
 - Strong focus on interactivity for exploring and modifying derived quantities

Critique: What Do You Think?



Spatial Axis Orientation

④ Axis Orientation

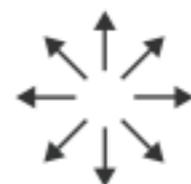
→ Rectilinear



→ Parallel



→ Radial



SPLOMs: Scatterplot Matrices

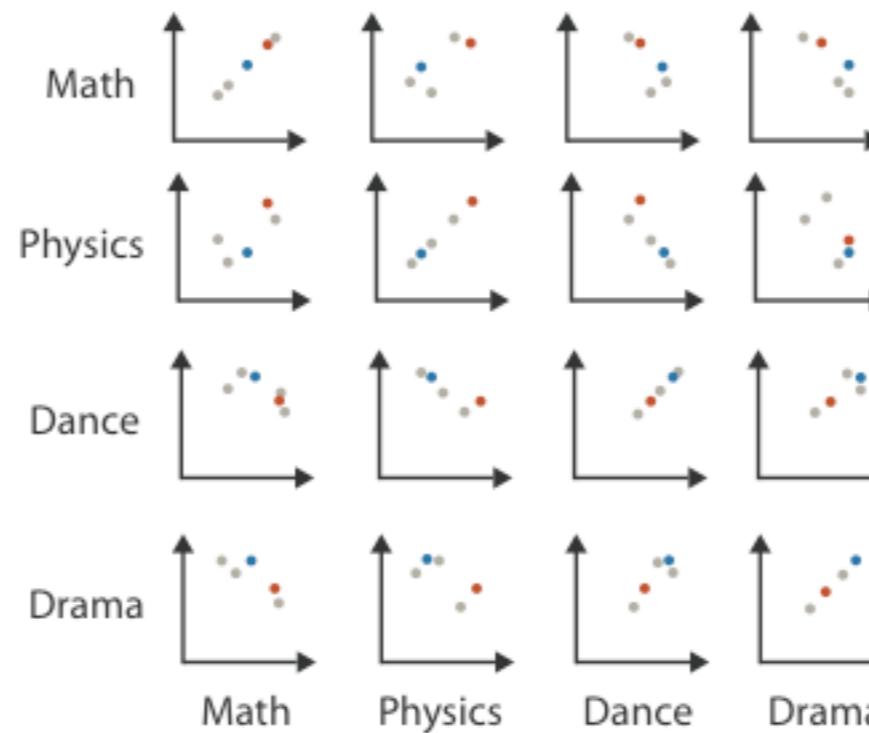
Nine Characteristics of
Abalone (Sea Snails)

Parallel Layouts

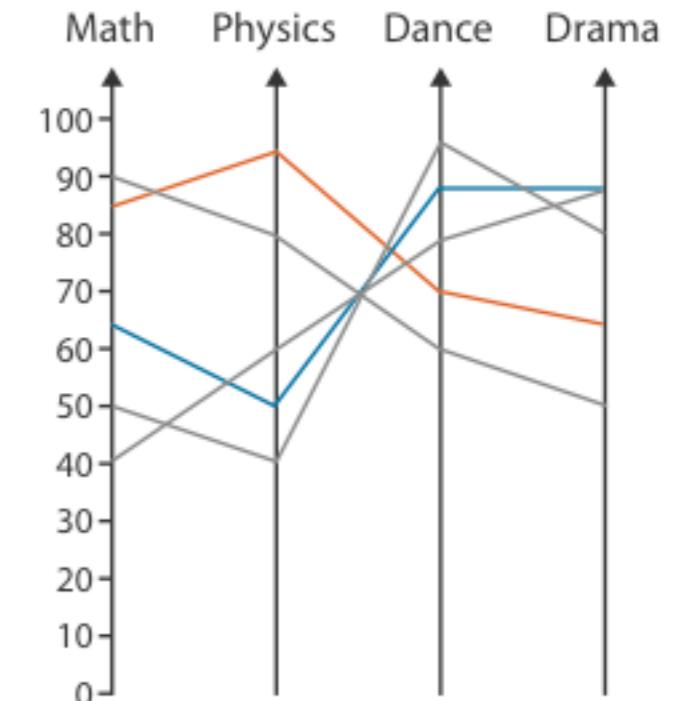
Table

Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

Scatterplot Matrix



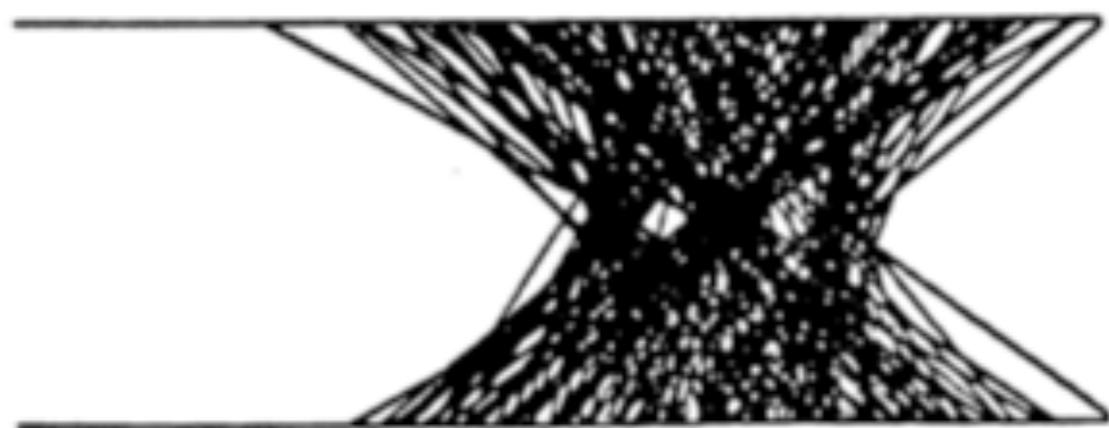
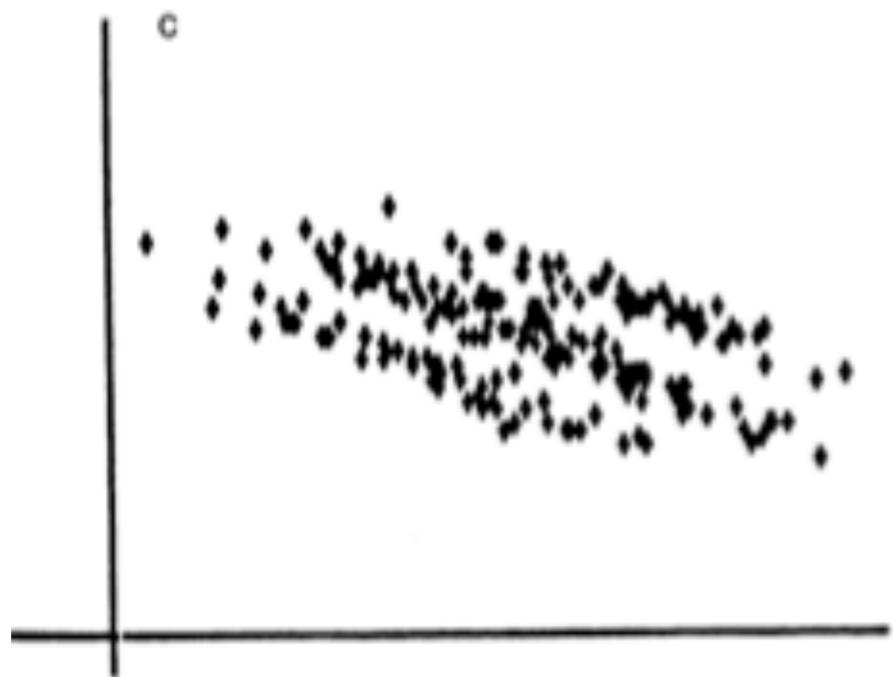
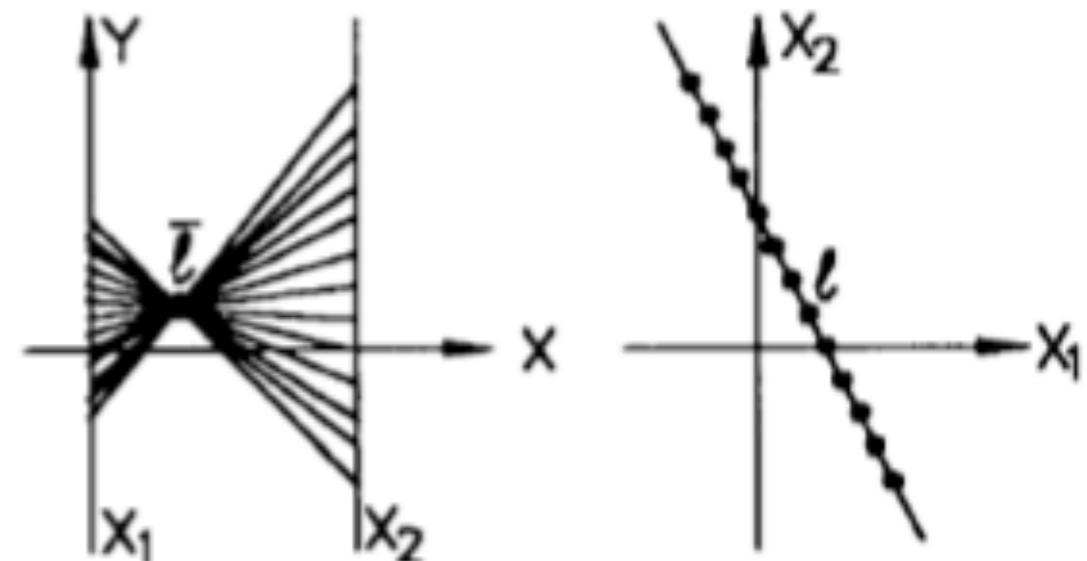
Parallel Coordinates



- Scatterplot limitation: visual representation with orthogonal axes can only show two attributes with spatial position channel
- Alternatively, use **parallel coordinates**: line up axes in parallel to show many attributes with position

Parallel Coordinates Duality

- Points becomes line strips
- Rotations are translations



Parallel Coordinates Tasks

- Goal: Show correlation
 - Positive correlation: straight lines
 - Negative correlation: all lines cross at a single point

