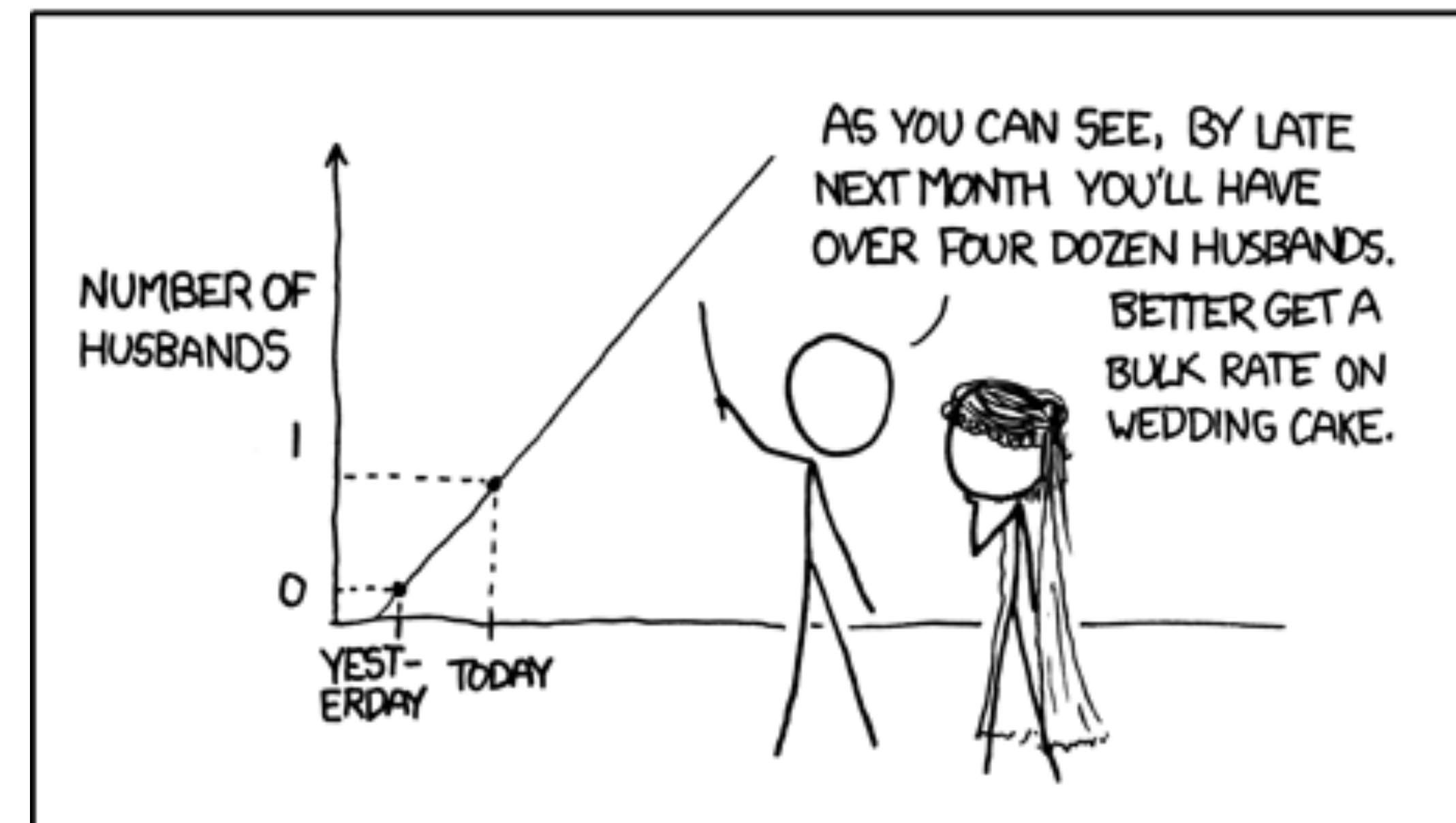


CS-5630 / CS-6630 Visualization for DataScience Tables

Alexander Lex
alex@sci.utah.edu



MY HOBBY: EXTRAPOLATING



Organizational

Review exam in my office hours starting Oct 29

HW Lab: Wed, 6pm, L110

Make sure to form your project teams!

If you can't find a team, e-mail me

Develop project idea

Set up your github repo

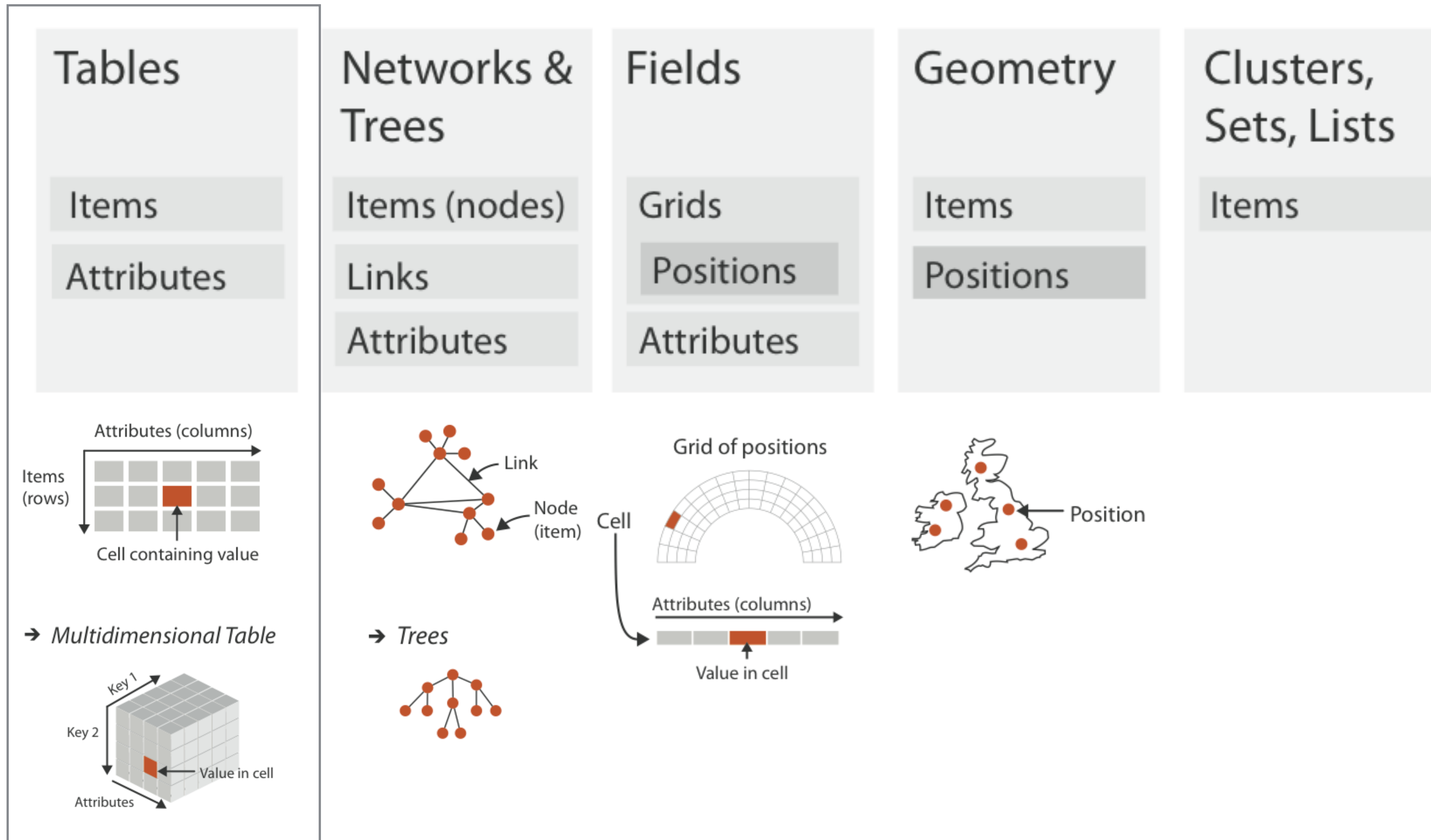
Proposal and HW6 due Oct 25

Peer review session (mandatory attendance) on Oct 29

Need to submit this info by Friday!

<https://forms.gle/aj6CwRRBNVnNzVqy5>

Dataset Types



Exercise: Sketch 2 Ways to Vis. Each Table

	Age	Best 100 m	Furthest Jump	Sex
Amy	16	13.2	5.2	F
Basil	18	12.4	4.2	F
Clara	14	14.1	2.5	F
Desmond	22	10.01	6.3	M
Charles	19	11.3	5.3	M

	BPM T1	BPM T2	BPM T3
Amy	90	130	150
Basil	70	110	109
Clara	60	140	141
Desmond	84	100	108
Charles	81	110	130

Scale of Tables

Need different approaches for “normal” and “high-dimensional” tables.

How many dimensions?

~50 – tractable with “just” vis

~1000 – need analytical methods

How many records?

~ 1000 – “just” vis is fine

>> 10,000 – need analytical methods

Homogeneity

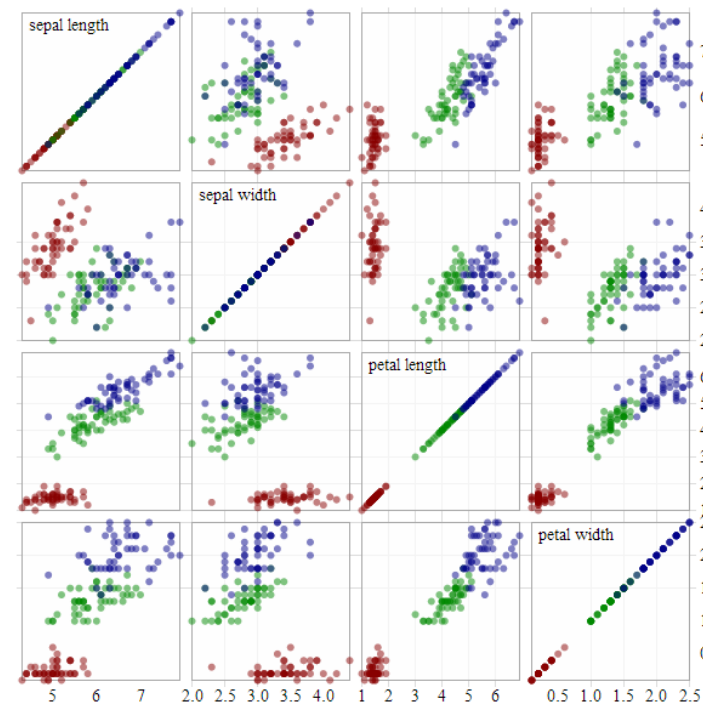
Same data type?

Same scales?

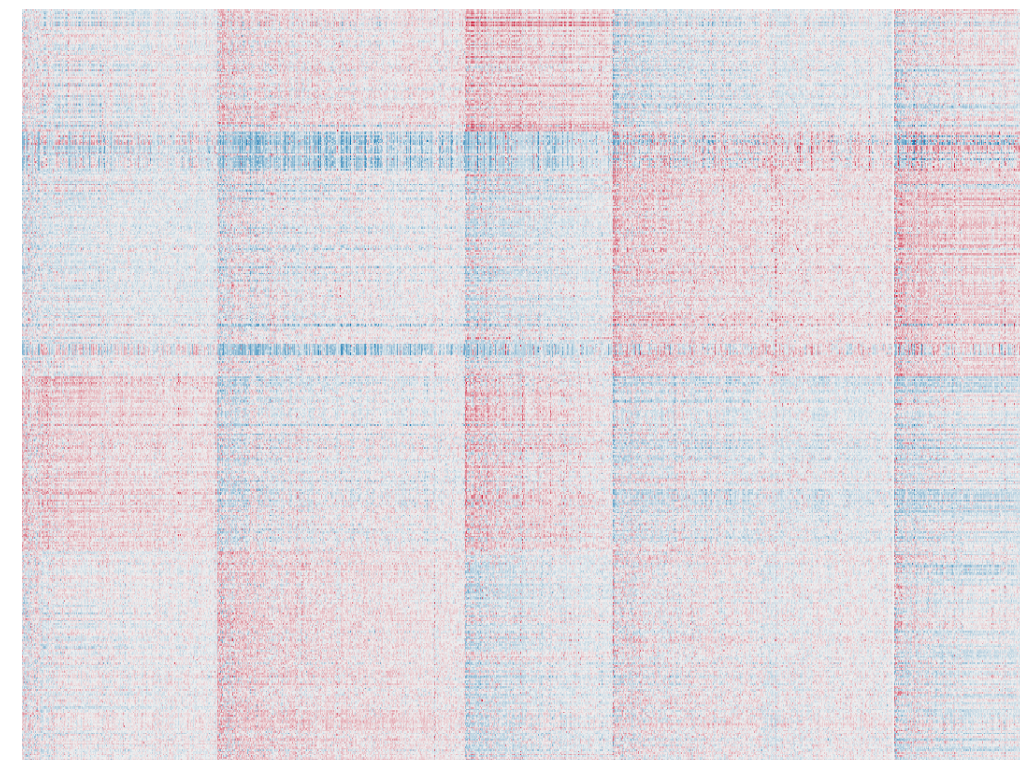
	Age	Gender	Height
Bob	25	M	181
Alice	22	F	185
Chris	19	M	175

	BPM 1	BPM 2	BPM 3
Bob	65	120	145
Alice	80	135	185
Chris	45	115	135

Analytic Component



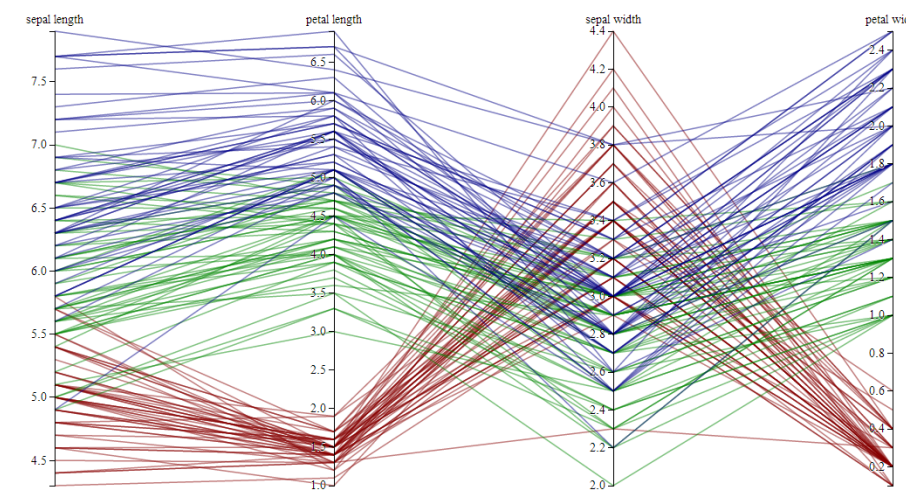
Scatterplot Matrices
[Bostock]



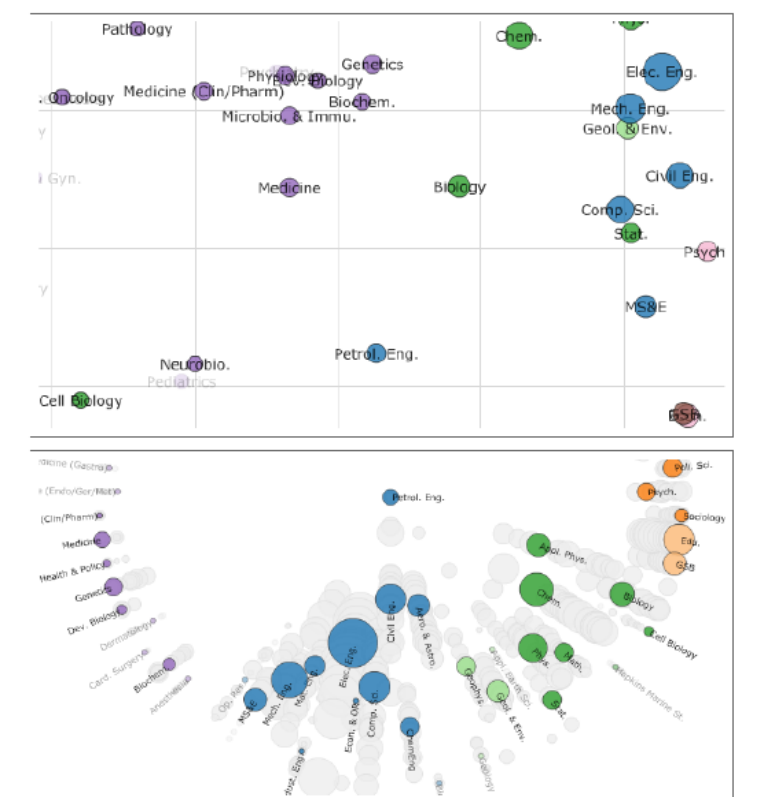
Pixel-based visualizations /
heat maps



Multidimensional Scaling
[Doerk 2011]



Parallel Coordinates
[Bostock]



[Chuang 2012]

no / little analytics

strong analytics
component

Techniques and Tasks

Magnitude

Distribution

Deviation

Correlation

Ranking

Part to whole

Change over Time

<https://github.com/ft-interactive/chart-doctor/tree/master/visual-vocabulary>
<https://gramener.github.io/visual-vocabulary-vega/#/Magnitude/>

Deviation
Deviation indicates how far a data point is from the mean. It is a measure of the spread of the data. Deviation can be calculated for each data point, and the average deviation is the standard deviation.

