

CS 571 - Data Visualization & Exploration

Visualizing Tabular Data

Instructor: Hamza Elhamdadi

UMassAmherst

Upcoming Dates

Tomorrow (Apr 25): Homework 4 Due

Today: Quiz 6 Released (all quizzes due May 2)

Apr 29: Final Group Activity

May 2:

- Homework 5 Due
- Project Screencast Submission Due

May 12: Final Project Submission Due

5 Dataset Types

Tables

Items

Attributes

Networks &
Trees

Items (nodes)

Links

Attributes

Fields

Grids

Positions

Attributes

Geometry

Items

Positions

Clusters,
Sets, Lists

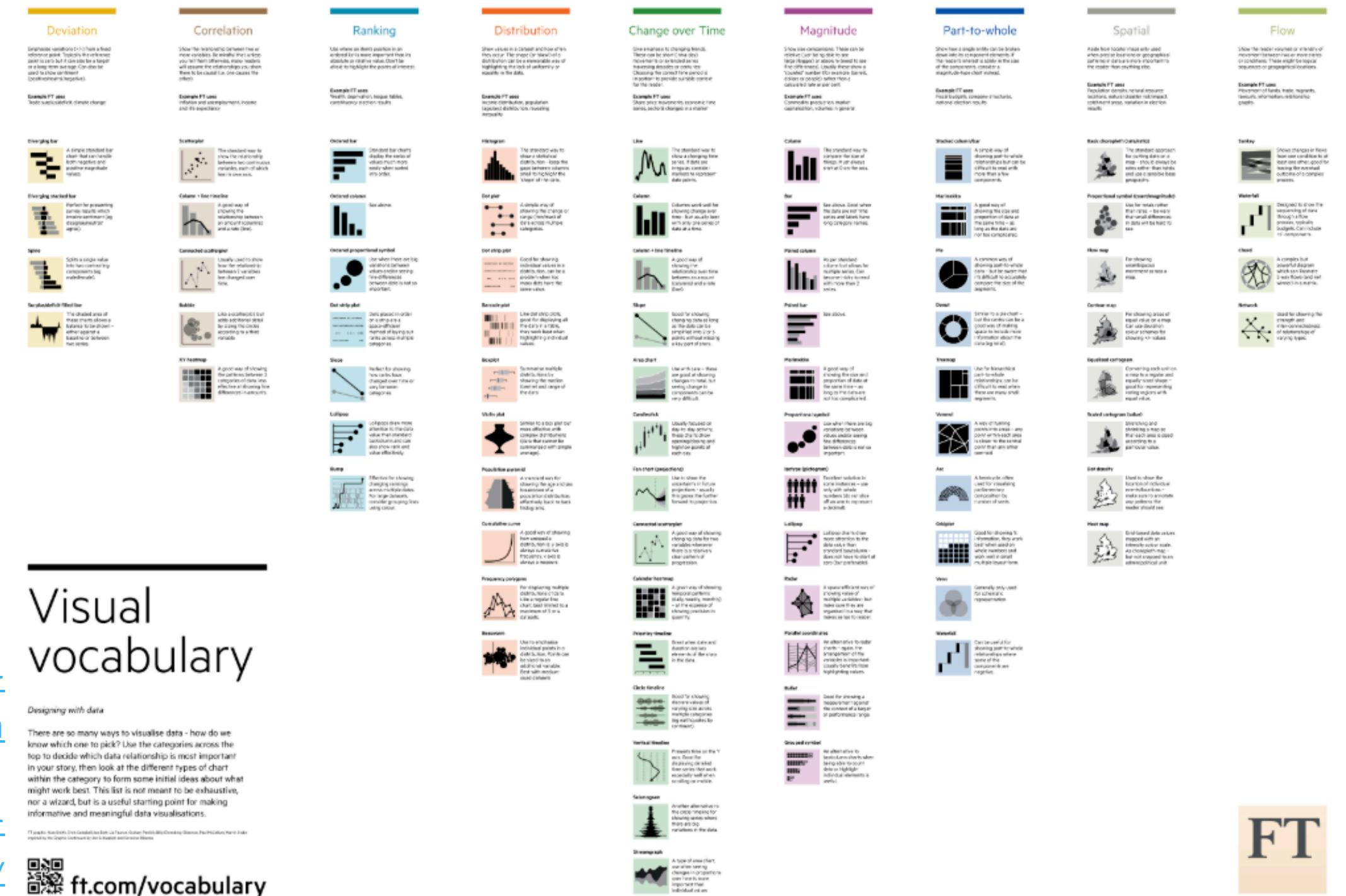
Items

7 Techniques for Visualizing Tabular Data

- Magnitude
- Part to Whole
- Distribution
- Deviation
- Change over Time
- Ranking
- Correlation

<https://gramener.github.io/visual-vocabulary-vega>

<https://github.com/Financial-Times/chart-doctor/tree/main/visual-vocabulary>





1. Magnitude

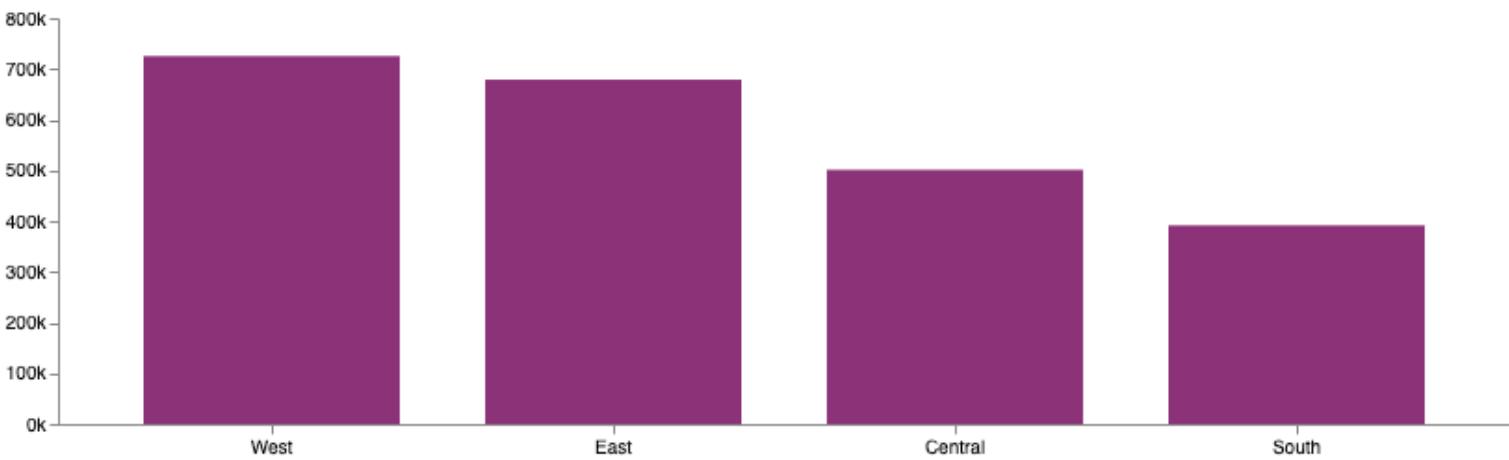
Magnitude

Show size comparisons

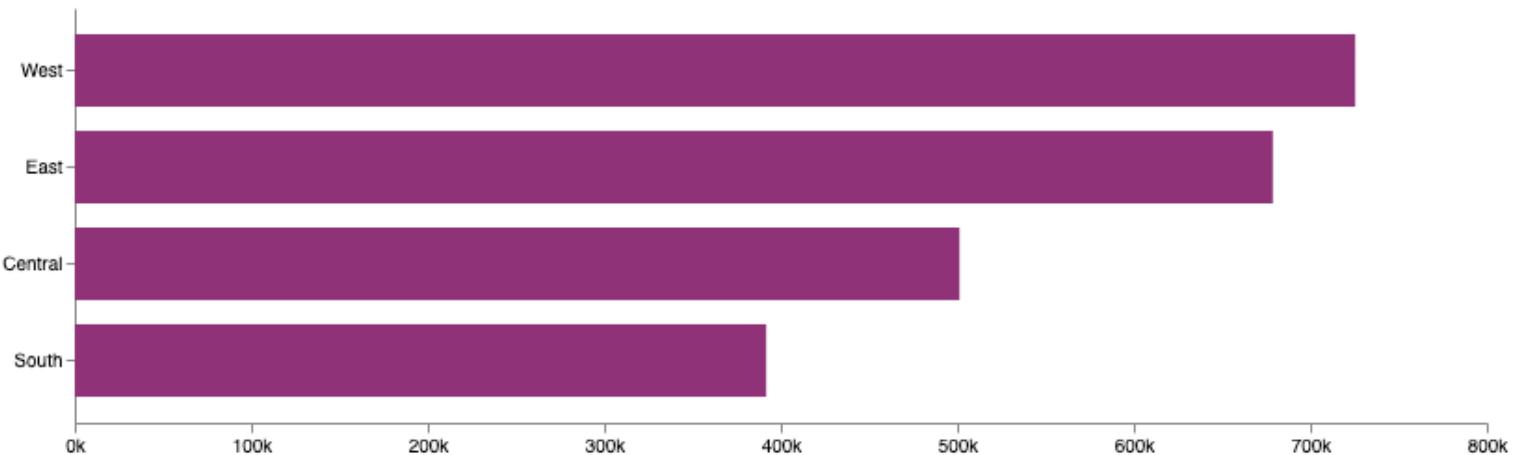
- relative or absolute
- usually for counts (not ratios or percents)

Bar Chart Variants

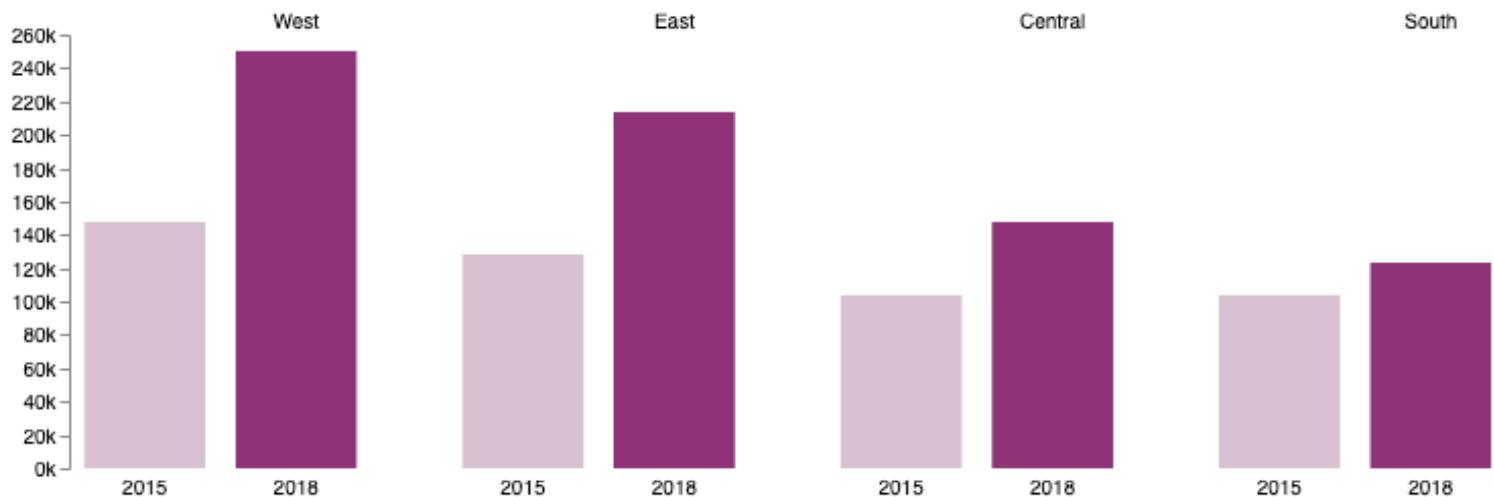
Vertical Bar (Column) Chart



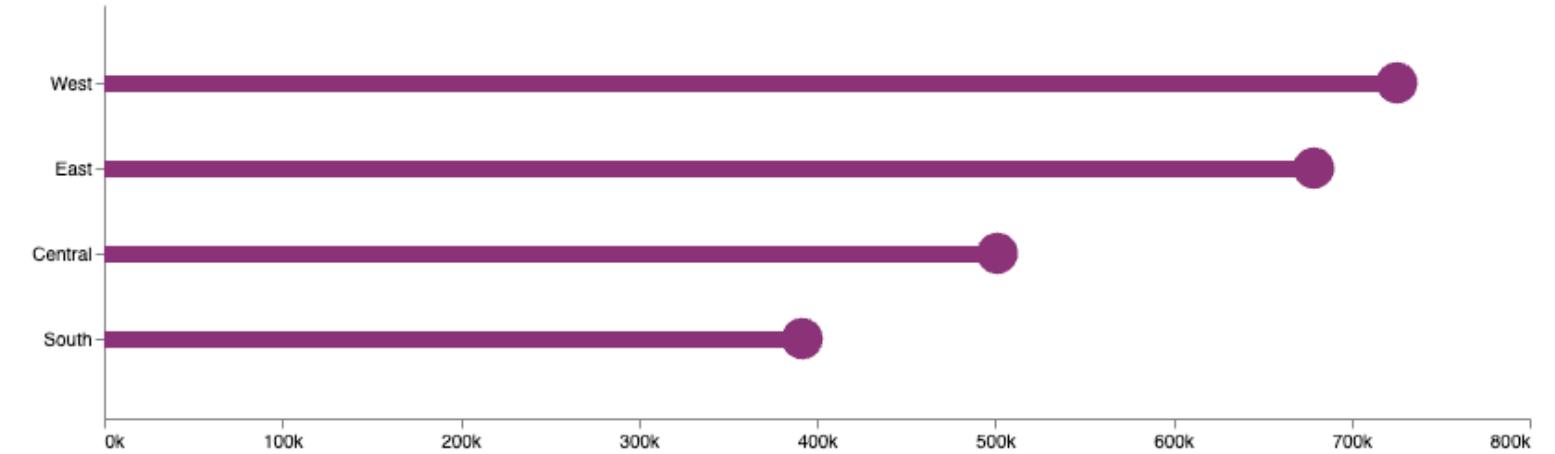
Horizontal Bar Chart



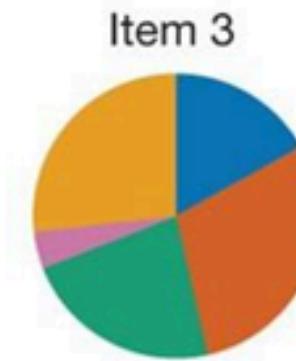
Grouped Bar Chart



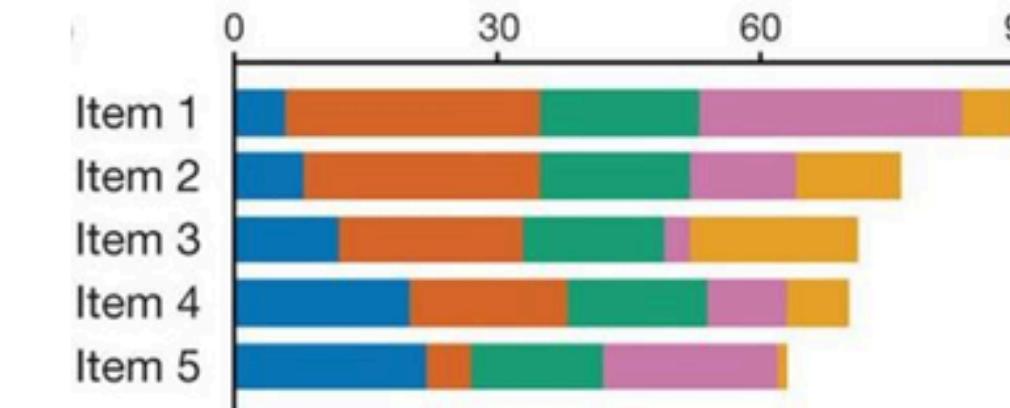
Lollipop Chart



Comparison of Bar Chart Variants (and Pie)

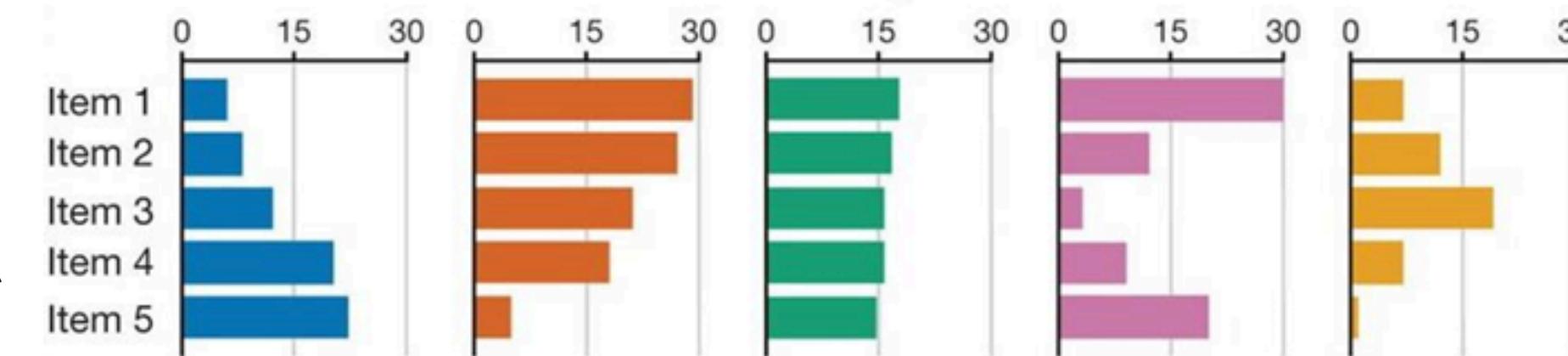


Pie Chart

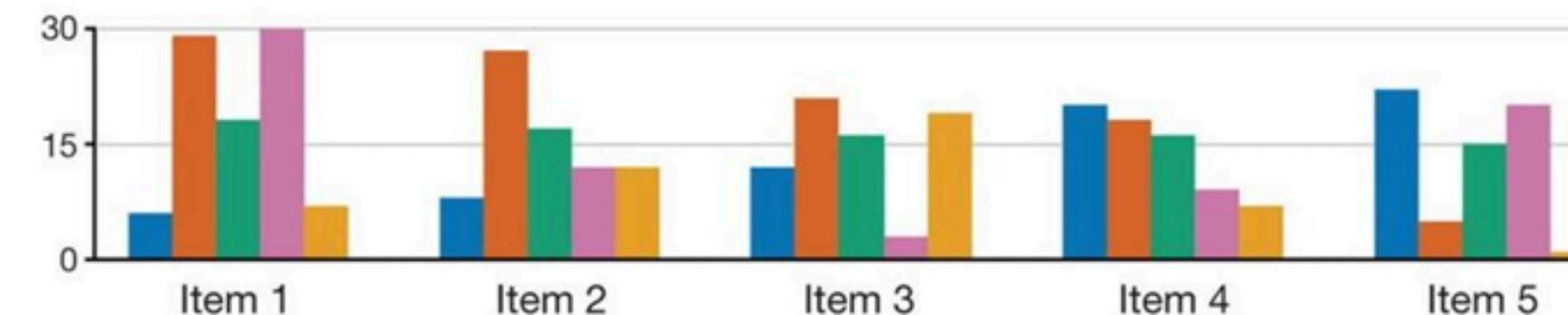


Stacked Bar Chart

Layered
Bar Chart



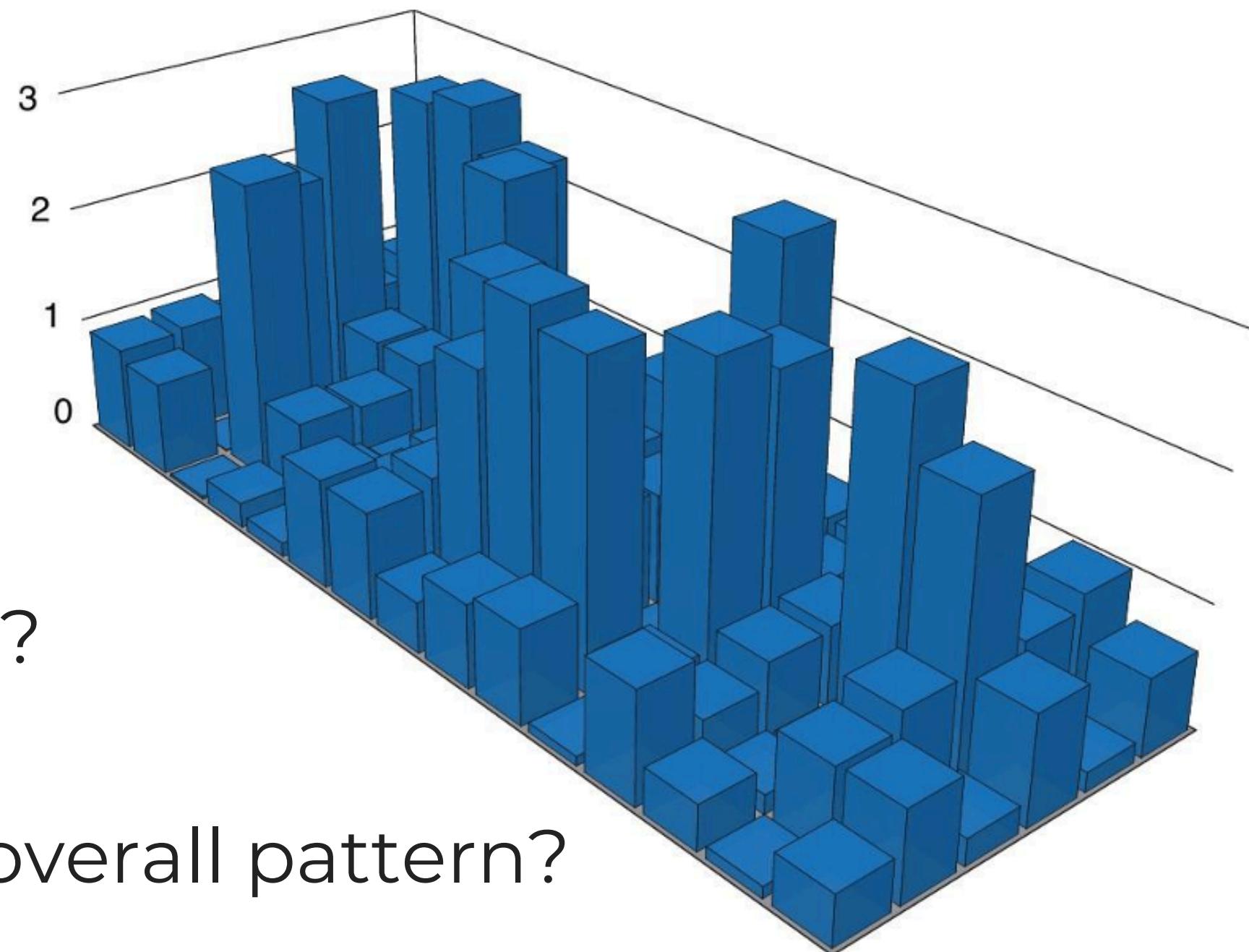
Grouped
Bar Chart



Small
Multiples

Two arrows point from the text "Small Multiples" to the "Stacked Bar Chart" and the "Grouped Bar Chart" sections. The "Stacked Bar Chart" shows how a single bar is divided into multiple segments. The "Grouped Bar Chart" shows how multiple bars are grouped together for each category.

3D Pitfall: Occlusion and Perspective

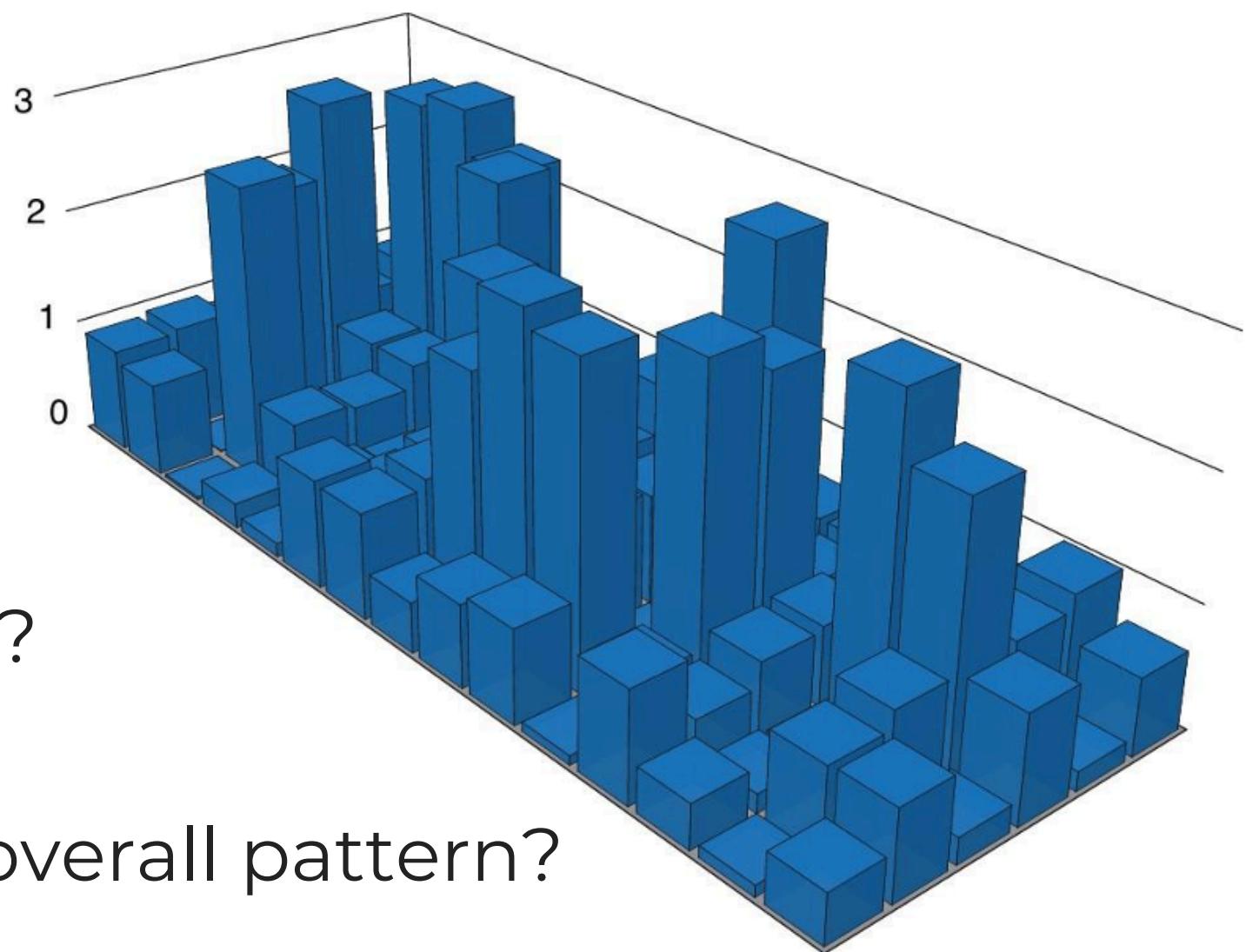


Which bar is tallest?

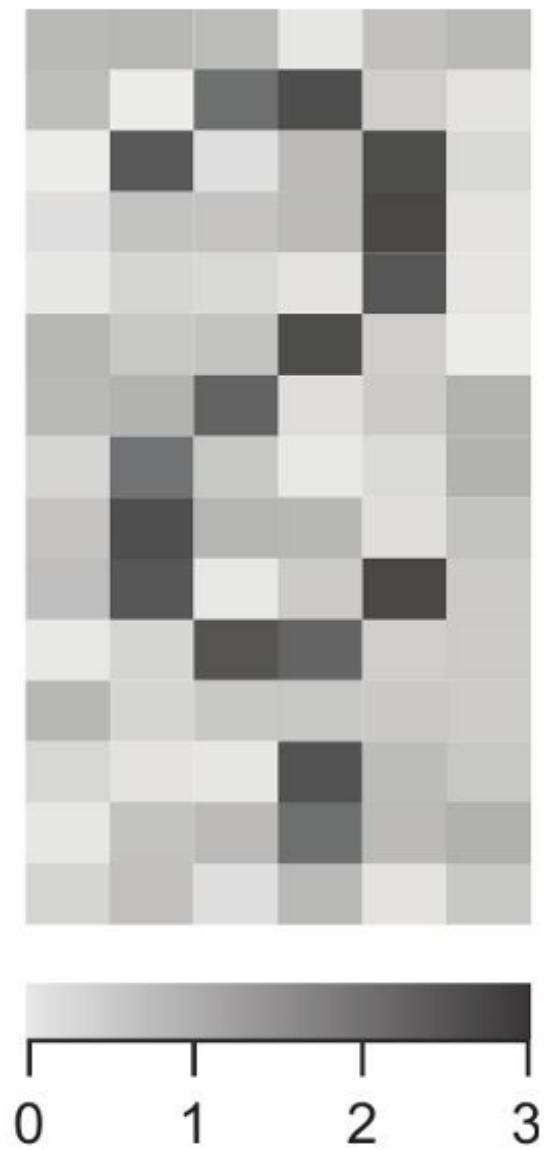
What is the data's overall pattern?