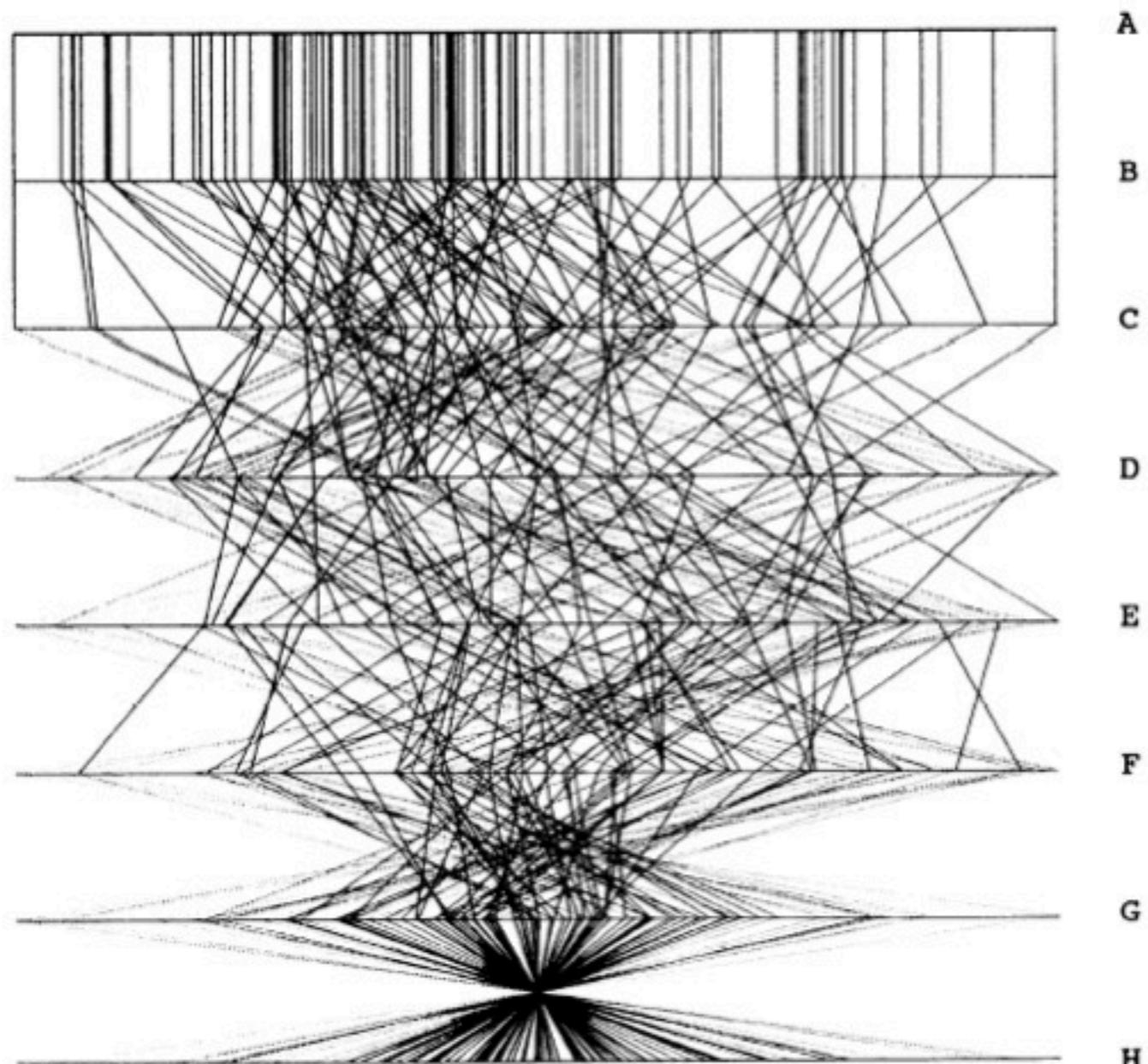
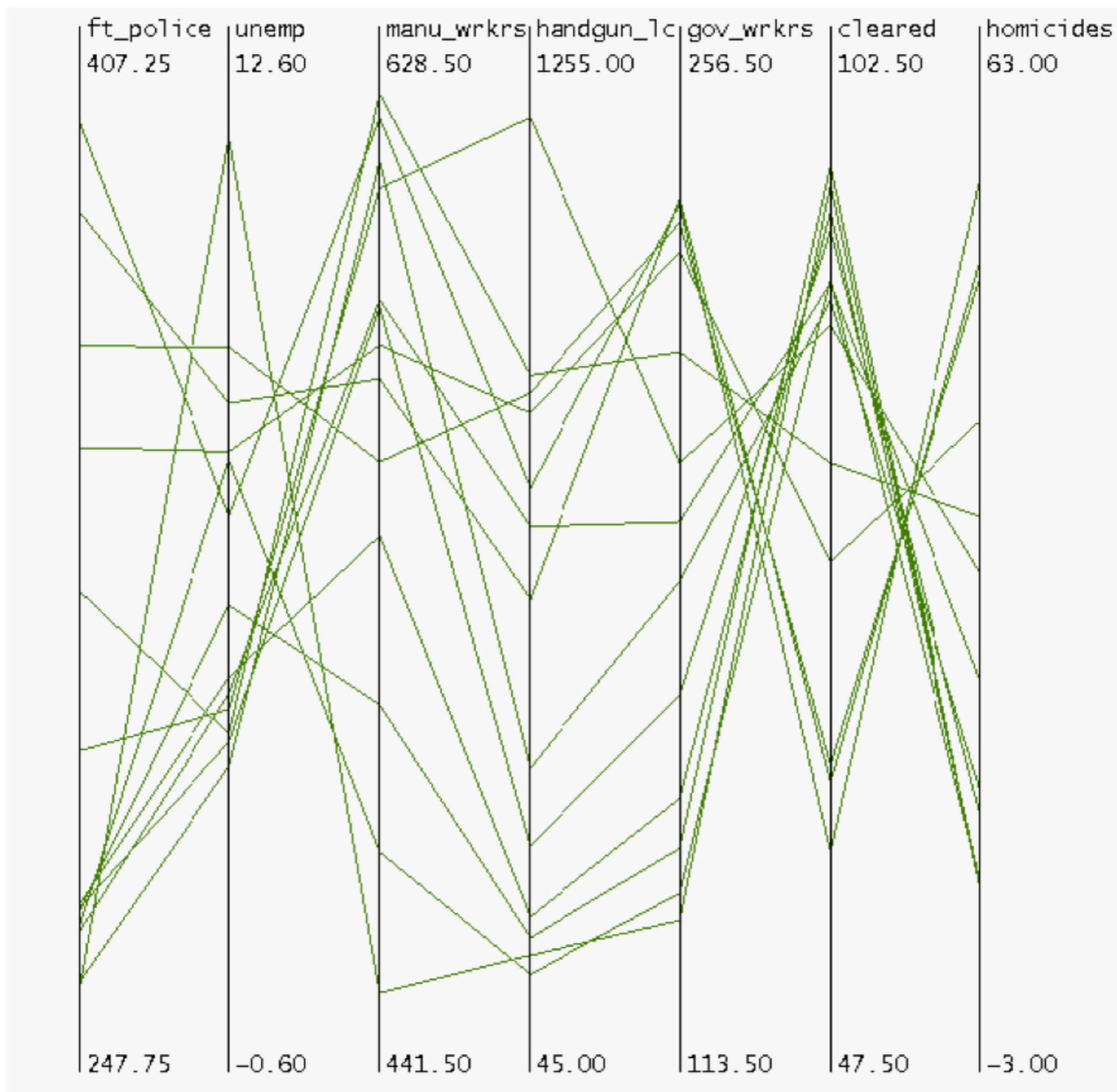


Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of $\rho = 1, .8, .2, 0, -.2, -.8, \text{ and } -1$.

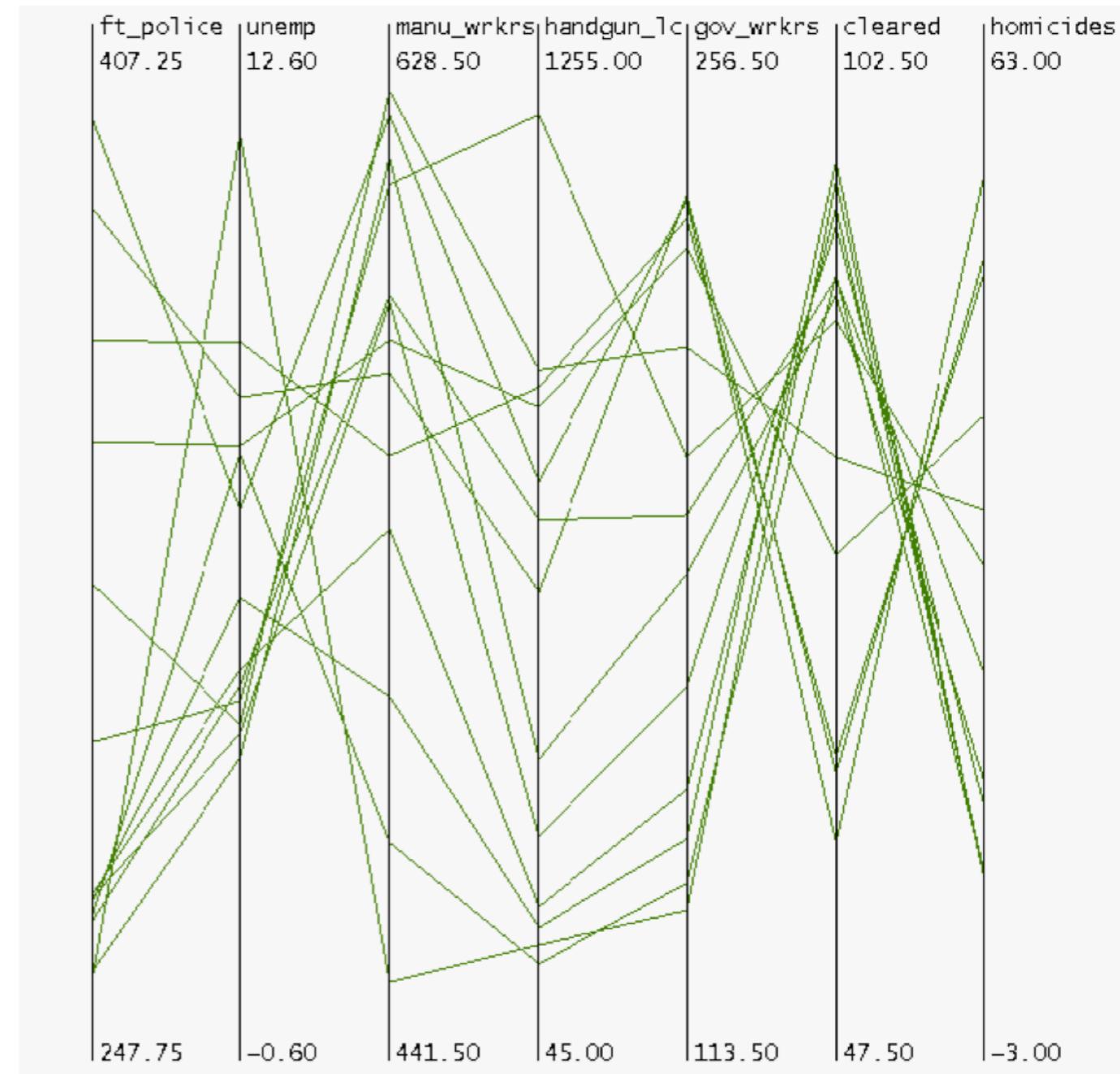


Do you see any correlations?