Correlation
Correlation indicates the relationship between two variables. It is a measure of how well the variables are related. Correlation can be positive or negative, and the strength of the relationship is indicated by the correlation coefficient.

Ranking
Ranking is a way of ordering data points based on their value. It is a simple and effective way to compare data points. Ranking can be used to identify the top or bottom performers in a group.

Distribution
Distribution shows the spread of data points. It is a way of visualizing the frequency of different values. Distribution can be used to identify patterns in the data, such as a normal distribution.

Change over Time
Change over time shows how a variable changes over a period of time. It is a way of visualizing trends and patterns in the data. Change over time can be used to identify growth or decline in a variable.

Magnitude
Magnitude shows the size of a variable. It is a way of visualizing the relative importance of different variables. Magnitude can be used to identify the most significant variables in a dataset.

Part-to-whole
Part-to-whole shows the relationship between a part and the whole. It is a way of visualizing the composition of a whole. Part-to-whole can be used to identify the most important parts of a whole.

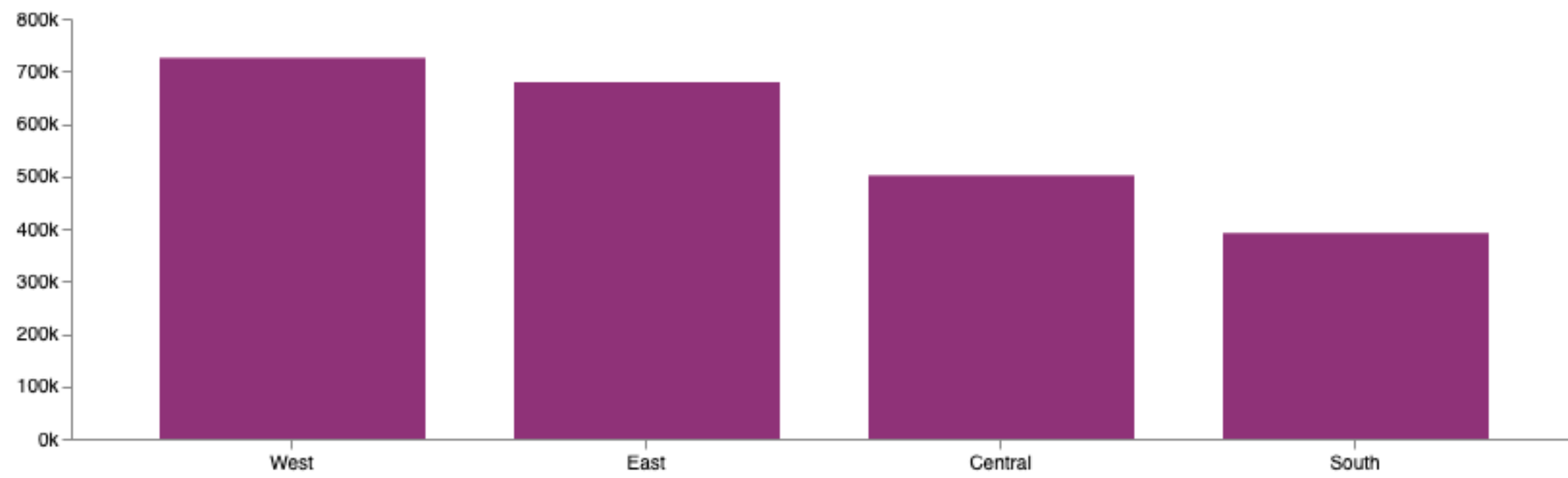
Spatial
Spatial shows the relationship between data points in space. It is a way of visualizing the geographic distribution of data. Spatial can be used to identify patterns in the data, such as a cluster of data points.

Flow
Flow shows the movement of data points over time. It is a way of visualizing the flow of information or resources. Flow can be used to identify the most important flows in a system.

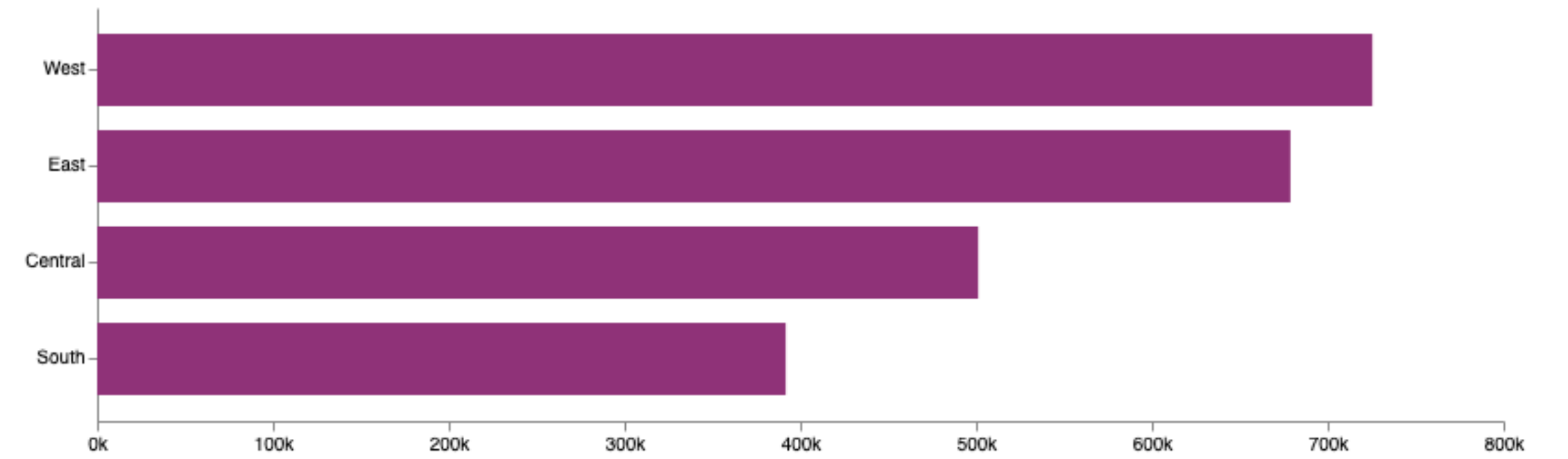
Visual vocabulary
Designing with data
There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart.