3D Pitfall: Occlusion and Perspective

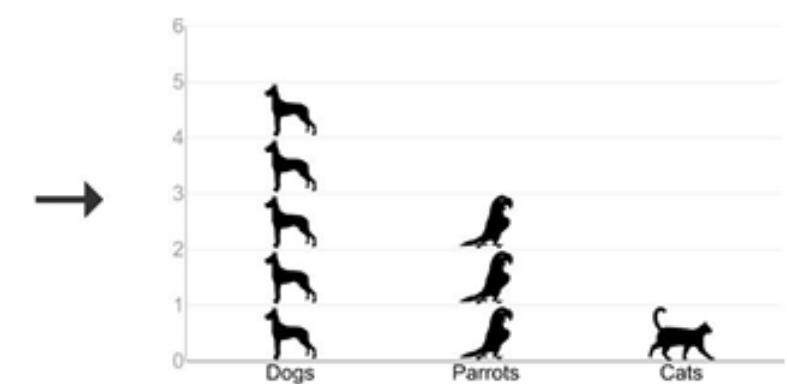
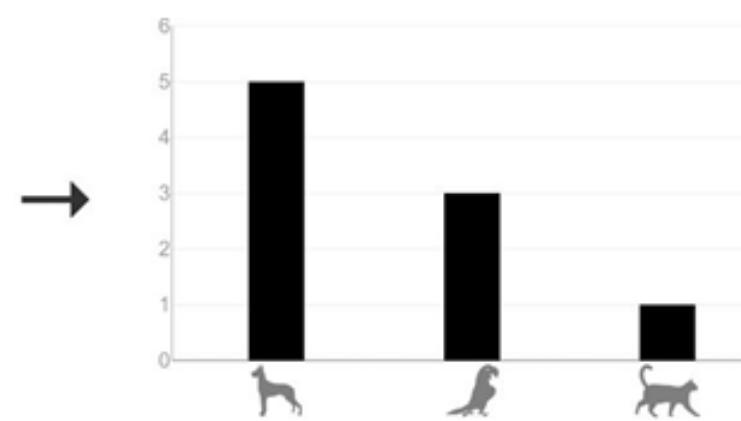
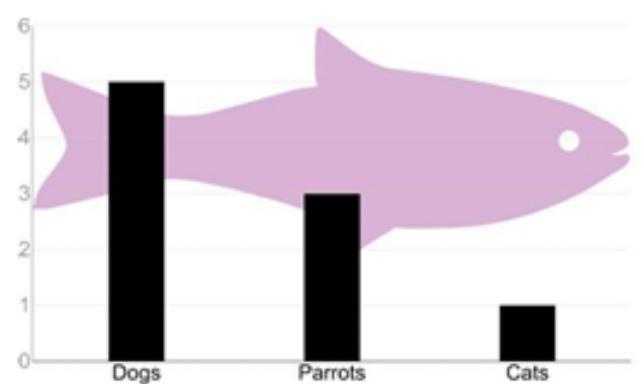
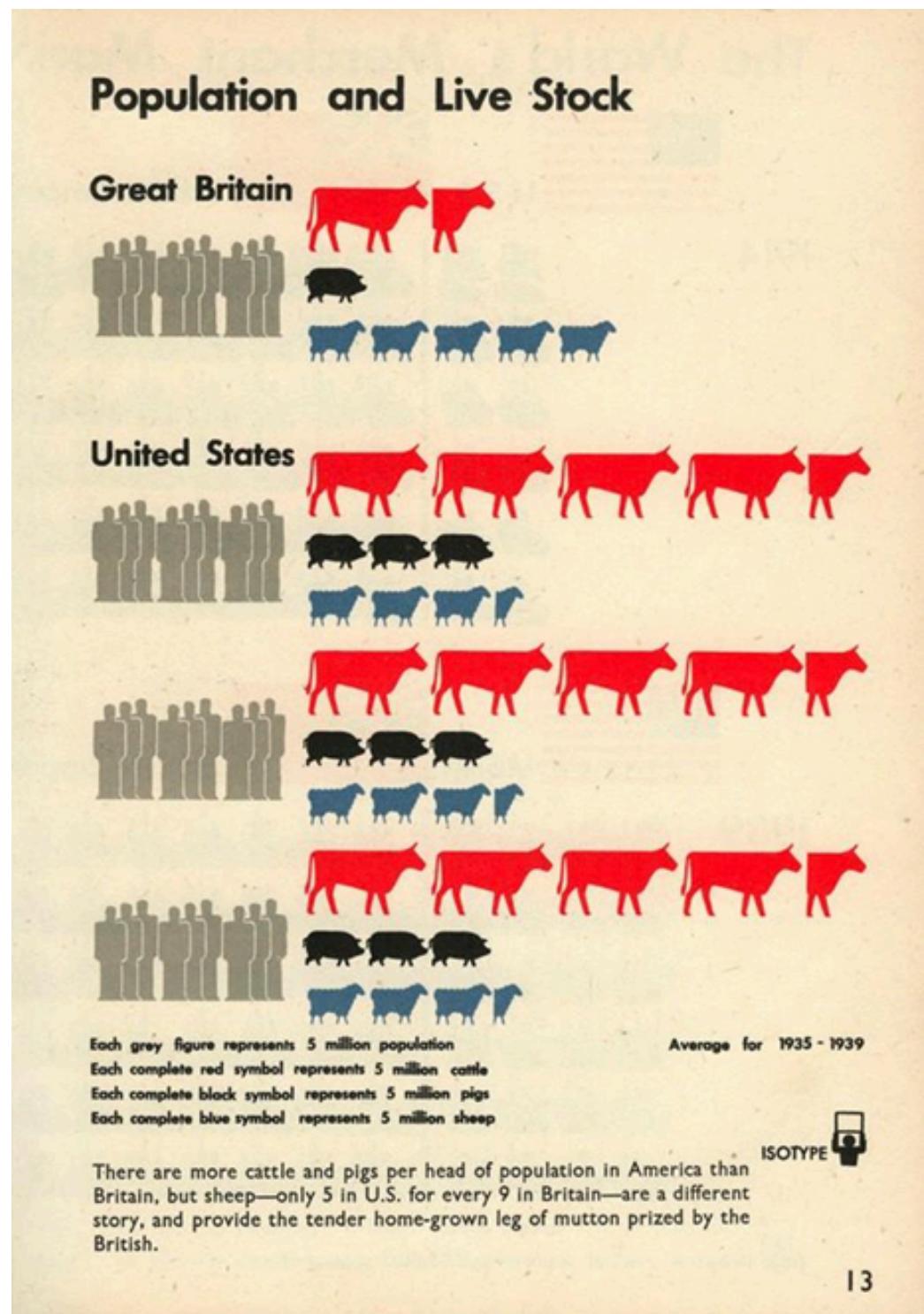
Which bar is tallest?



What is the data's overall pattern?



IsoType Visualizations



<http://steveharoz.com/research/isotype/>

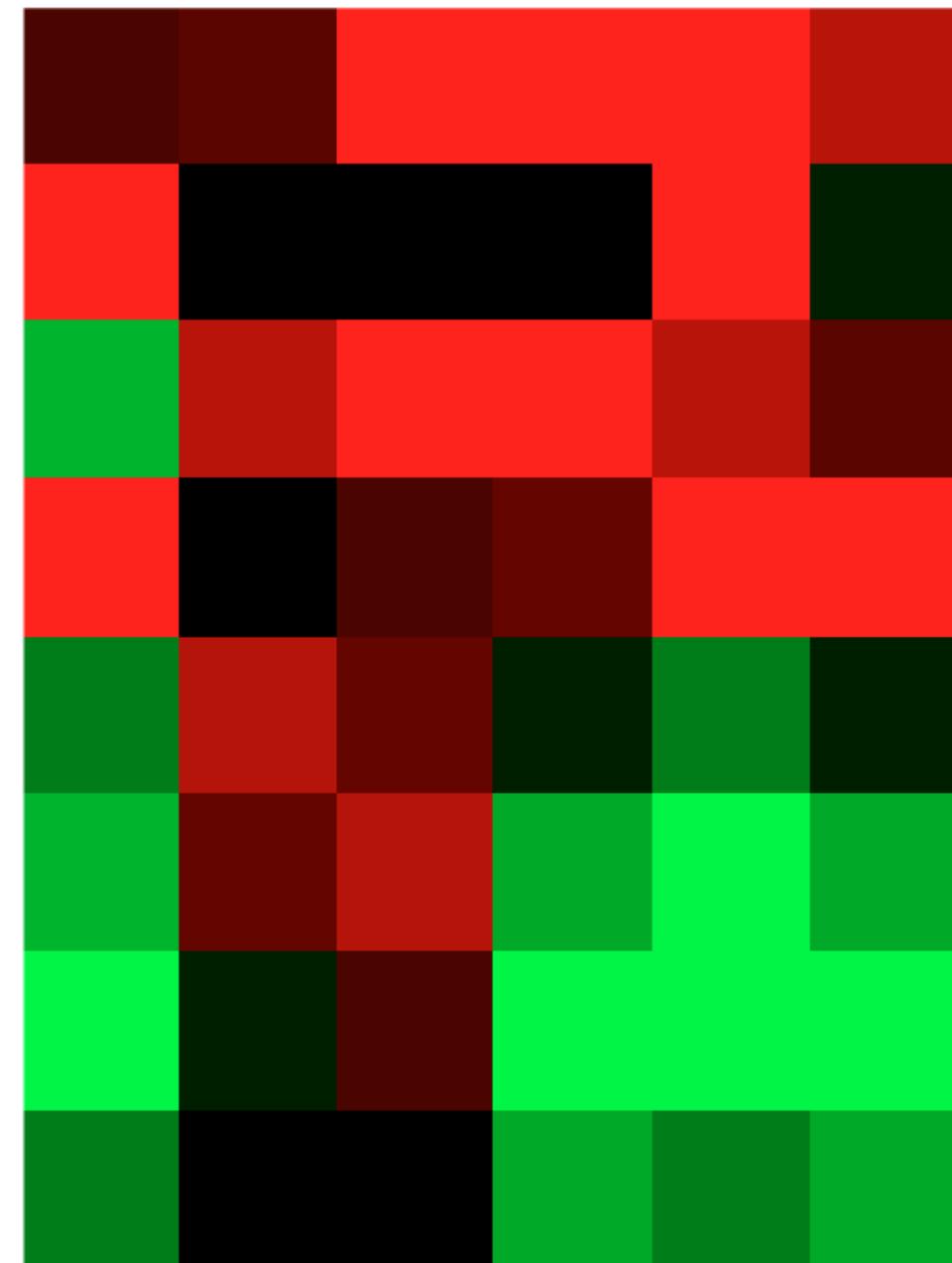
Heatmaps

- matrix layout using keys
- values encoded using color

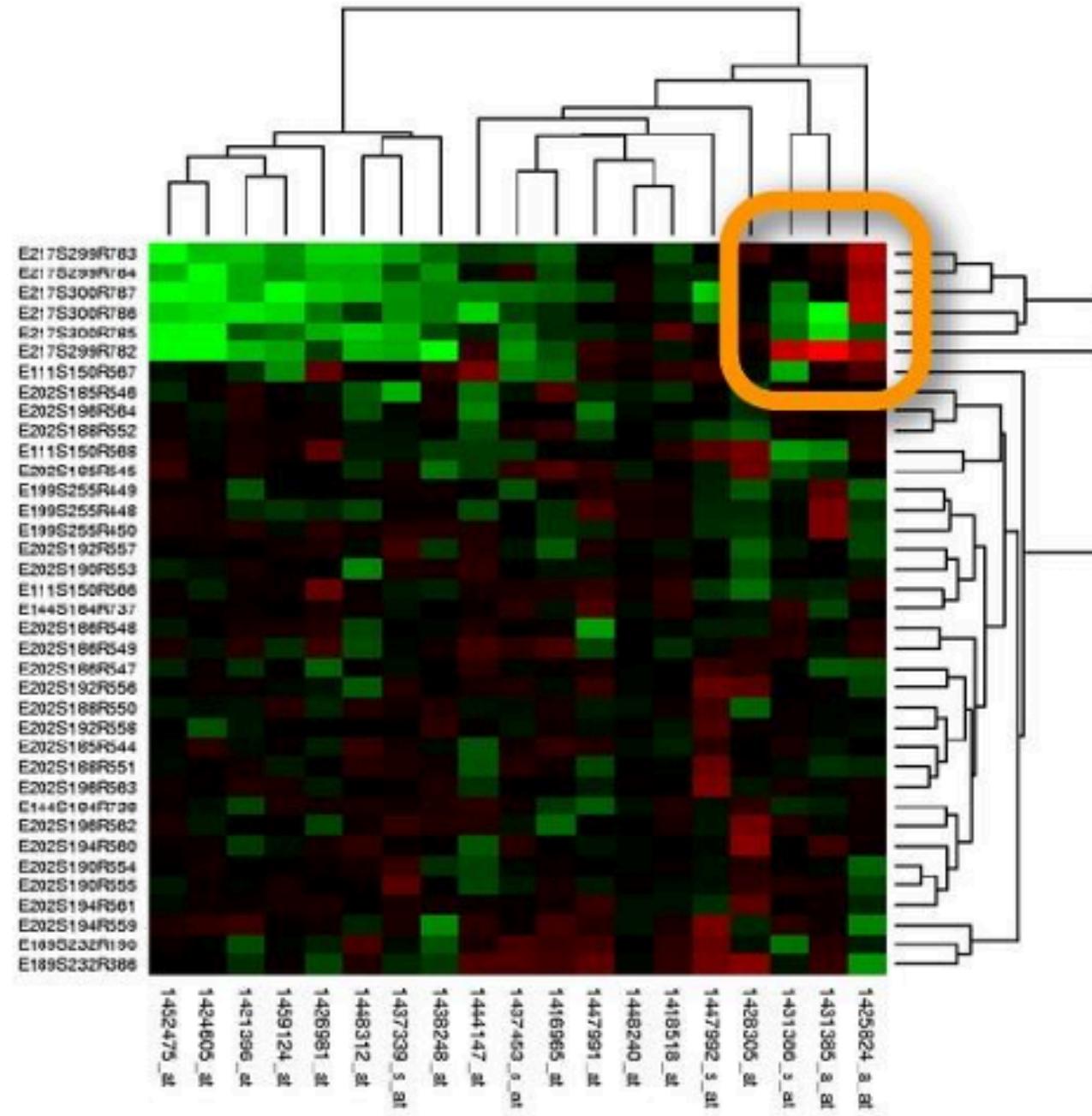
0.2	0.4				0.8
	0	0	0		
0.7	0.8			0.8	0.6
	0	0.2	0.5		
0.5	0.8	0.5	0.3	0.5	0.8
0.7	0.5	0.8	0.7		
	0.3	0.4			
0.5	0	0	0.7	0.5	0.3

Heatmaps

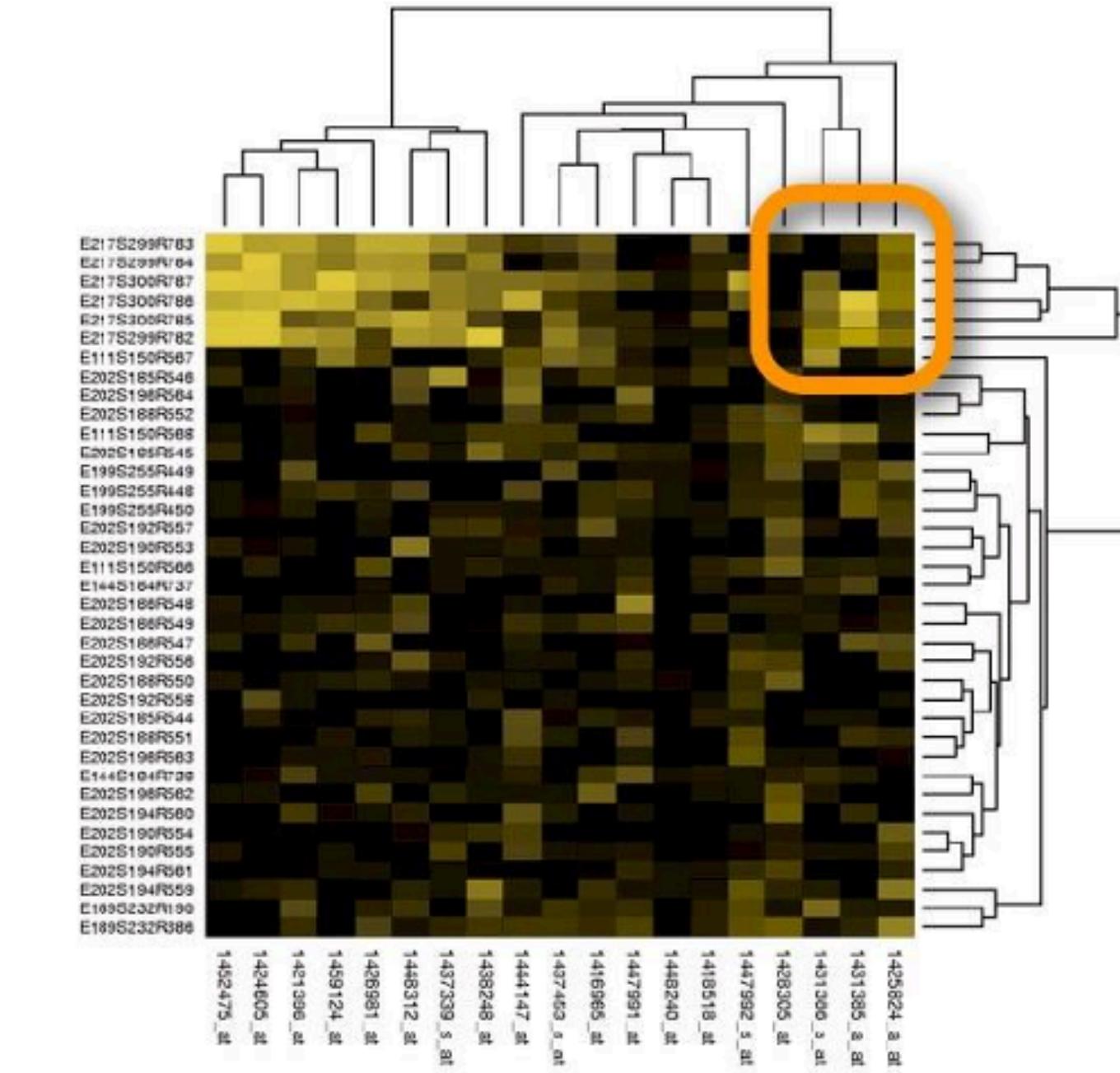
- matrix layout using keys
- values encoded using color



Heatmaps: Bad Color Mapping

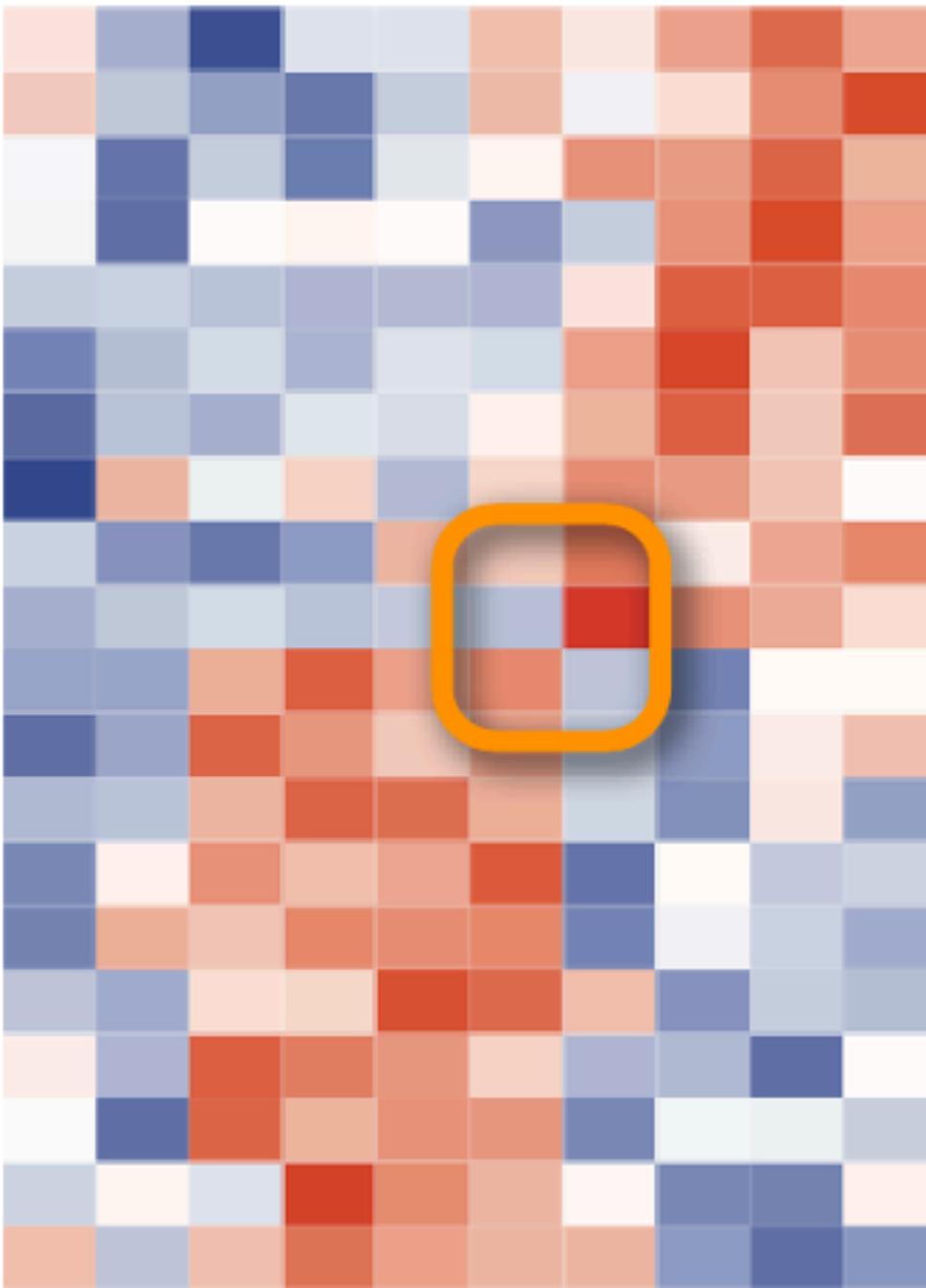


Normal Vision

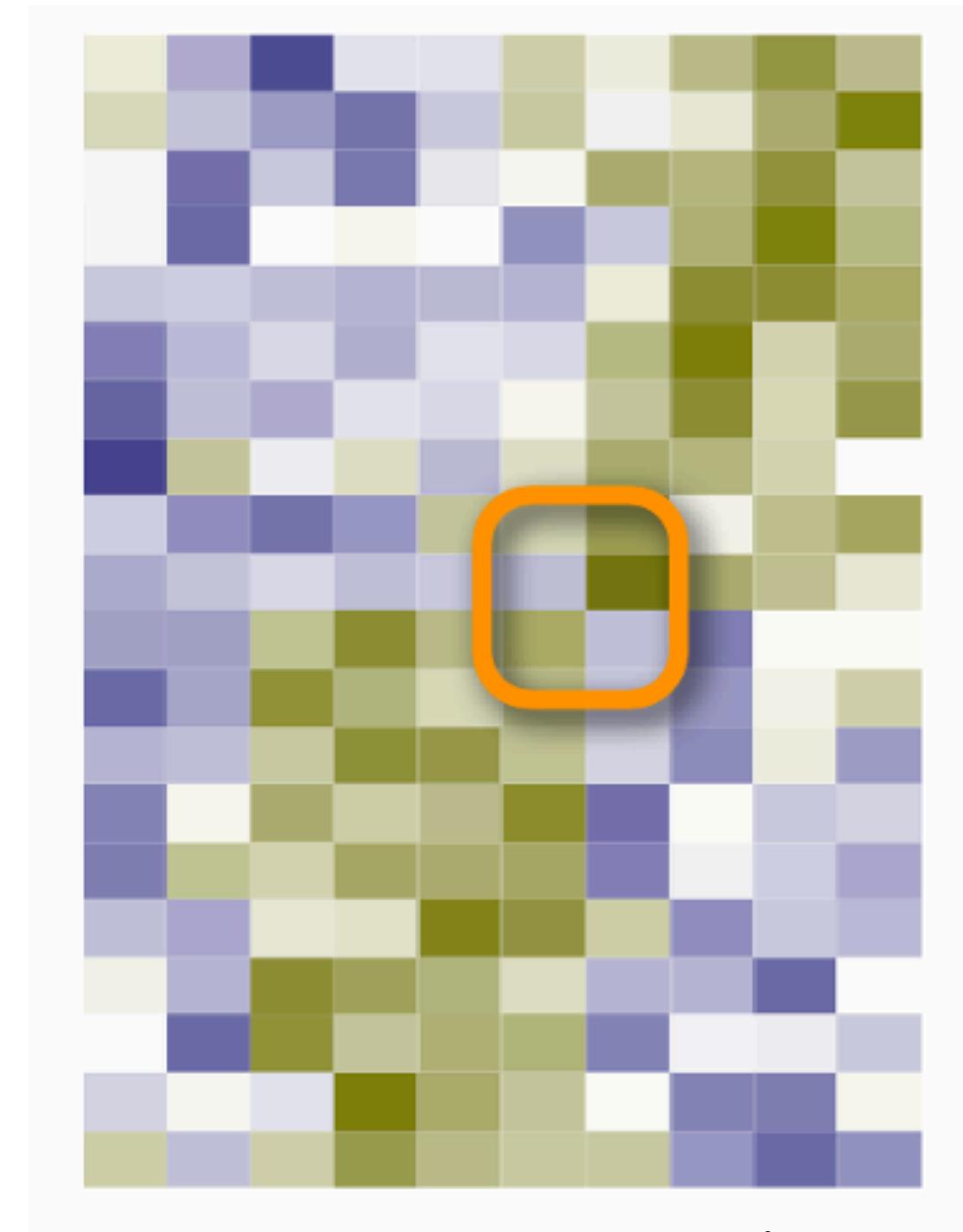


Deutanopia
(Red-Green Colorblindness)

Heatmaps: Good Color Mapping

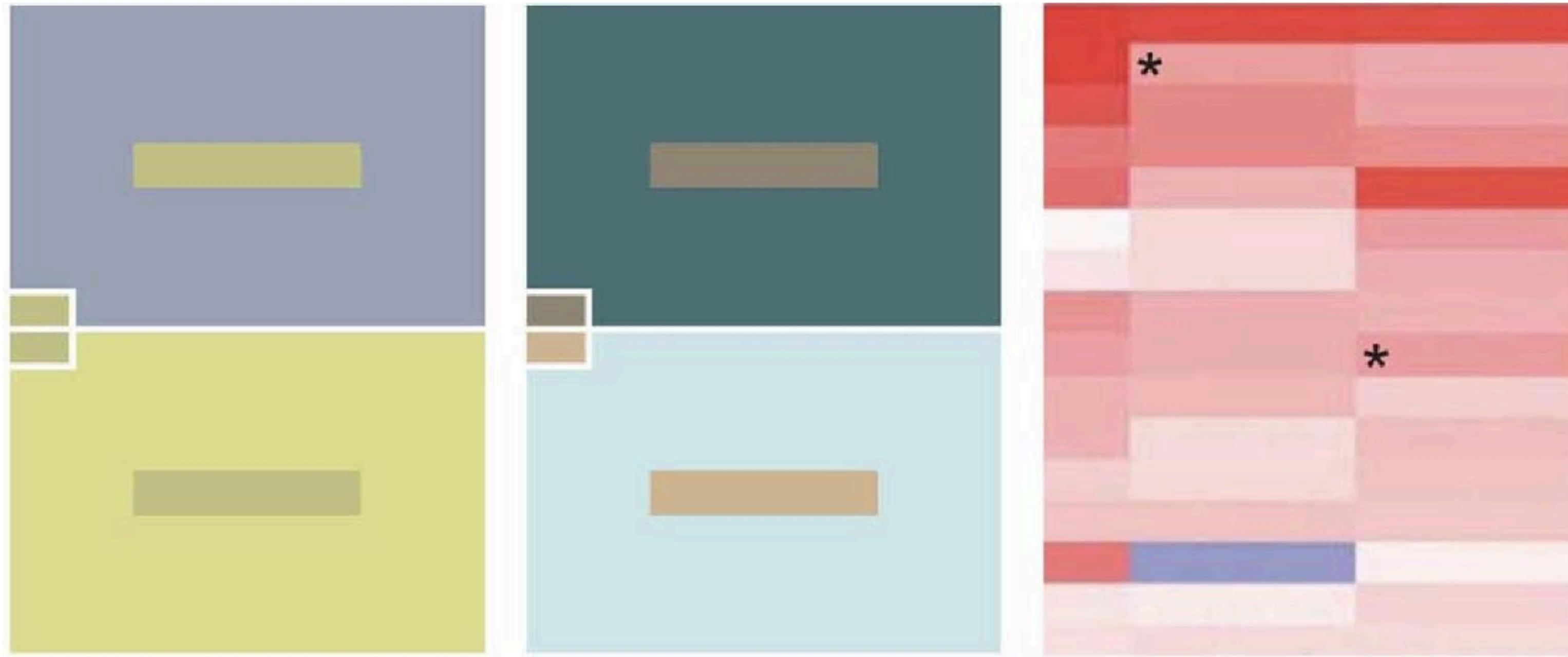


Normal Vision



Deuteranopia
(Red-Green Colorblindness)