Fua et al. 1999

PC Tasks

- Visible patterns only between neighboring axis pairs
- How to pick axis order?
- Usual solution: reorderable axes, interactive exploration
 - Same weakness as many other techniques
 - Downside: human-powered search



Another Example

- The Dataset:
 - Production data for 473 batches of a VLSI chip
 - 16 process parameters:
 - X1: The yield = % of produced chips that are useful
 - X2: The quality of the produced chips (speed)
 - X3-X12: 10 types of defects (zero defects shown at top)
 - X13-X16: 4 physical parameters
 - The Objective: Raise the yield (X1) and maintain high quality (X2)

PC Example

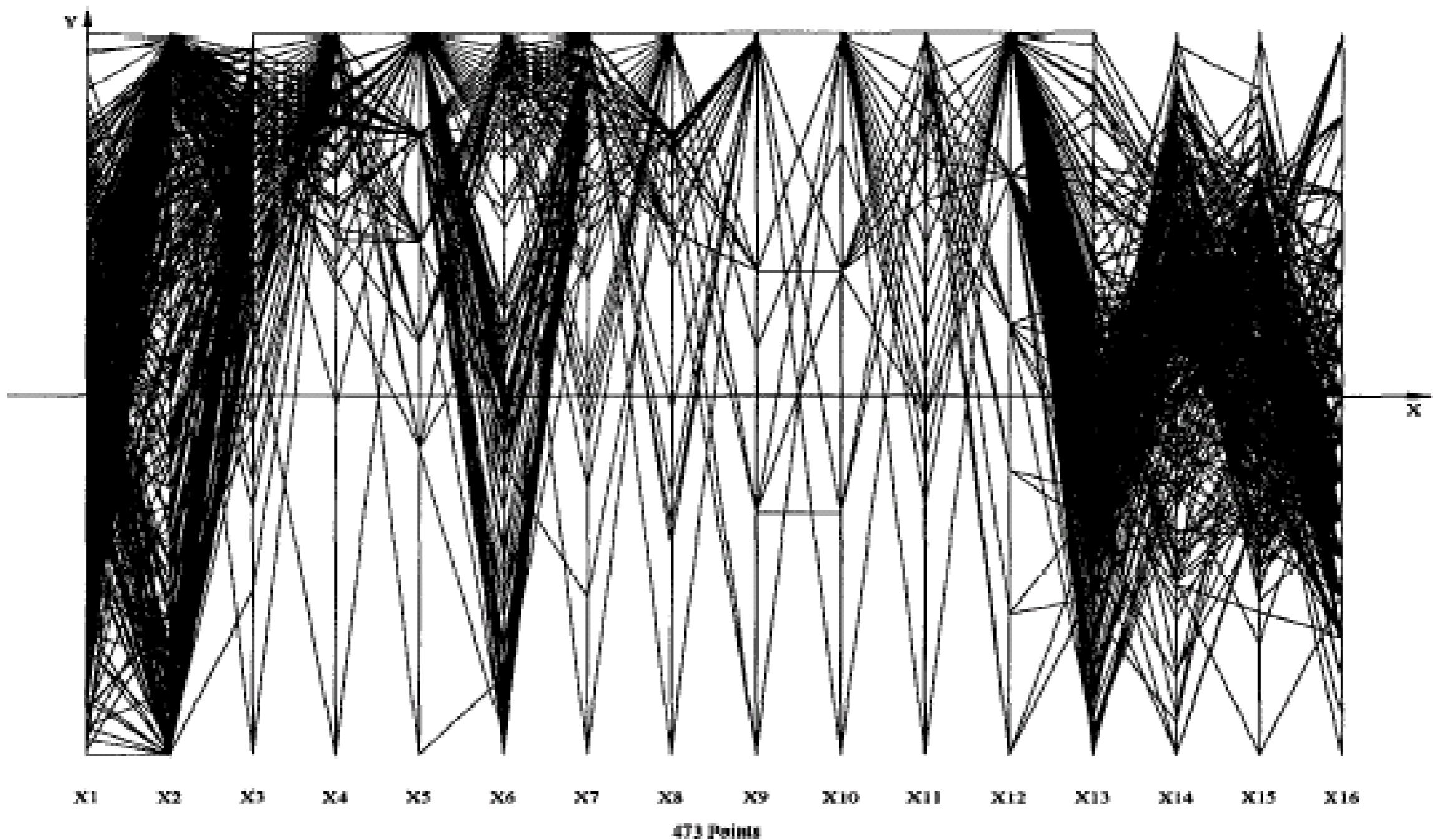


Figure 1: The full dataset consisting of 473 batches

Inselberg, 1997

Nutrient Explorer

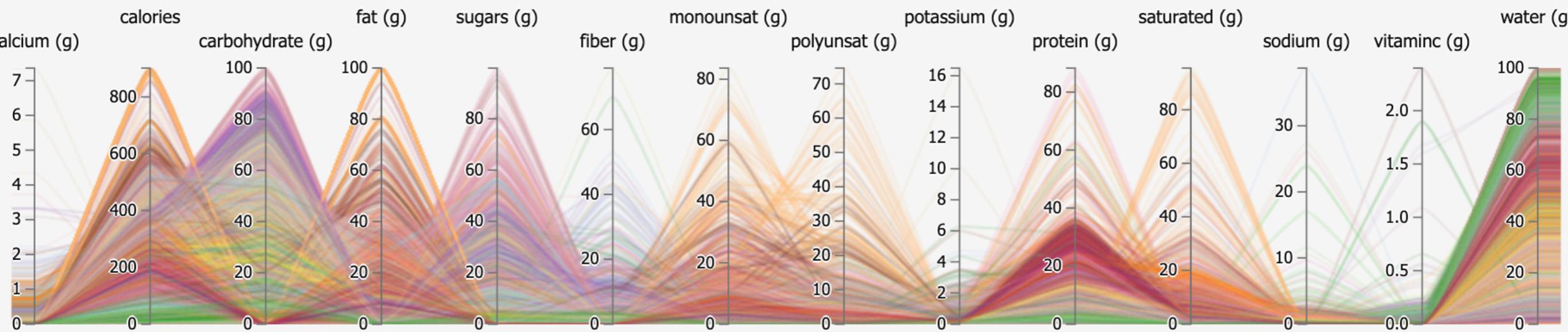
Keep Exclude Export

7637/7637

Lines at 13.6% opacity.

Hide Ticks

Dark



What is this?

A multidimensional explorer of [nutrient data](#) from the [USDA](#).

The parallel coordinates displays the nutrient content of foods in the database across 14 dimensions, colored by food group.

[Never heard of parallel coordinates? Read this tutorial.](#)

[Brush](#) the visualization to update other charts on this page.

Let me know what you think on [Reddit](#).

Controls

Brush: Drag vertically along an axis.

Remove Brush: Tap the axis background.

Reorder Axes: Drag a label horizontally.

Invert Axis: Tap an axis label.

Remove Axis: Drag axis label to the left edge.

[Credits & License](#)