Magnitude

Bar Chart Variants



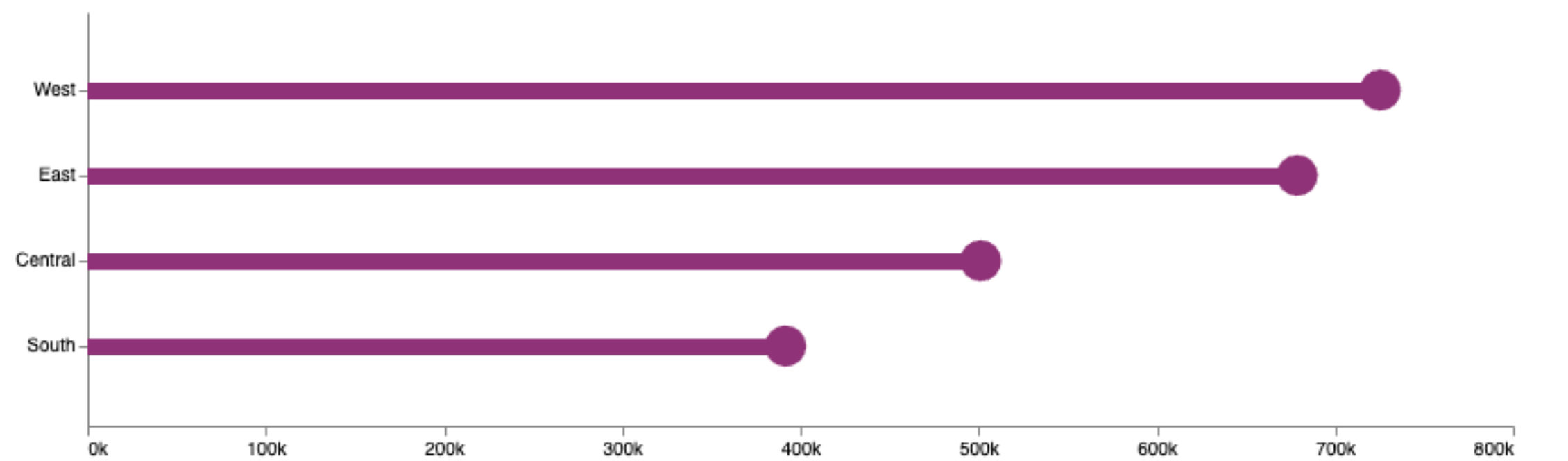
Vertical Bar Chart / Column Chart



Horizontal Bar Chart



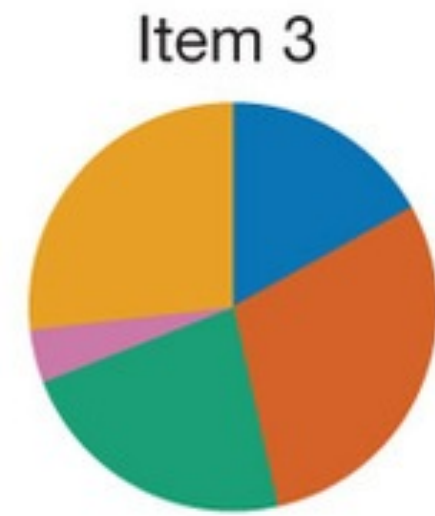
Grouped Bar Chart



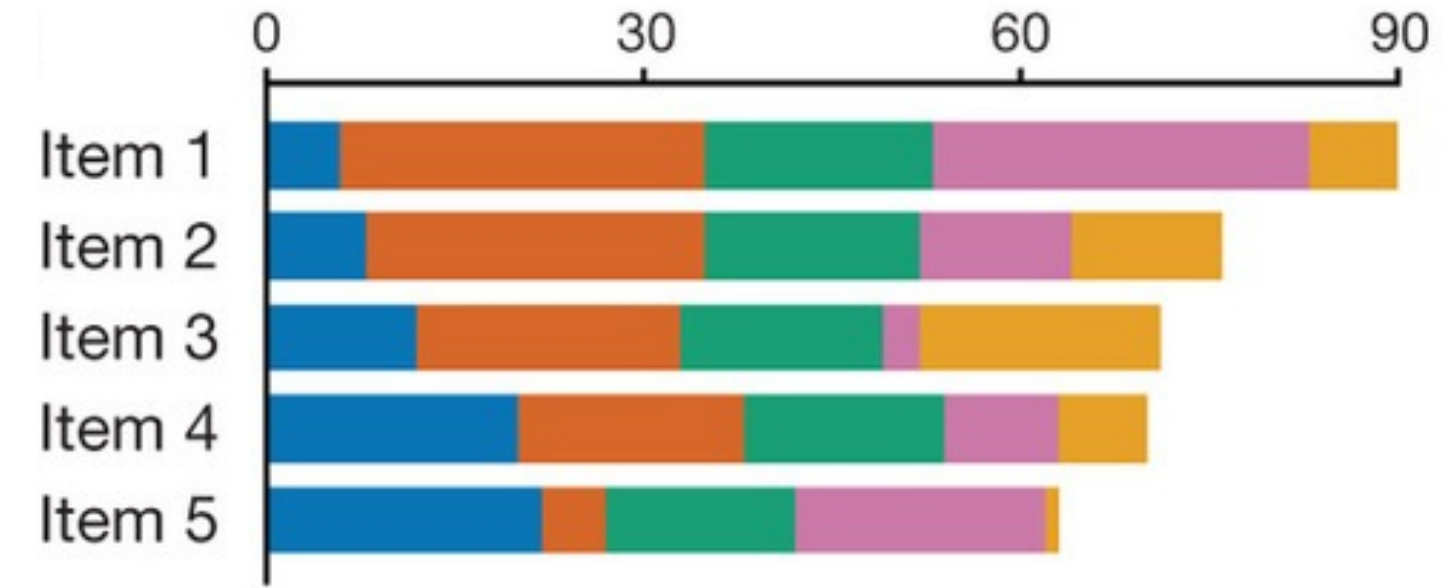
Lollipop Chart

Comparison of bar chart types

- Category 1 ●
- Category 2 ●
- Category 3 ●
- Category 4 ●
- Category 5 ●

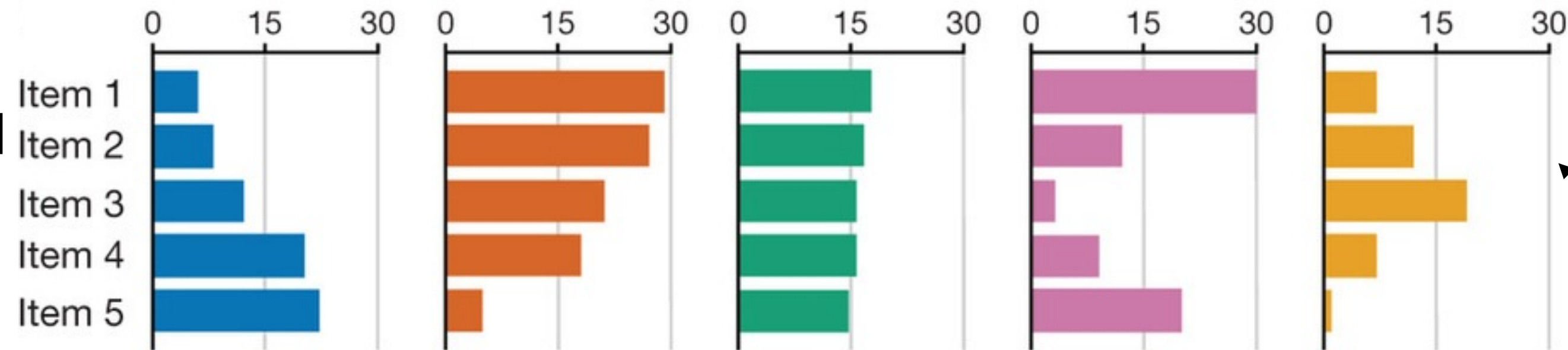


Pie Chart

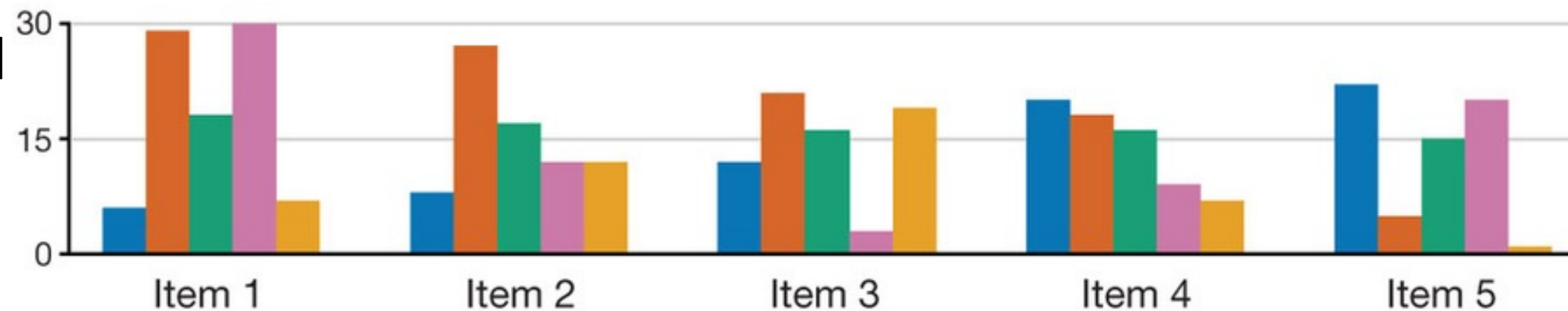


Stacked bar chart

Layered Bar Chart



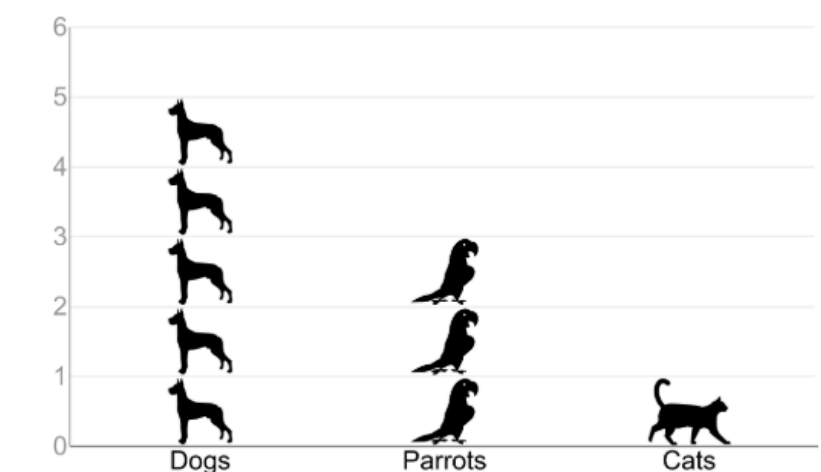
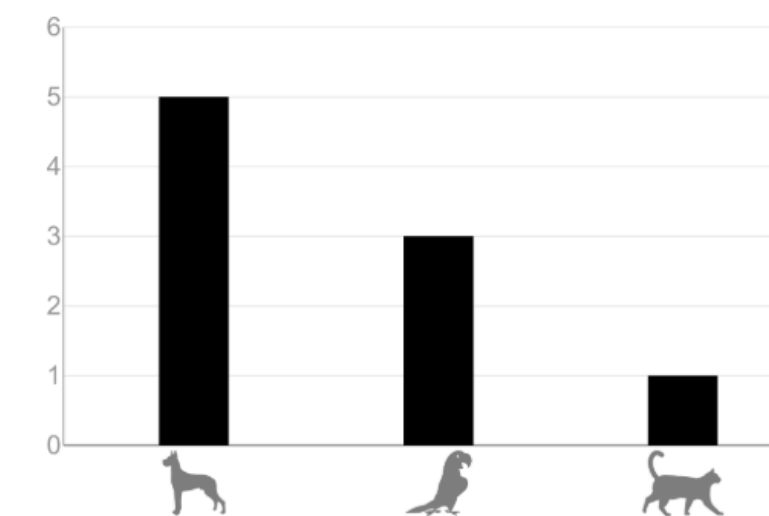
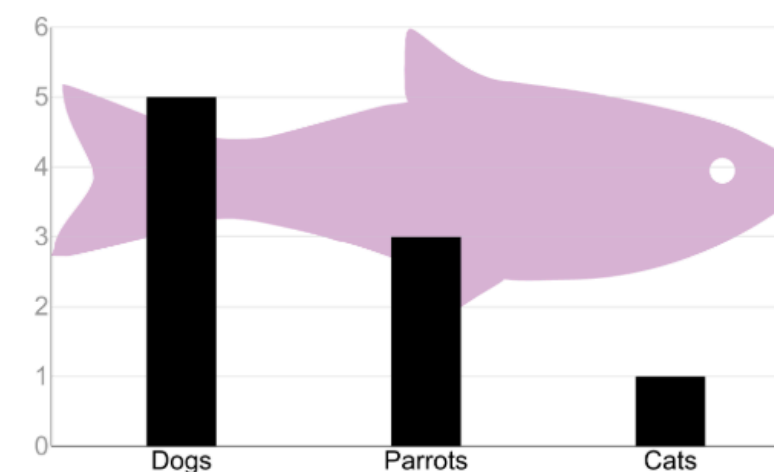
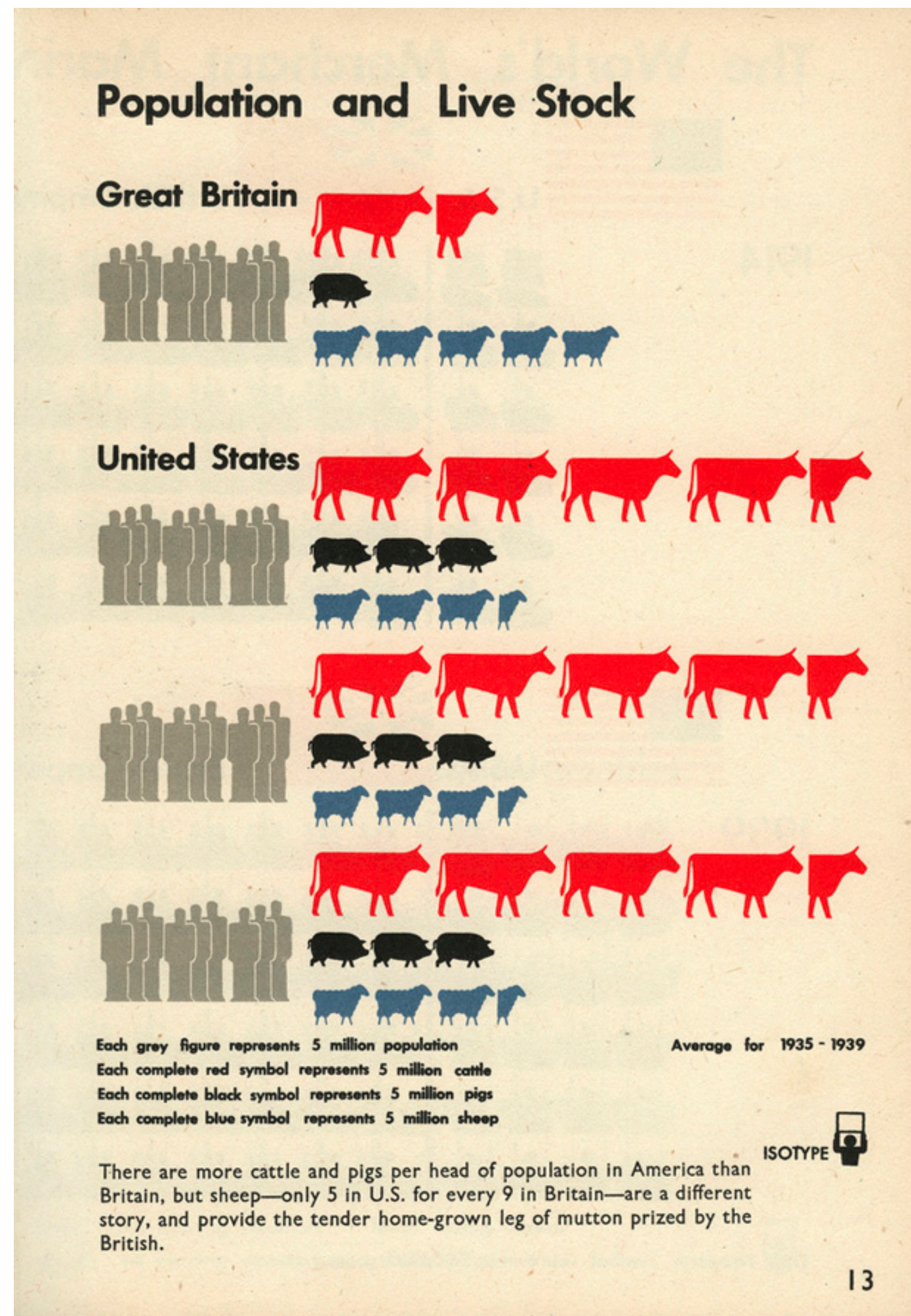
Grouped Bar Chart



Small Multiples

Rank	Player	Current League	Current Club	Position	Foot	# Age	# Height	Goals	Games
...
1	Dmitri Poloz	RUS - Premier Liga	Zenit St. Petersburg	W	right	38.00	160.00	1	1
2	Emiliano Rigoni	RUS - Premier Liga	Zenit St. Petersburg	W	both	38.00	160.00	1	1
3	Sebastián Driussi	RUS - Premier Liga	Zenit St. Petersburg	CF	both	38.00	160.00	1	1
4	Aleksandr Kokorin	RUS - Premier Liga	Zenit St. Petersburg	CF	right	38.00	160.00	1	1
5	Anton Zabolotnyi	RUS - Premier Liga	Zenit St. Petersburg	CF	right	38.00	160.00	1	1
6	Quincy Promes	RUS - Premier Liga	Spartak Moscow	W	both	38.00	160.00	1	1
7	Pedro Rocha	RUS - Premier Liga	Spartak Moscow	W	right	38.00	160.00	1	1
8	Lorenzo Melgarejo	RUS - Premier Liga	Spartak Moscow	W	left	38.00	160.00	1	1
9	Zelimkhan Bakaev	RUS - Premier Liga	Spartak 2 Moscow	W	left	38.00	160.00	1	1
10	Luiz Adriano	RUS - Premier Liga	Spartak Moscow	CF	right	38.00	160.00	1	1
11	Zé Luís	RUS - Premier Liga	Spartak Moscow	CF	left	38.00	160.00	1	1
12	Ahmed Musa	RUS - Premier Liga	CSKA Moscow	CF	both	38.00	160.00	1	1
13	Fedor Chalov	RUS - Premier Liga	CSKA Moscow	CF	right	38.00	160.00	1	1
14	Timur Zhamaletdinov	RUS - Premier Liga	CSKA Moscow	CF	right	38.00	160.00	1	1
15	Wanderson	RUS - Premier Liga	FK Krasnodar	W	right	38.00	160.00	1	1
16	Joãozinho	RUS - Premier Liga	FK Krasnodar	W	left	38.00	160.00	1	1
17	Andrei Ivan	RUS - Premier Liga	FK Krasnodar	W	right	38.00	160.00	1	1
18	Ricardo Laborde	RUS - Premier Liga	FK Krasnodar	W	right	38.00	160.00	1	1
19	Magomed-Shapi Suleyn	RUS - Premier Liga	FK Krasnodar	W	left	38.00	160.00	1	1
20	Fedor Smolov	RUS - Premier Liga	FK Krasnodar	CF	right	38.00	160.00	1	1
21	Ivan Ignatjev	RUS - Premier Liga	FK Krasnodar	CF	right	38.00	160.00	1	1
22	Alan Kasaev	RUS - Premier Liga	Lokomotiv Moscow	W	right	38.00	160.00	1	1
23	Jefferson Farfán	RUS - Premier Liga	Lokomotiv Moscow	W	right	38.00	160.00	1	1
24	Arshak Koryan	RUS - Premier Liga	Lokomotiv Moscow	W	right	38.00	160.00	1	1
25	Éder	RUS - Premier Liga	Lokomotiv Moscow	CF	both	38.00	160.00	1	1
26	Ari	RUS - Premier Liga	Lokomotiv Moscow	CF	both	38.00	160.00	1	1
27	Gökdeniz Karadeniz	RUS - Premier Liga	Rubin Kazan	W	right	38.00	160.00	1	1
28	Rifat Zhemaletdinov	RUS - Premier Liga	Rubin Kazan	W	right	38.00	160.00	1	1
29	Sardar Azmoun	RUS - Premier Liga	Rubin Kazan	CF	both	38.00	160.00	1	1
30	Léo Jabá	RUS - Premier Liga	Akhmat Grozny	W	right	38.00	160.00	1	1
31	Bernard Berisha	RUS - Premier Liga	Akhmat Grozny	W	right	38.00	160.00	1	1
32	Magomed Mitrishev	RUS - Premier Liga	Akhmat Grozny	W	right	38.00	160.00	1	1
33	Odise Roshi	RUS - Premier Liga	Akhmat Grozny	W	right	38.00	160.00	1	1

IsoType Visualization



Part of Whole

Stacked Bar Chart

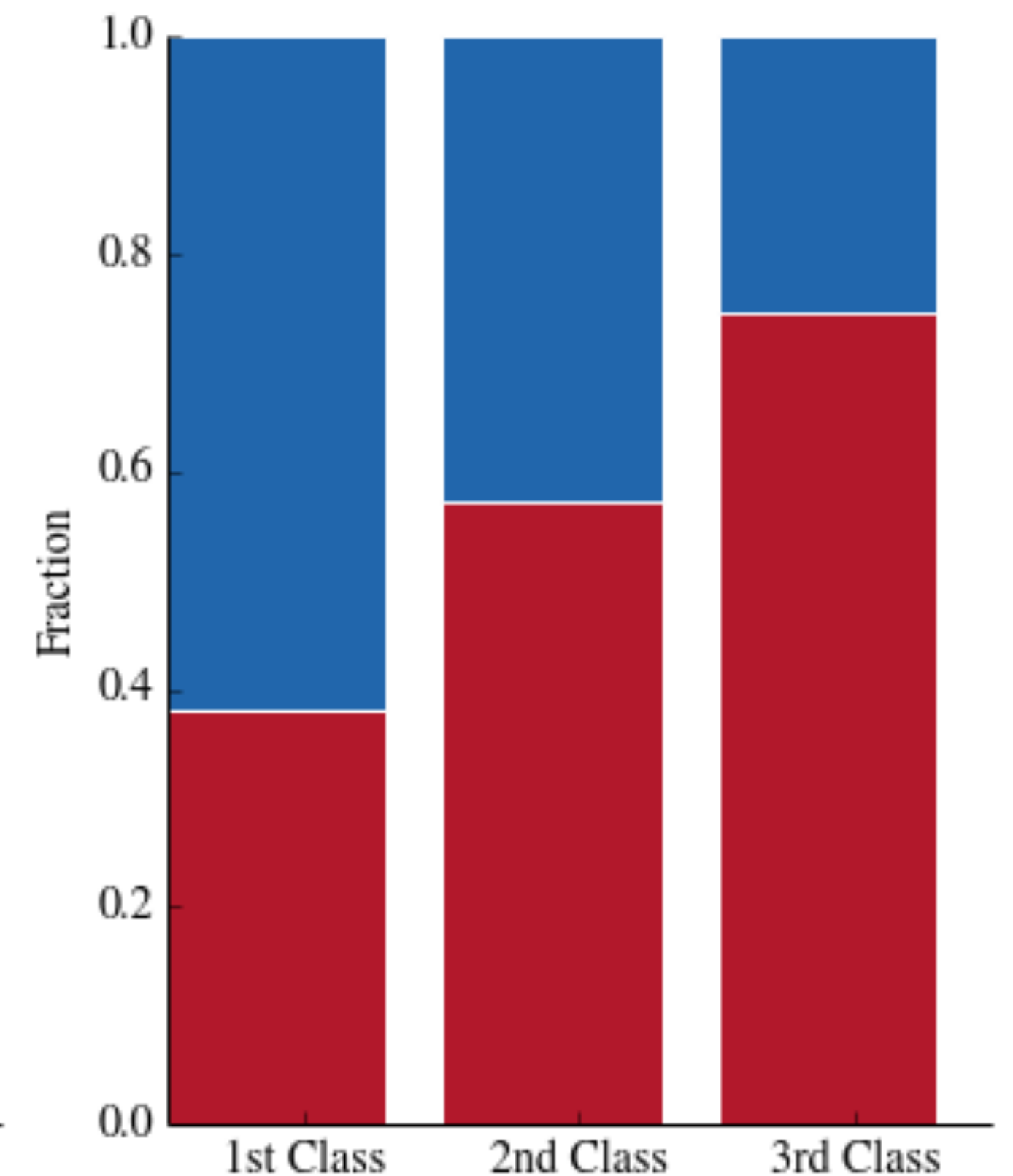
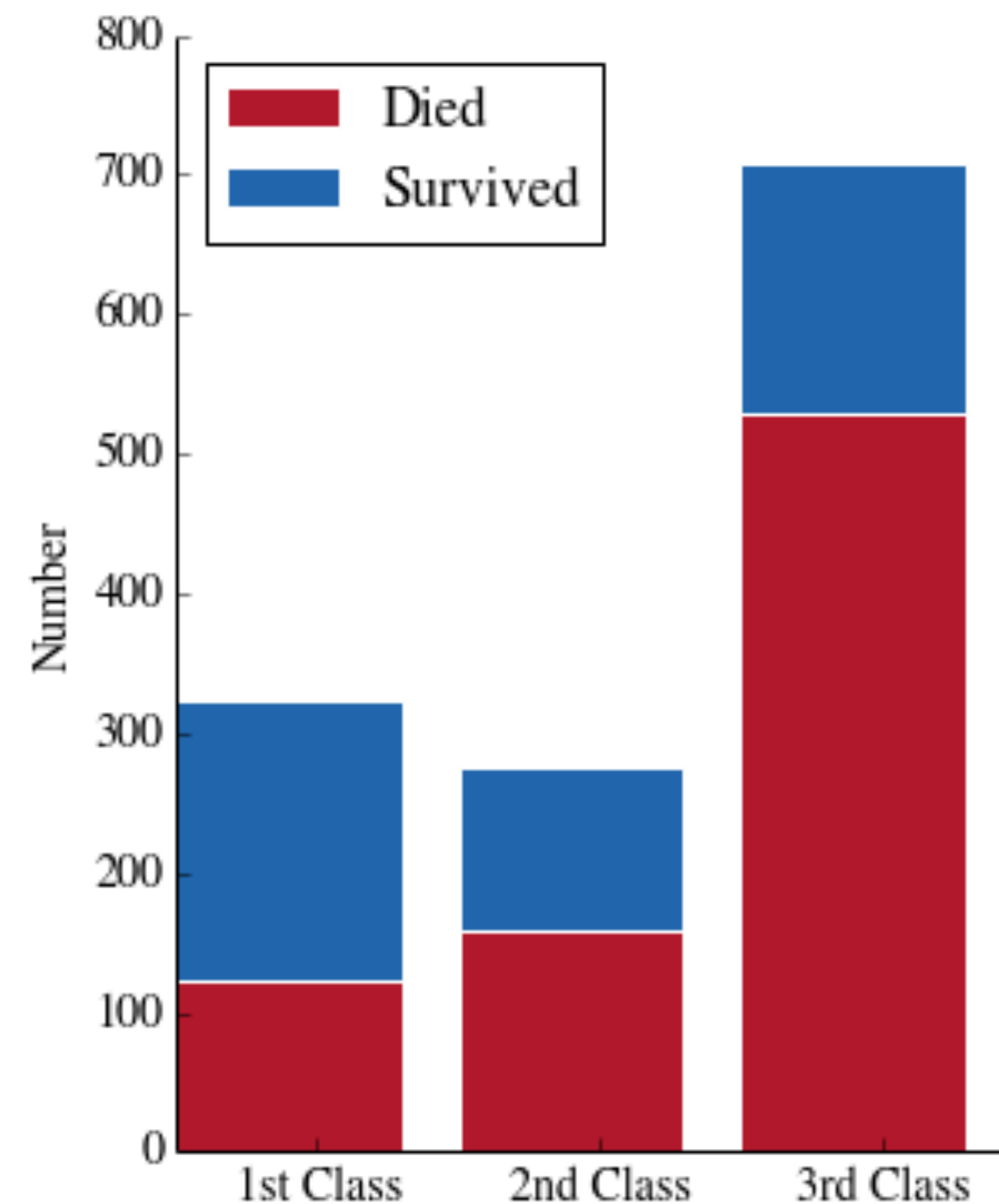
Keys: Class, Survival