Heatmaps: Color is Relative





2. Part to Whole

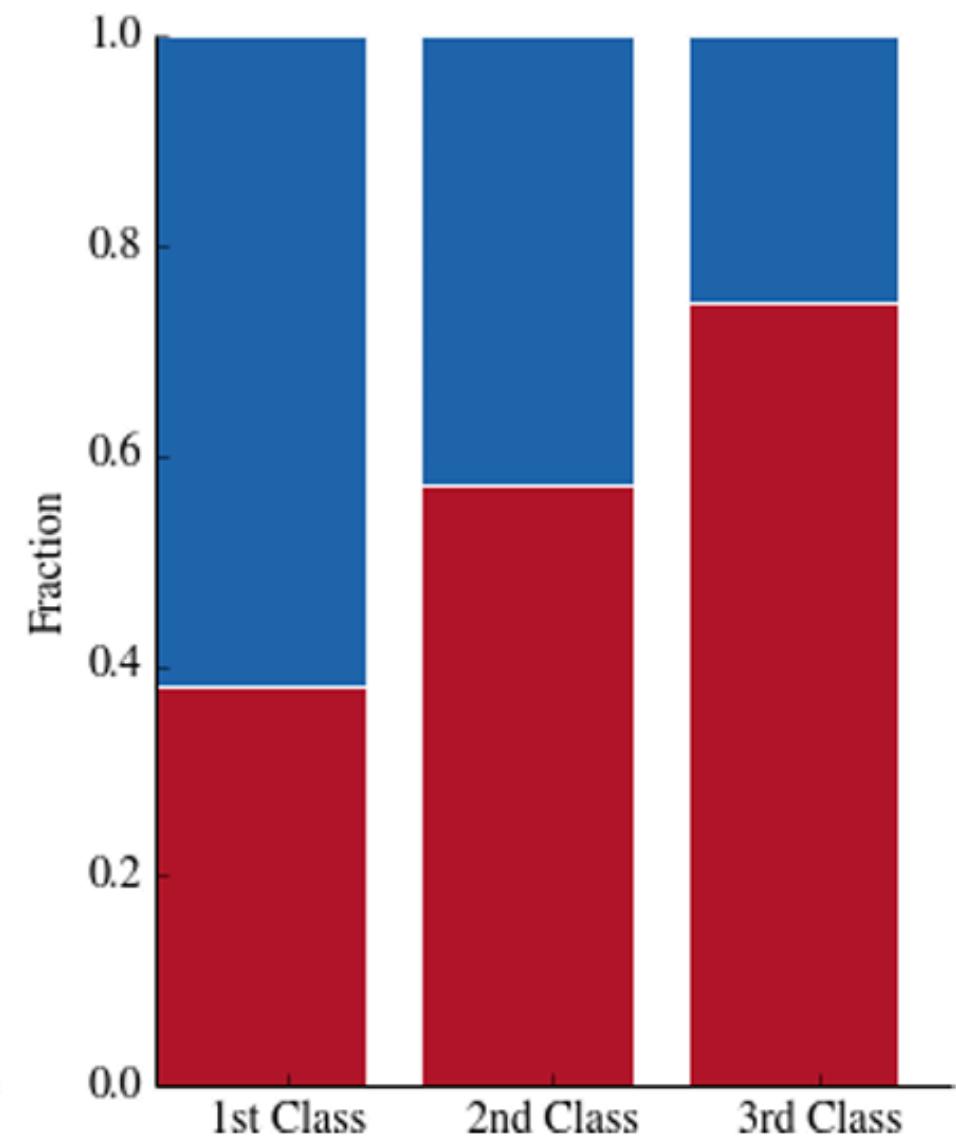
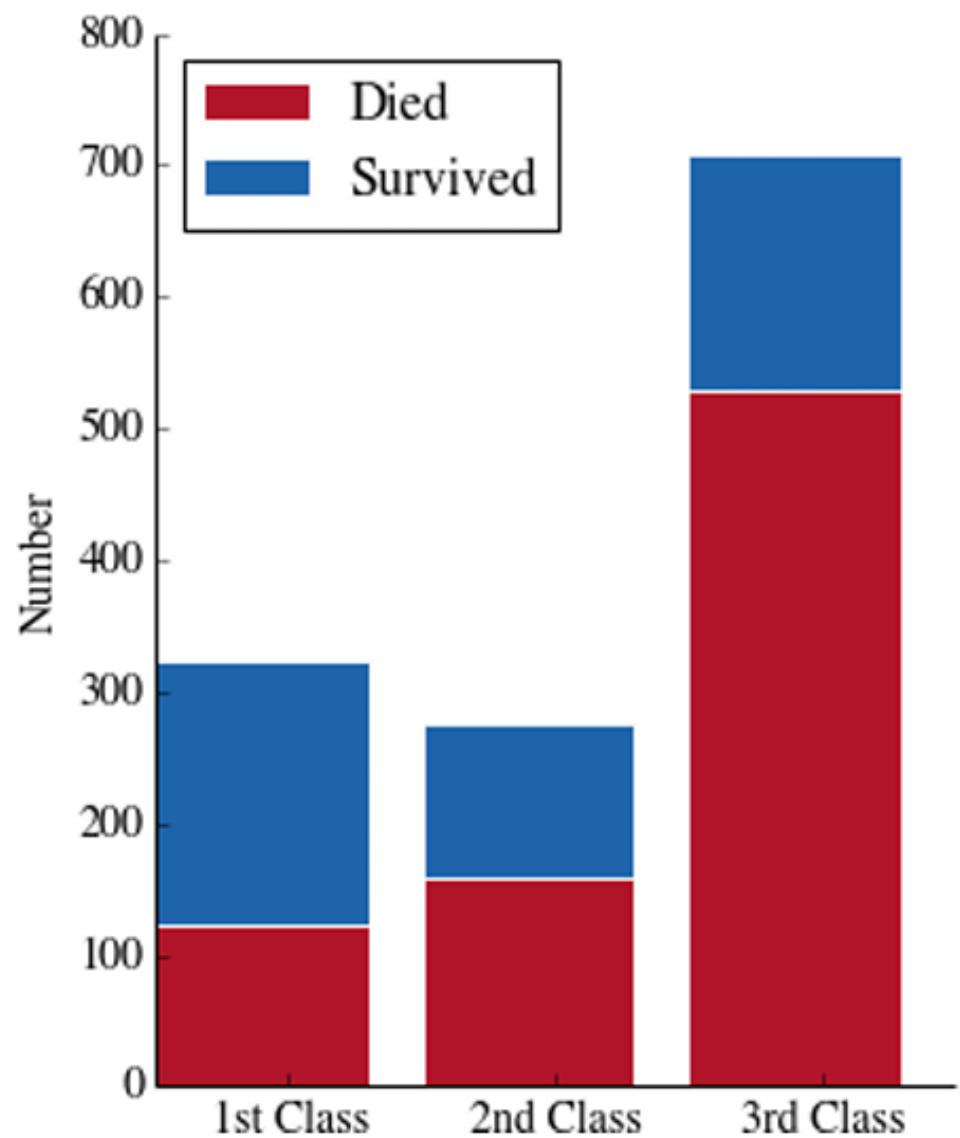
Part to Whole

Show how a single entity can be broken down into its component elements

- useful if you care about the relationship between the components and the whole

Stacked Bar Charts

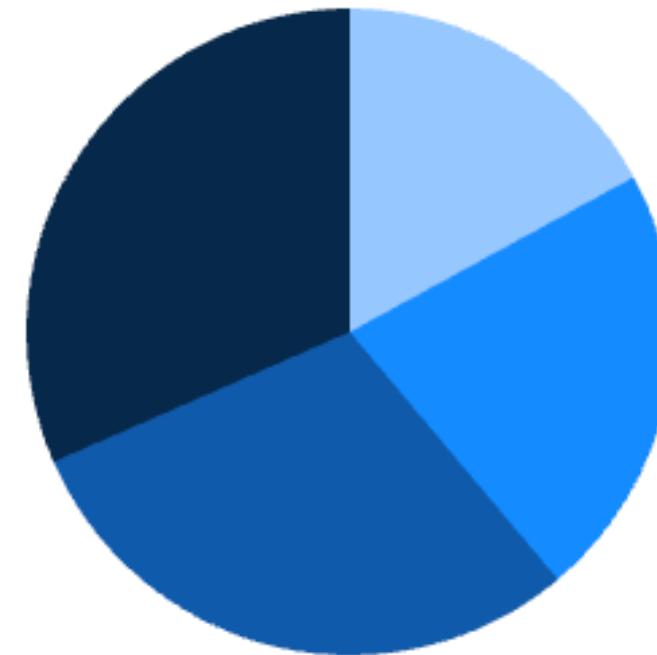
- absolute values (left)
- proportions (right)



Pie and Donut Charts

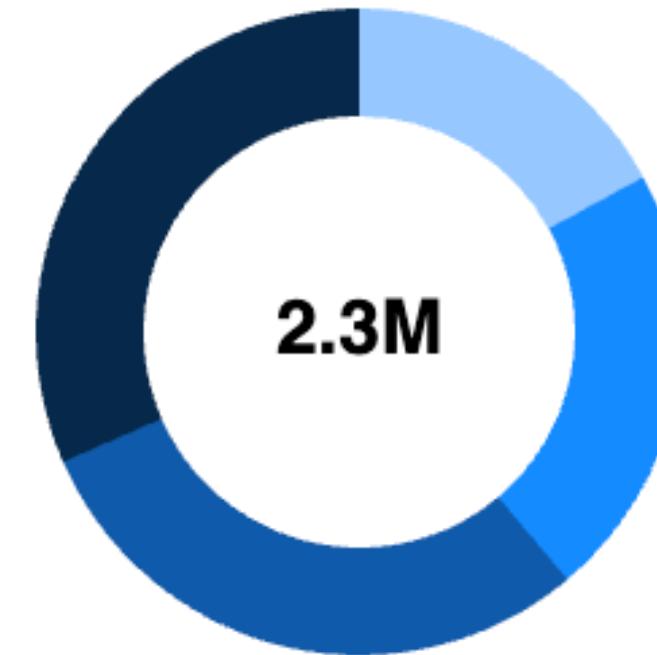
Pie chart:

- angle/size comparisons
- comparison accuracy lower than with stacked bar chart



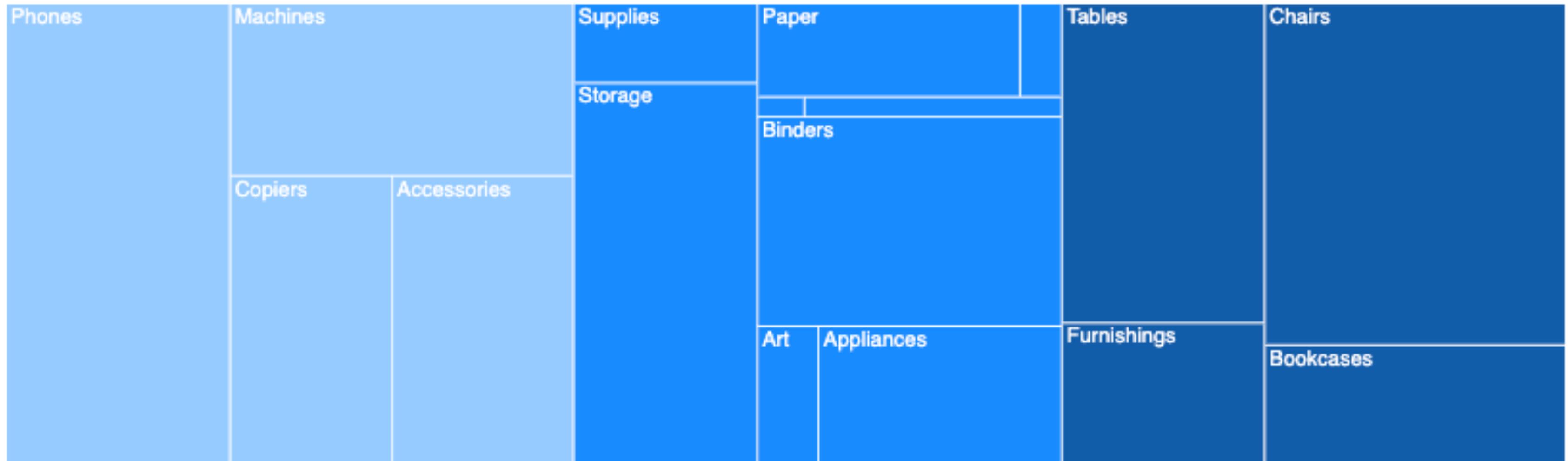
Donut chart:

- pie chart with the center omitted
- higher data-ink ratio
- room for additional info in center



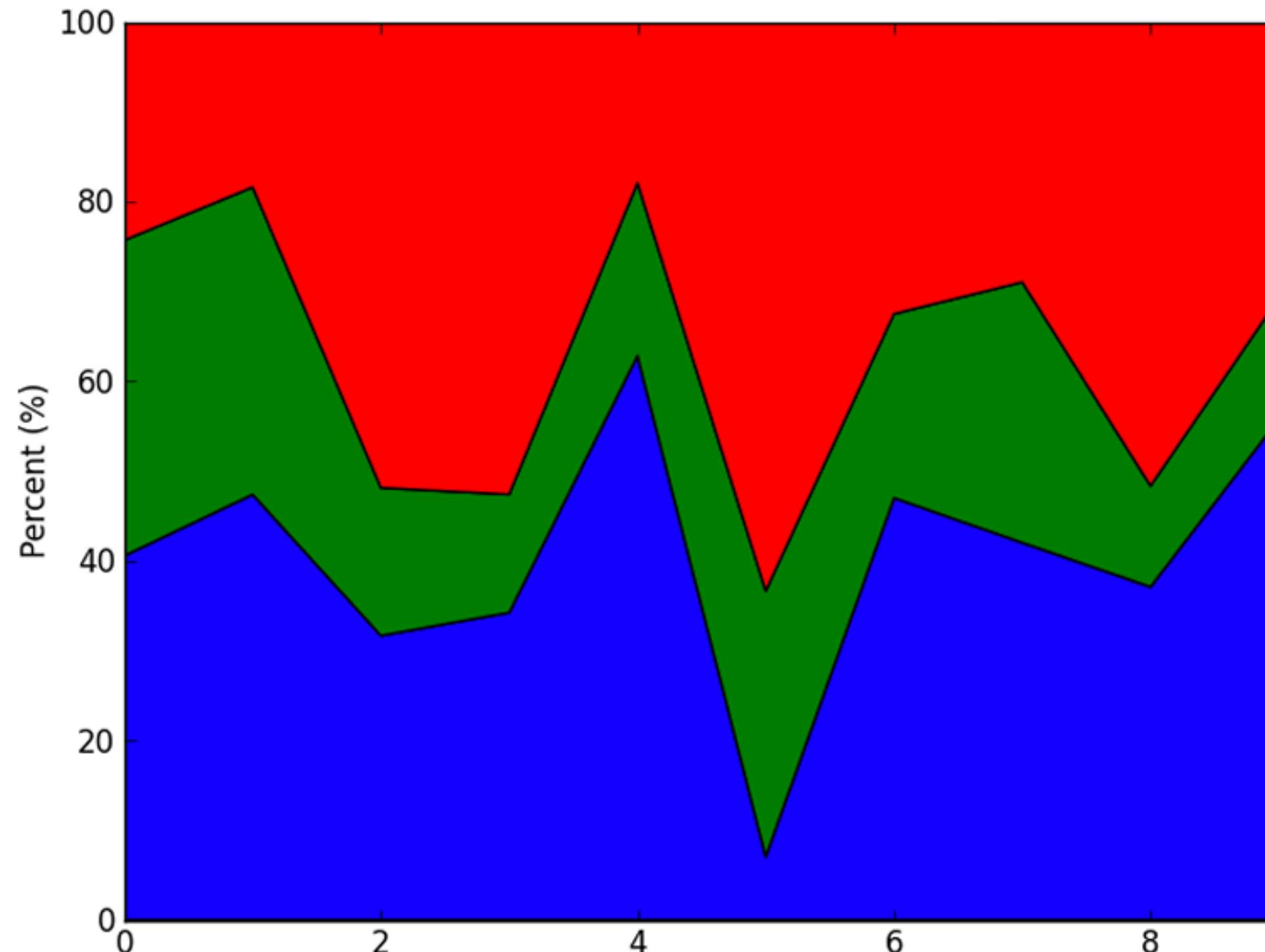
TreeMap

- useful for hierarchical part-to-whole relationships
- can get unwieldy if there are too many small segments



Stacked Area Chart

Part of Whole for Time-Series Data



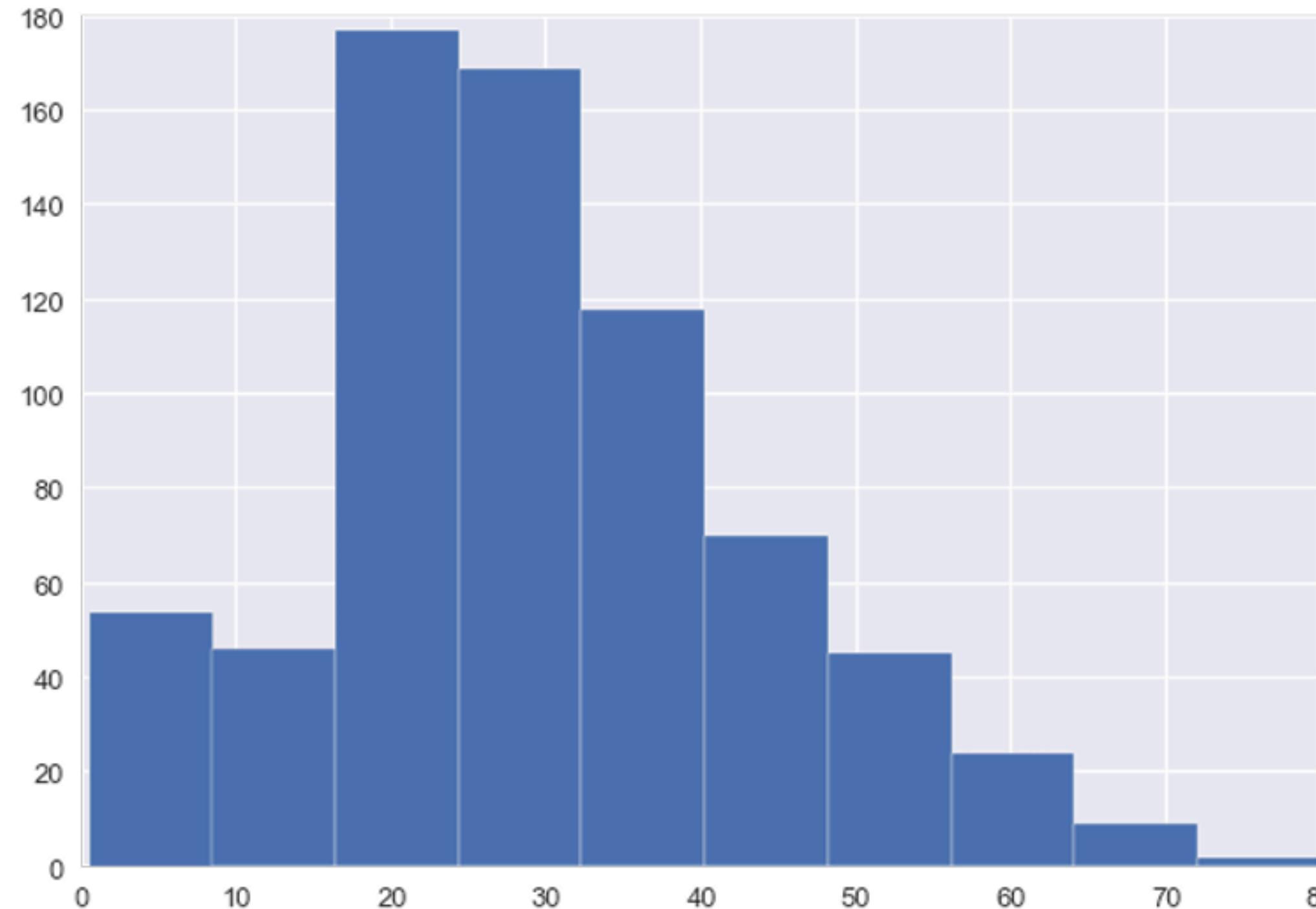


3. Distribution

Distribution

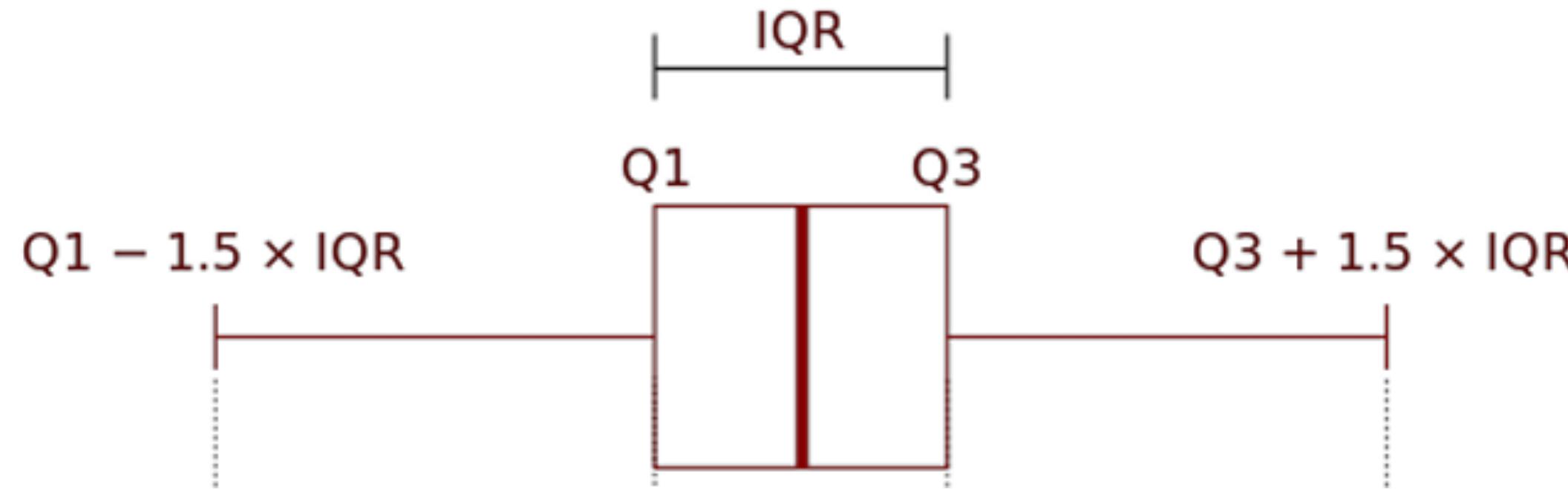
Show how often specific values in a dataset occur

Histogram

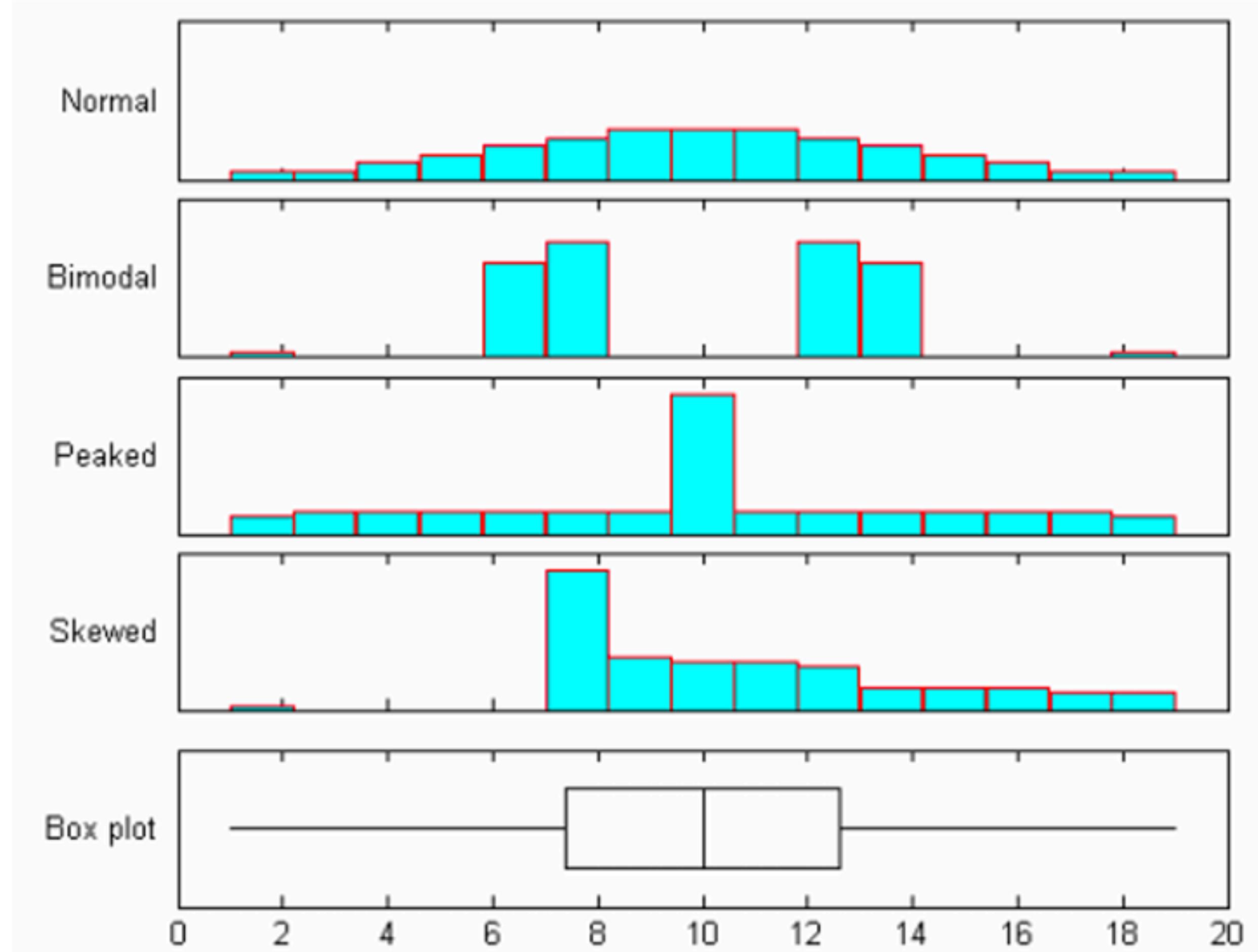


Box Plots

- also referred to as box-and-whisker plots
- can be misleading for more complex distributions
 - especially bi- or multi-modal distributions

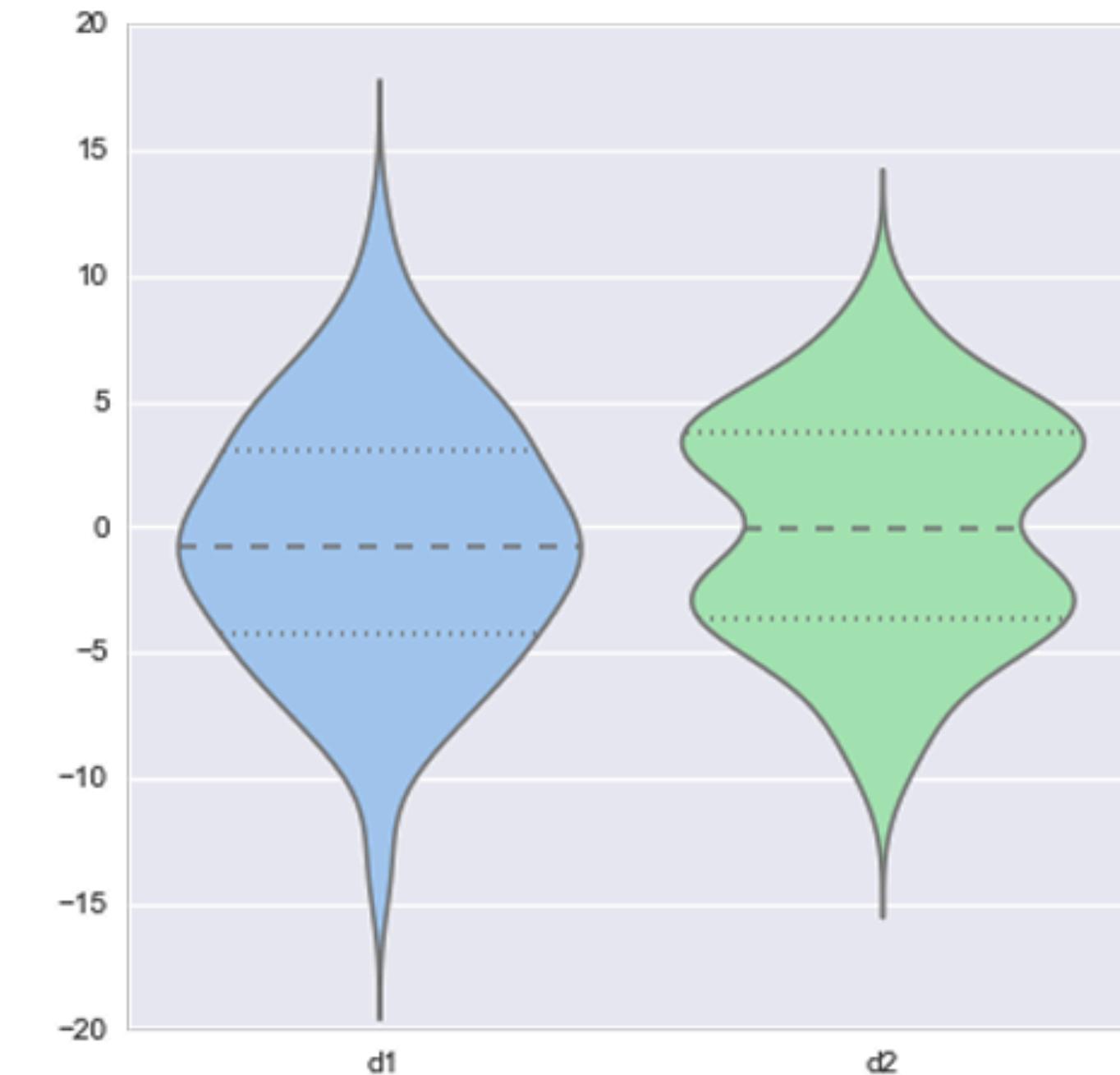
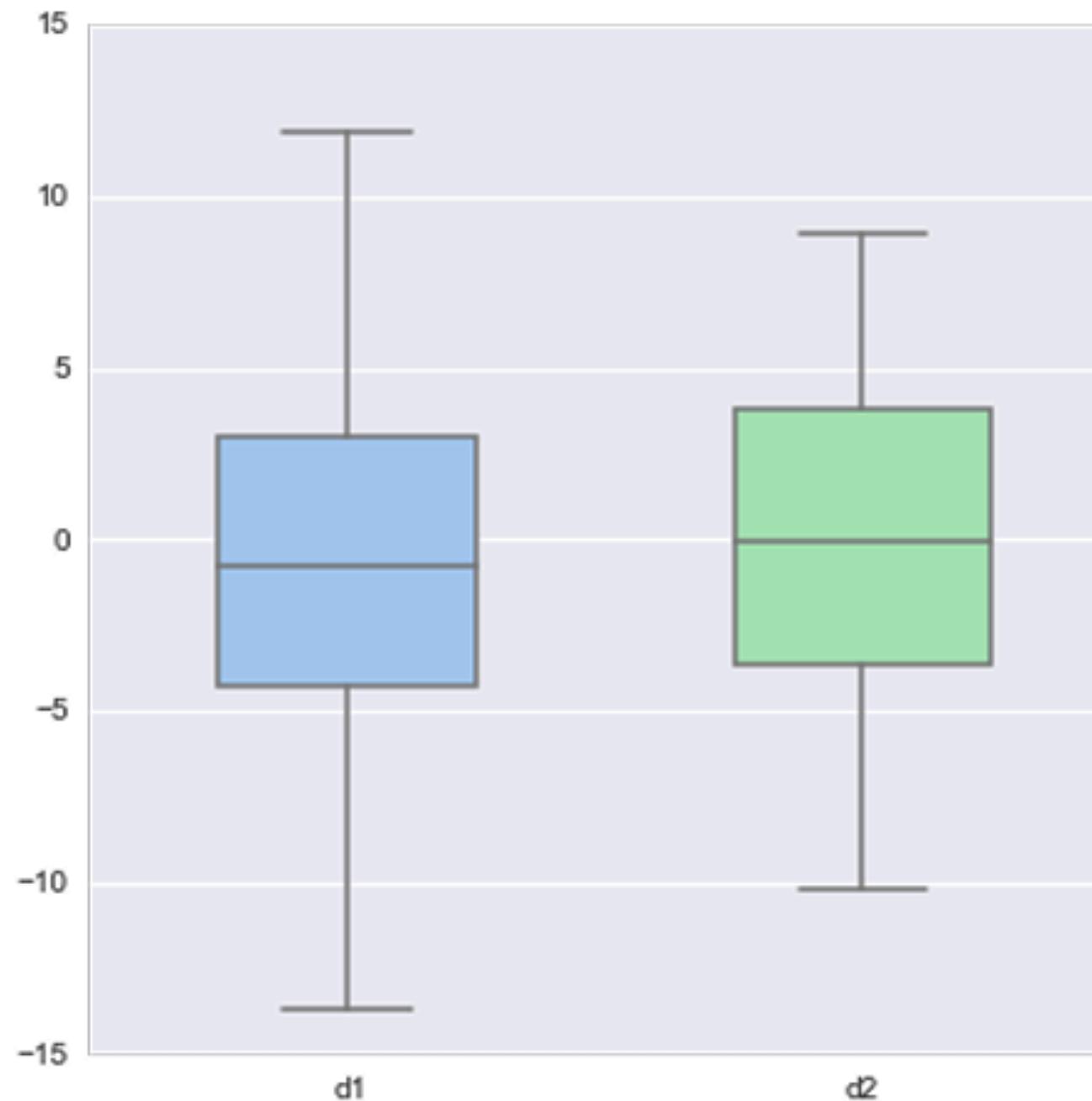