Food Groups

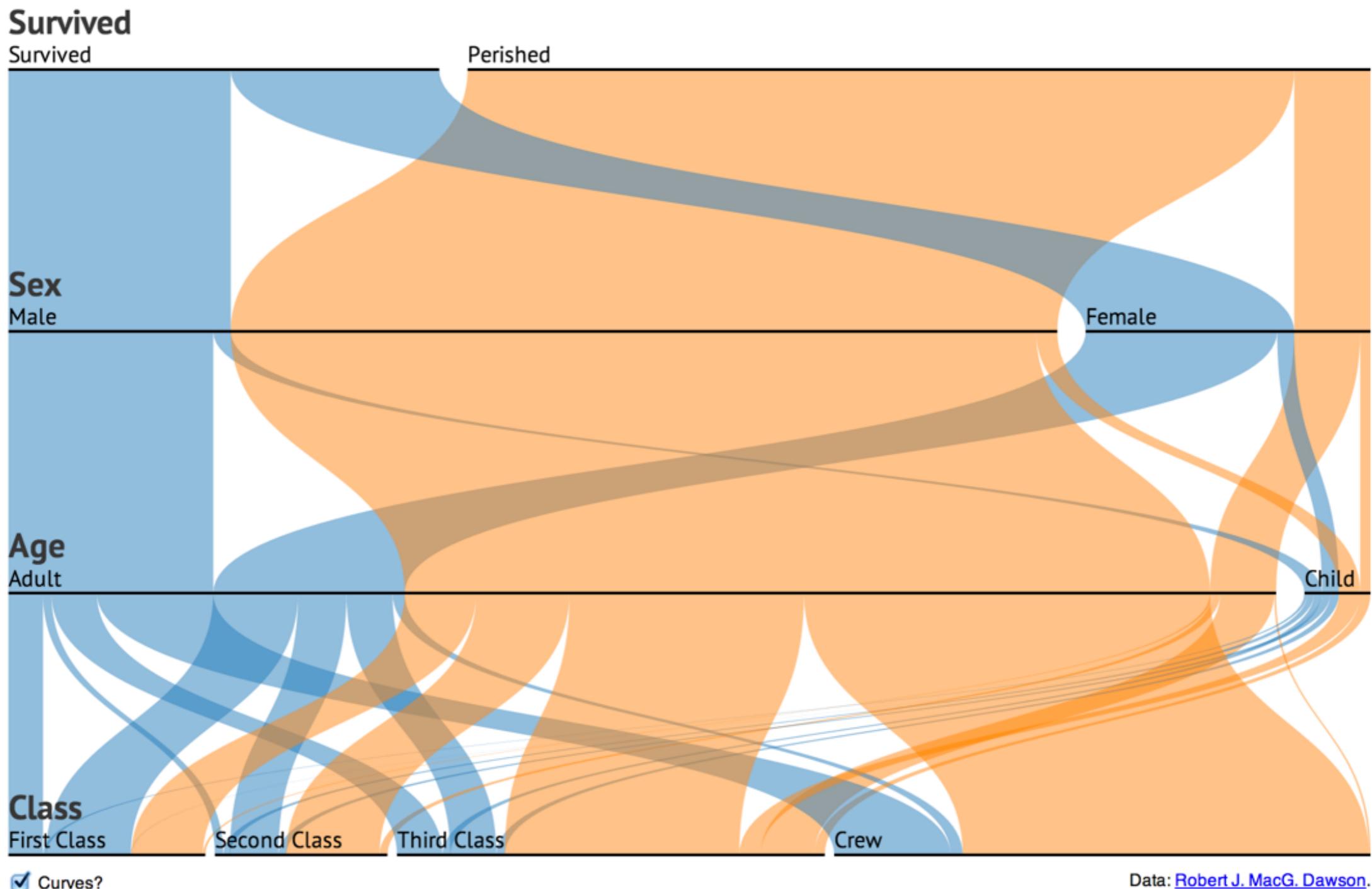
- 327 Baby Foods
- 496 Baked Products
- 619 Beef Products
- 278 Beverages
- 403 Breakfast Cereals
- 183 Cereal Grains and Pasta
- 239 Dairy and Egg Products
- 165 Ethnic Foods
- 365 Fast Foods
- 215 Fats and Oils
- 255 Finfish and Shellfish Products
- 328 Fruits and Fruit Juices
- 345 Lamb, Veal, and Game Products
- 365 Legumes and Legume Products
- 57 Meals, Entrees, and Sidedishes
- 128 Nut and Seed Products
- 328 Pork Products
- 381 Poultry Products
- 51 Restaurant Foods
- 234 Sausages and Luncheon Meats

Sample of 25 entries

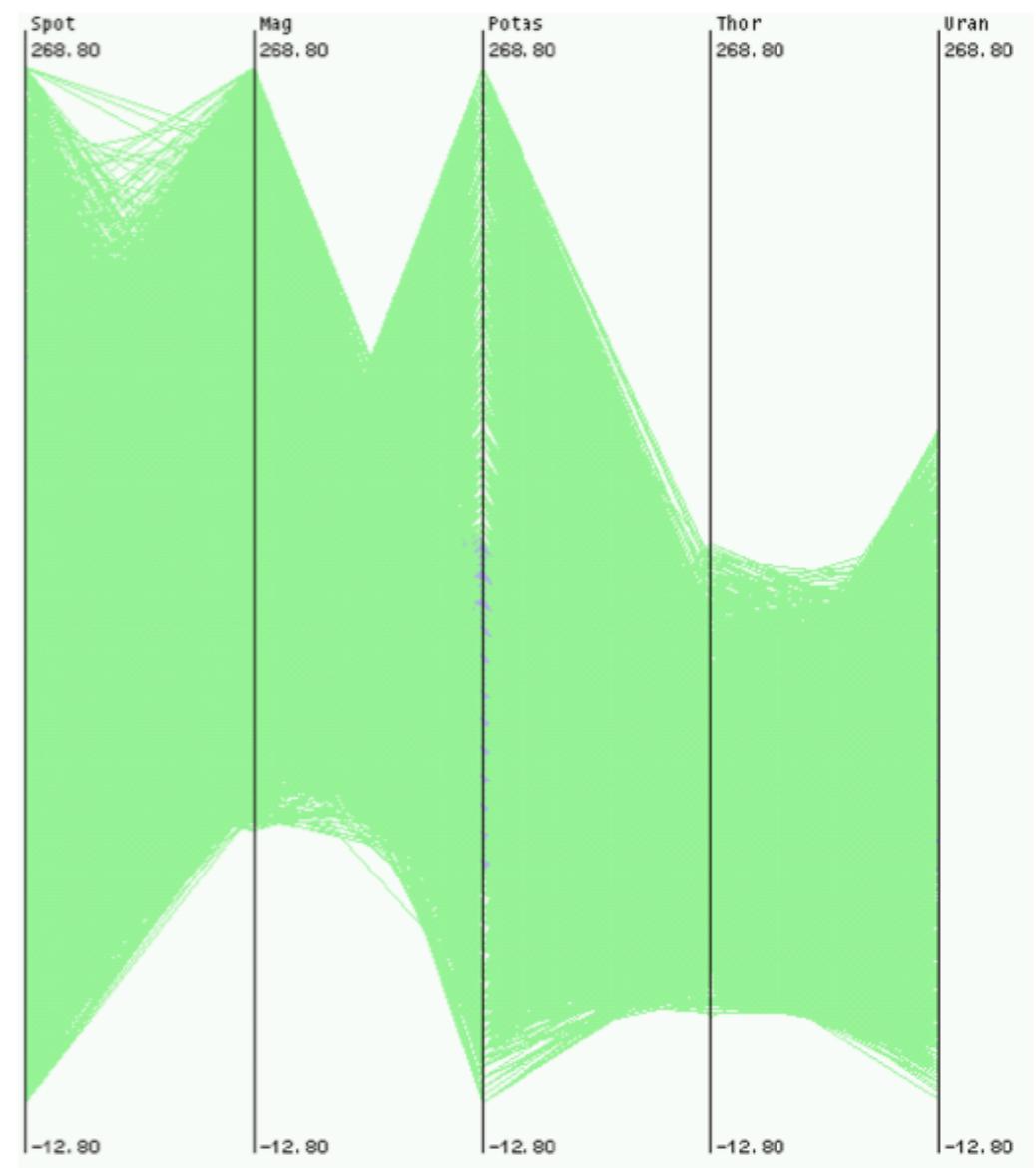
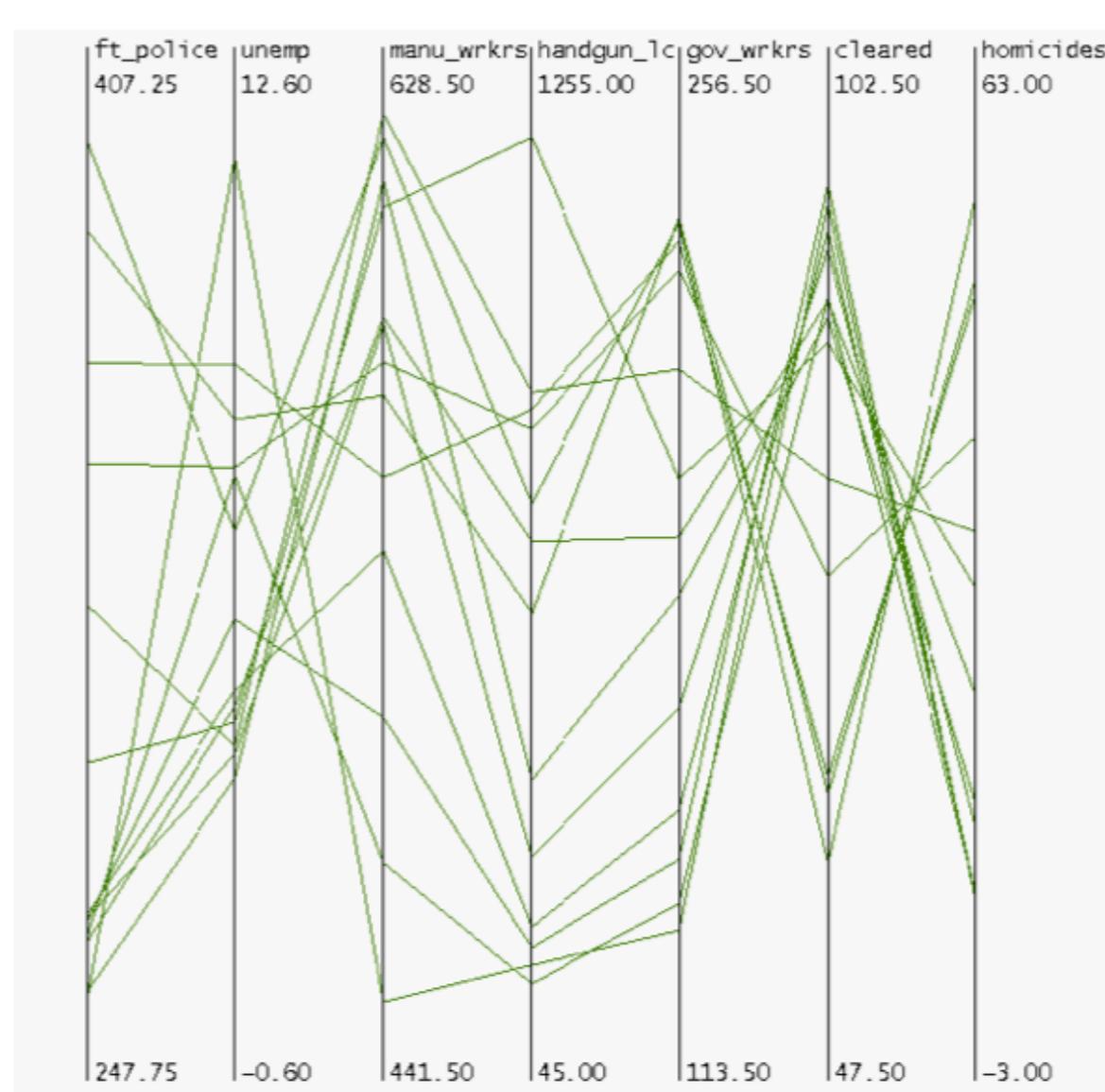
- Alcoholic beverage, beer, light, BUDWEISER SELECT
- Beans, french, mature seeds, cooked, boiled, with salt
- Beef, chuck, arm pot roast, separable lean only, trimmed to 0% fat, all grades
- Beef, short loin, t-bone steak, separable lean and fat, trimmings removed, all grades
- Beef, top sirloin, steak, separable lean only, trimmed to 0% fat, all grades
- Cereals, KASHI, KASHI GO LEAN Hot Cereal, Truly Vanilla
- Cheese, port de salut
- Chicken, broilers or fryers, back, meat and skin, cooked, with bones
- Chicken, broilers or fryers, wing, meat and skin, cooked, with bones
- Cowpeas (Blackeyes), immature seeds, frozen, unprepared
- Crackers, milk
- Fast foods, burrito, with beans, cheese, and chili pepper
- Ice creams, BREYERS, All Natural Light French Chocolate
- Lamb, domestic, rib, separable lean and fat, trimmed to 0% fat, all grades
- Lemonade, frozen concentrate, white, prepared with water
- Nuts, almonds, dry roasted, with salt added
- Peas, green, canned, no salt added, solids and liquids
- Pickles, cucumber, sweet (includes bread and butter pickles)
- Pork, fresh, loin, top loin (chops), boneless, separable lean only
- Pork, fresh, variety meats and by-products, pancreas, raw