Class is spatial

Survival is color

Left: absolute values

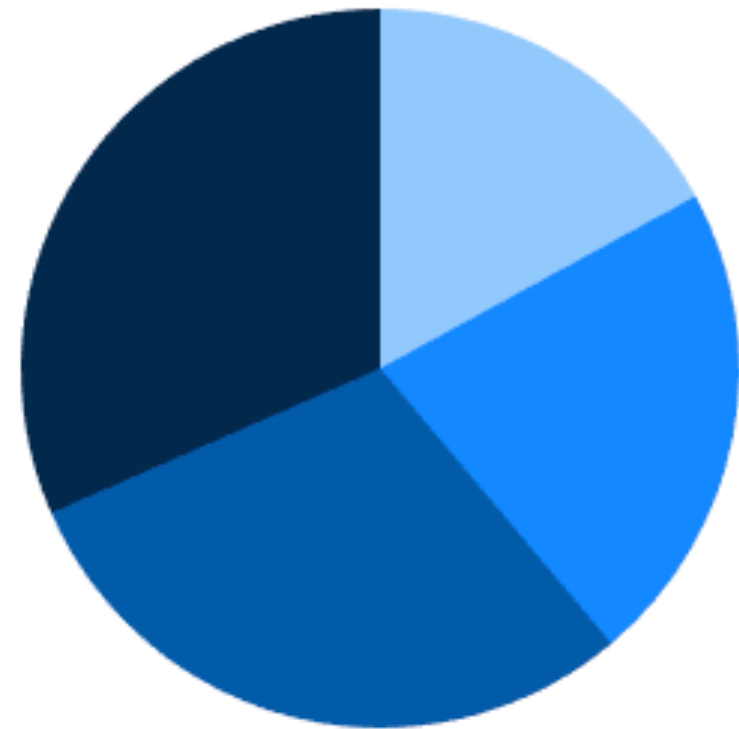
Right: proportional values



Pie and Donut Charts

Pie

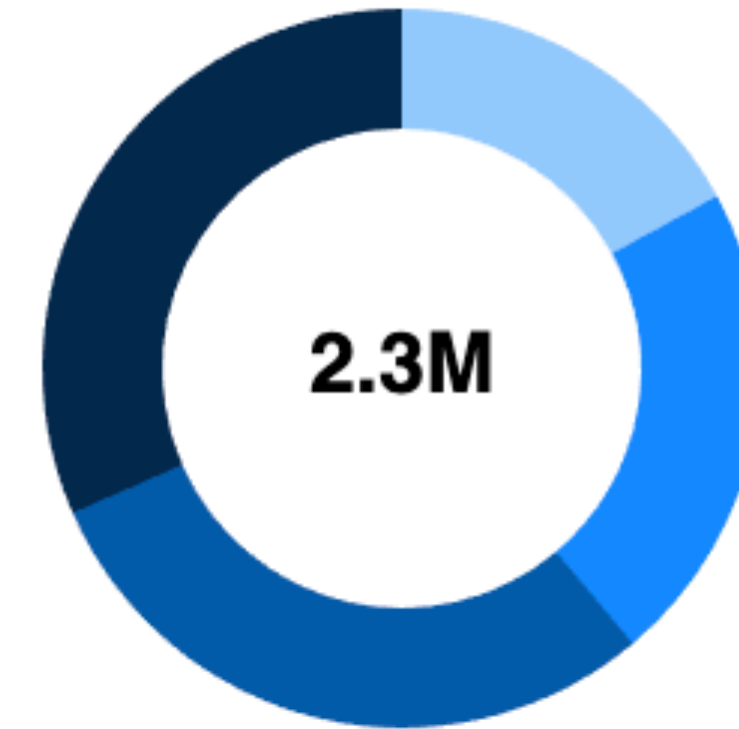
A common way of showing part-to-whole data - but be aware that it's difficult to accurately compare the size of the segments.



Edit

Donut

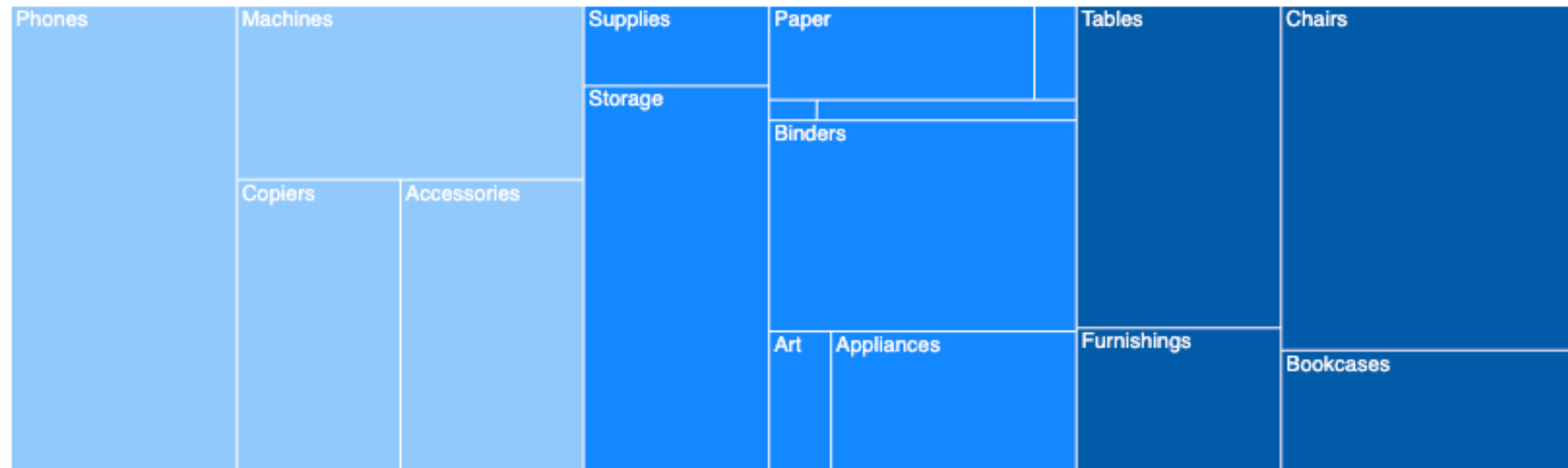
Similar to a pie chart - but the centre can be a good way of making space to include more information about the data (eg. total)



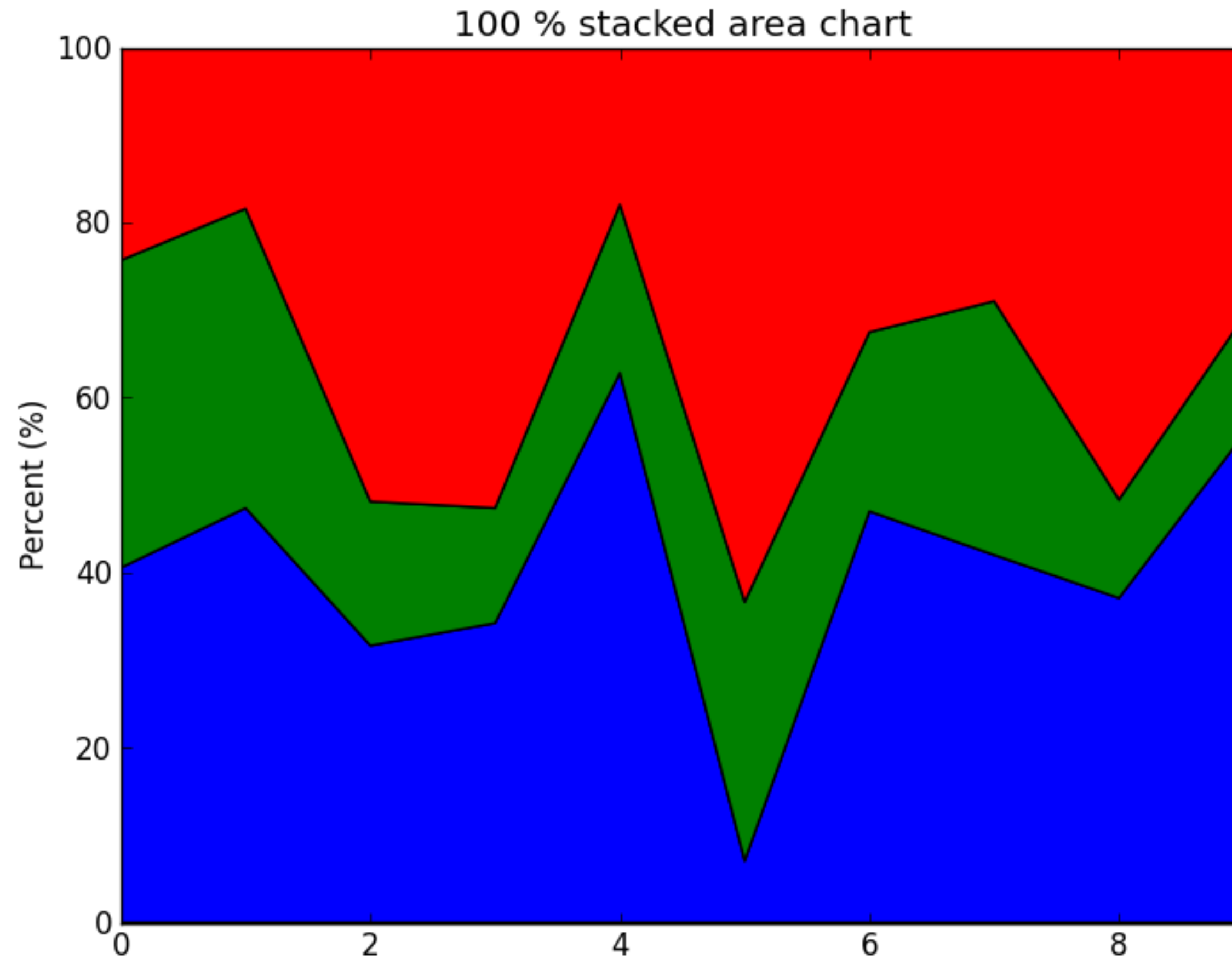
TreeMap

Treemap

Use for hierarchical part-to-whole relationships; can be difficult to read when there are many small segments



Part of Whole for Time Series



Distribution

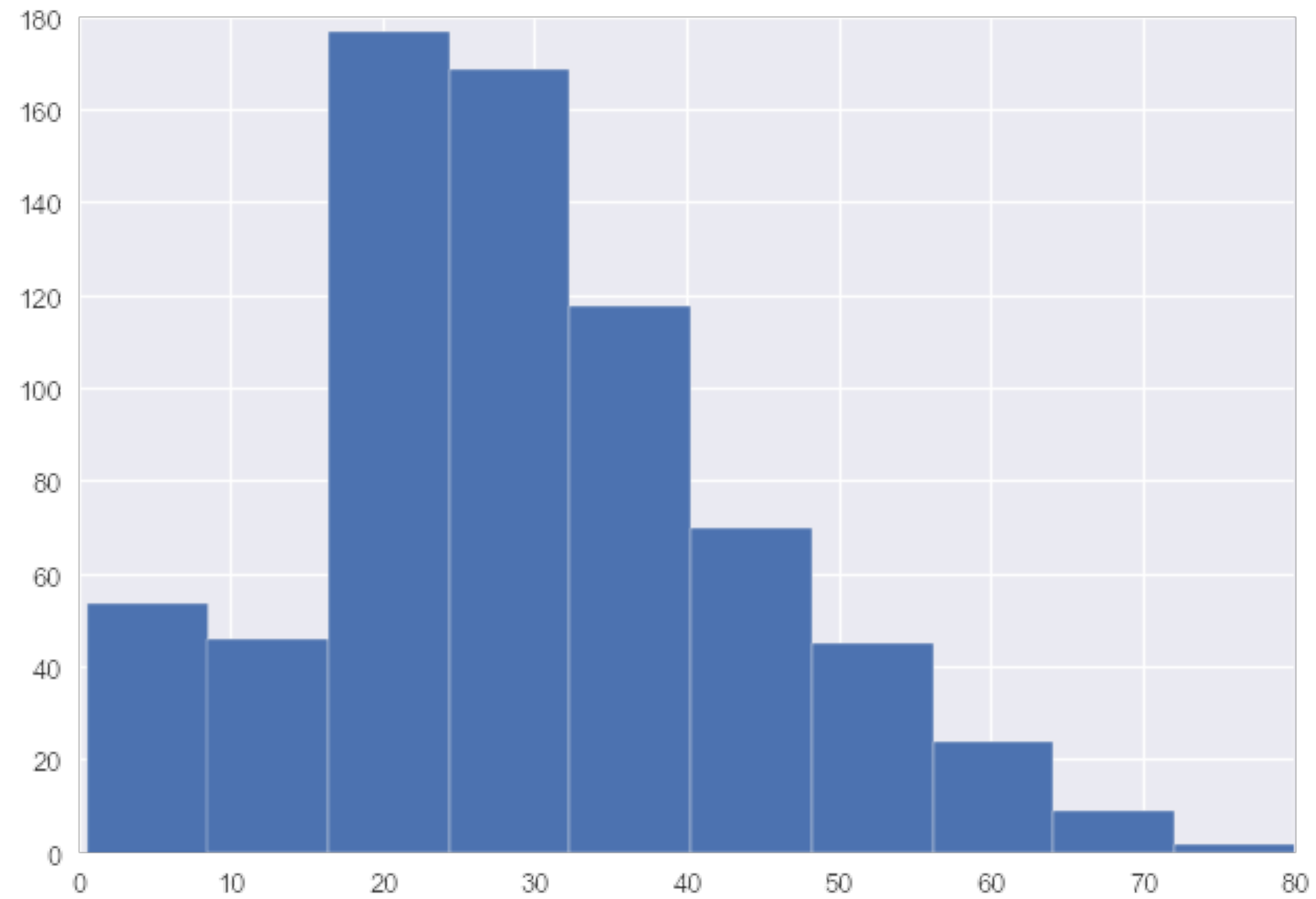
Aggregating Large Data Vectors

Instead of showing all data points, show a data's distribution

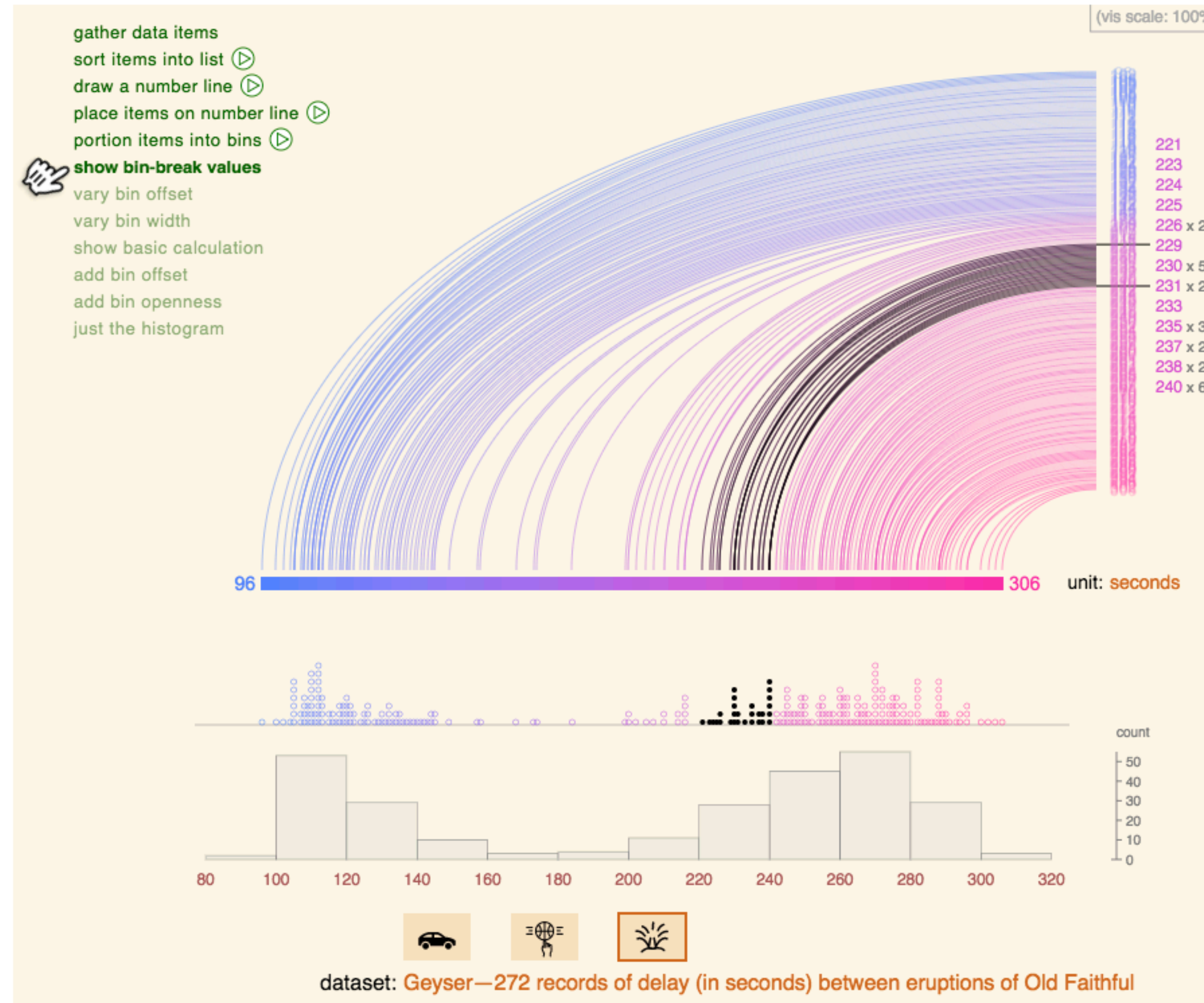
Pro: compact representation

Con: Works only if data is “well behaved” for the type of distribution visualization.