Box Plots



Violin Plots

- box plot + probability density function





4. Deviation

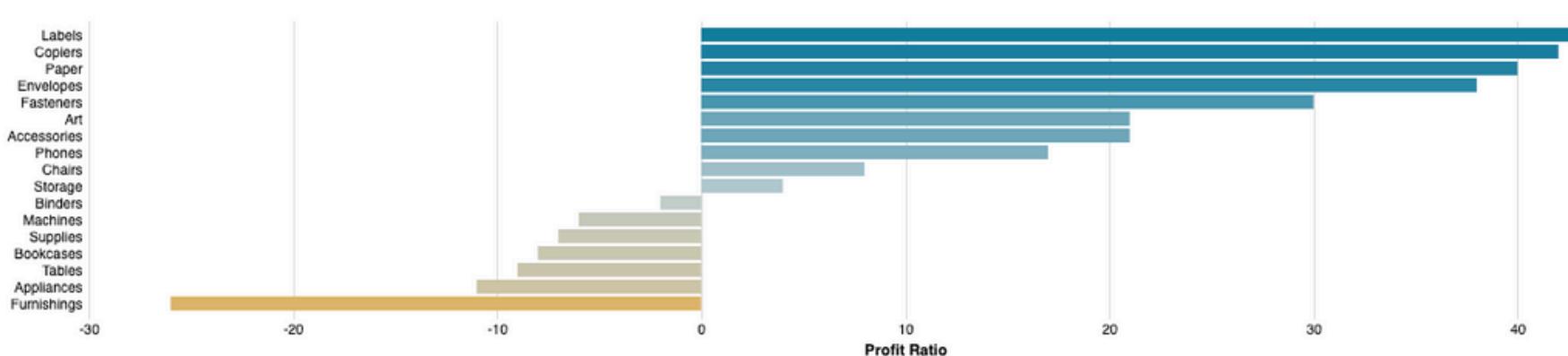
Deviation

Emphasize variation from a fixed reference point

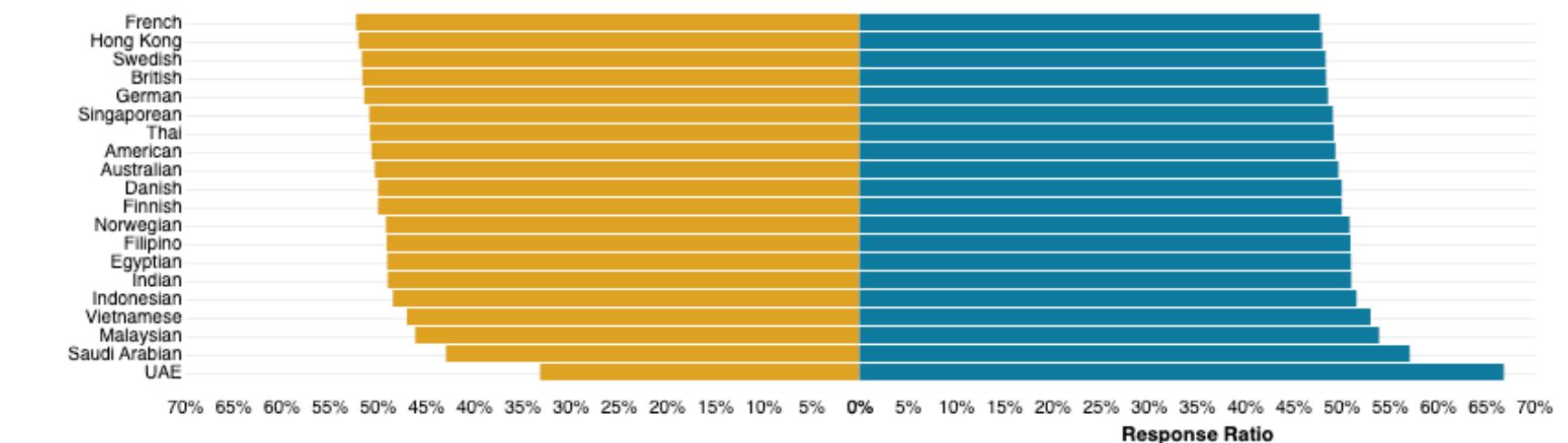
- reference point can be 0, or an average, etc.
- useful for showing sentiment

Comparison to Reference Point

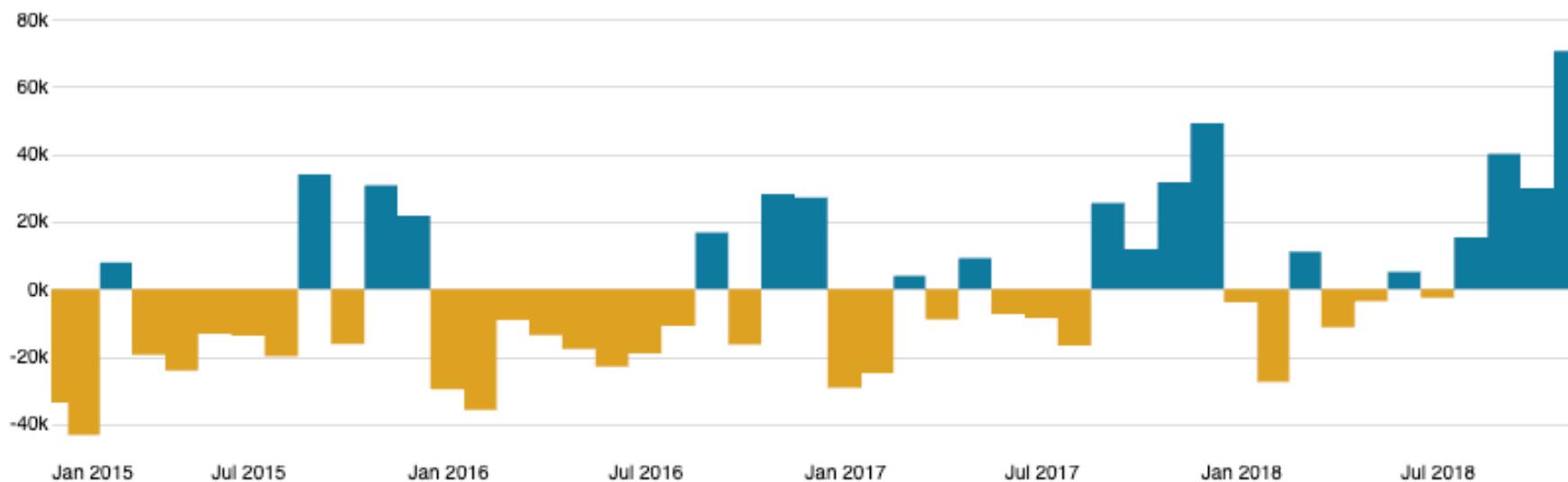
Diverging Bar Chart



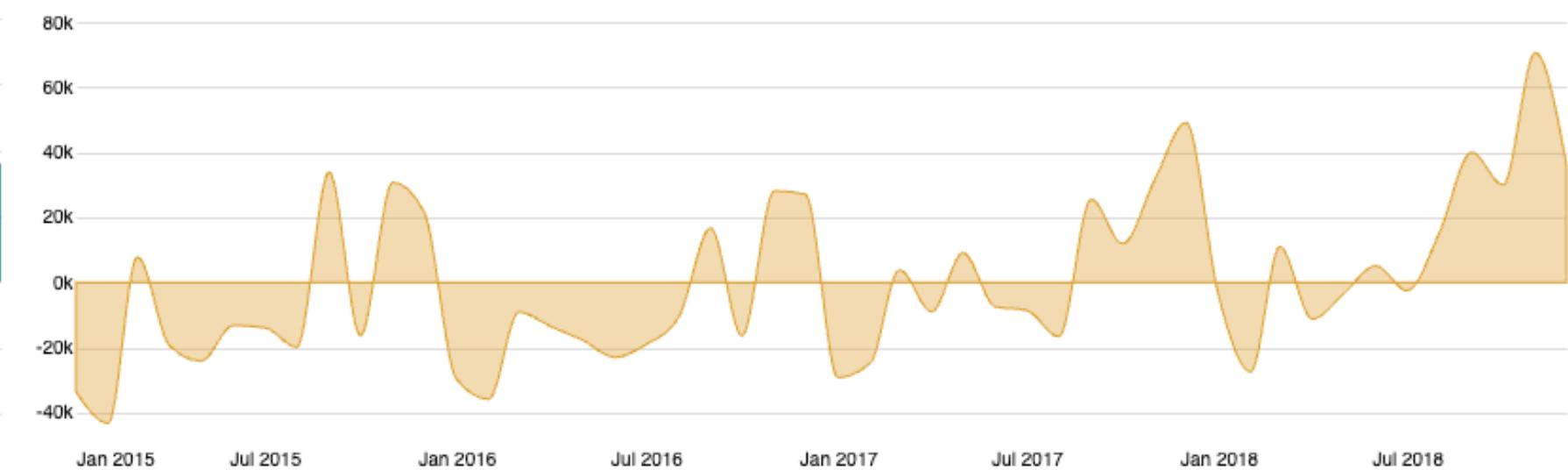
Juxtaposing Two Variables



Surplus/Deficit Bar Chart



Surplus/Deficit Area Chart





5. Change over Time

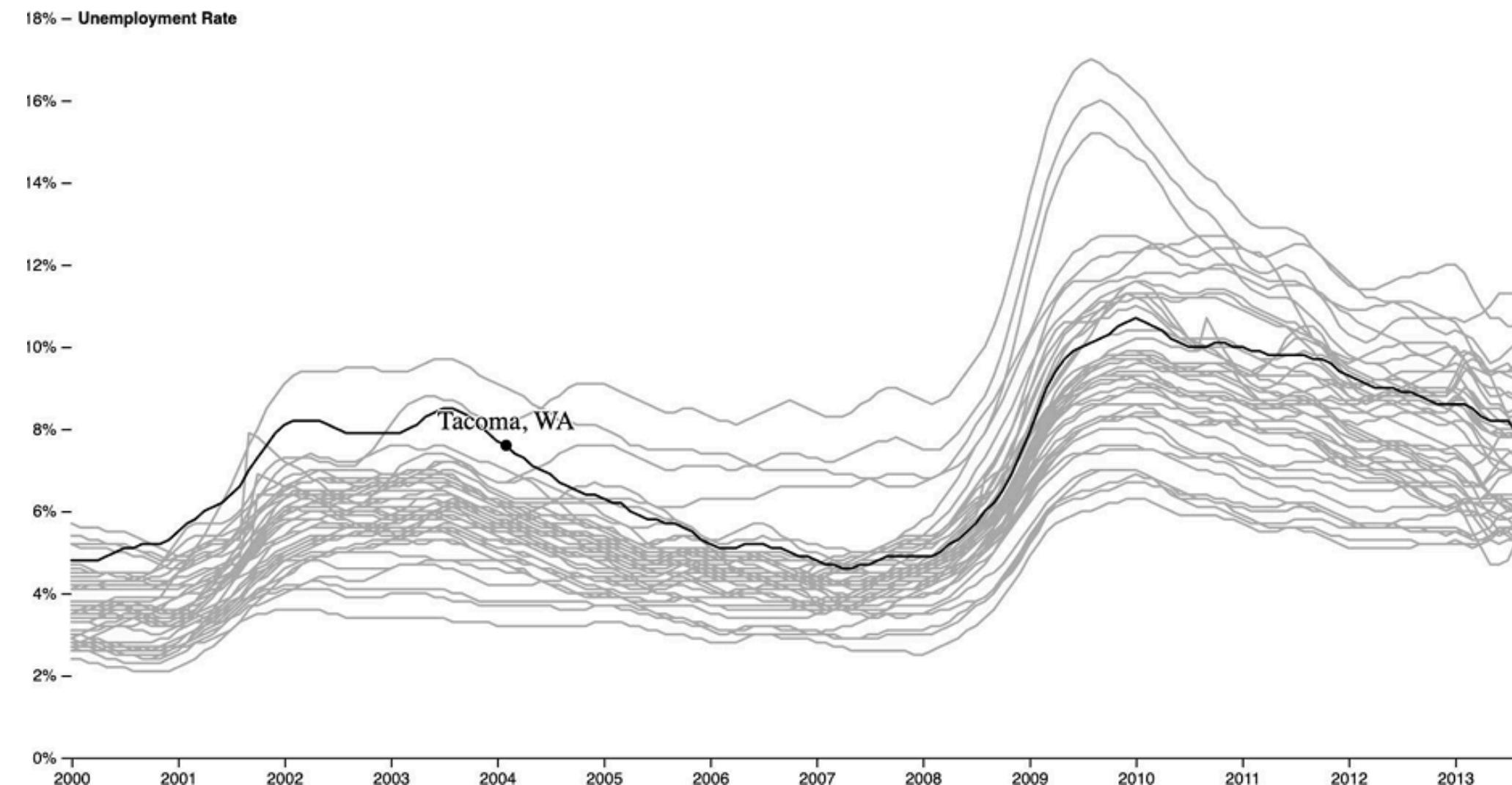
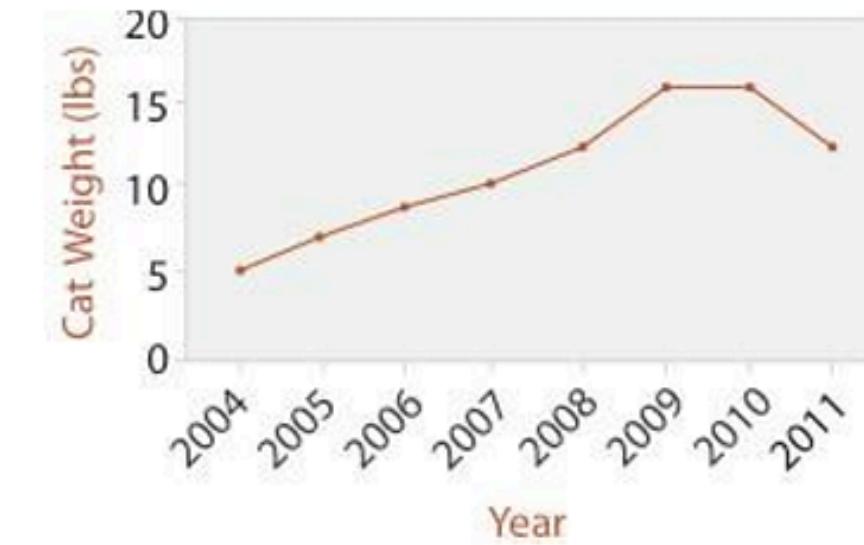
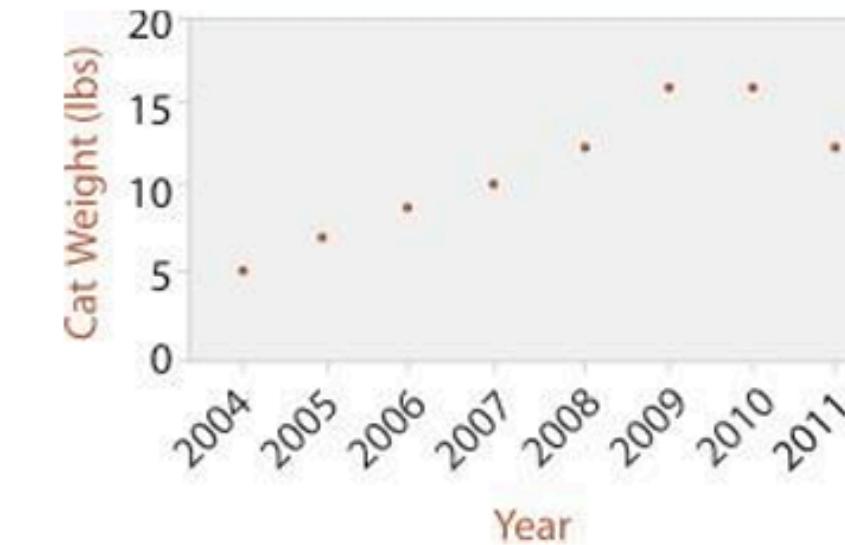
Change over Time

Emphasize changing trends

- important to choose the correct time period to provide enough context

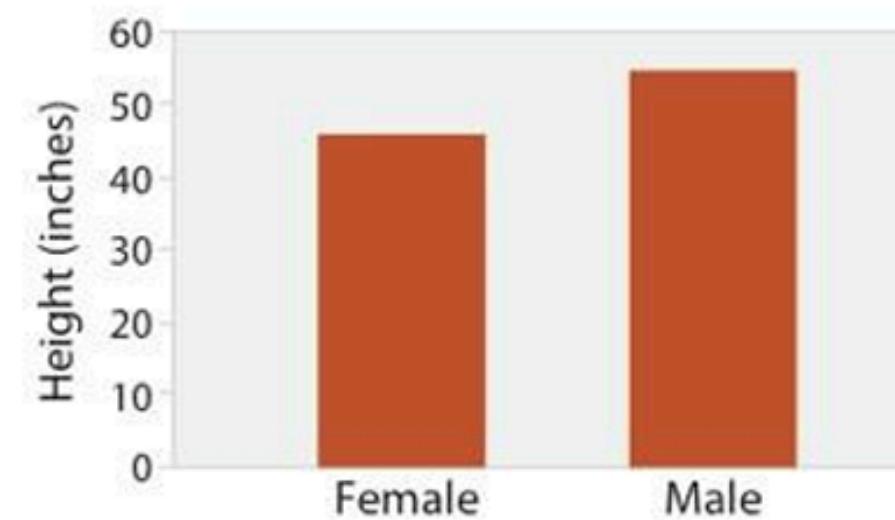
Line Chart

- simple, familiar, accurate
- fairly scalable

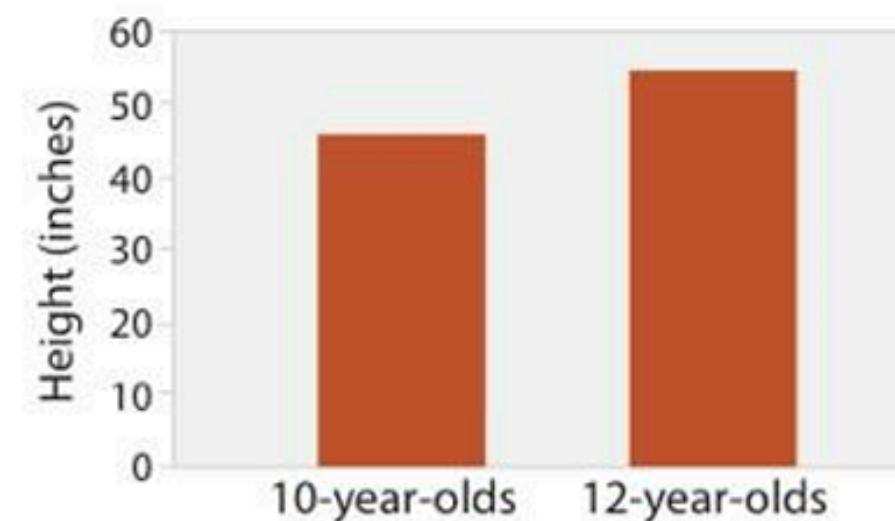
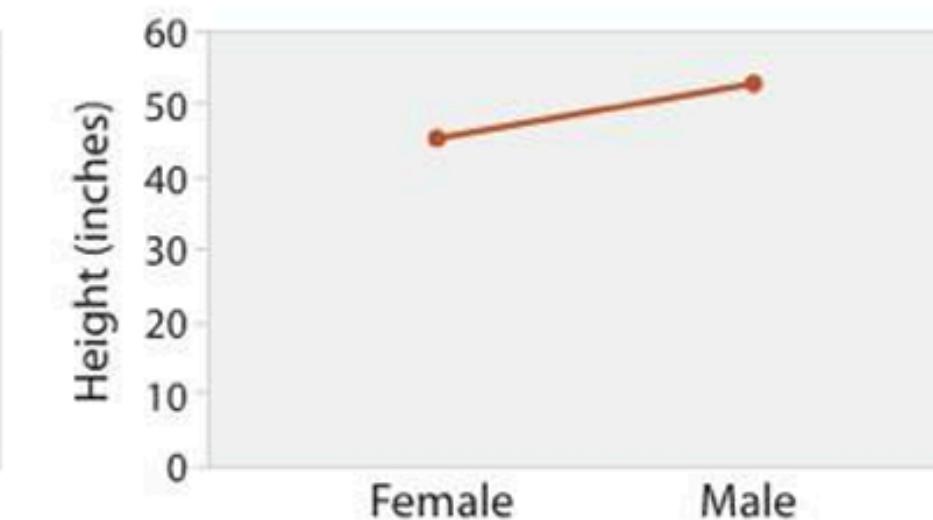


Line Chart: Bad for Categorical Attributes

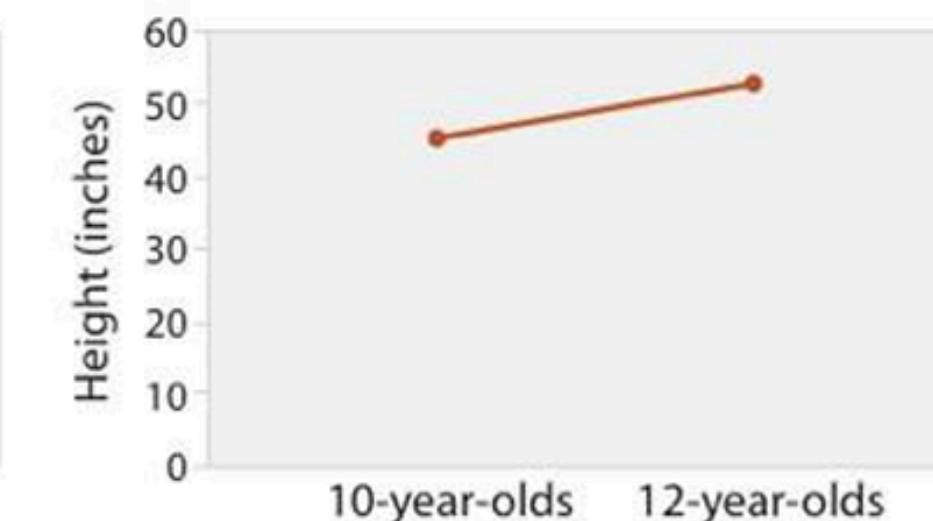
ok: “Men are taller than women (on average)”



bad: “The more male a person is, the taller he/she is”



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



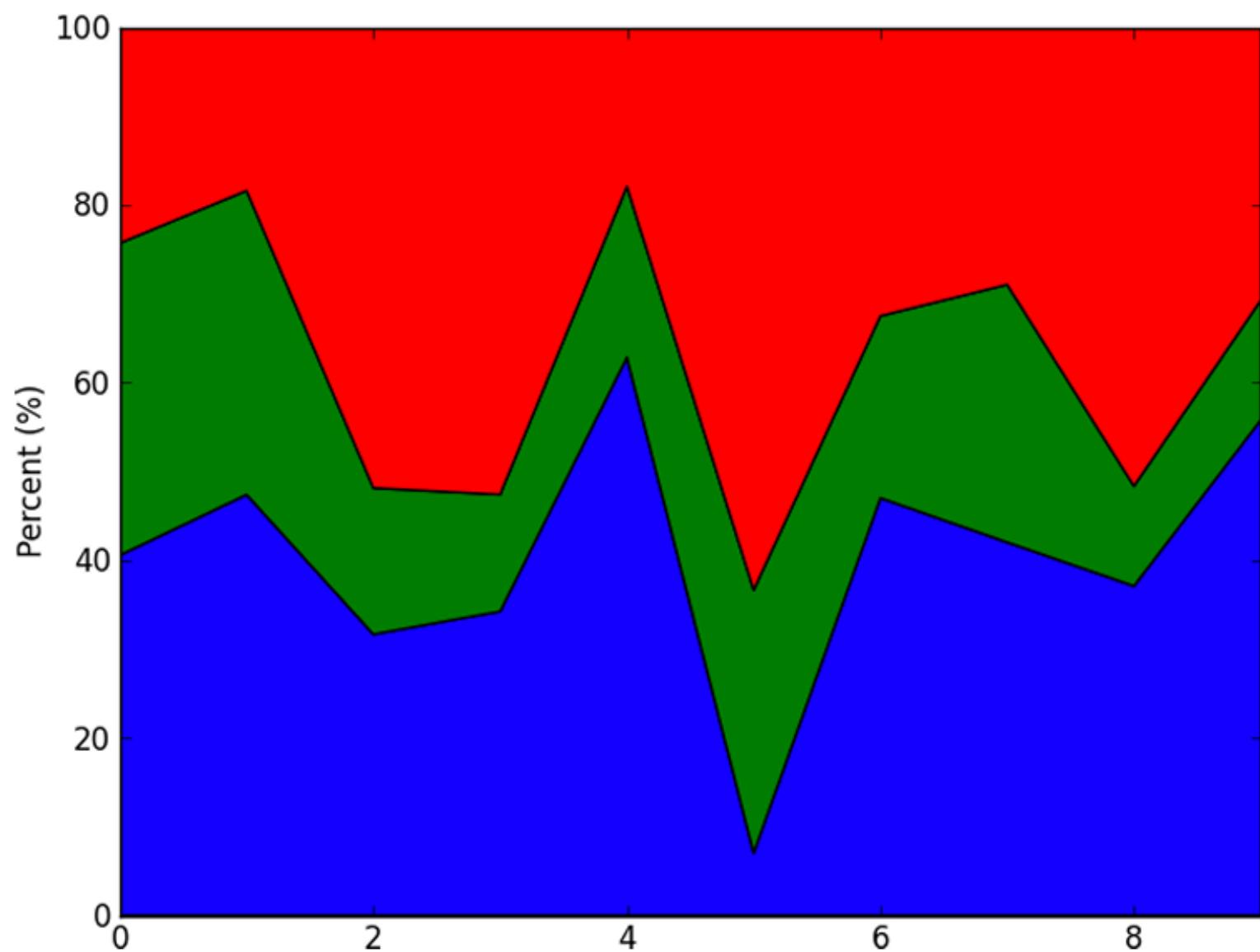
ok: “Height increases with age”