Parallel Sets

Titanic Survivors

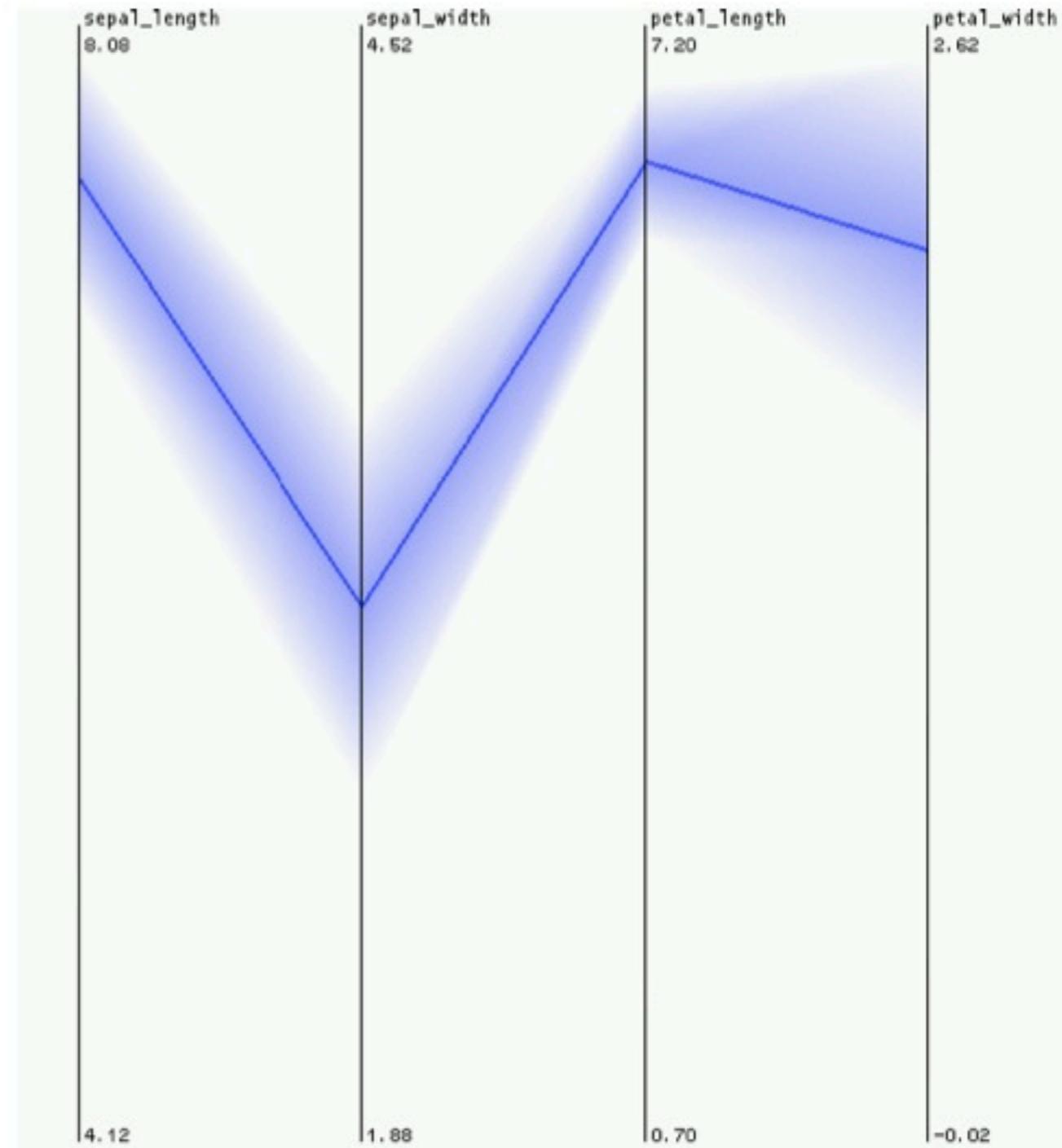


Hierarchical Parallel Coordinates



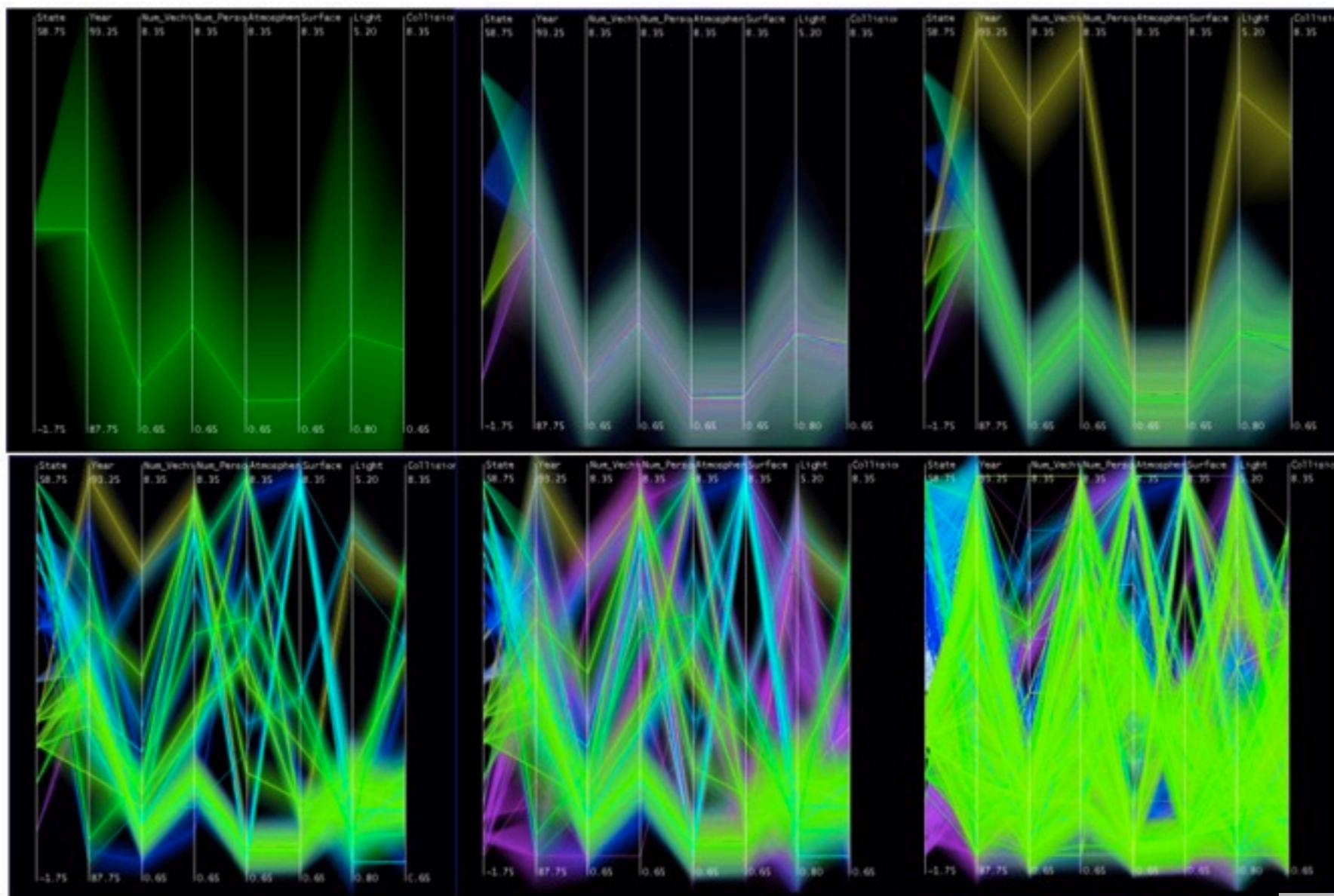
HPC: Encoding Derived Data

- Visual representation:
variable-width opacity
bands
 - Show whole cluster, not
just single item
 - Min / max: spatial
position
 - Cluster density:
transparency
 - Mean: opaque