What's a histogram?



Histograms Explained



<http://tinlizzie.org/histograms/>

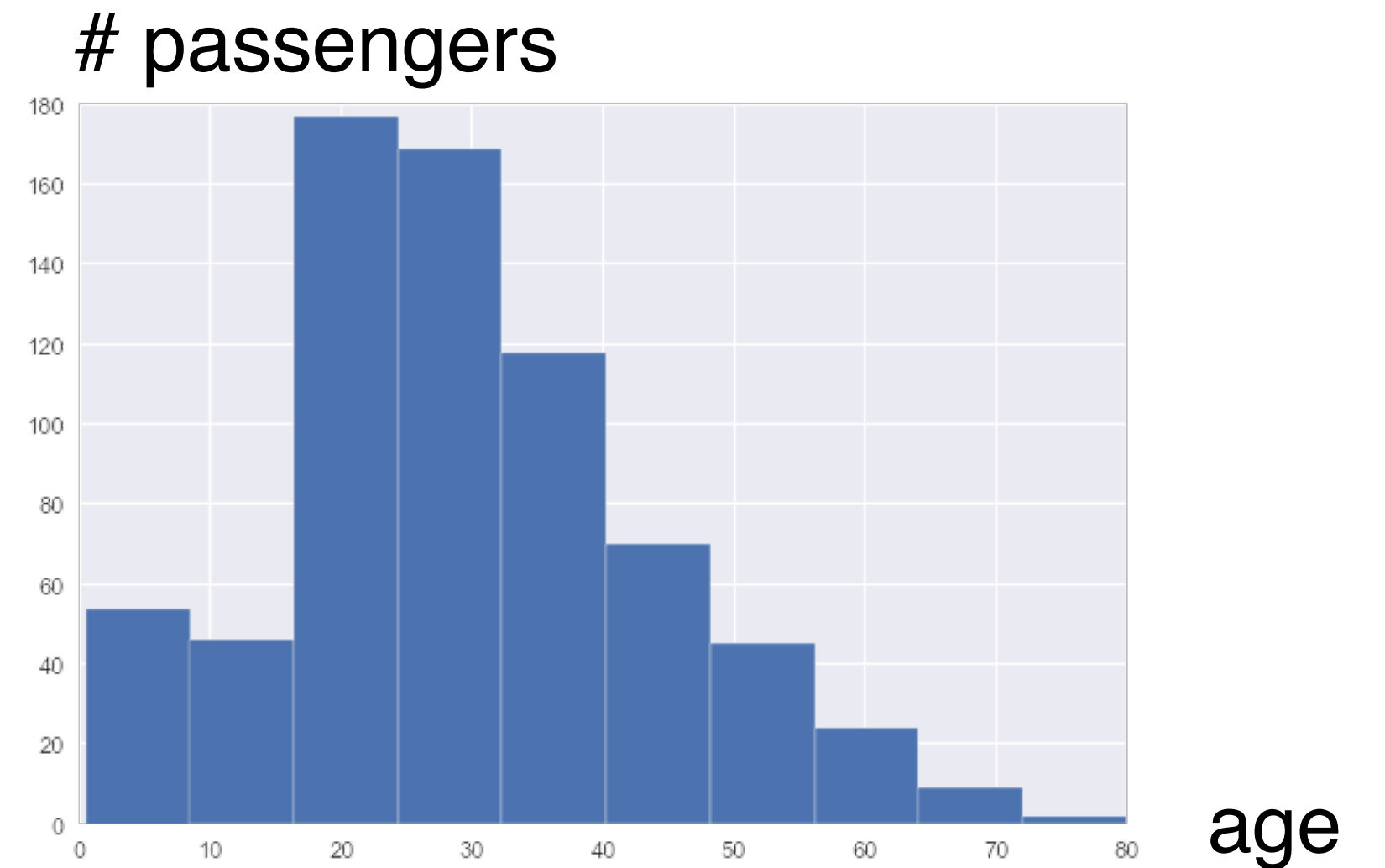
Histogram

Good #bins hard to predict
make interactive!

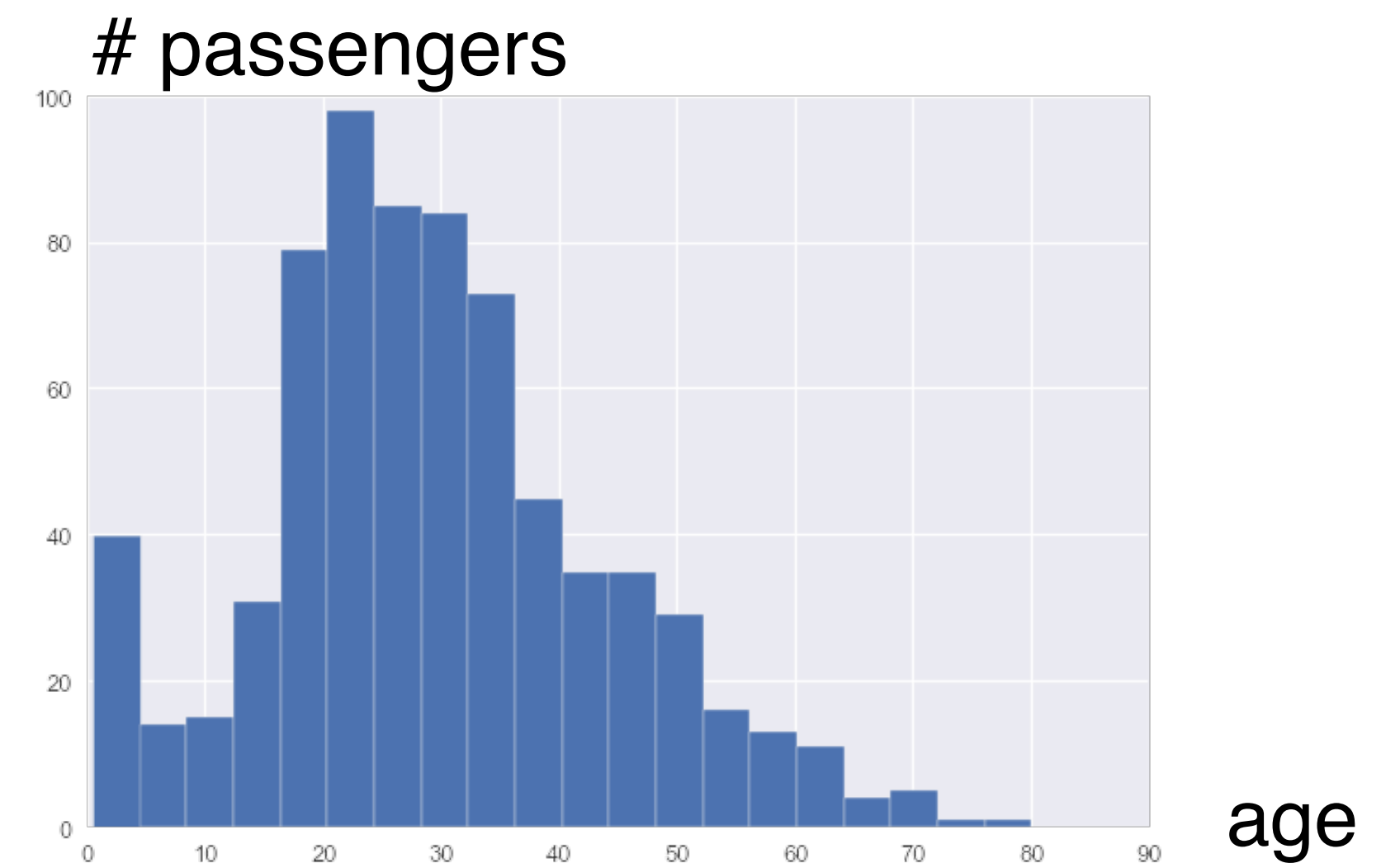
rules of thumb:

$$\#bins = \sqrt{n}$$

$$\#bins = \log_2(n) + 1$$

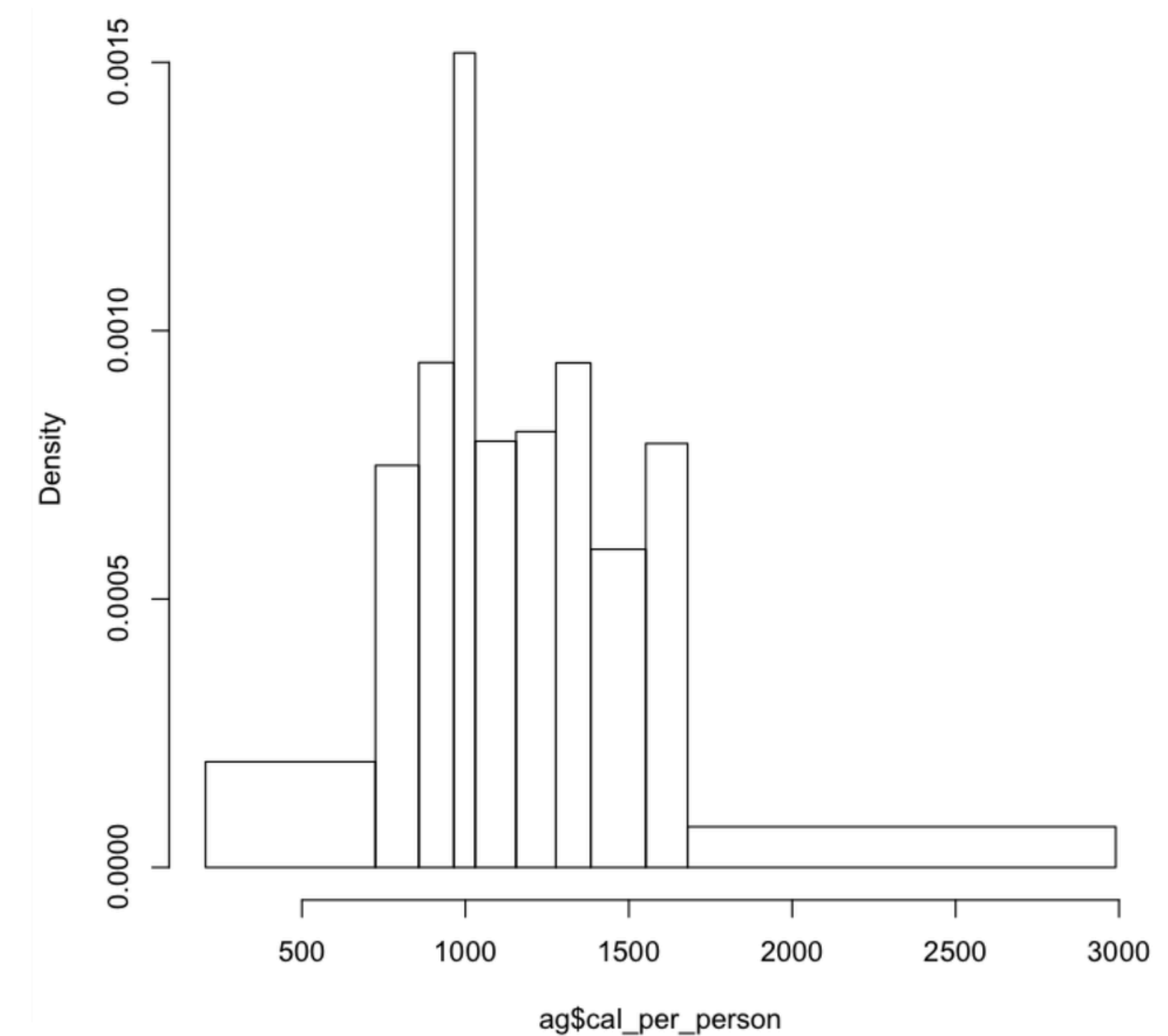
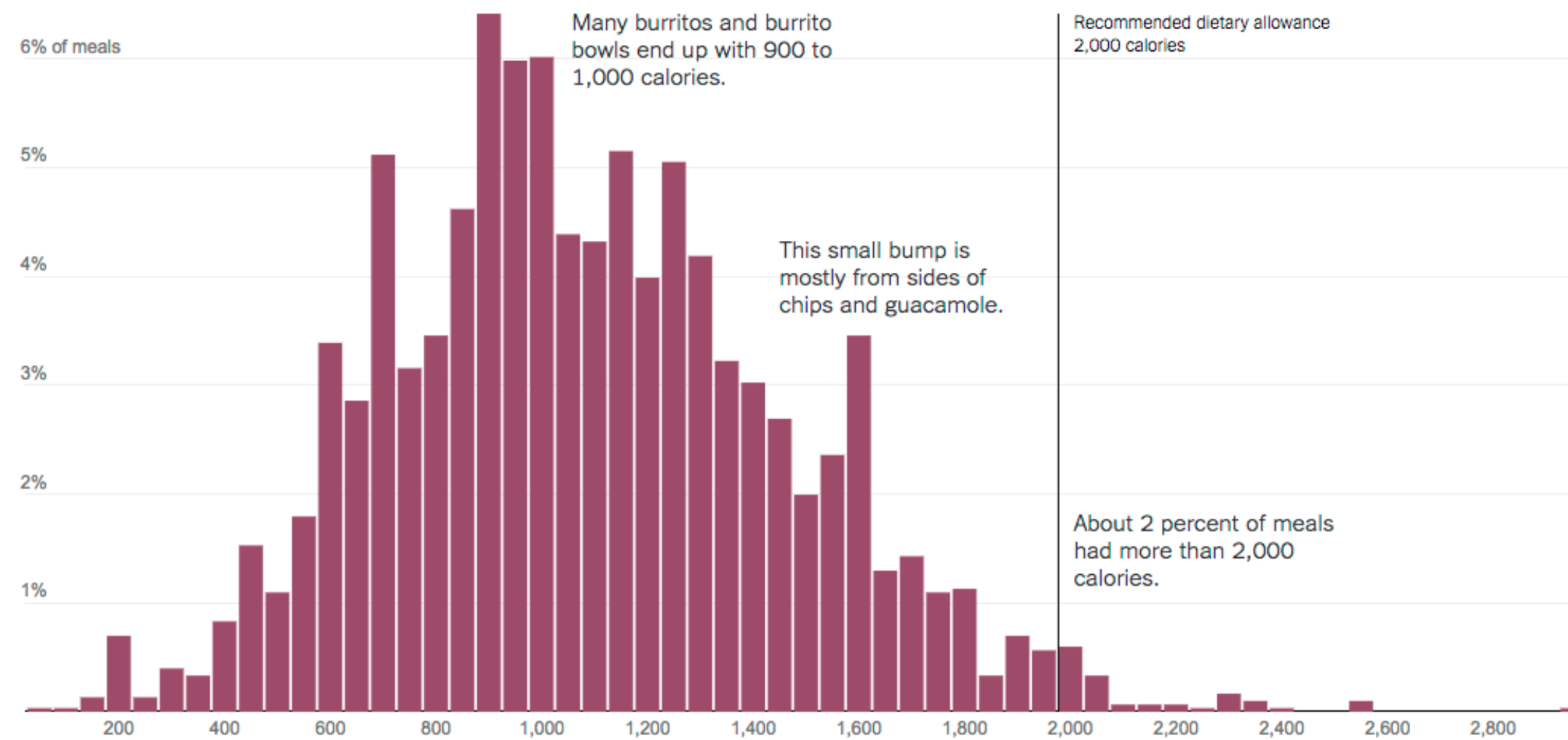