Stacked Area Chart

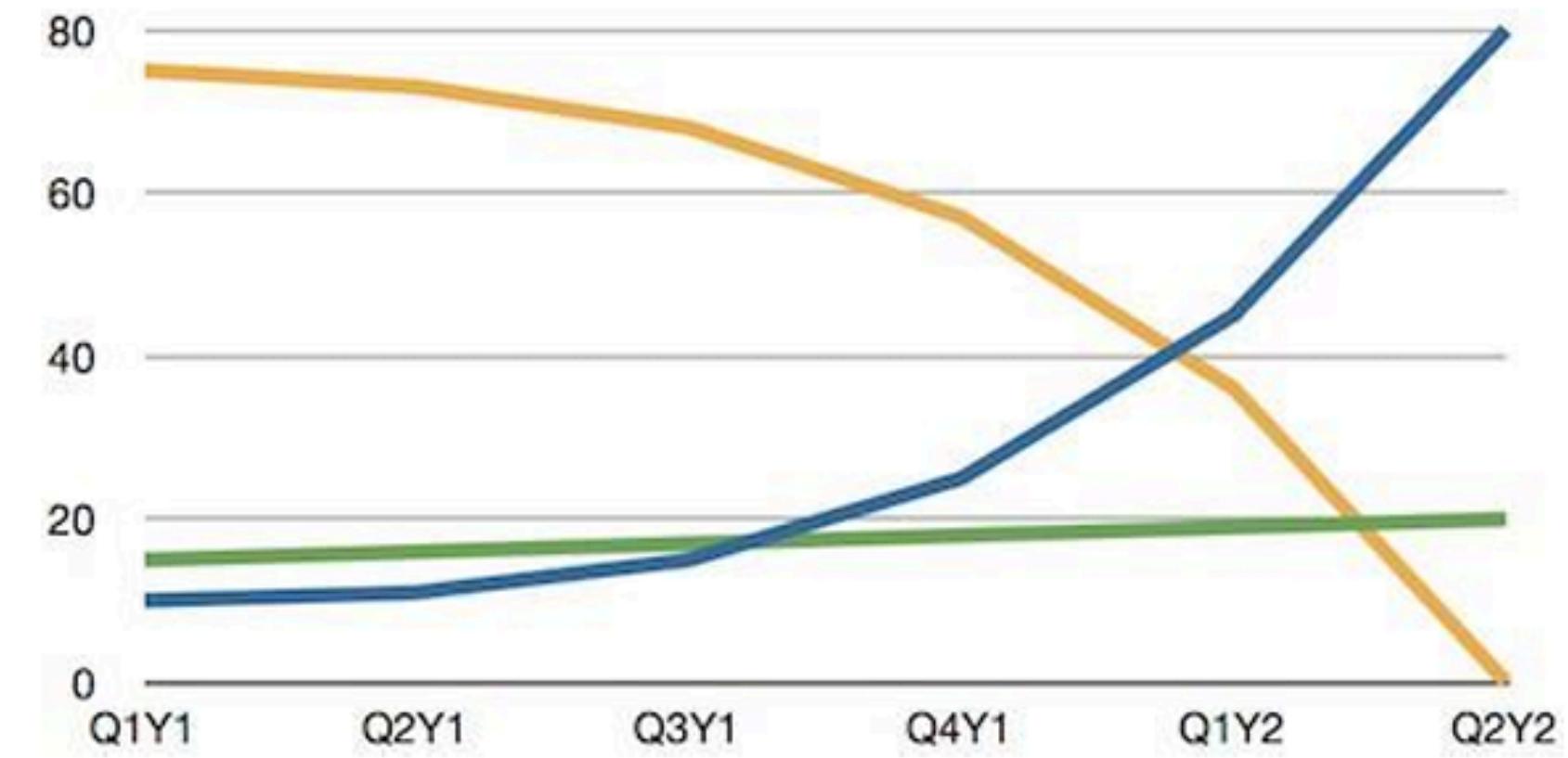
Absolute Totals



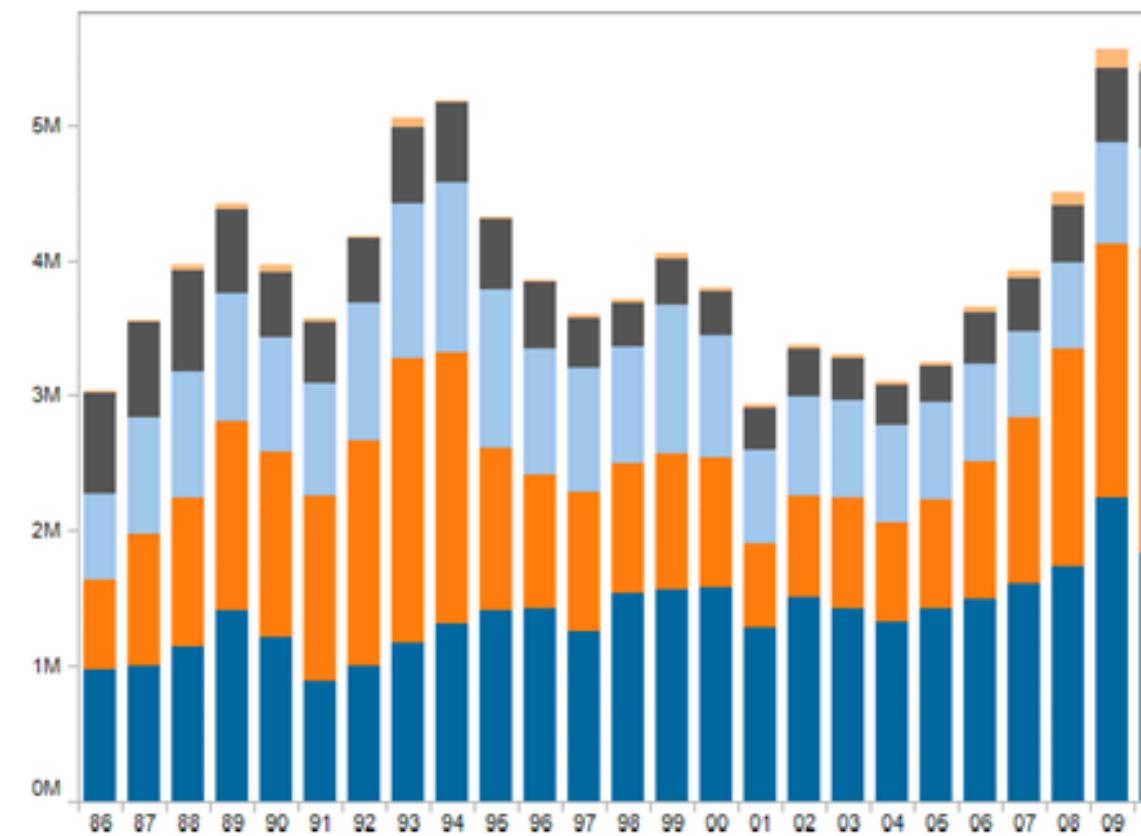
Percentages



Stacked Area Chart vs. Line Charts

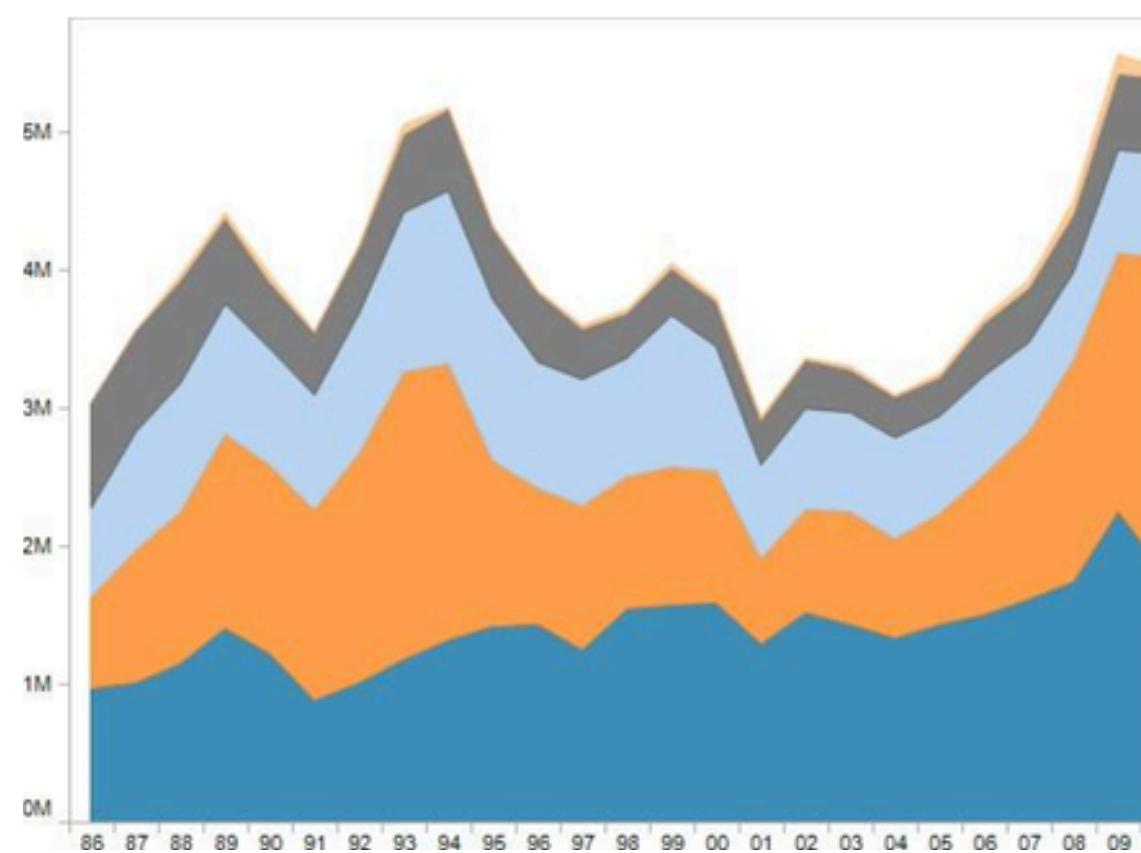


Trends with Different Charts



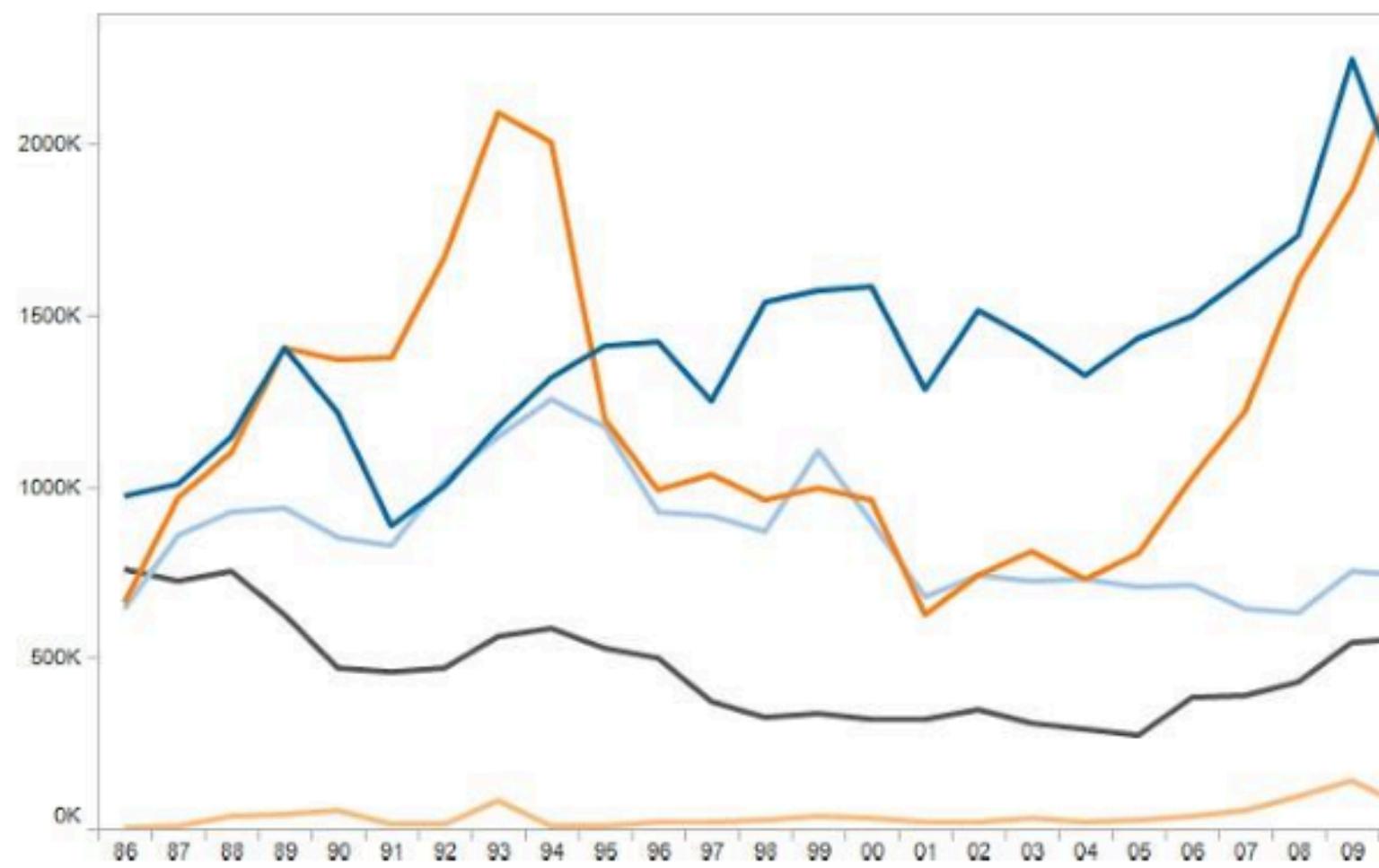
Weapon

- Misc
- Revolvers
- Shotguns
- Pistols
- Rifles



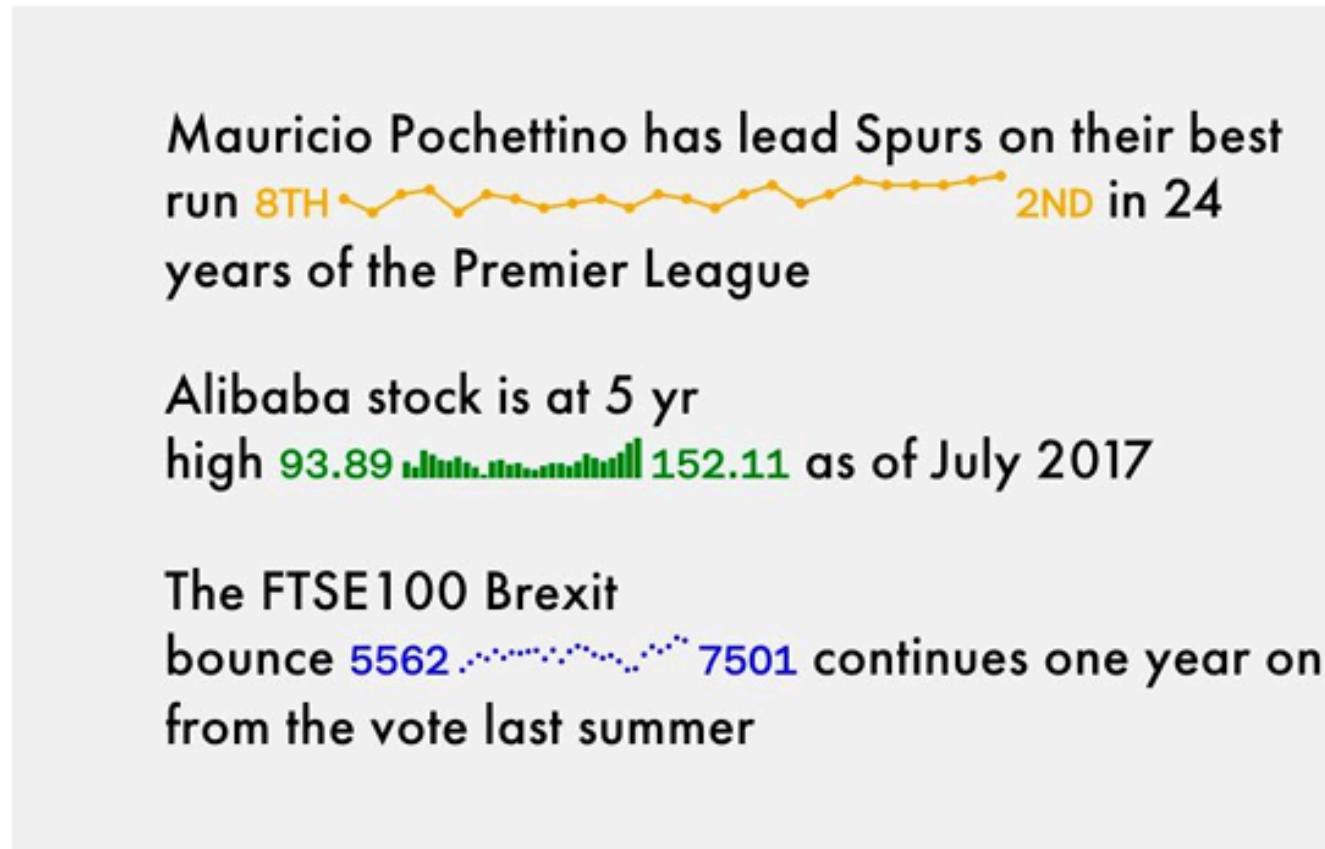
Weapon

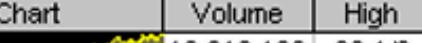
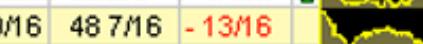
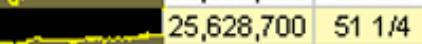
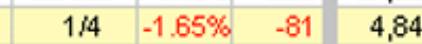
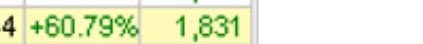
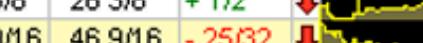
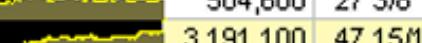
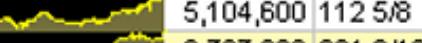
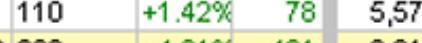
- Misc
- Revolvers
- Shotguns
- Pistols
- Rifles



Sparklines

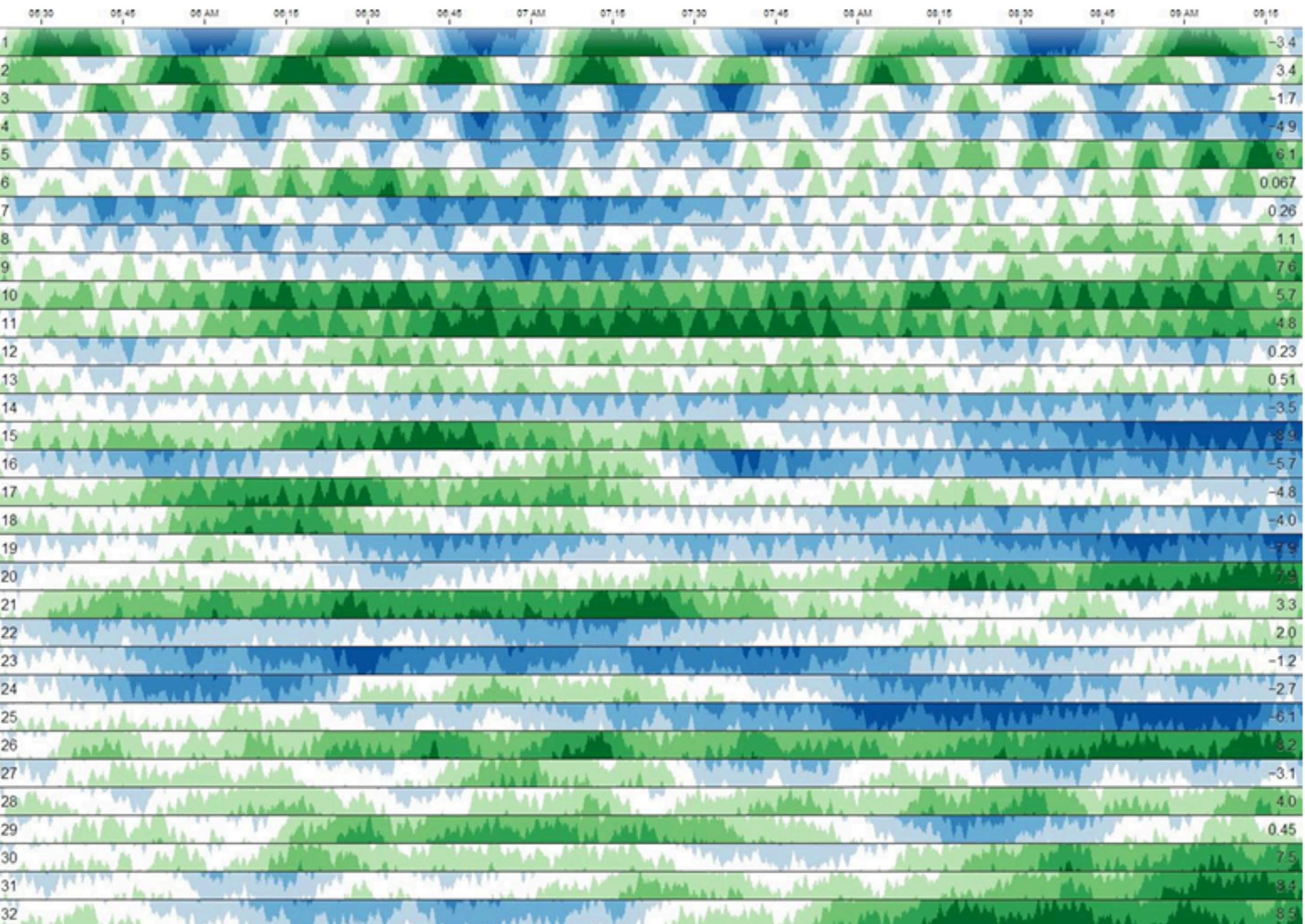
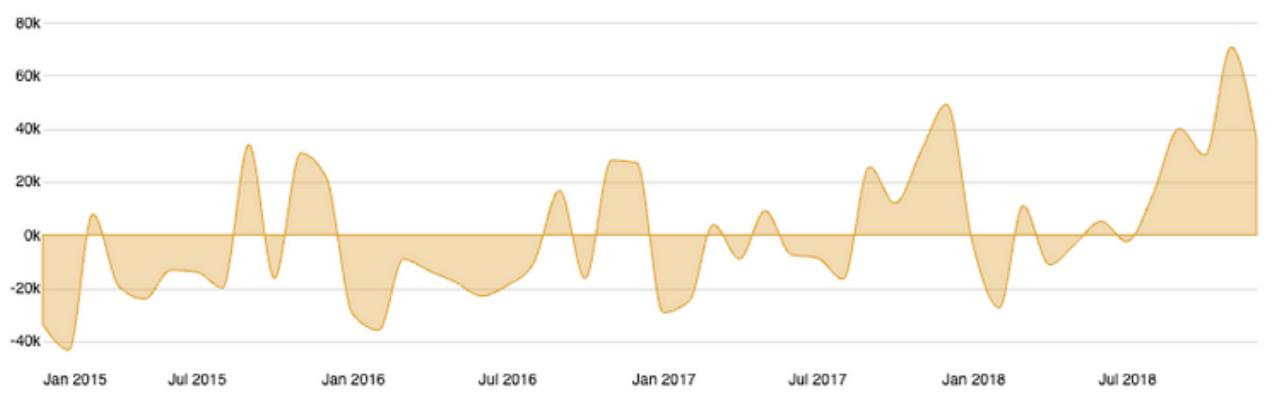
- small line charts
- can be embedded in text or in a spreadsheet/table



Symbol	Bid	Ask	Last	Change	T	Chart	Volume	High	Low	Value Change	Value	Gain
DELL	89 3/4	89 13/16	89 3/4	+ 1 1/4	↑		10,310,100	90 1/8	88 1/2	+1.41%	250	17,950 +273.72% 13,147
CPQ	48 7/16	48 9/16	48 7/16	- 13/16	↓		25,628,700	51 1/4	1/4	-1.65%	-81	4,844 +60.79% 1,831
SDTI	26 1/4	26 3/8	26 3/8	+ 1/2	↓		504,600	27 3/8	25 5/8	+1.93%	250	13,188 +133.15% 7,531
COMS	46 1/2	46 9/16	46 9/16	- 25/32	↓		3,191,100	47 15/16	45 3/4	-1.65%	-102	6,053 +29.79% 1,389
LU	111 5/8	111 11/16	111 9/16	+ 1 9/16	↑		5,104,600	112 5/8	110	+1.42%	78	5,578 +22.76% 1,034
YHOO	368 1/16	368 1/2	368 1/2	+ 17 1/4	↓		3,787,800	381 3/16	280	+4.91%	431	9,213 -0.41% -38
AOL	162 13/16	163	163	+ 8	↑		10,008,500	164	158 1/2	+5.16%	280	5,705 +73.06% 2,408
CMGI	97 3/8	97 1/2	97 1/2	+ 5 7/8	↓		1,323,800	98 1/2	93	+6.41%	705	11,700 +186.76% 7,620
SPLN	33 13/16	33 15/16	33 13/16	+ 7/16	↓		300,200	34 3/4	33 5/8	+1.31%	88	6,763 +94.60% 3,288
BEAS	13 1/2	13 5/8	13 5/8	- 7/16	↓		389,200	14 1/4	13 1/8	-3.11%	-44	1,363 -9.17% -138
GNET	102	103 3/16	101 5/16	+ 6 1/8	↑		307,600	108	97	+6.43%	613	10,131 +130.26% 5,731
RNWK	67	67 1/4	67	+ 2 3/4	↓		1,233,900	69	64 15/16	+4.28%	275	6,700 +79.87% 2,975
MSFT	173 1/8	173 1/4	173 5/16	+ 1 3/4	↓		13,284,500	174 7/16	170	+1.02%	175	17,331 +54.74% 6,131
INTC	133 3/4	133 13/16	133 13/16	- 3 1/8	↓		8,094,300	137 1/2	133 3/8	-2.28%	-625	26,763 +65.20% 10,563
TOTAL					↑		205,302	80,993	+1.63%	2,293	143,280 +79.41% 63,377	

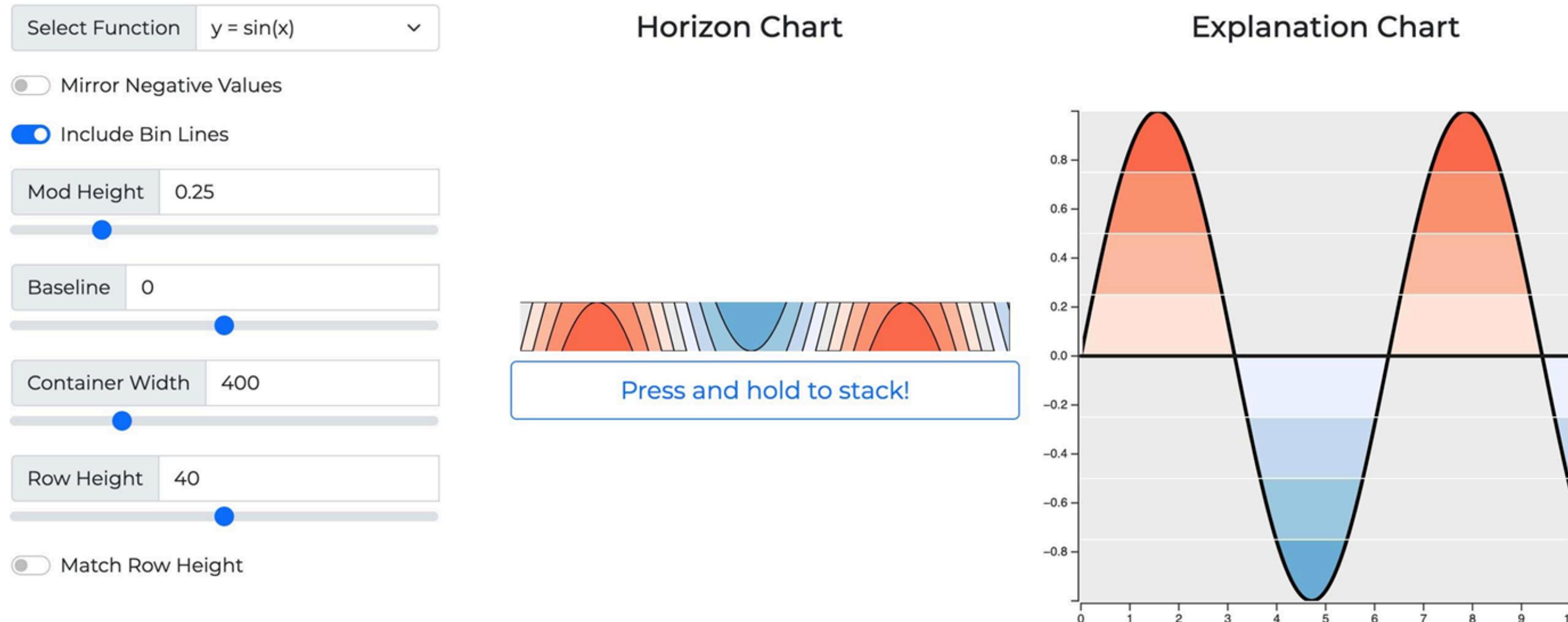
Horizon Charts

Alternative to diverging area chart



Horizon Charts

A Horizon Chart is a specialized type of chart for time series data. It is especially useful for showing data with large amplitudes in a short vertical space. The idea was introduced by Saito et al. in [Two-Tone Pseudo Coloring: Compact Visualization for One-Dimensional Data](#). Panopticon commercialized and coined the term [Horizon Chart](#). Like any novel visualization, one downside is the cost for your audience to learn and understand that chart. Therefore, I have built this interactive visualization to help make it easier to understand how Horizon Charts work.



<https://www.horizon-chart-explanation.devinlange.com/>

Connected Scatter Plot

- Two Attributes + Time
- Labels are important

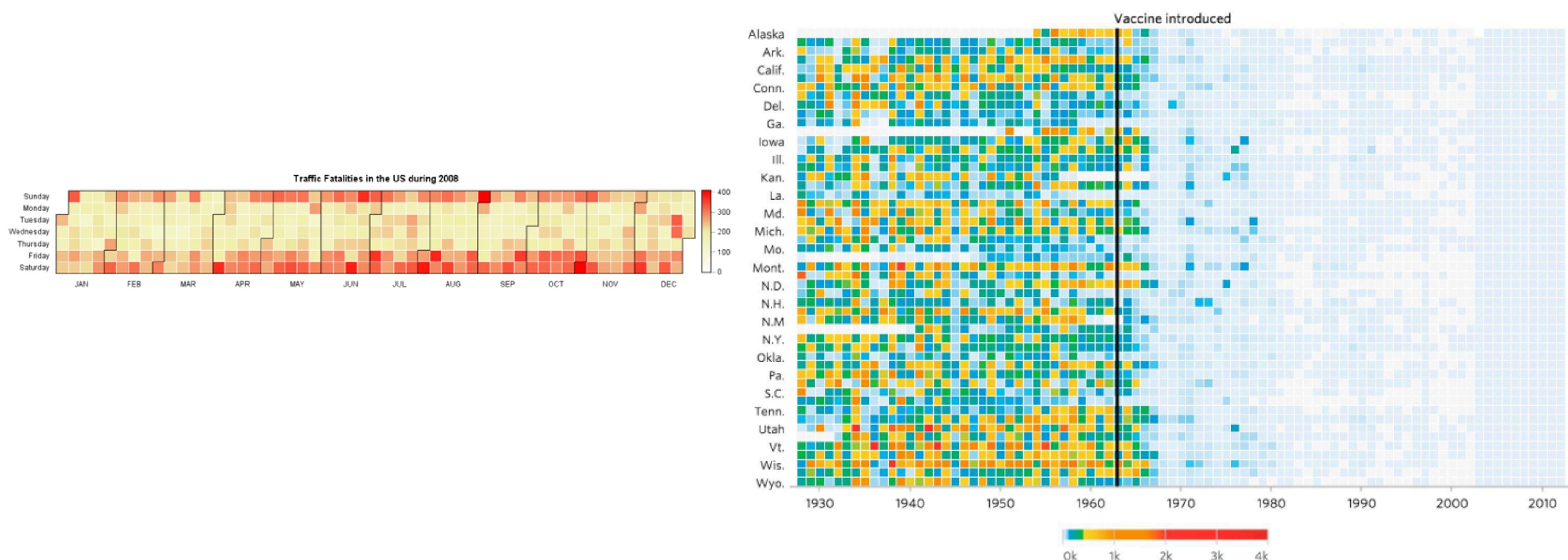
<https://archive.nytimes.com/www.nytimes.com/interactive/2012/09/17/science/driving-safety-in-fits-and-starts.html>



Calendar Heatmaps

The heat maps below show number of cases per 100,000 people.