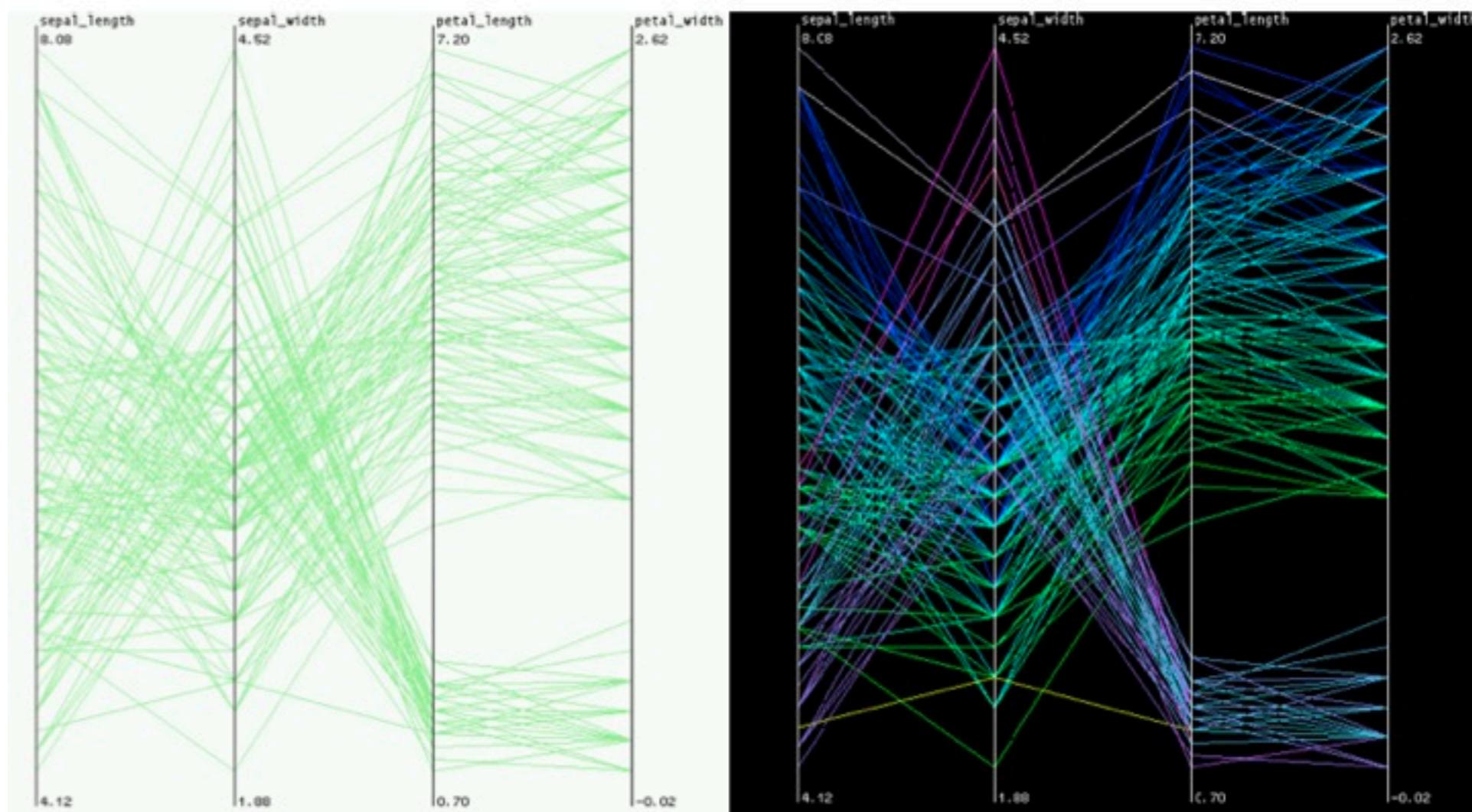
HPC: Interaction

- Intuitively change level of detail to navigate cluster hierarchy



HPC: Encoding Derived Data

- Visual encoding: color based on cluster proximity in the derived attribute
 - Resolves ambiguity from crossing, clarifies structure



HPC: Magnification Interaction

- Dimensional zooming: use all available space
 - Linked views to show true extent
 - Overview + detail to maintain context

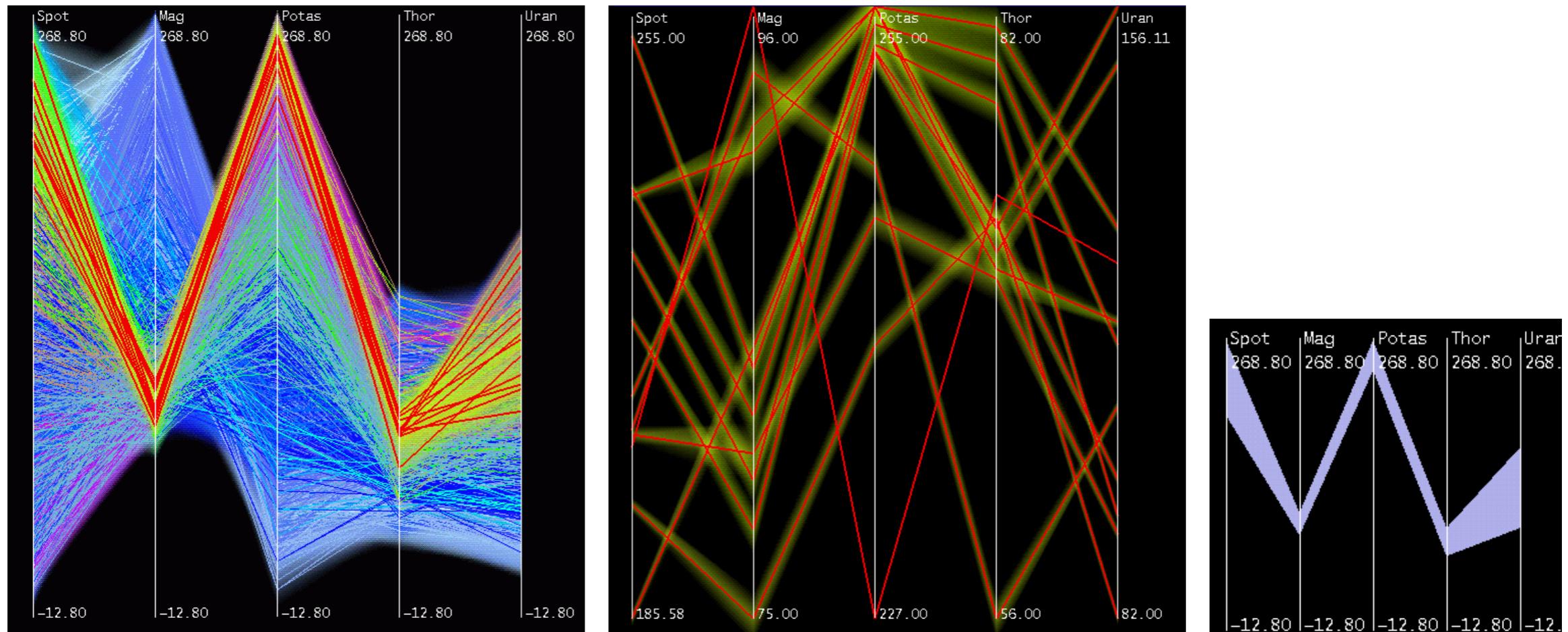
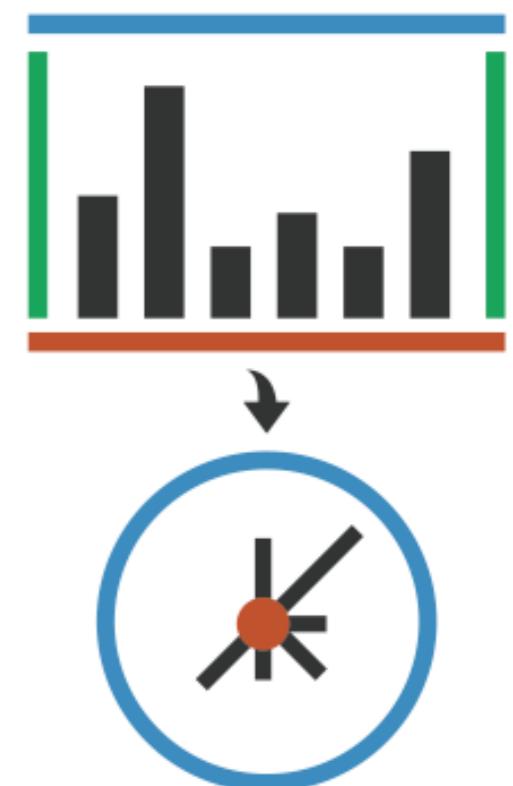
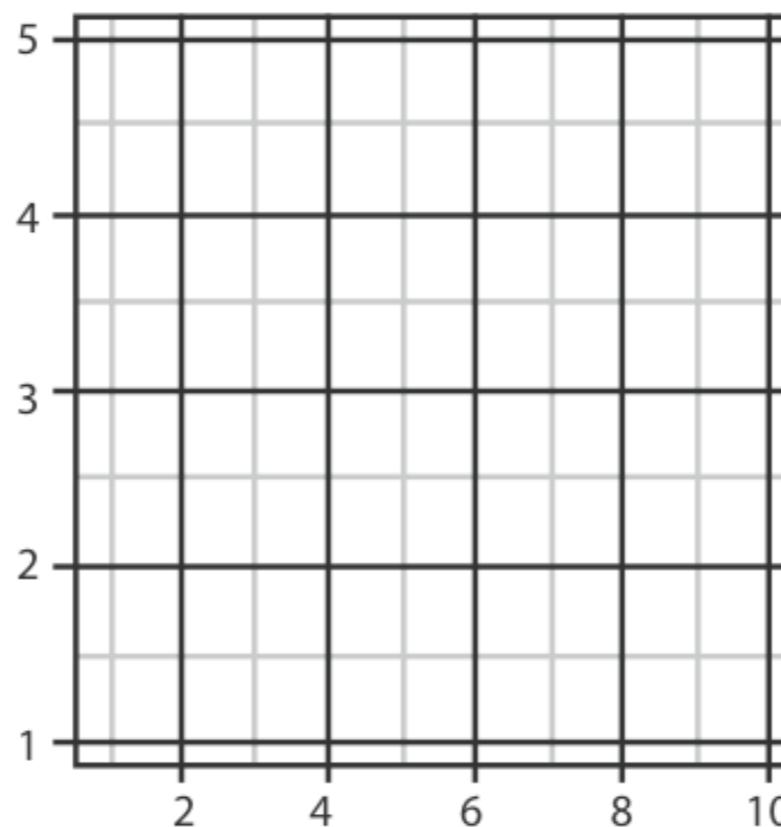
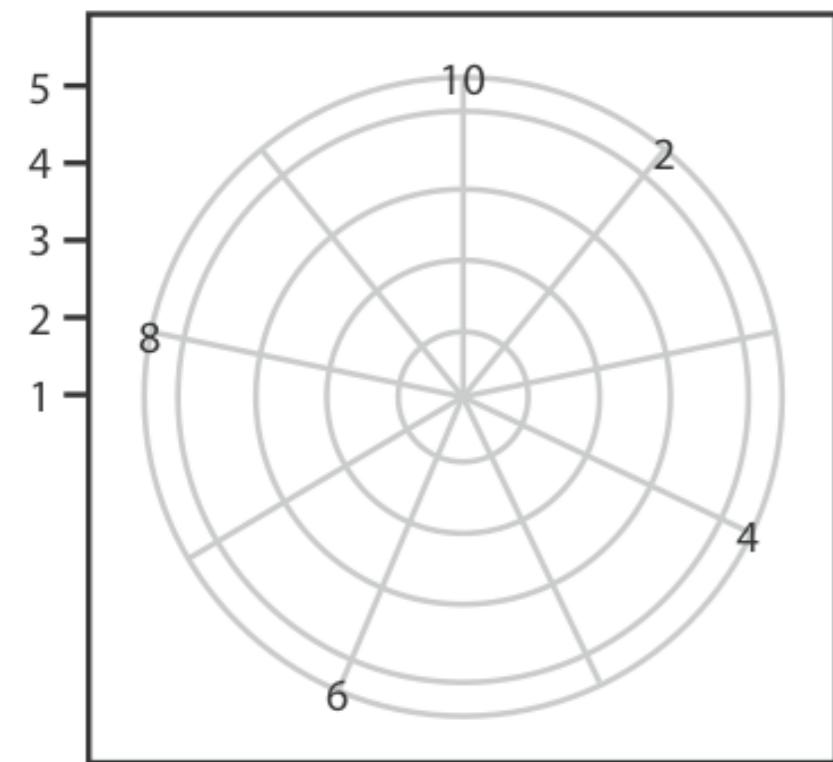