10 Bins



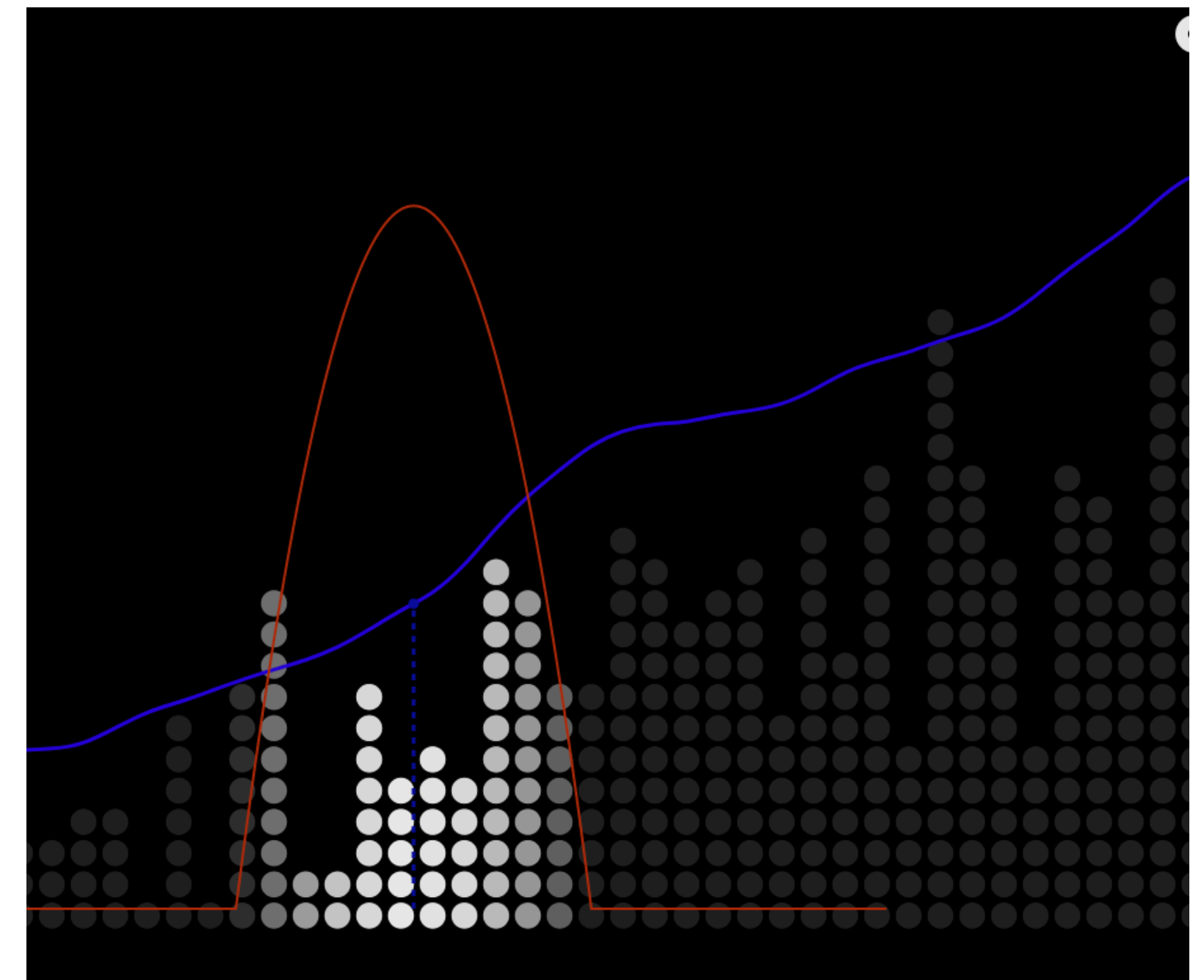
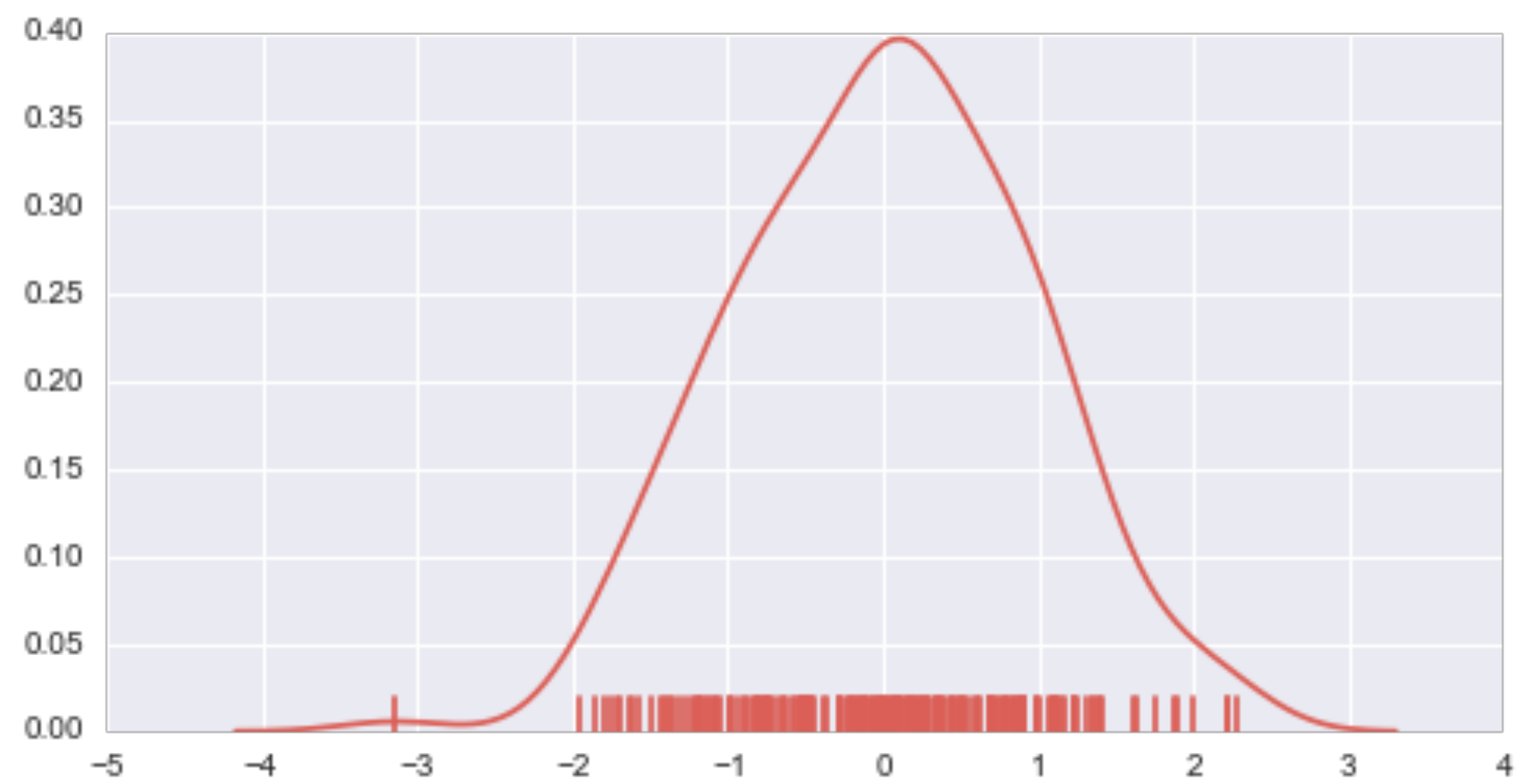
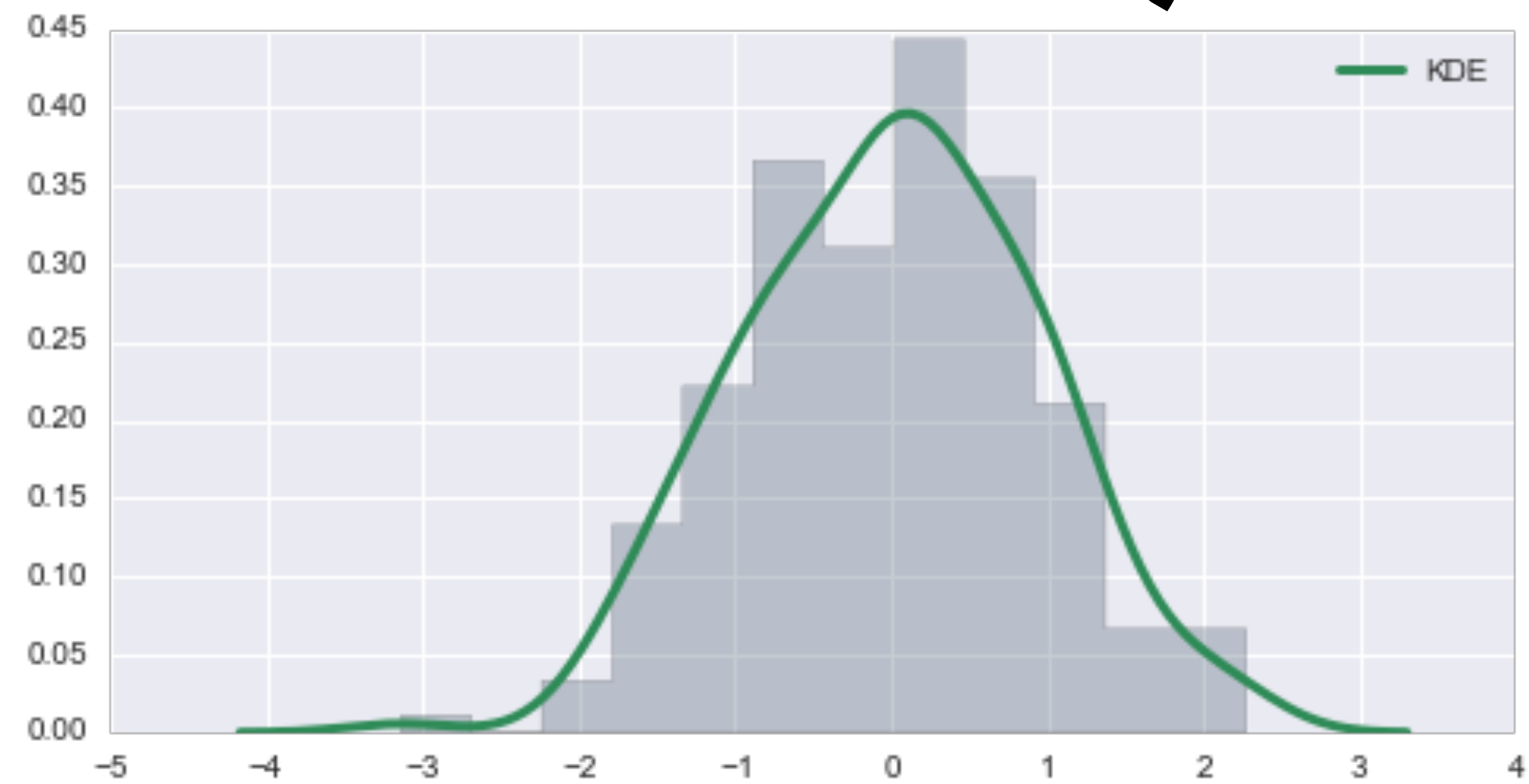
20 Bins

Unequal Bin Width



Can be useful if data is much sparser in some areas than others
Show density as area, not height.

Density Plots (Kernel Density Estimation)



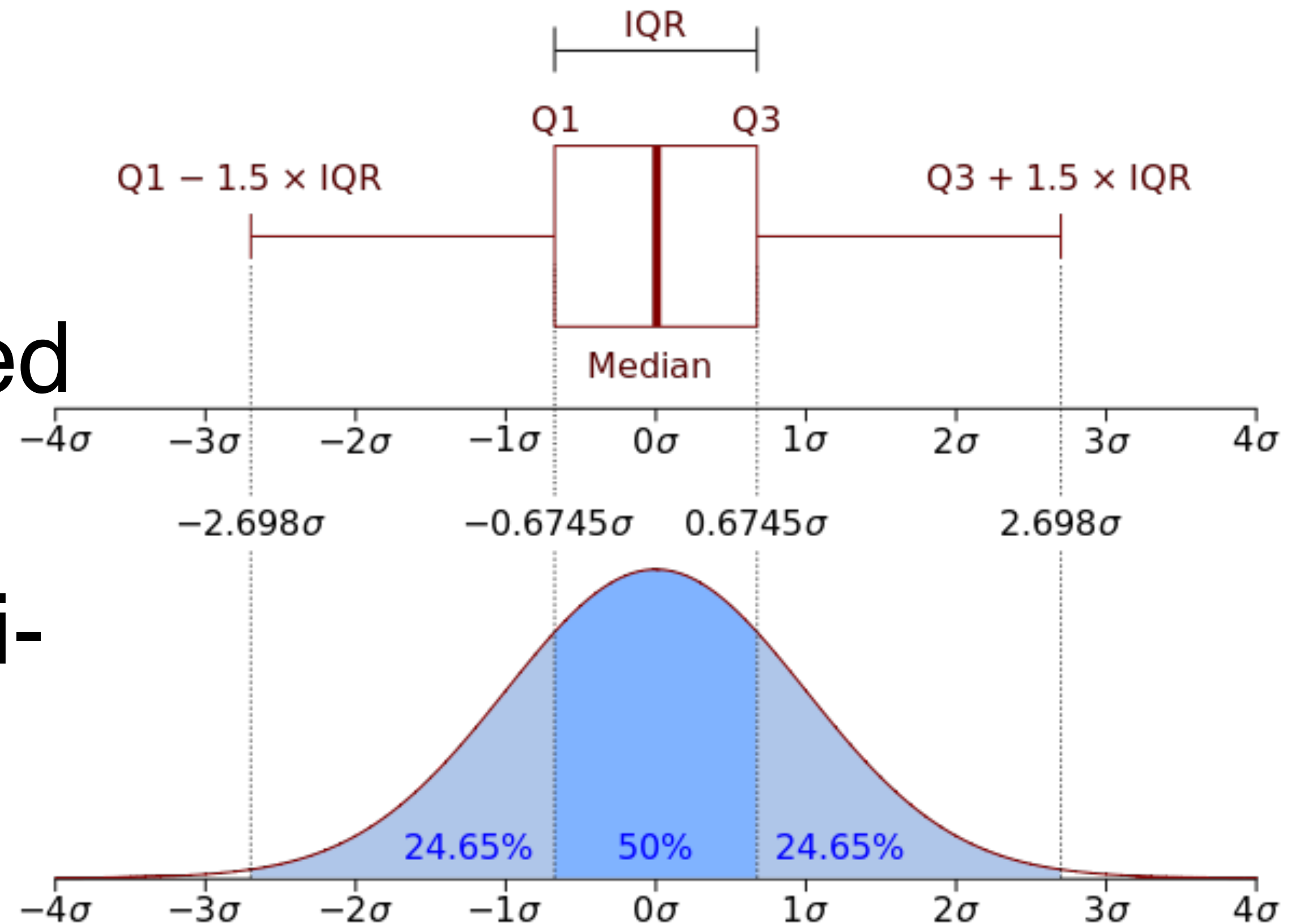
Box Plots

aka Box-and-Whisker Plot

Show outliers as points!