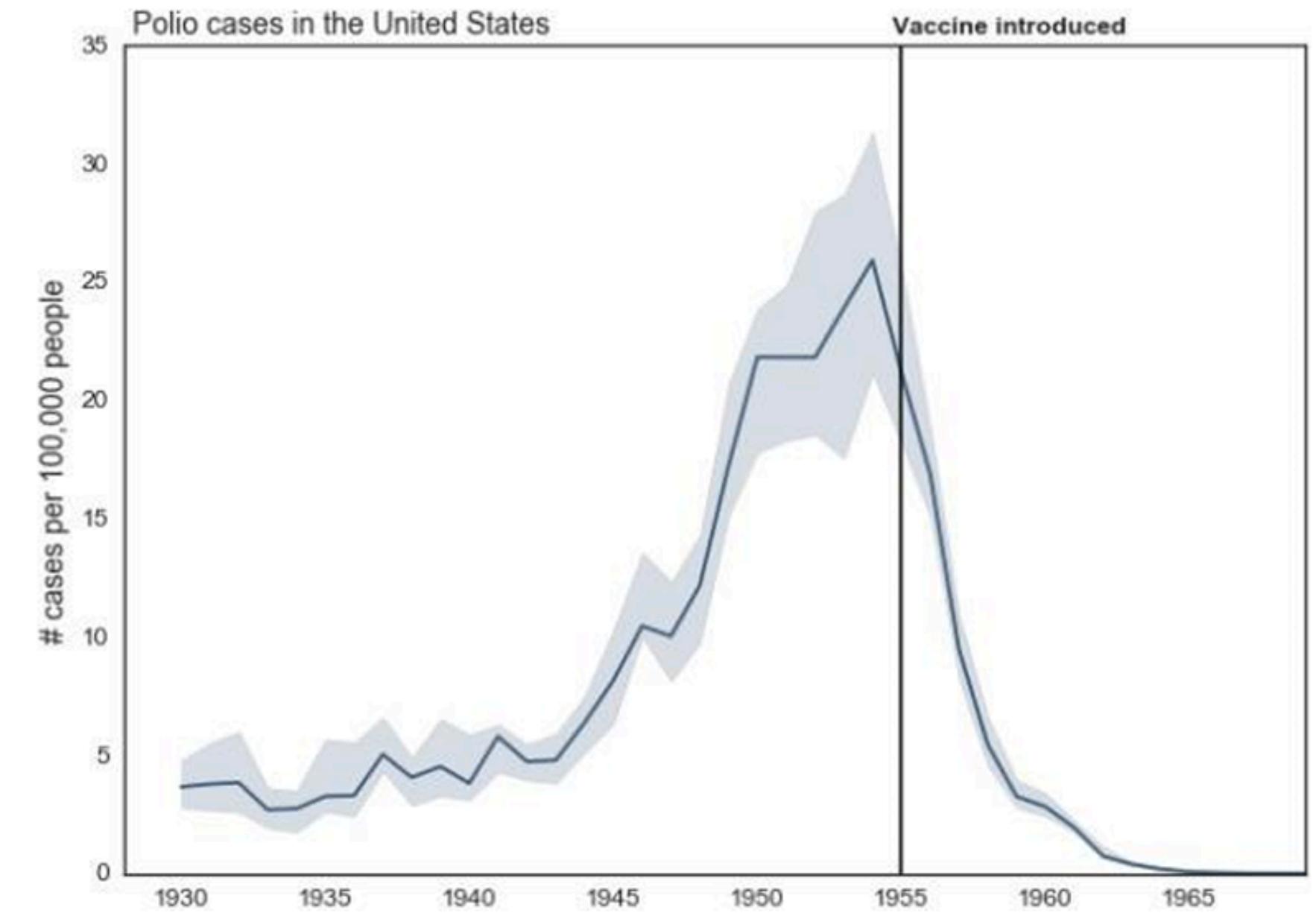
Measles



Note: CDC data from 2003-2012 comes from its Summary of Notifiable Diseases, which publishes yearly rather than weekly and counts confirmed cases as opposed to provisional ones.

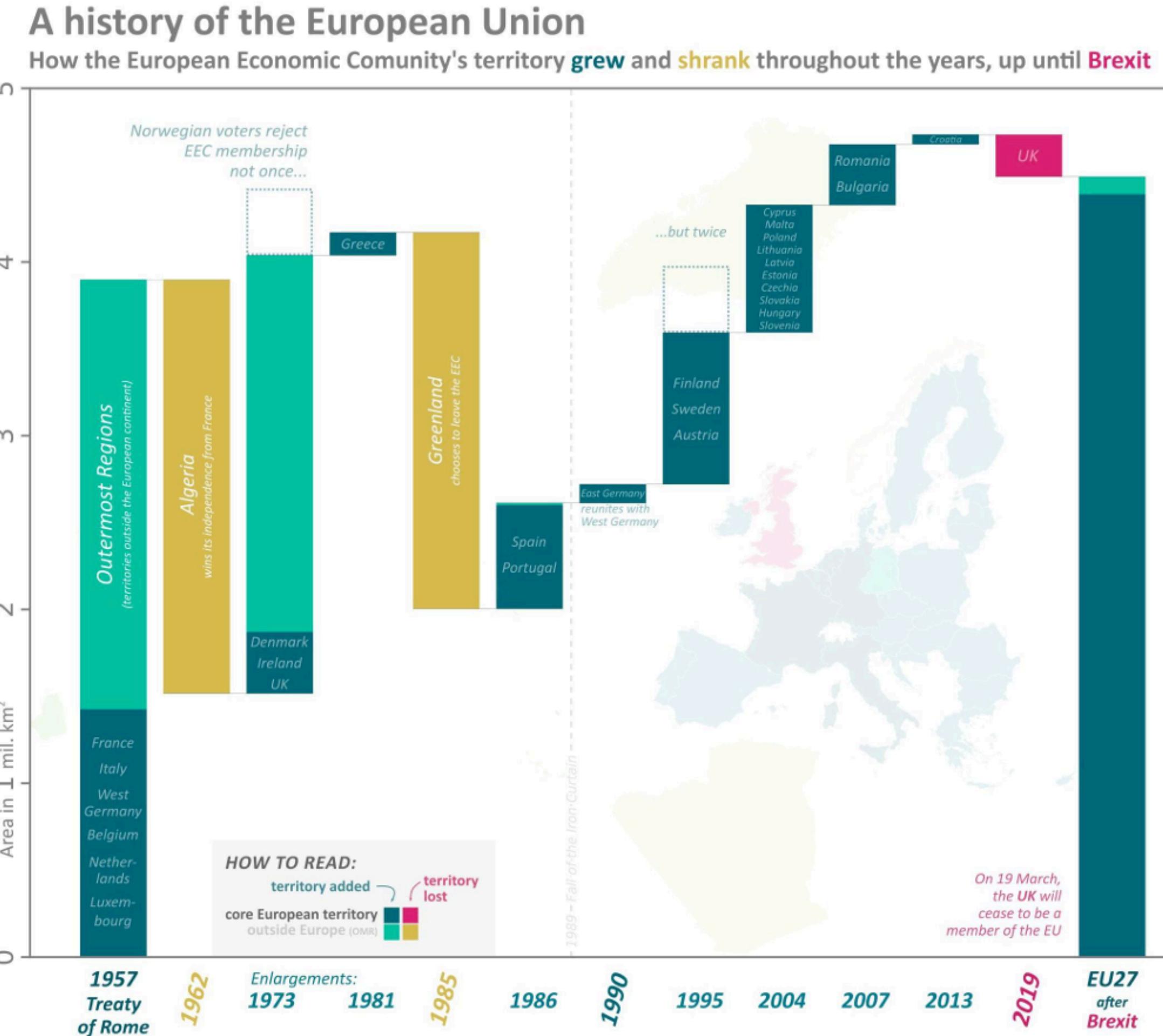
Calendar Heatmaps

Sometimes you can show too much data...



Waterfall Chart

Useful for showing part-of-whole over time

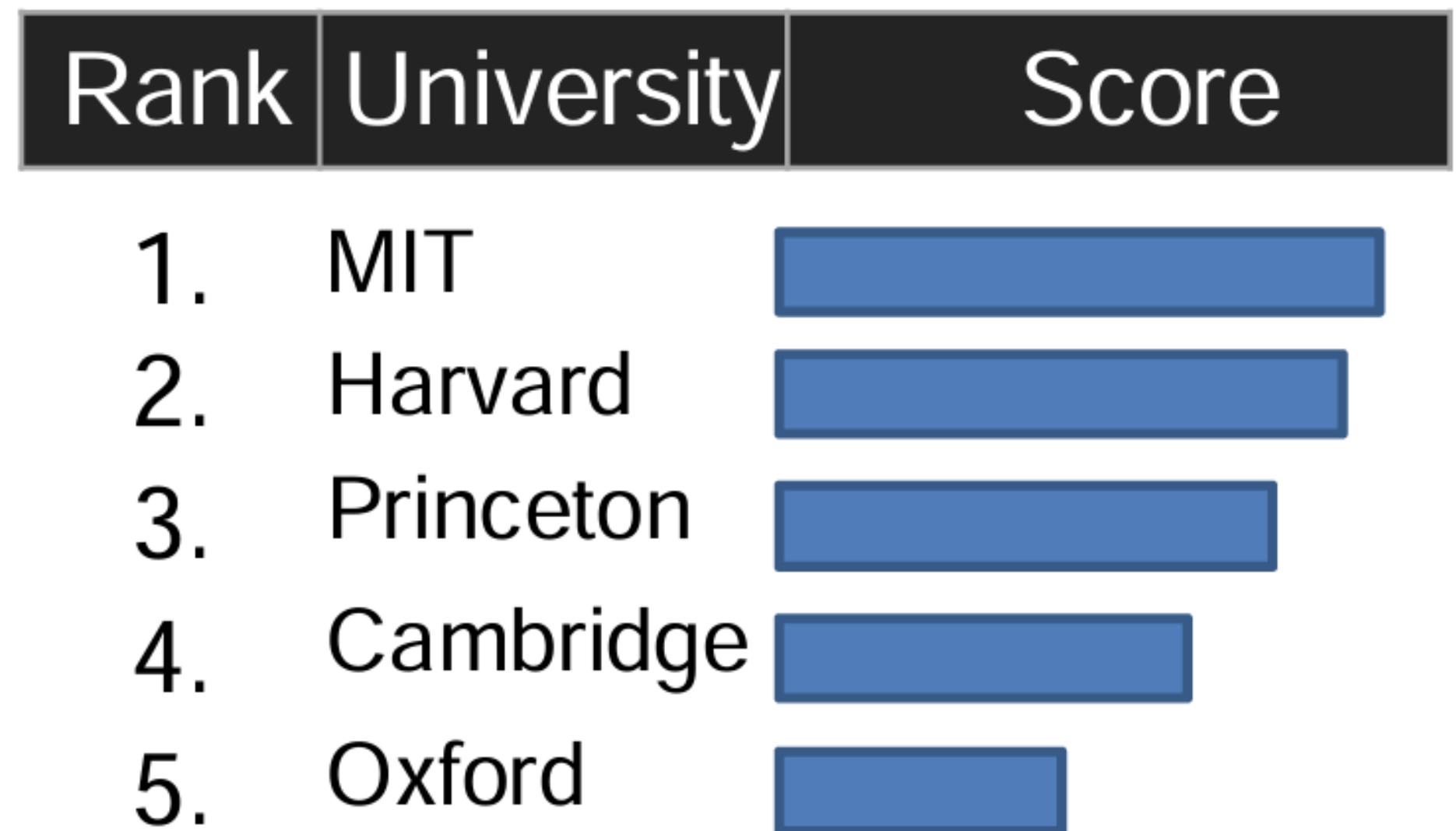




6. Ranking

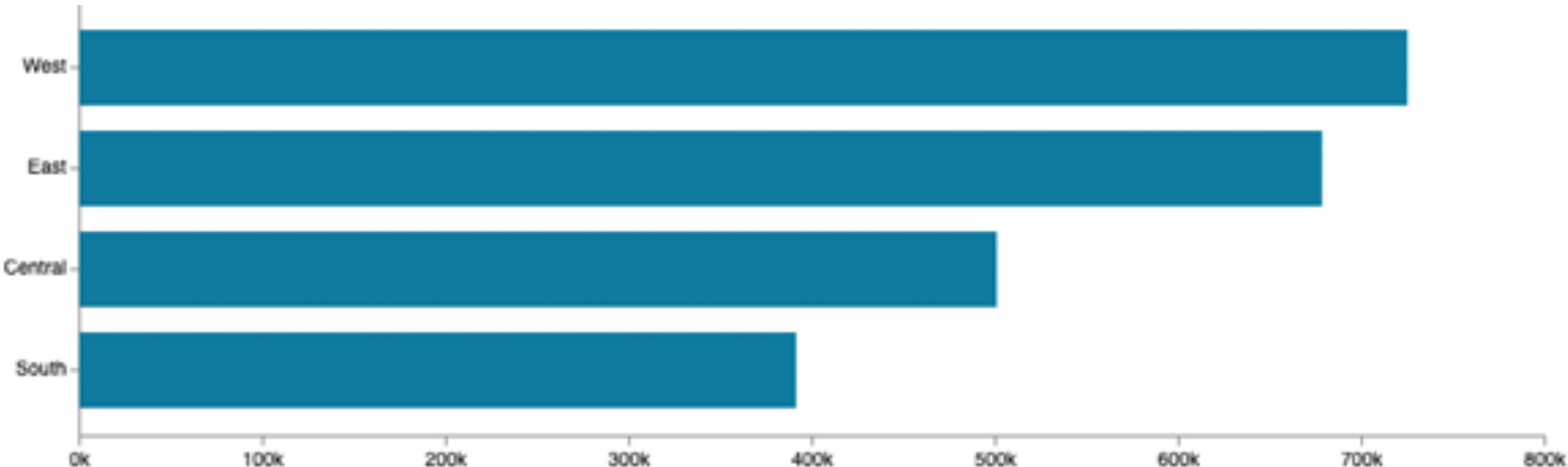
Ranking

When position in an ordered list is more important than the an item's absolute or relative value



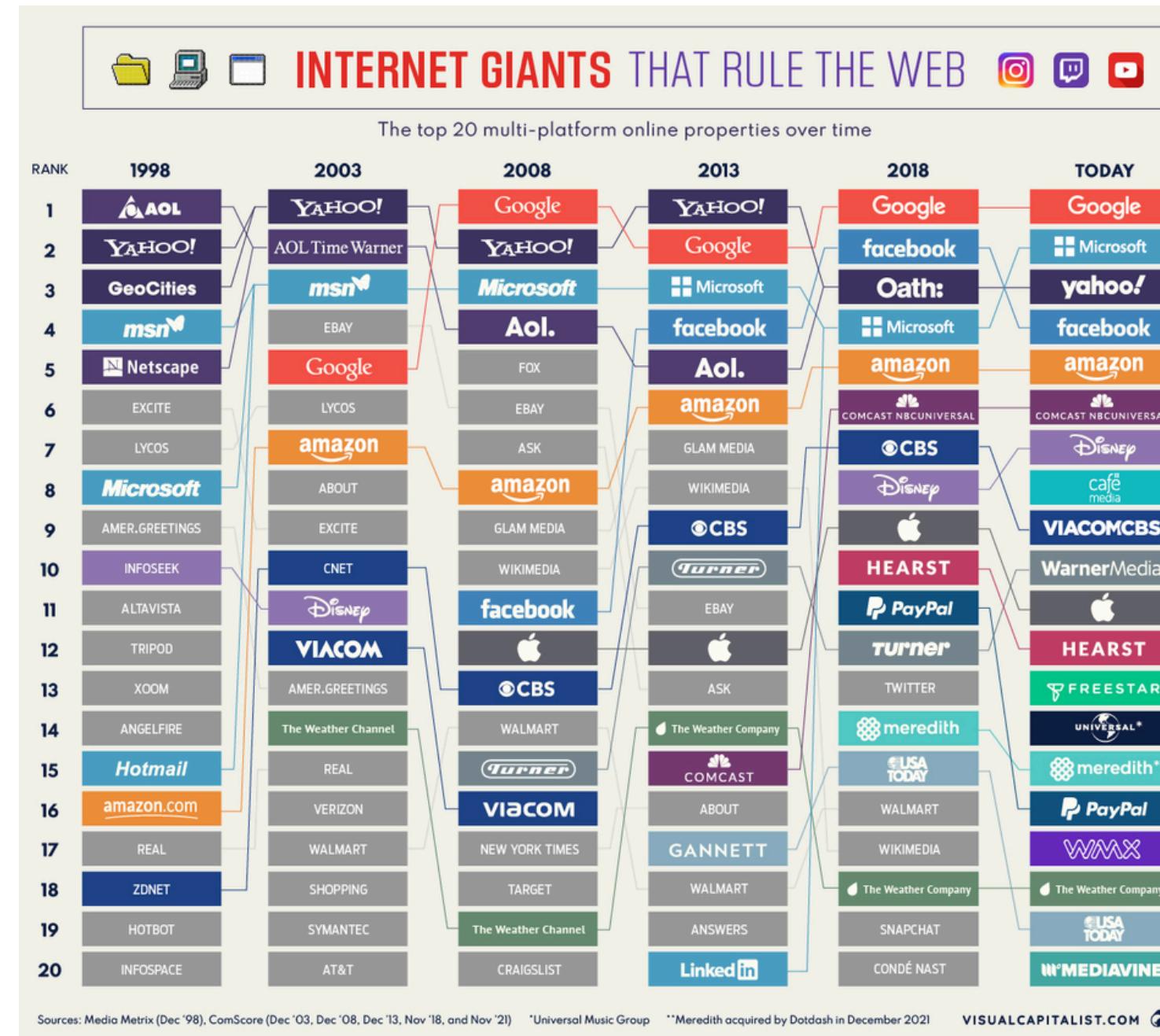
Common Methods for Visualizing Ranking

Magnitude + Sorting

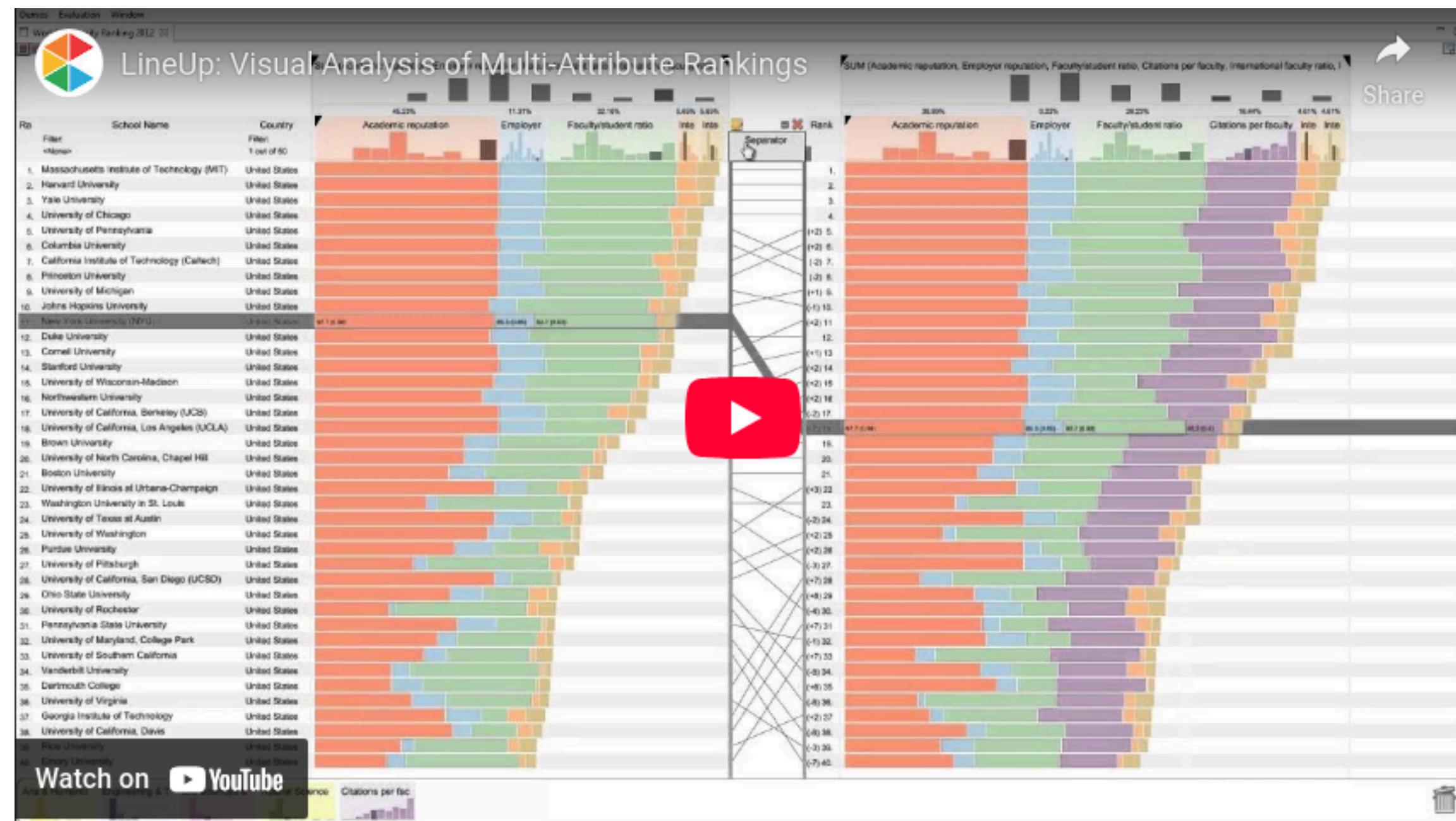


Common Methods for Visualizing Ranking

Bump Chart (useful for rankings over time)



LineUp



<https://jku-vds-lab.at/tools/lineup/>



7. Correlation

Correlation

Show relationship between two or more variables

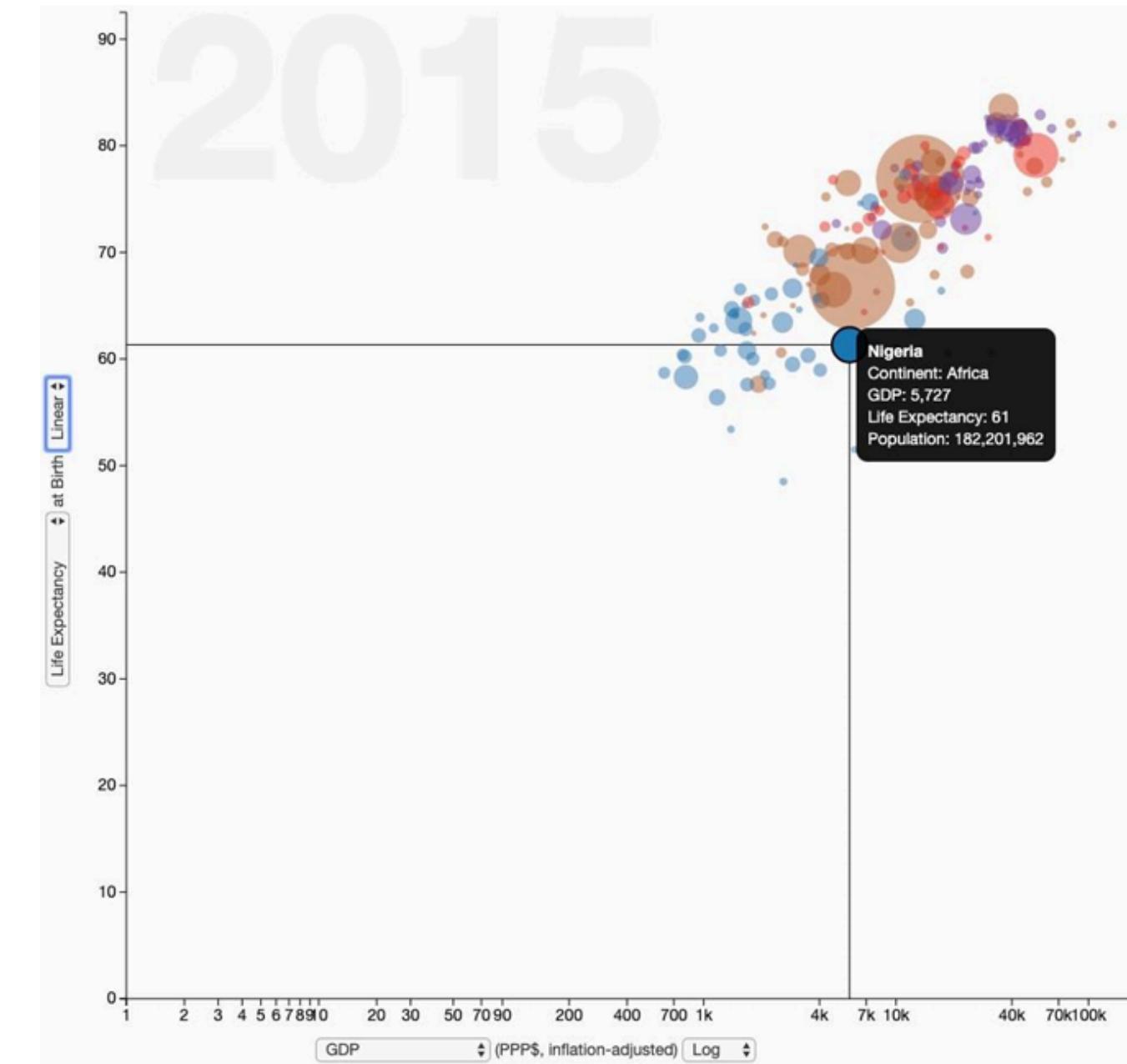
- Careful! Users will likely assume that visualized correlations are causal unless explicitly told otherwise