Figure 8: The image in the middle shows a magnified view of the brushed region indicated by the red lines in the leftmost image and accompanying mini-map that captures the location of the brush with respect to the entire data space. (See Color Plates).

Radial Layouts

Radial Layouts Use Polar Coordinates

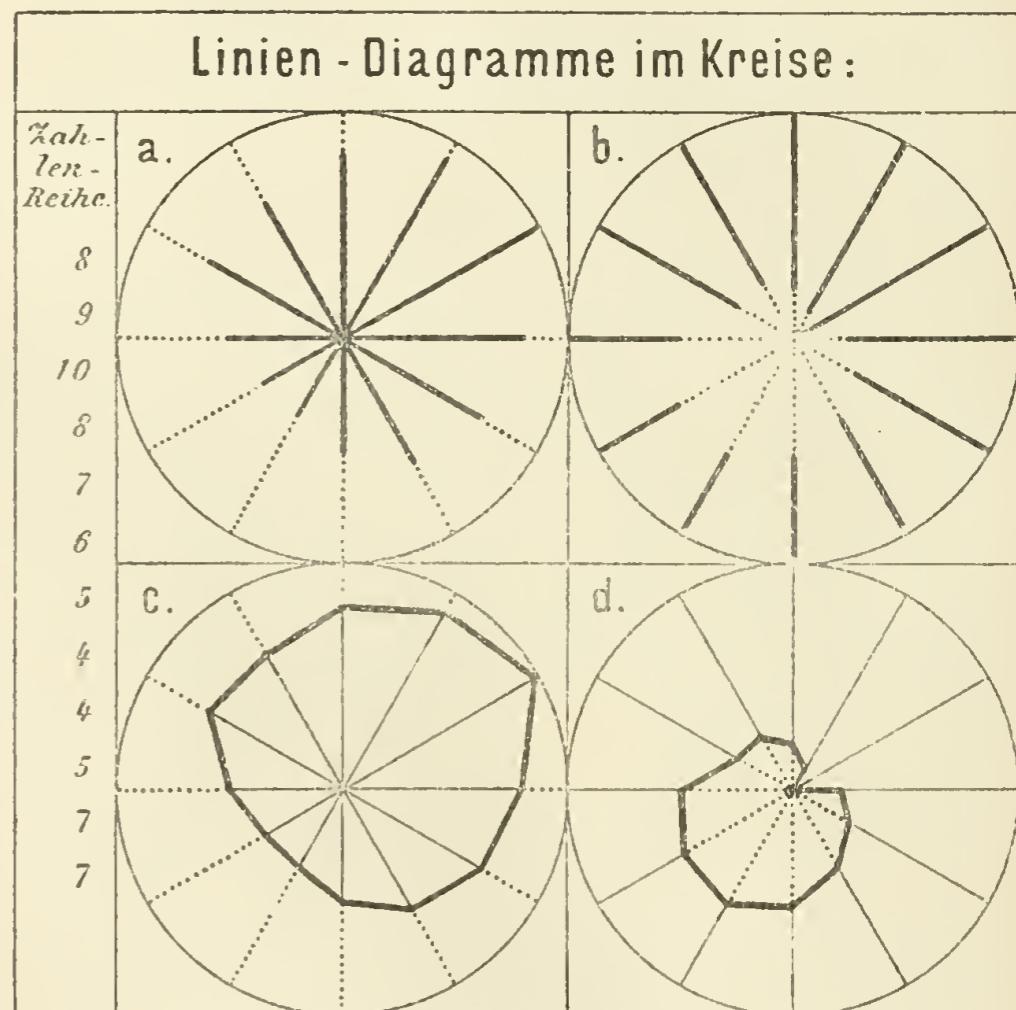


Radar Plot / Star Graph

- “Parallel” dimensions in polar coordinate space
- Best if same units apply to each axis

Zahlenergebnissen proportional ist. Auch können Verlängerungen der Radien über die Peripherie hinaus hiezu benutzt werden. Zweckmäßig wird auch hier die lineare Verbindung der Endpunkte der betreffenden Geraden vorgenommen.

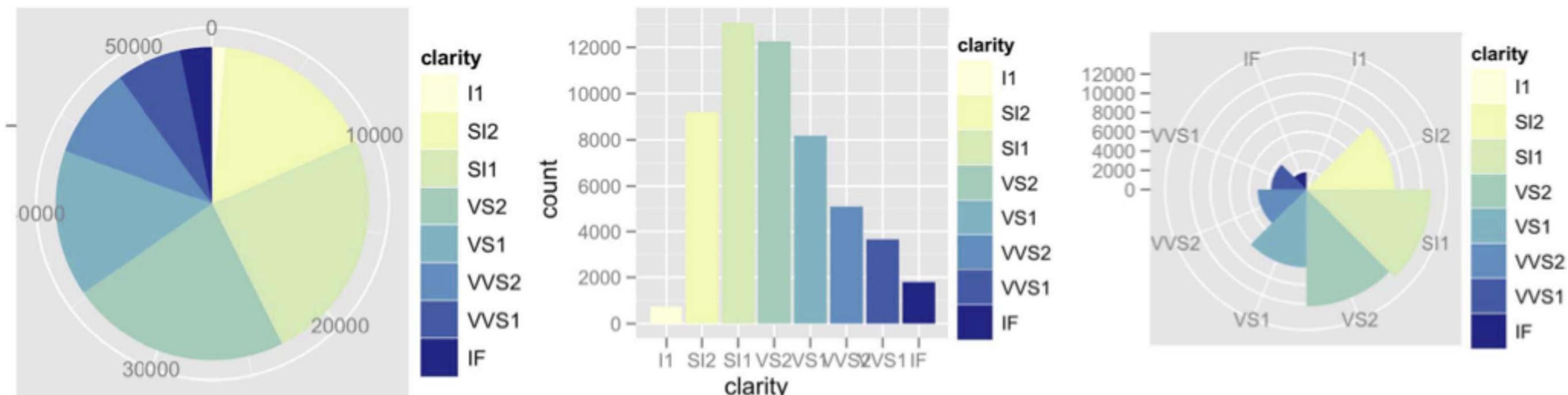
Beispiele von Linien-Diagrammen im Kreise sind in der folgenden Fig. 4 gegeben. Bei a und c bildet der Mittelpunkt, bei b und d die Peripherie den Ausgangspunkt der



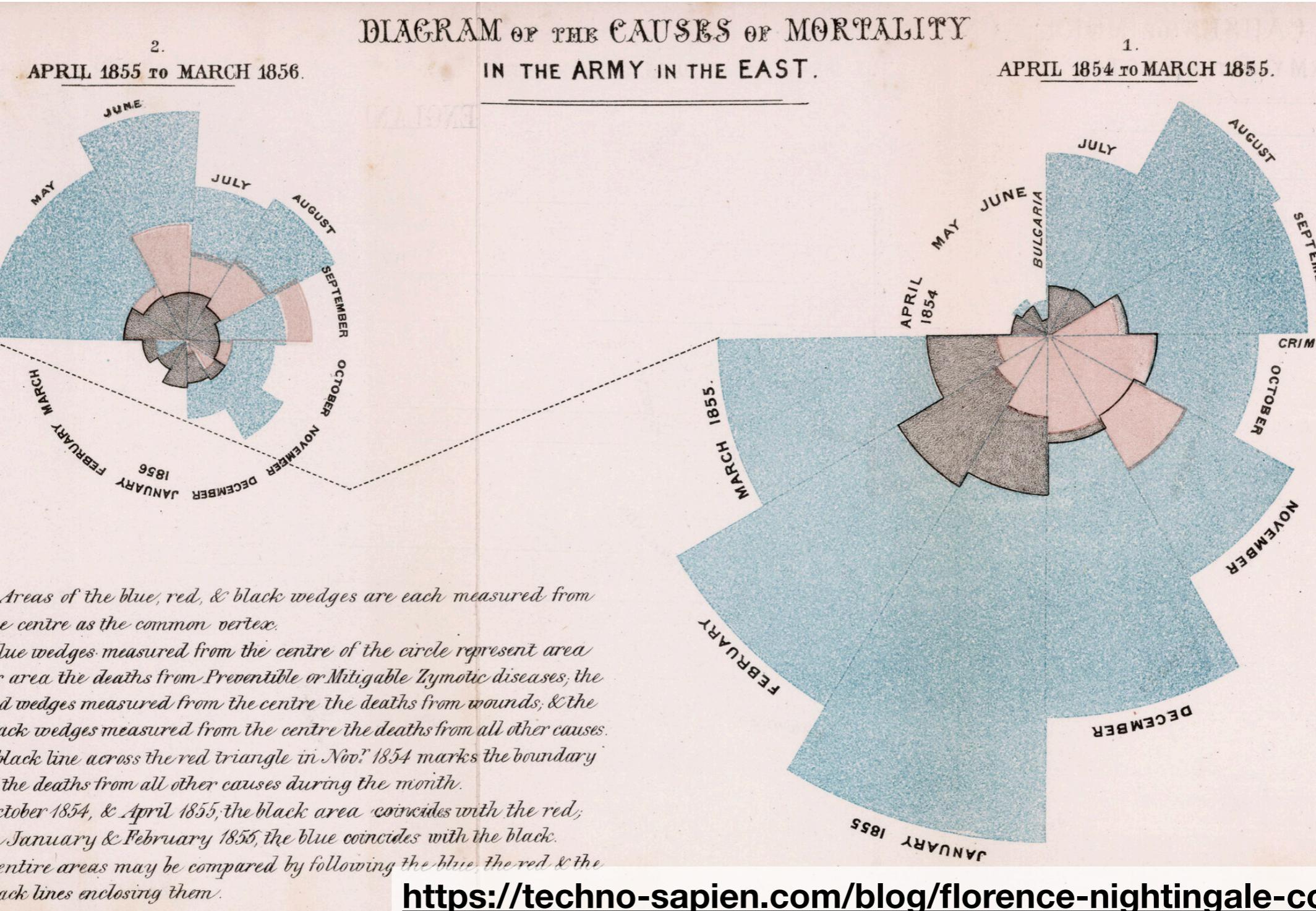
Figur 4.