Bad for non-normal distributed data

Especially bad for bi- or multi-modal distributions



One Boxplot, Four Distributions

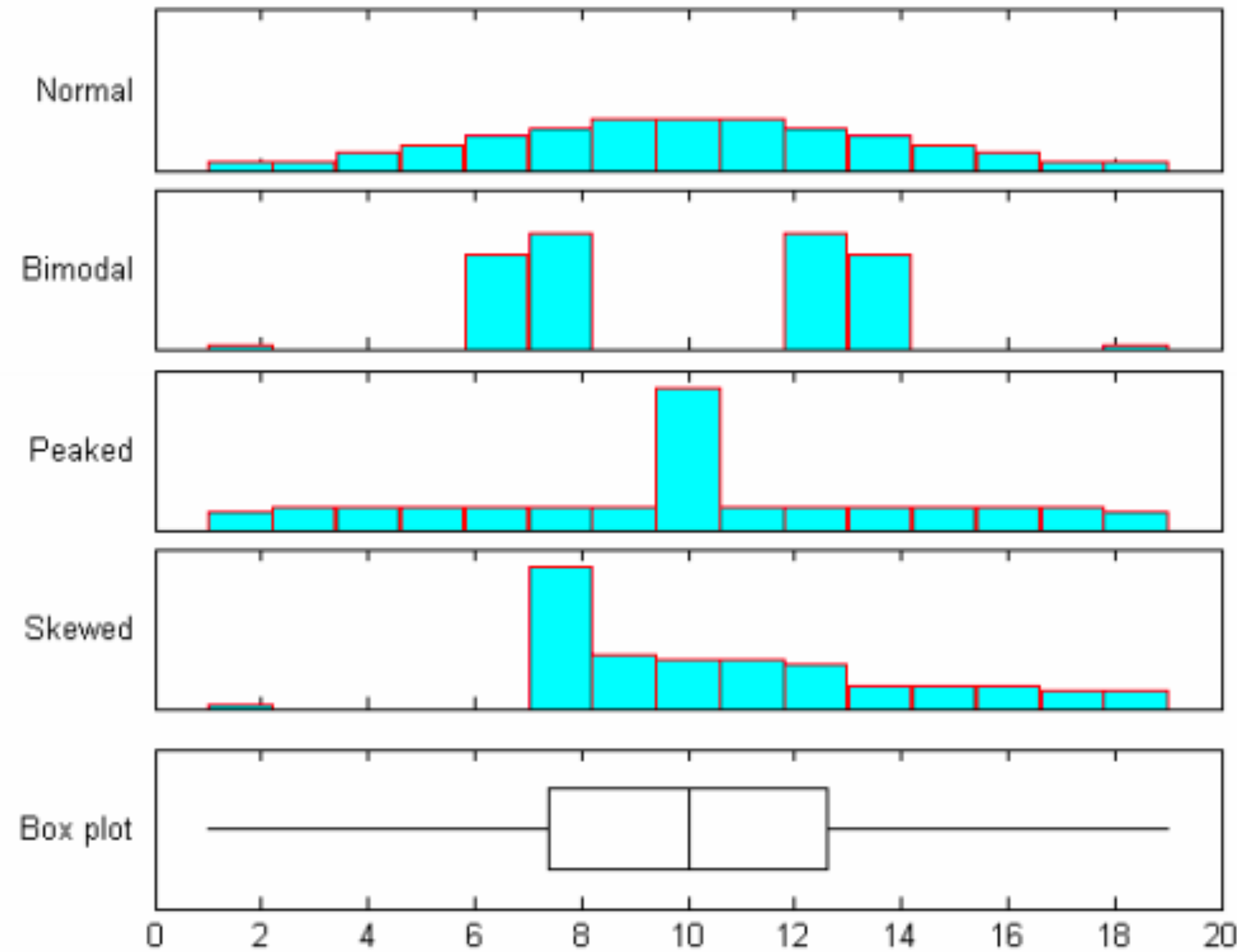


Figure 1: Histograms and box plot: four samples each of size 100

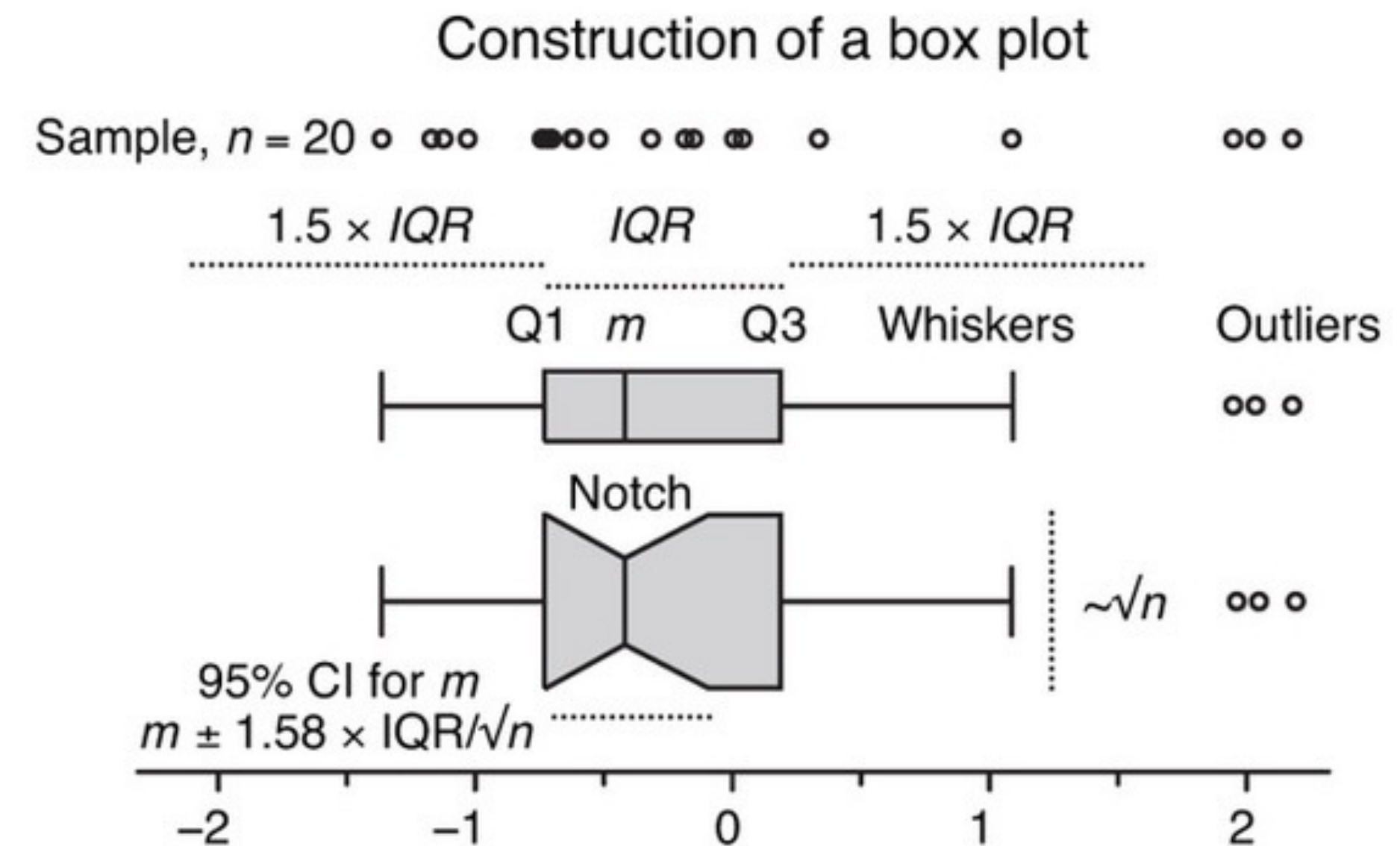
Notched Box Plots

Notch shows

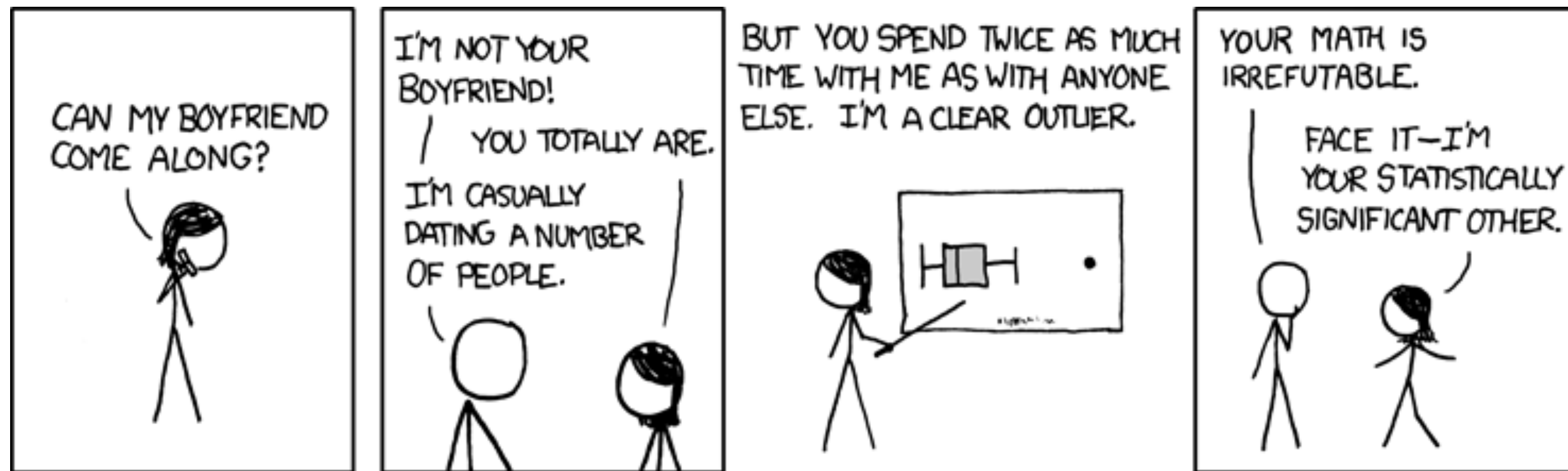
$m \pm 1.58 \times IQR/\sqrt{n}$

-> 95% Confidence Intervall

A guide to statistical significance.