Axis-Based Techniques

- Scatter Plot
- Scatter Plot Matrix (SPLOM)
- Parallel Coordinates

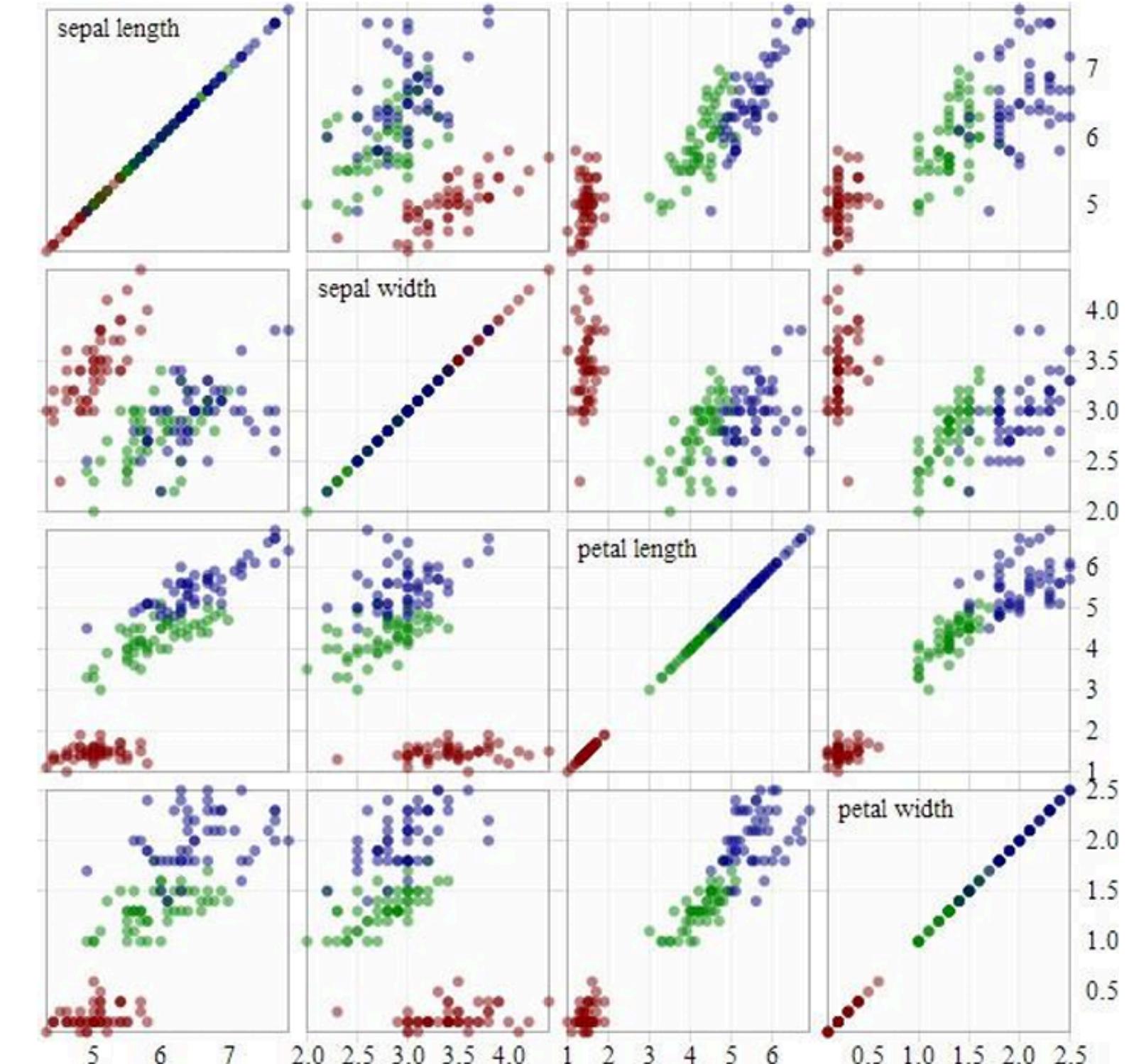
Axis-Based Techniques: Scatter Plots

- Two orthogonal axes visualizing one attribute each



Axis-Based Techniques: Scatter Plot Matrix

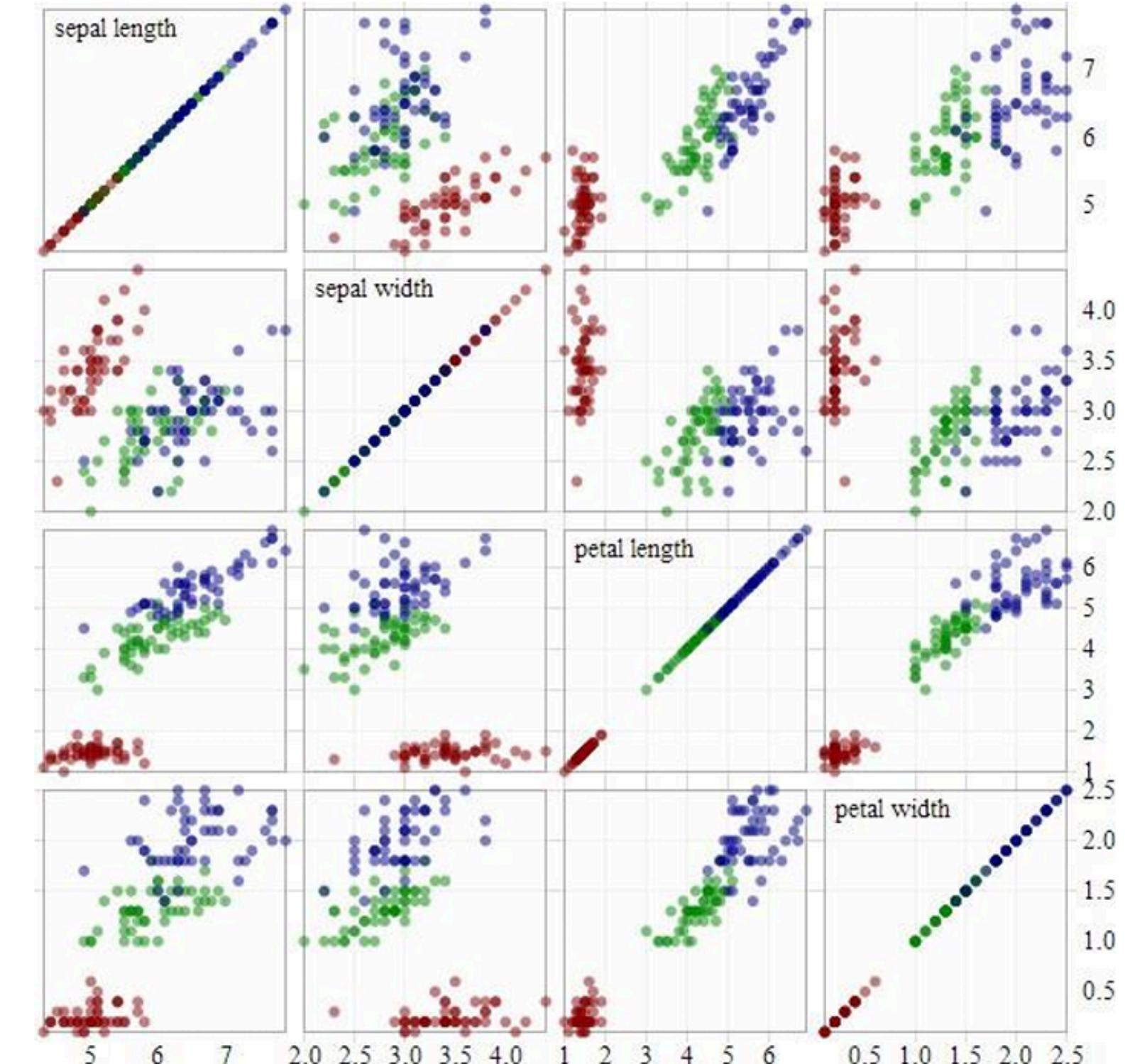
- Matrix of size $d * d$
- Each row/column is one dimension
- Each cell plots a scatter plot of two dimensions



Axis-Based Techniques: Scatter Plot Matrix

Considerations:

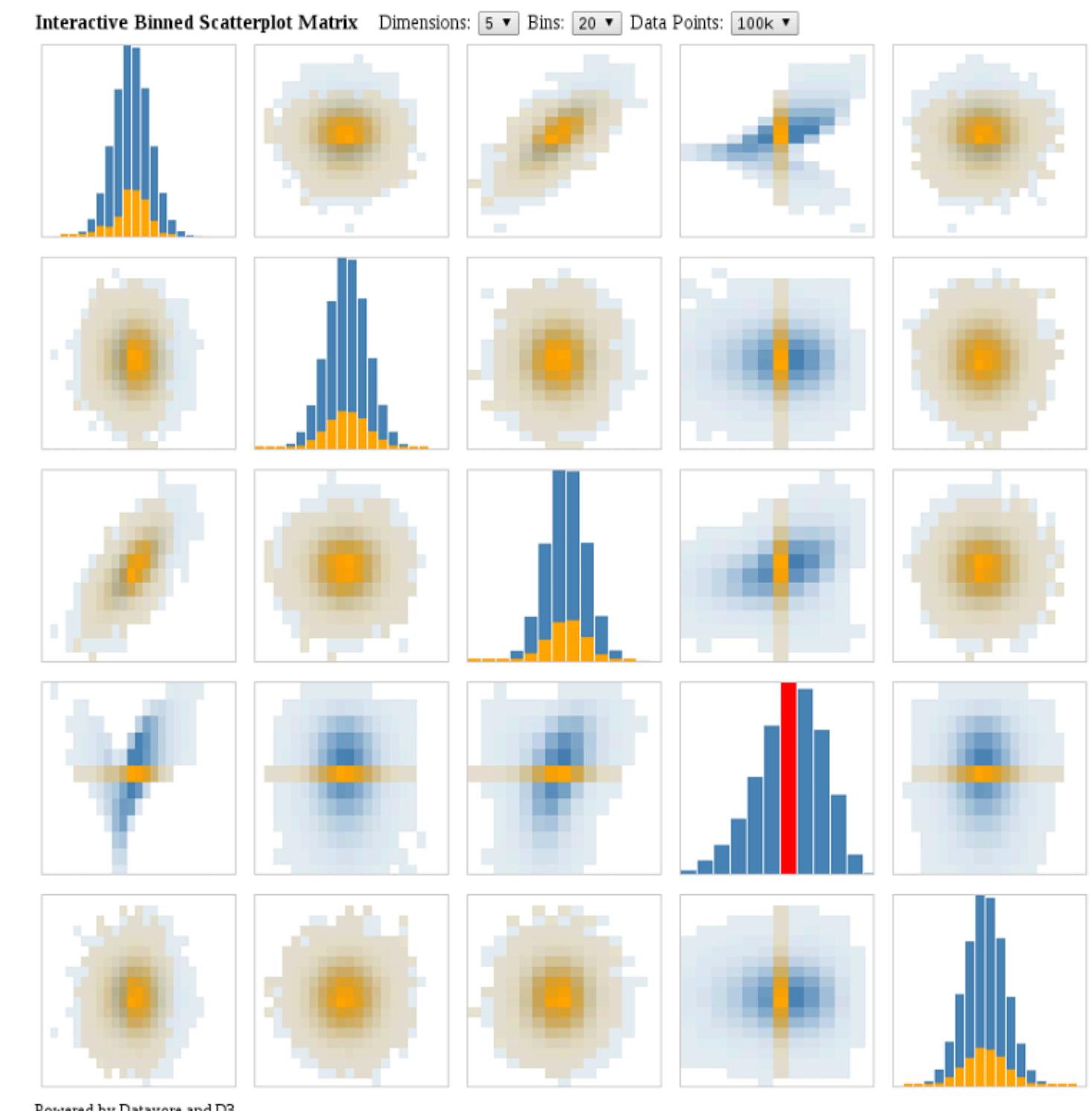
- Limited Scalability (max ~20 attributes)
- brush and link is necessary



Axis-Based Techniques: Scatter Plot Matrix

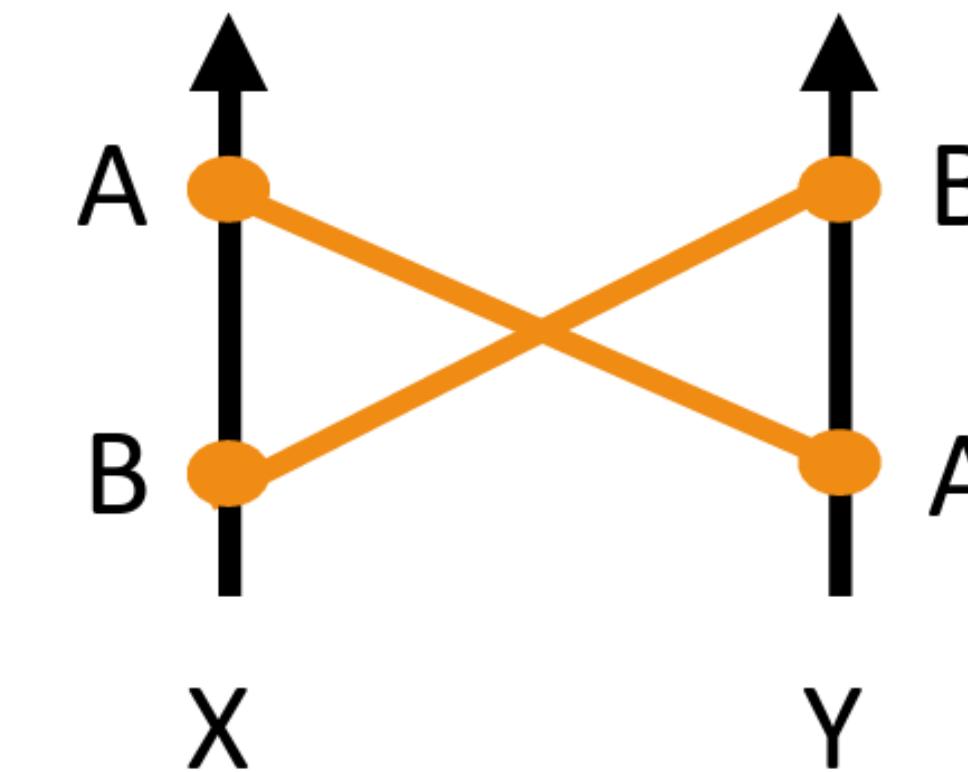
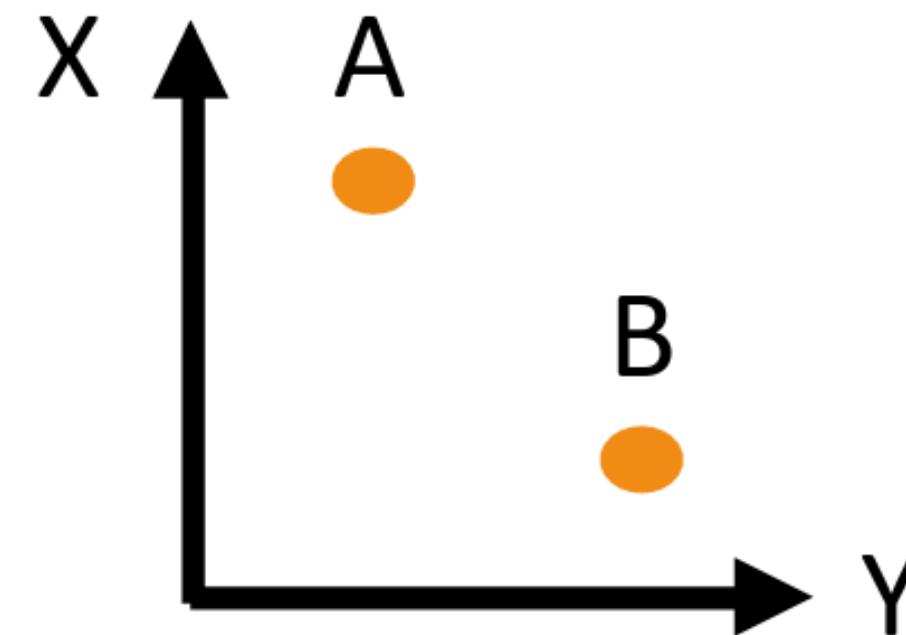
Scatter Plot Matrix + Heatmap

- aggregation

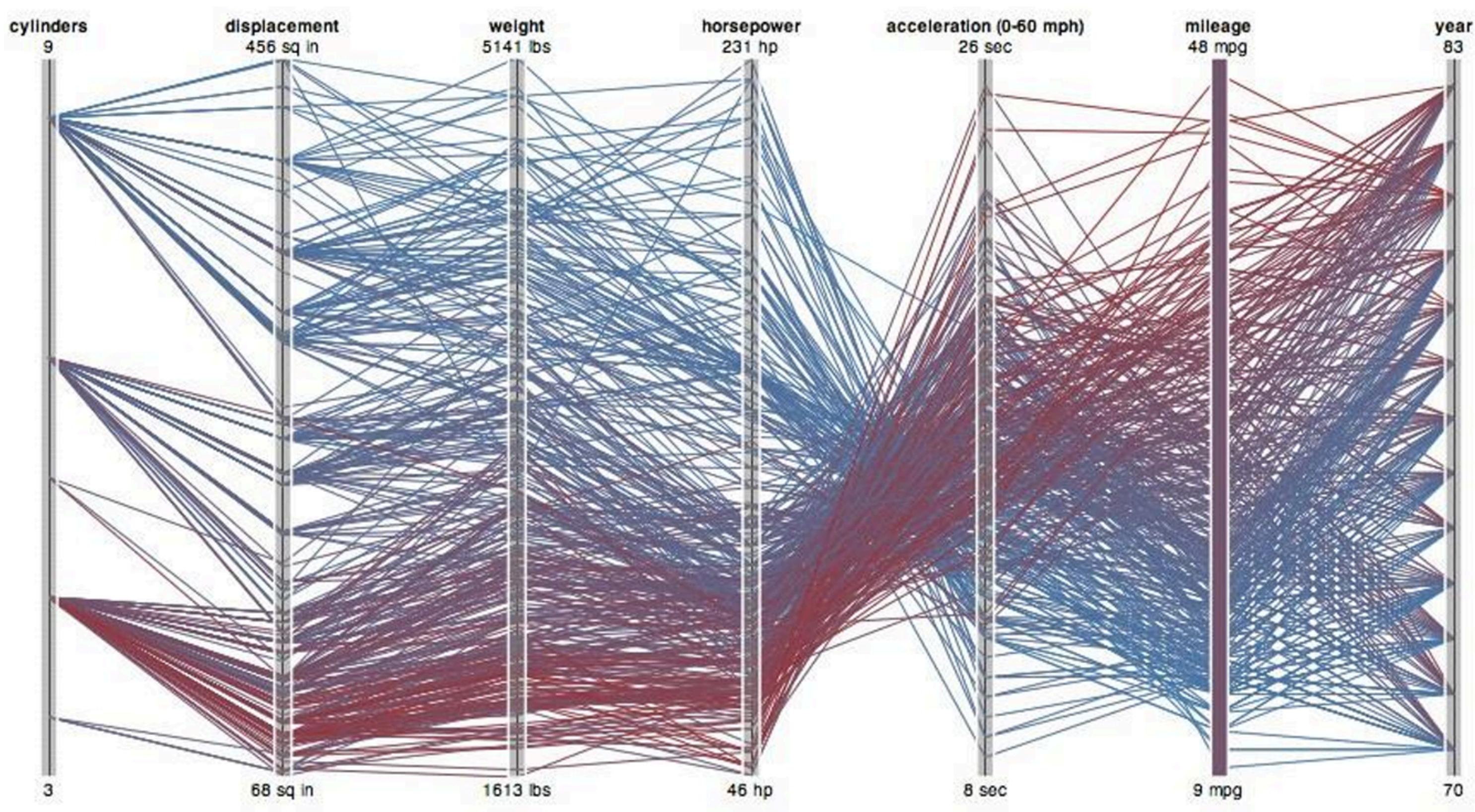


Axis-Based Techniques: Parallel Coordinates

- Each vertical axis represents a different attribute
- Lines connecting axes represent items
- Suitable for:
 - all tabular data types
 - heterogenous data



Axis-Based Techniques: Parallel Coordinates



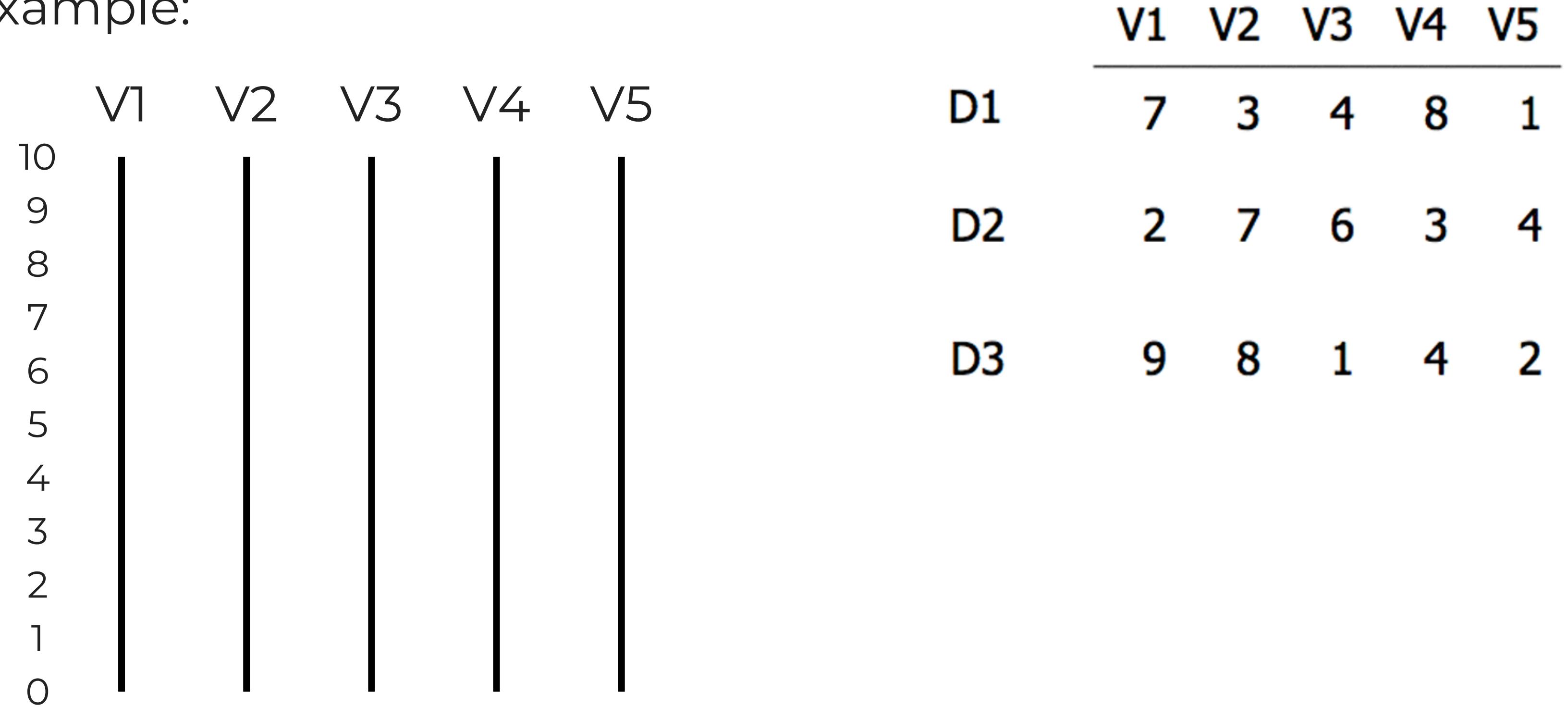
Axis-Based Techniques: Parallel Coordinates

Example:

	V1	V2	V3	V4	V5
D1	7	3	4	8	1
D2	2	7	6	3	4
D3	9	8	1	4	2

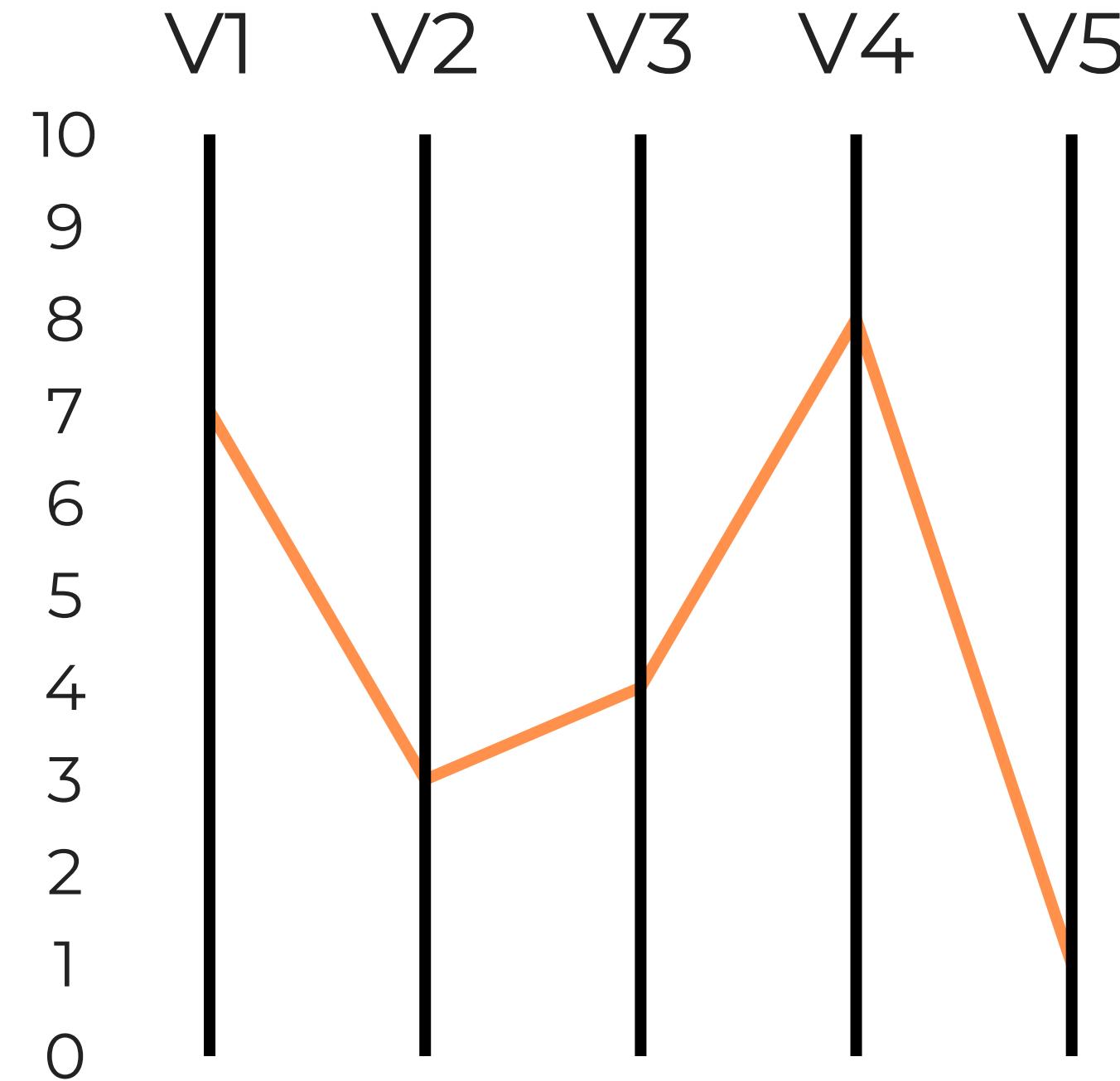
Axis-Based Techniques: Parallel Coordinates

Example:



Axis-Based Techniques: Parallel Coordinates

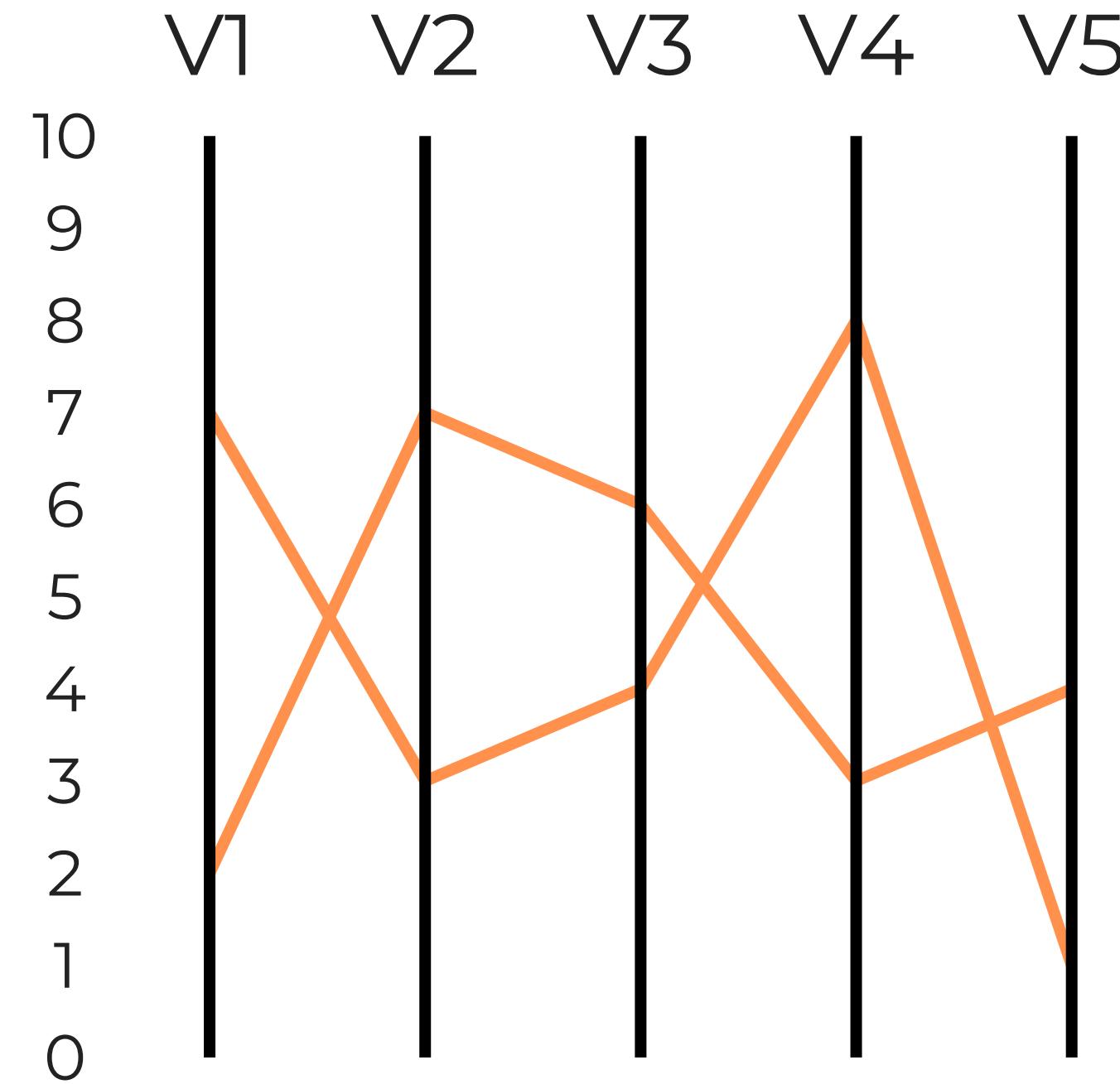
Example:



	V1	V2	V3	V4	V5
D1	7	3	4	8	1
D2	2	7	6	3	4
D3	9	8	1	4	2

Axis-Based Techniques: Parallel Coordinates

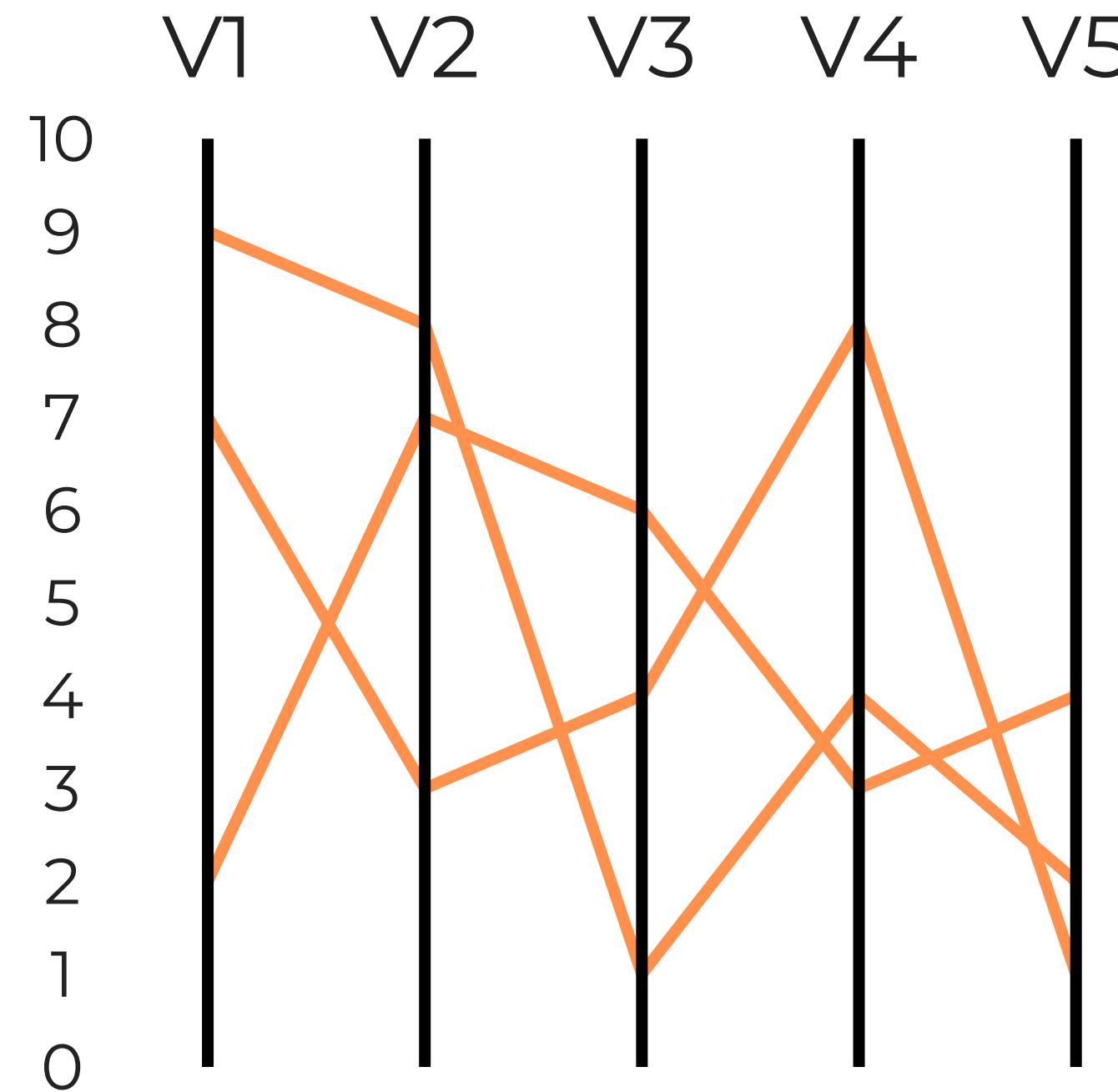
Example:



	V1	V2	V3	V4	V5
D1	7	3	4	8	1
D2	2	7	6	3	4
D3	9	8	1	4	2

Axis-Based Techniques: Parallel Coordinates

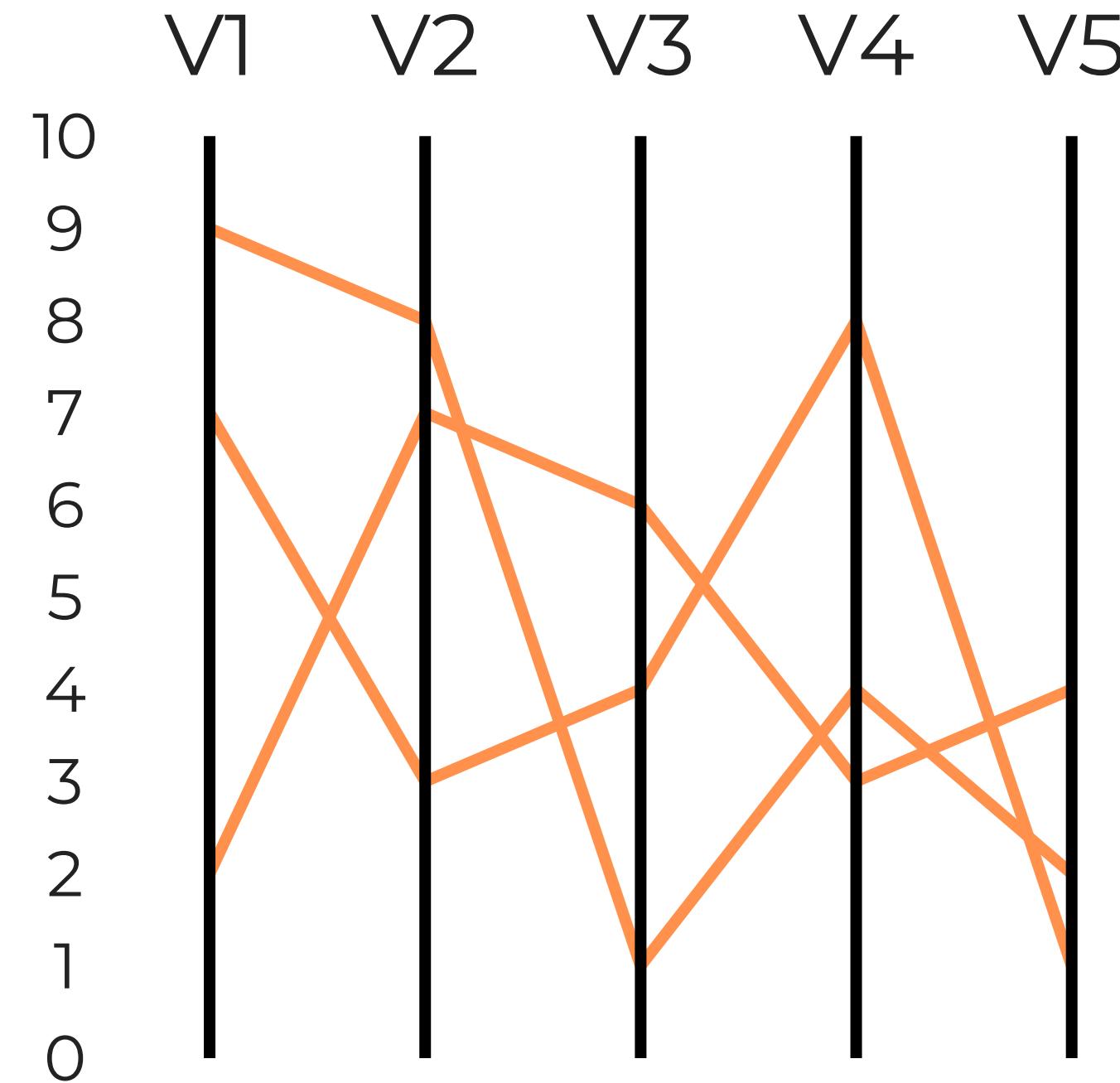
Example:



	V1	V2	V3	V4	V5
D1	7	3	4	8	1
D2	2	7	6	3	4
D3	9	8	1	4	2

Axis-Based Techniques: Parallel Coordinates

Example:



	V1	V2	V3	V4	V5
D1	7	3	4	8	1
D2	2	7	6	3	4
D3	9	8	1	4	2

Axis-Based Techniques: Parallel Coordinates

Reading Correlation from Parallel Coordinates:

- Positive Correlation = Parallel Lines
- Negative Correlation = Intersecting Lines

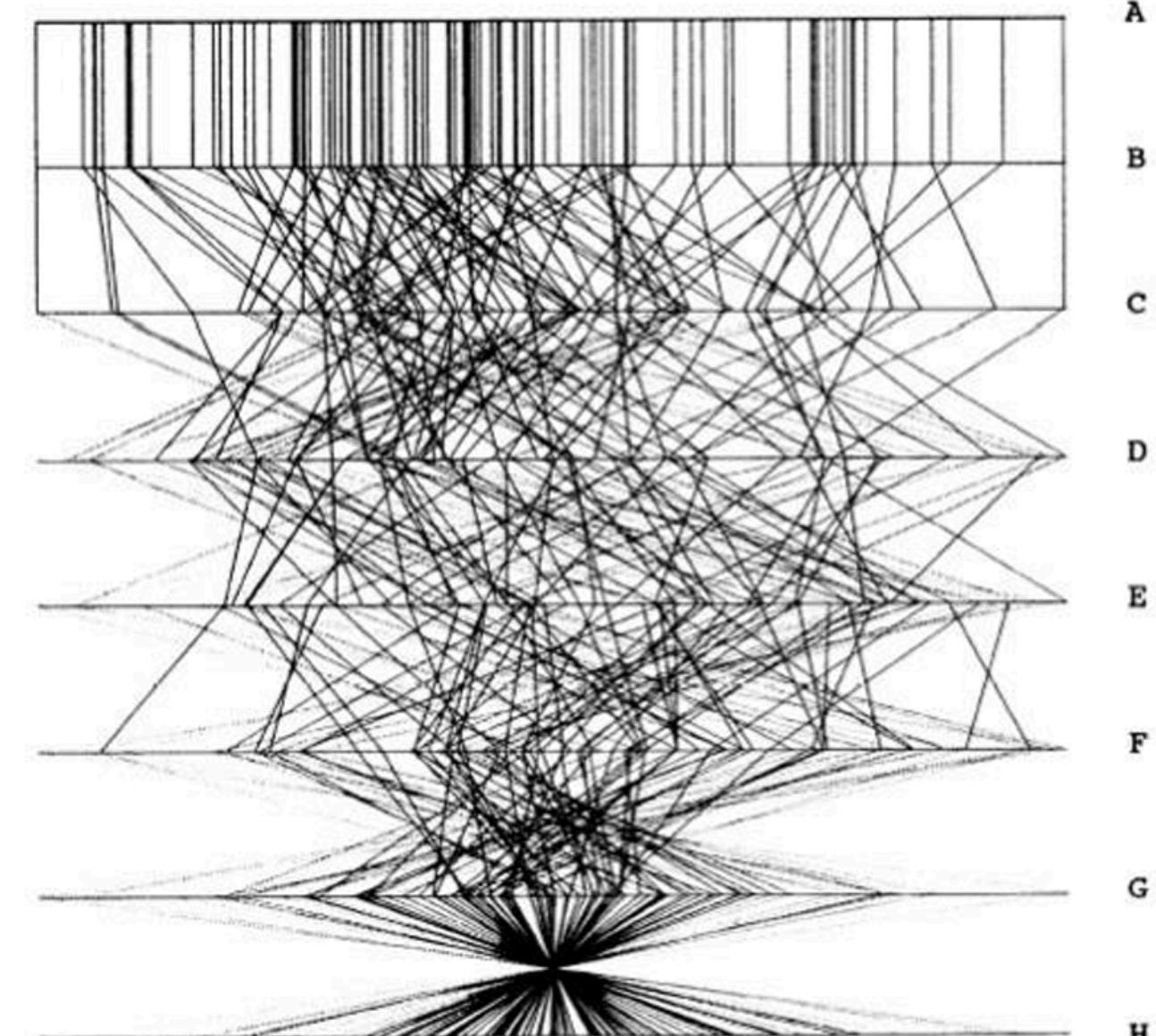
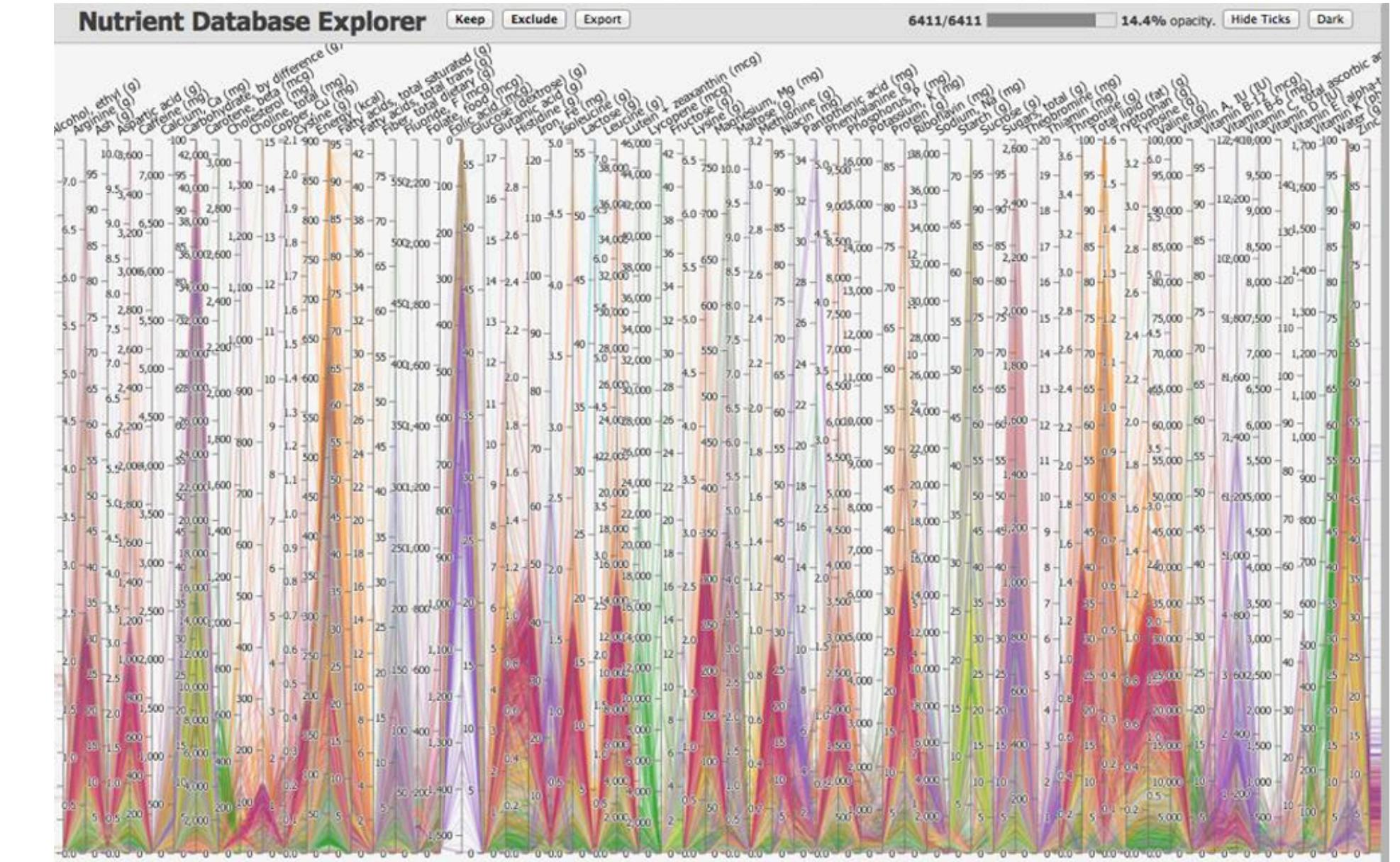


Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of $\rho = 1, .8, .2, 0, -.2, -.8$, and -1 .

Axis-Based Techniques: Parallel Coordinates

Limitations:

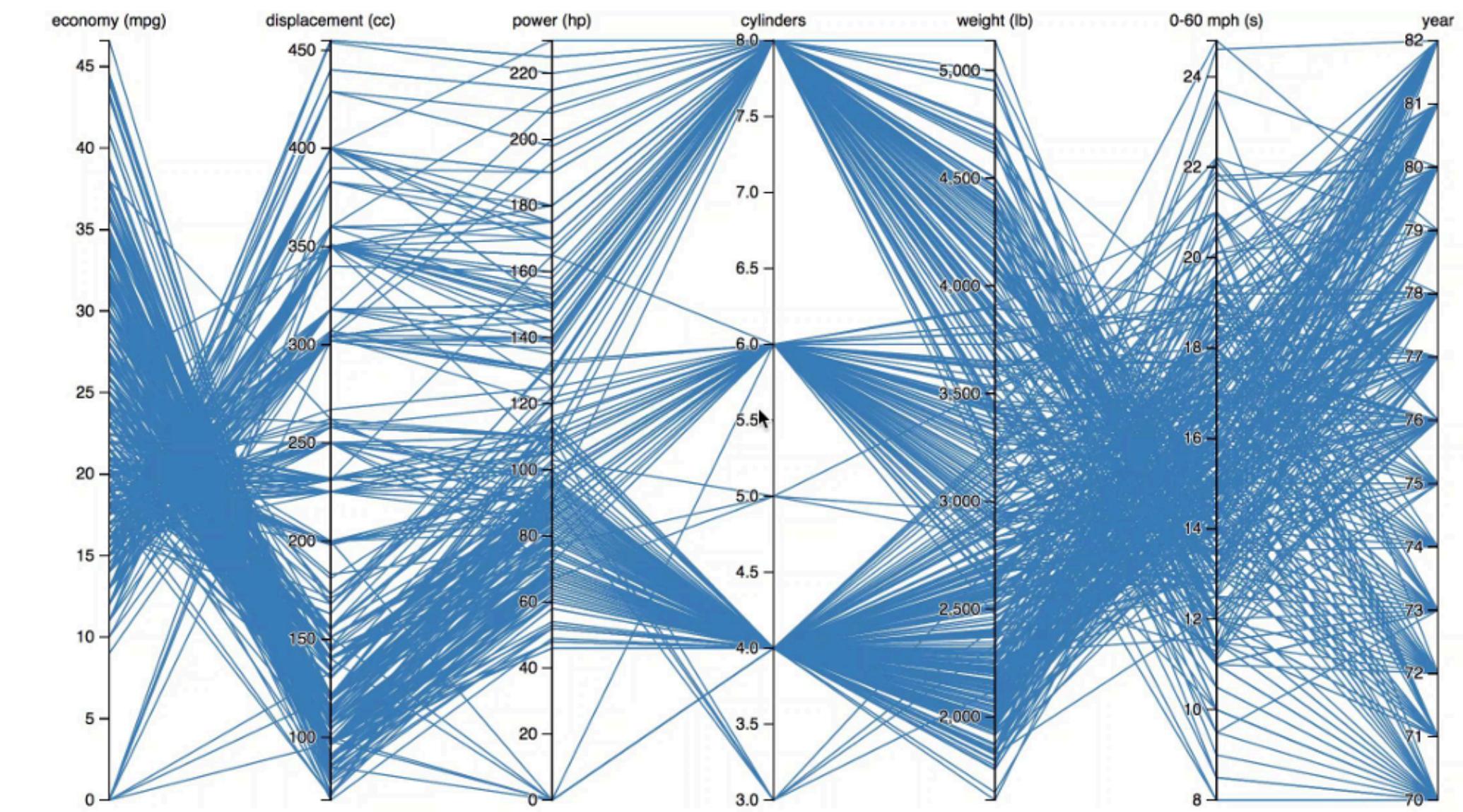
- Not very scalable to many attributes



Axis-Based Techniques: Parallel Coordinates

Limitations:

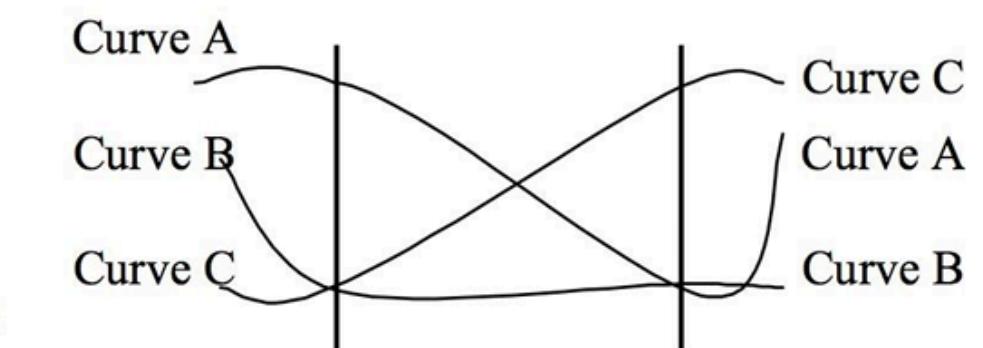
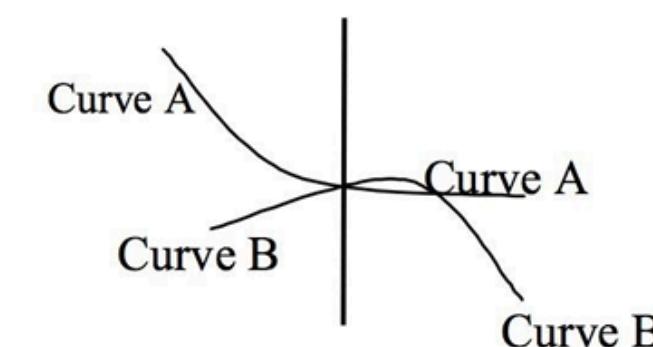
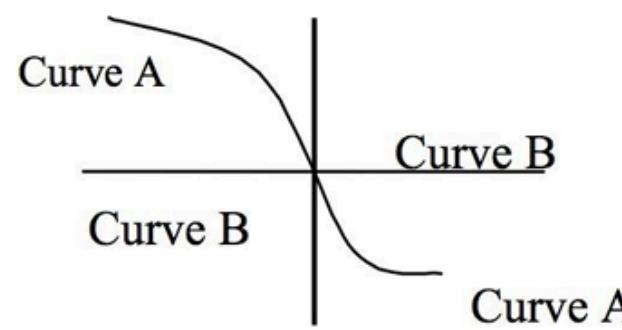
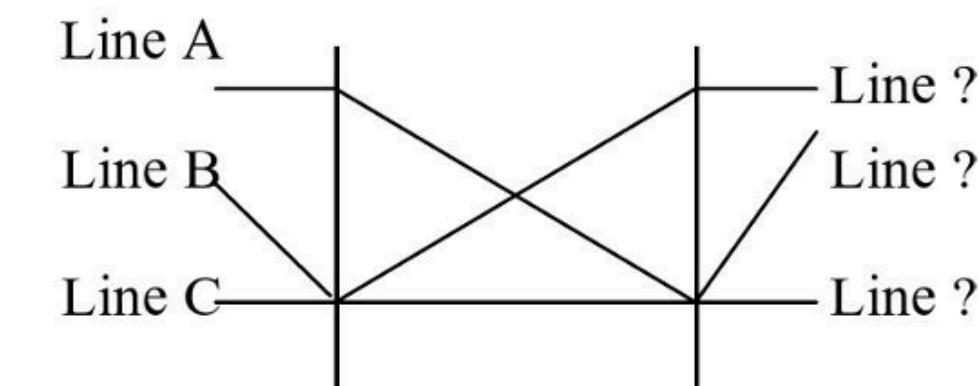
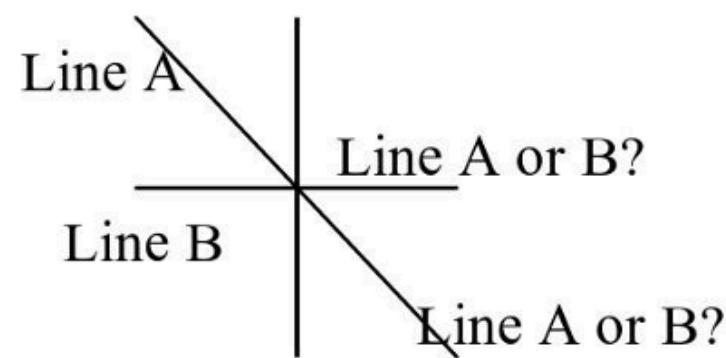
- Correlations only comparable between adjacent axes
 - let user change axes



Axis-Based Techniques: Parallel Coordinates

Limitations:

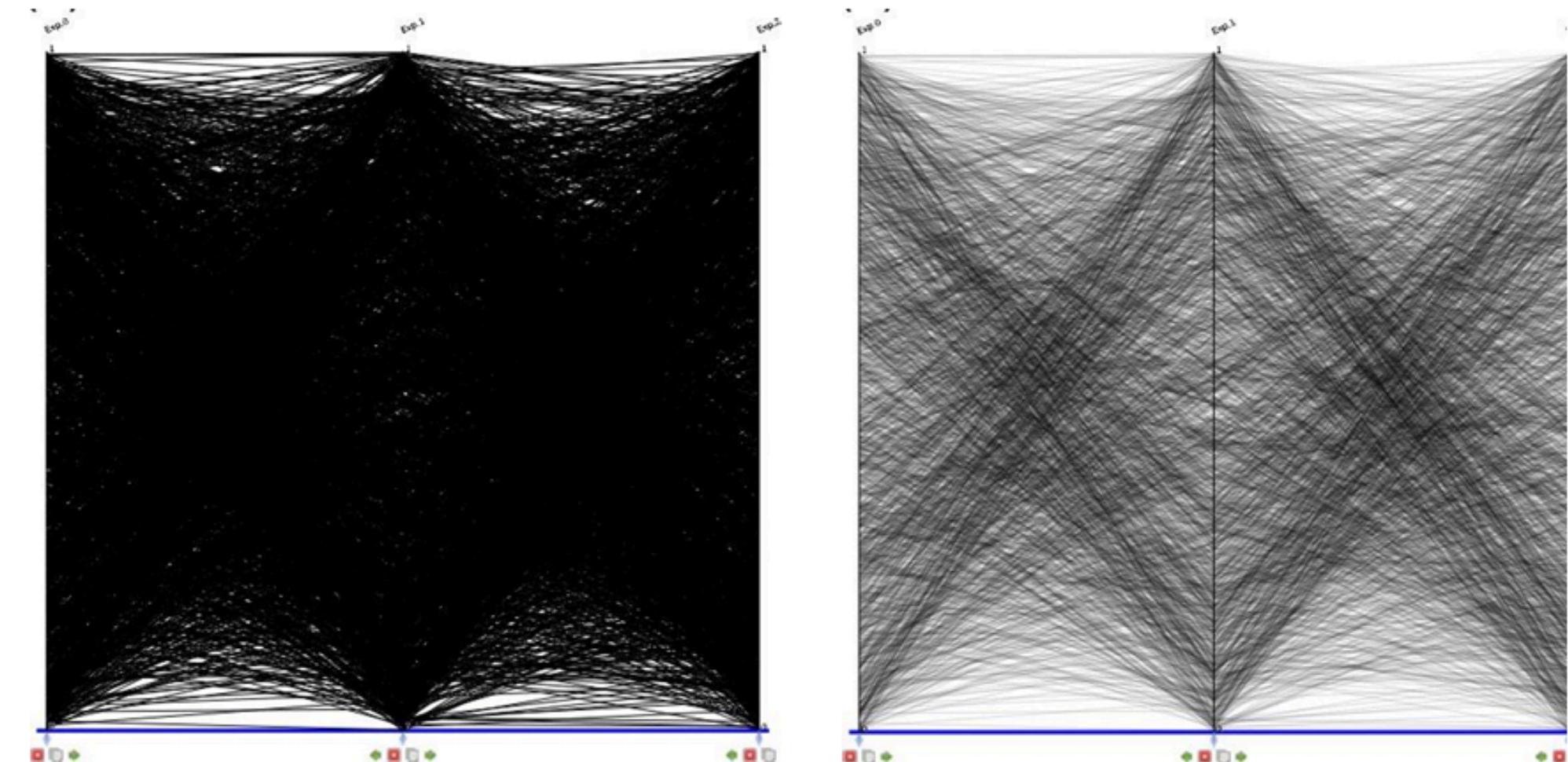
- Ambiguity
 - use curves



Axis-Based Techniques: Parallel Coordinates

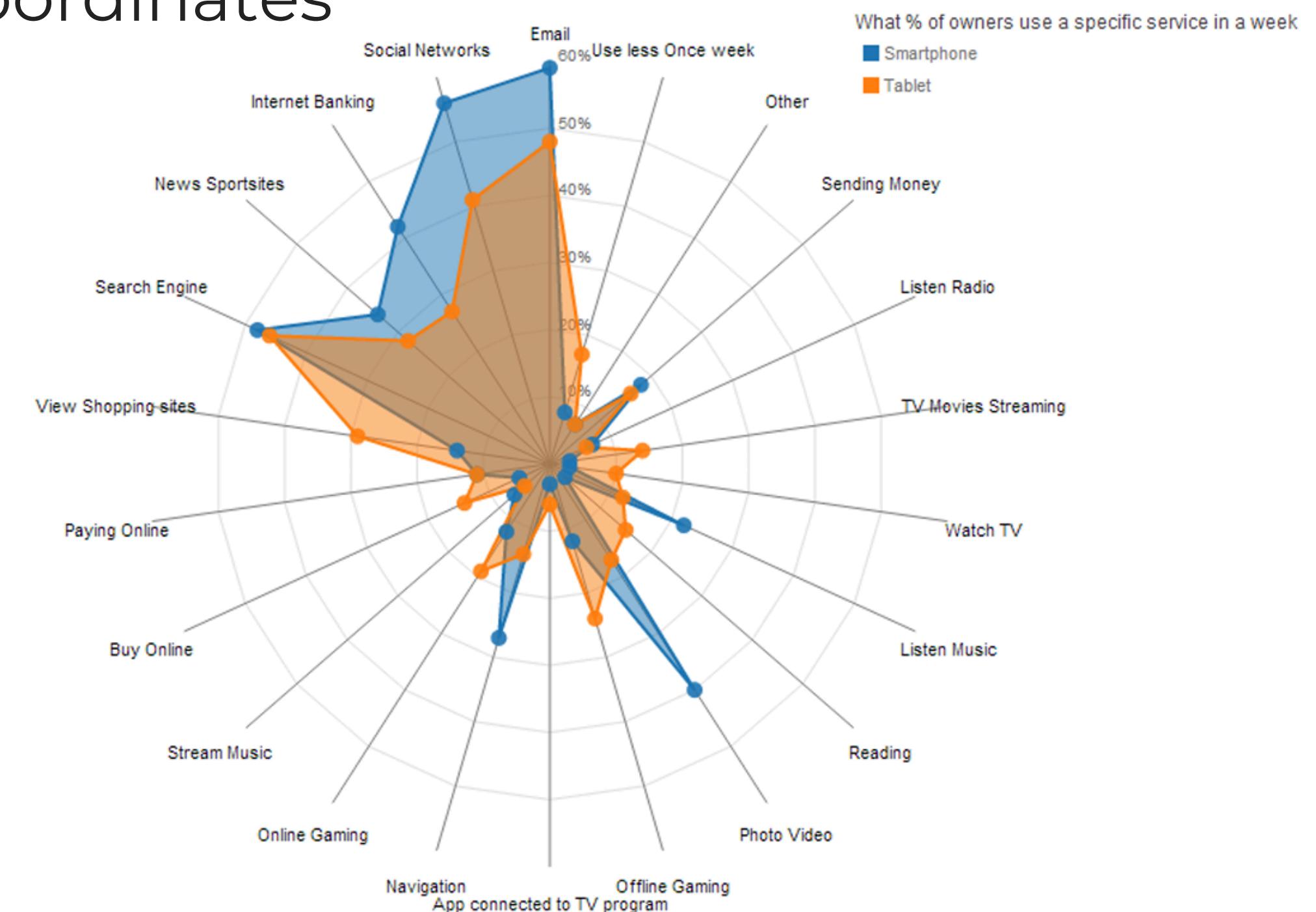
Limitations:

- Not very scalable to many items
 - use transparency, clustering, sampling



Axis-Based Techniques: Star Plot

Radial version of parallel coordinates



FIN

Upcoming Dates

Tomorrow (Apr 25): Homework 4 Due

Today: Quiz 6 Released (all quizzes due May 2)

Apr 29: Final Group Activity

May 2:

- Homework 5 Due
- Project Screencast Submission Due

May 12: Final Project Submission Due