Geraden, welche als Radientheile von differenter Größe die Zahlenverschiedenheiten der statistischen Reihe darstellen. Bei a und b ist die Veranschaulichung lediglich durch

Pie Charts: Take Care with Accuracy



Better: Florence Nightingale's "Coxcomb"



Better: Florence Nightingale's “Coxcomb”



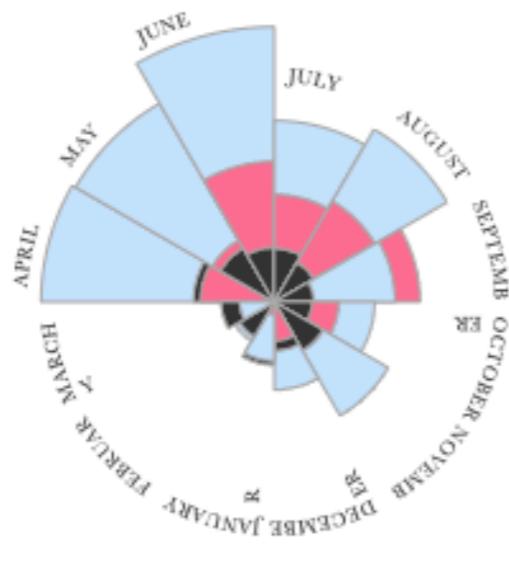
Athan's Block 5926740
Updated March 18, 2020

[Popular](#) / [About](#)

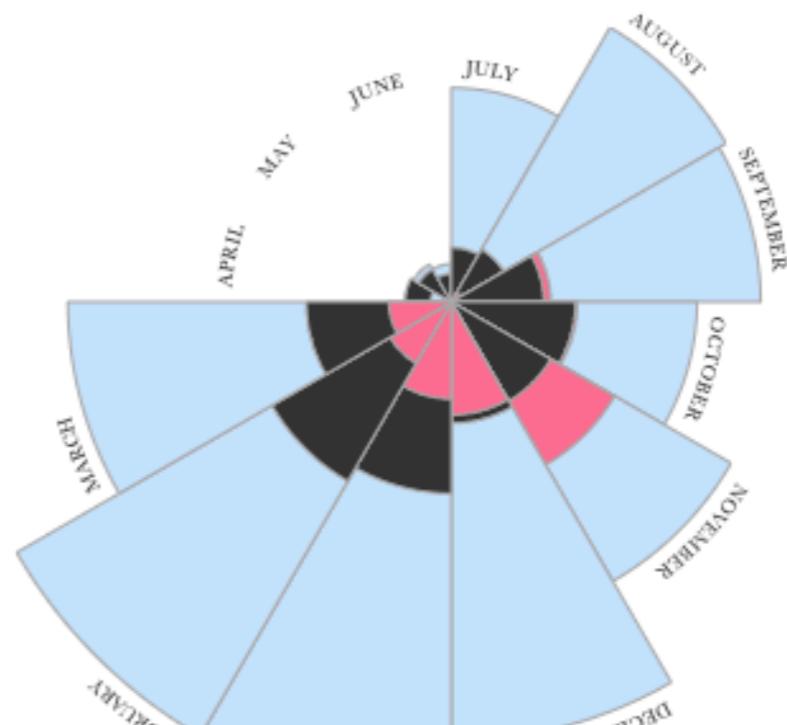
Nightingale's Rose + D3.js

DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY OF THE EAST

APRIL 1855 - MARCH 1856



APRIL 1854 - MARCH 1855



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex

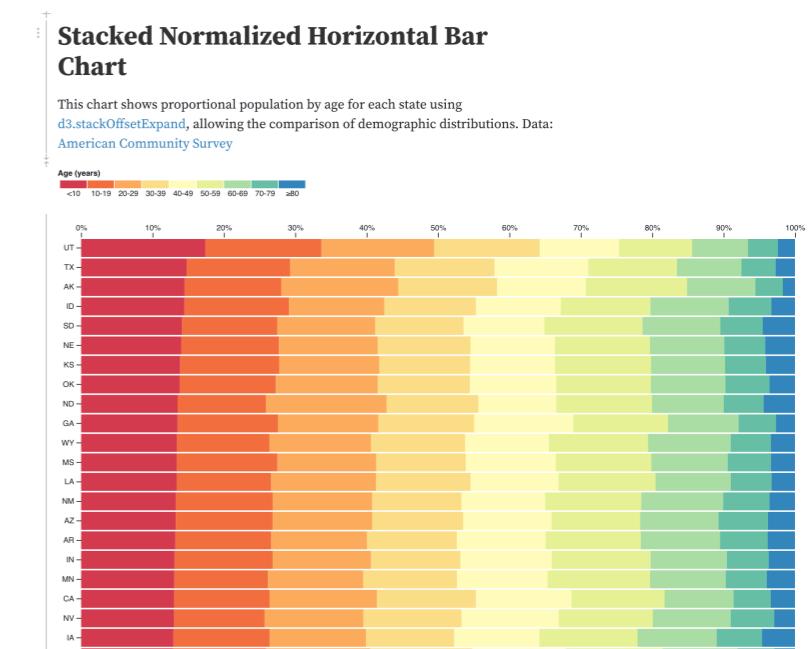
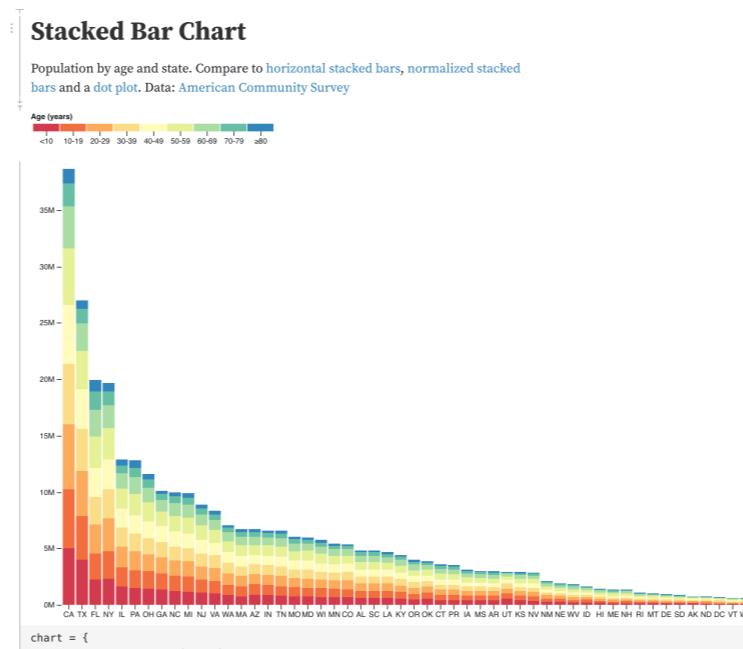
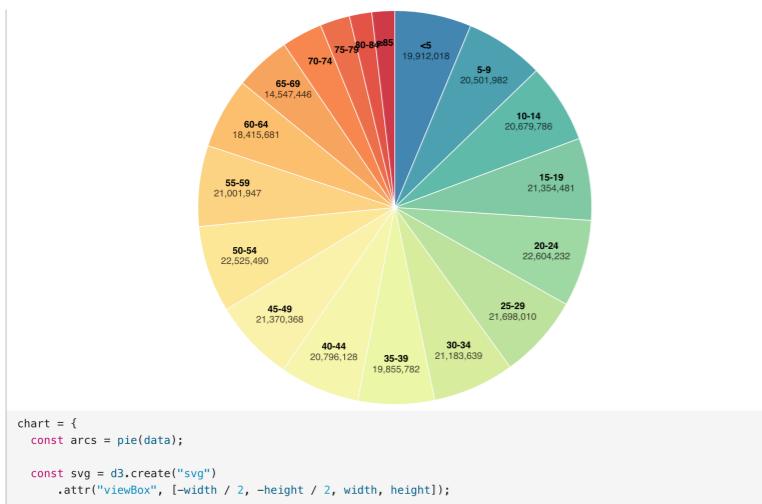
The blue wedges measured from the centre of the circle represent area for

Pie Charts: Contributions of Parts to a Whole

Pie Chart

This chart shows the estimated population by age in the United States as of 2015. The total estimated population is 316,515,021. Compare to a [donut chart](#).

Data: American Community Survey



<https://observablehq.com/@d3/pie-chart>

<https://observablehq.com/@d3/stacked-bar-chart>

<https://observablehq.com/@d3/stacked-normalized-horizontal-bar>

Lec13 Reading

- Munzner, Ch. 9
- Treevis.net: A Tree Visualization Reference. Hans-Jorg Schulz, IEEE CG&A, 31(6), pp. 11-15, Nov.-Dec. 2011.
 - See also: A Visual Bibliography of Tree Visualization. InfoVis 2010 Best Poster.

Reminder

Assignment 03

Assigned: Monday, February 20

Due: Monday, March 13, 4:59:59 pm

Project Milestone 02

Assigned: Wednesday, February 22

Due: Wednesday, March 29, 4:59:59 pm