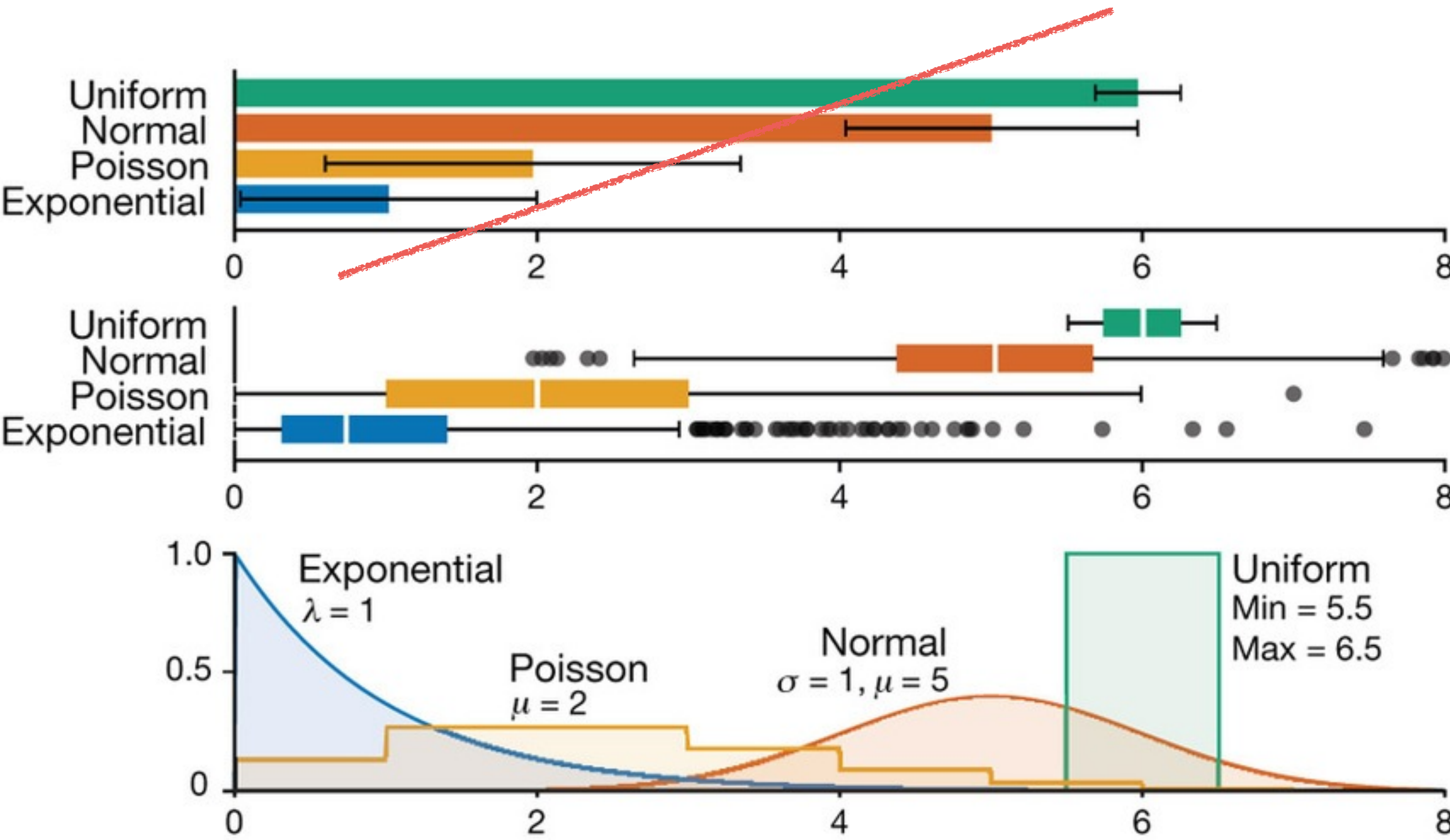
Box(and Whisker) Plots



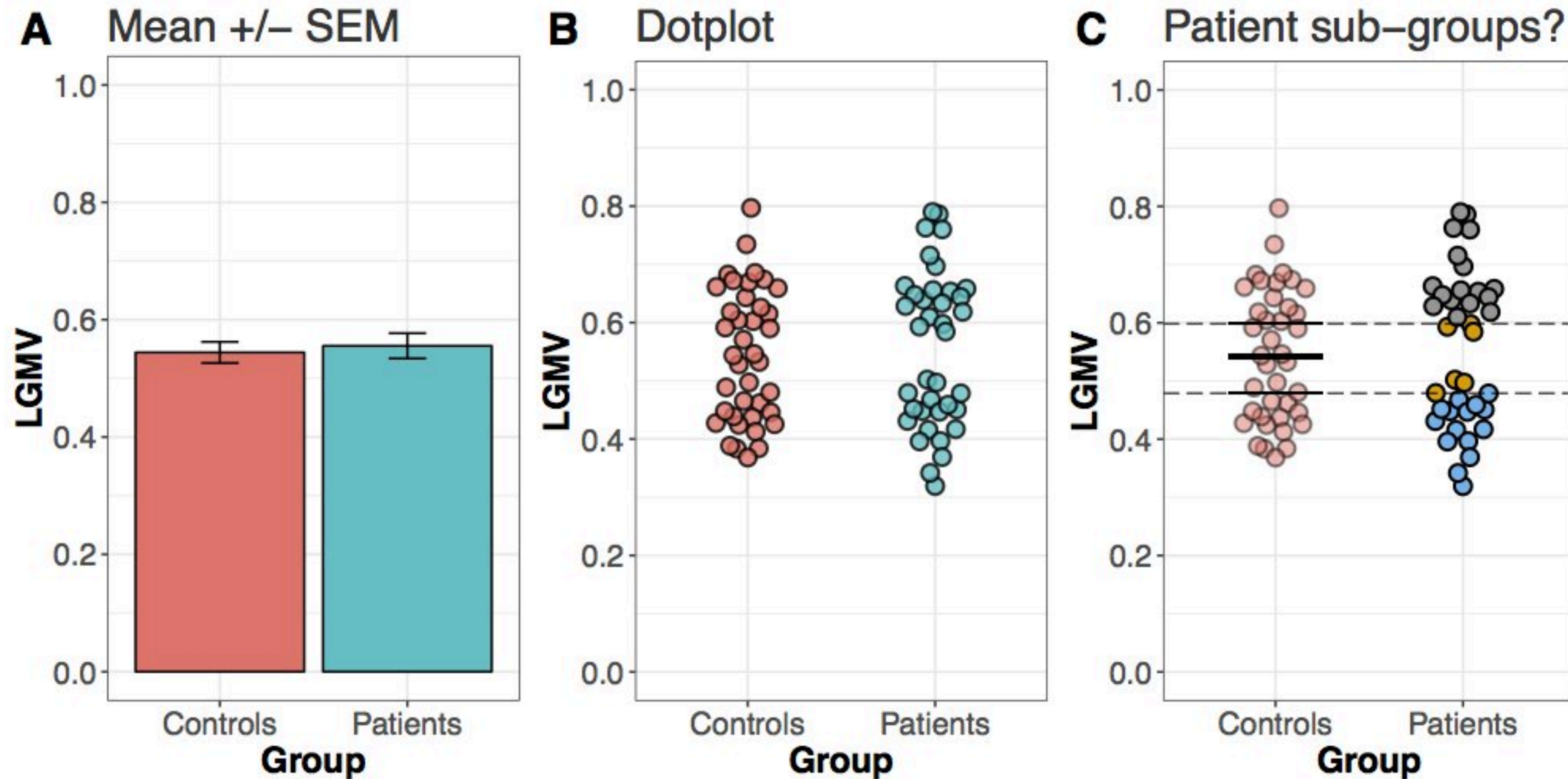
<http://xkcd.com/539/>



Comparison

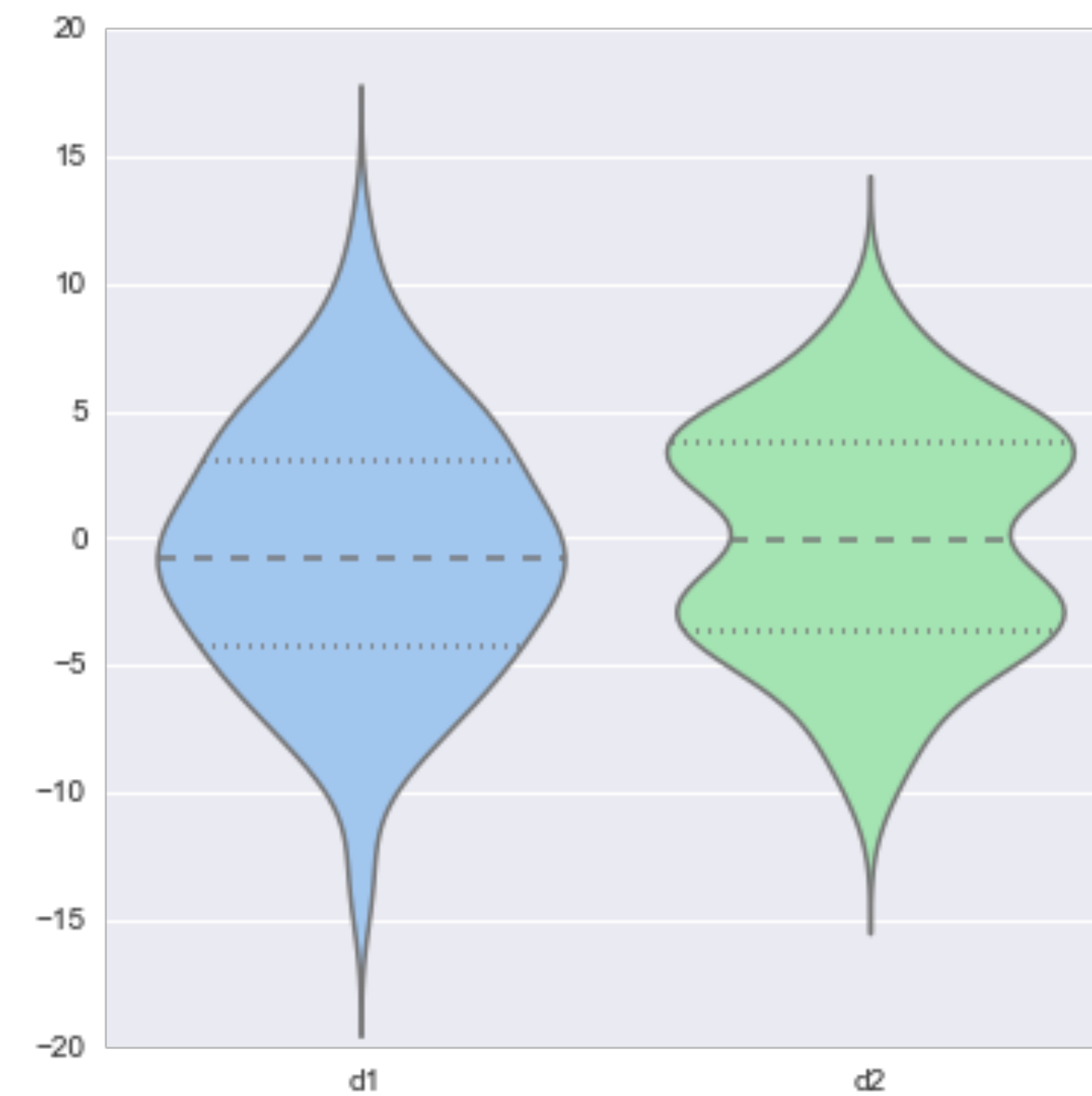
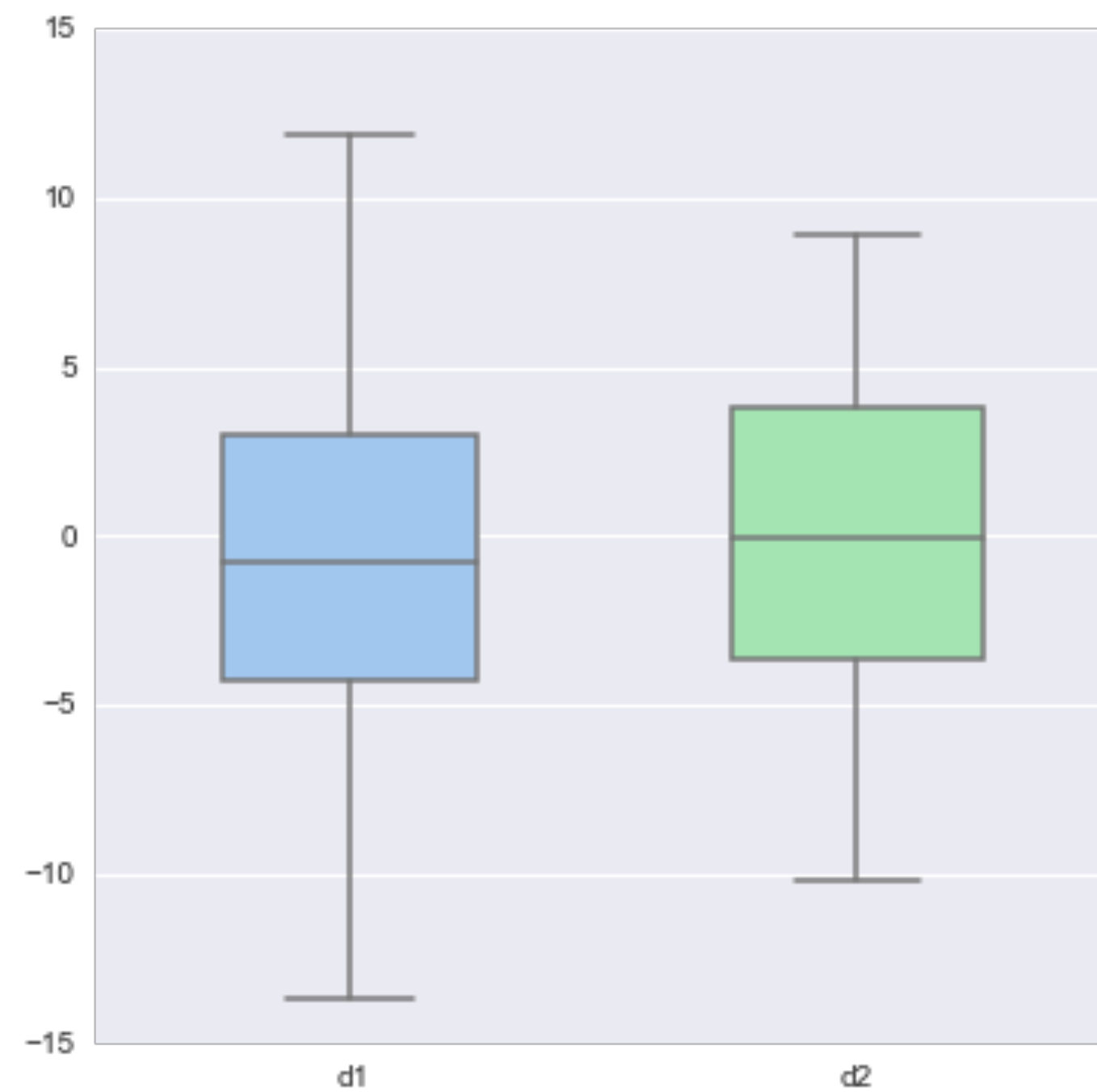


Bar Charts vs Dot Plots

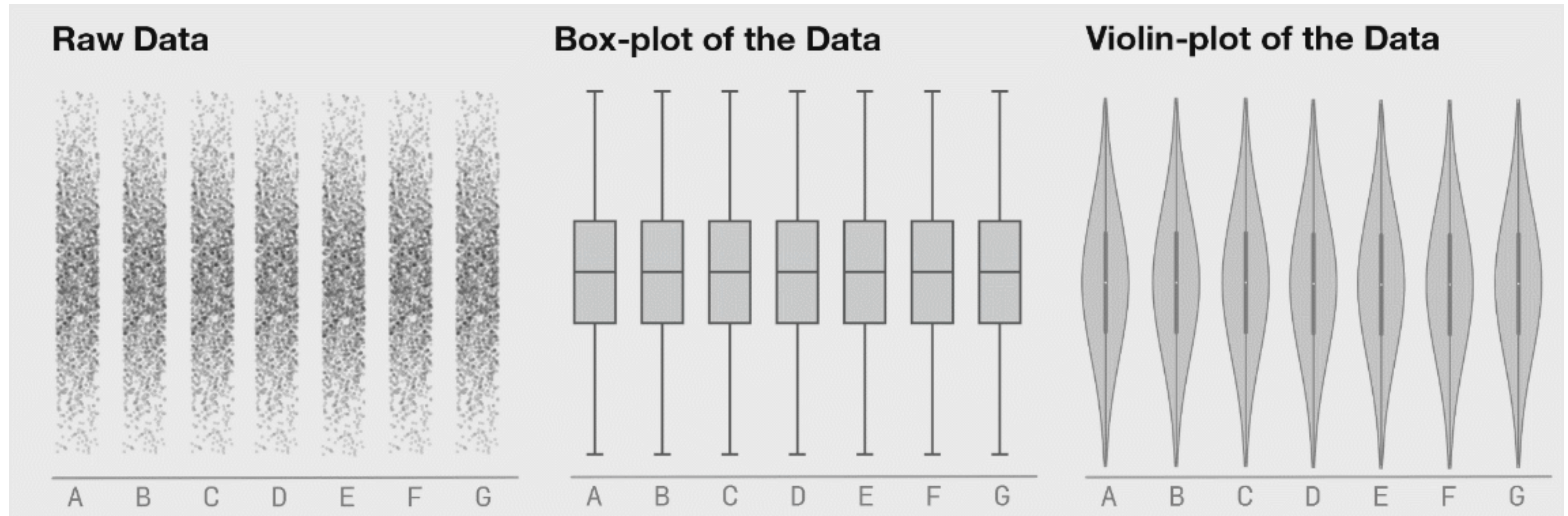


Violin Plot

= Box Plot + Probability Density Function

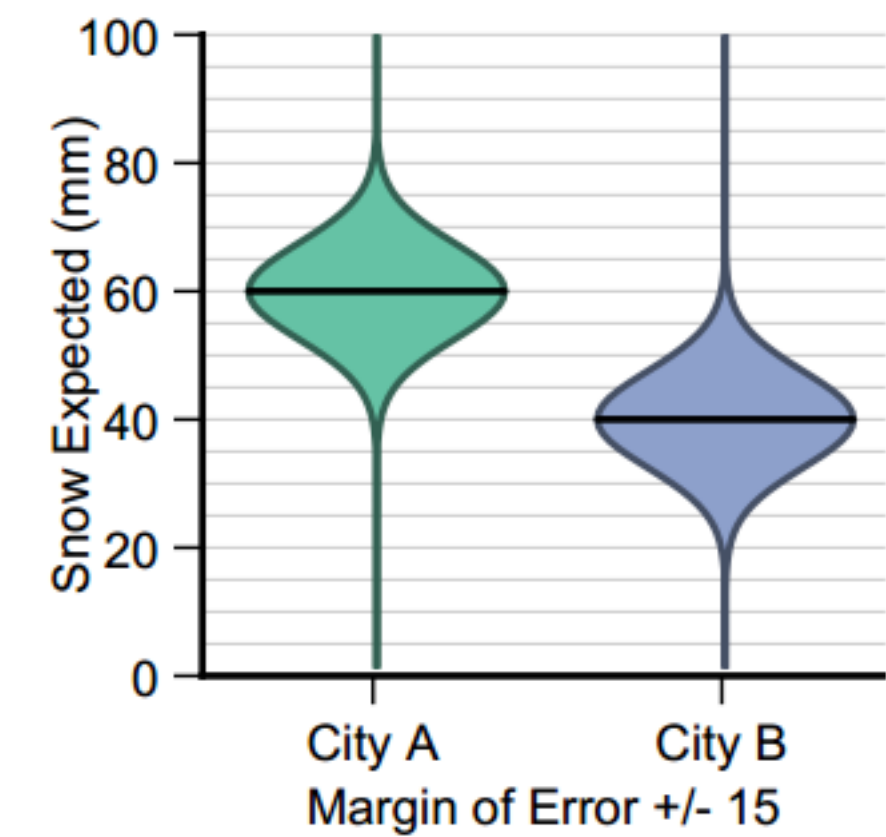
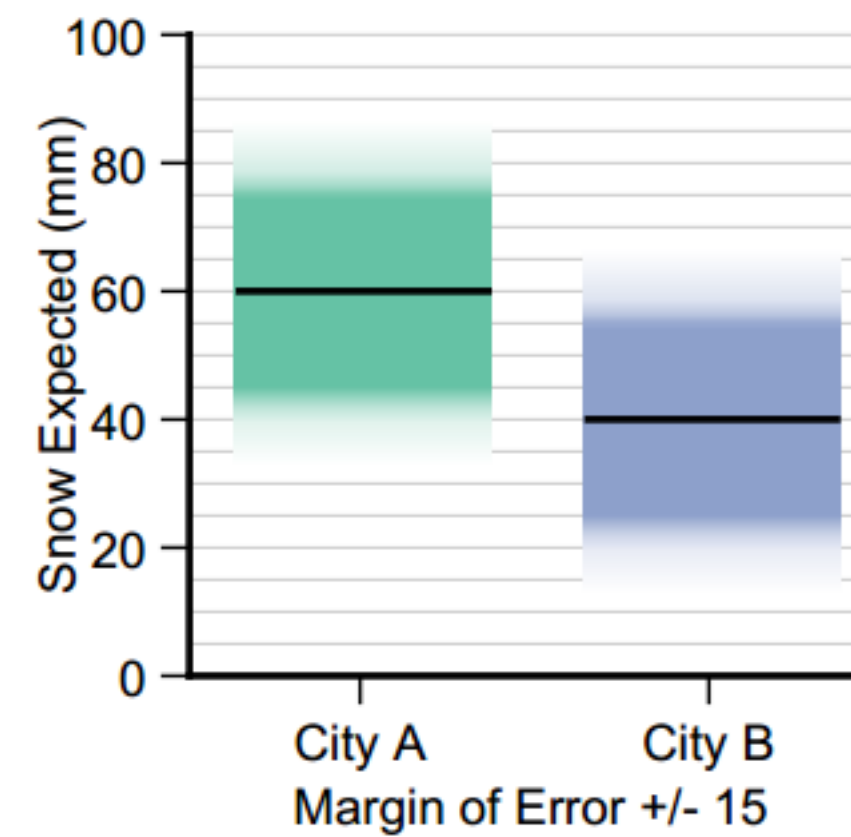
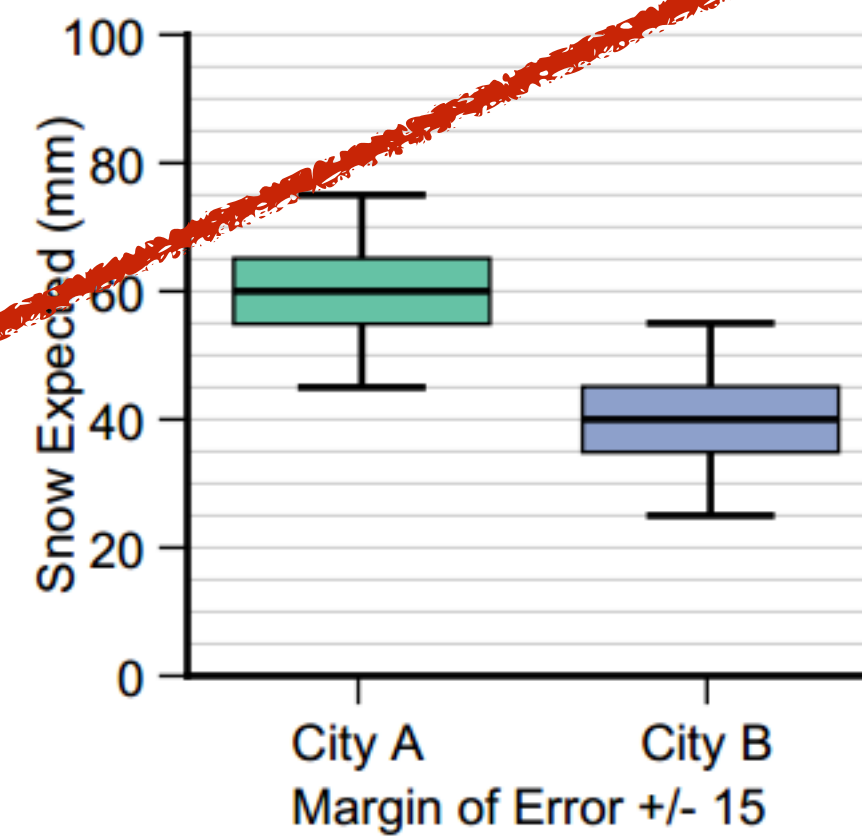
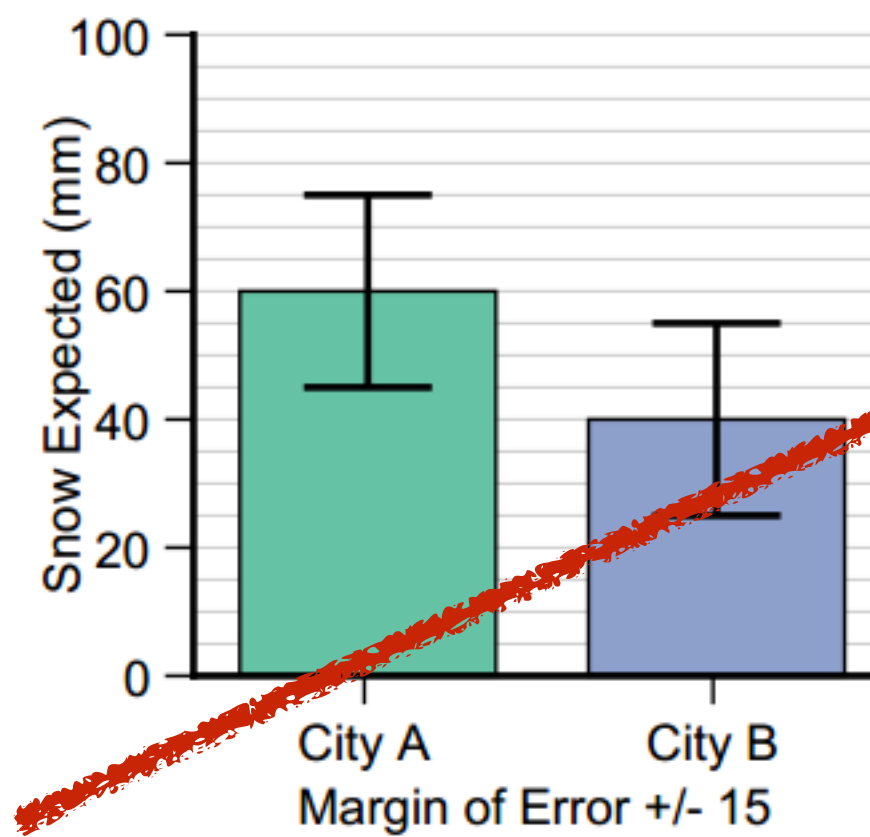


Different Distributions



Showing Expected Values & Uncertainty

NOT a distribution!



Error Bars Considered Harmful:
Exploring Alternate Encodings for Mean and Error
Michael Correll, and Michael Gleicher

One of these things is not like the other...

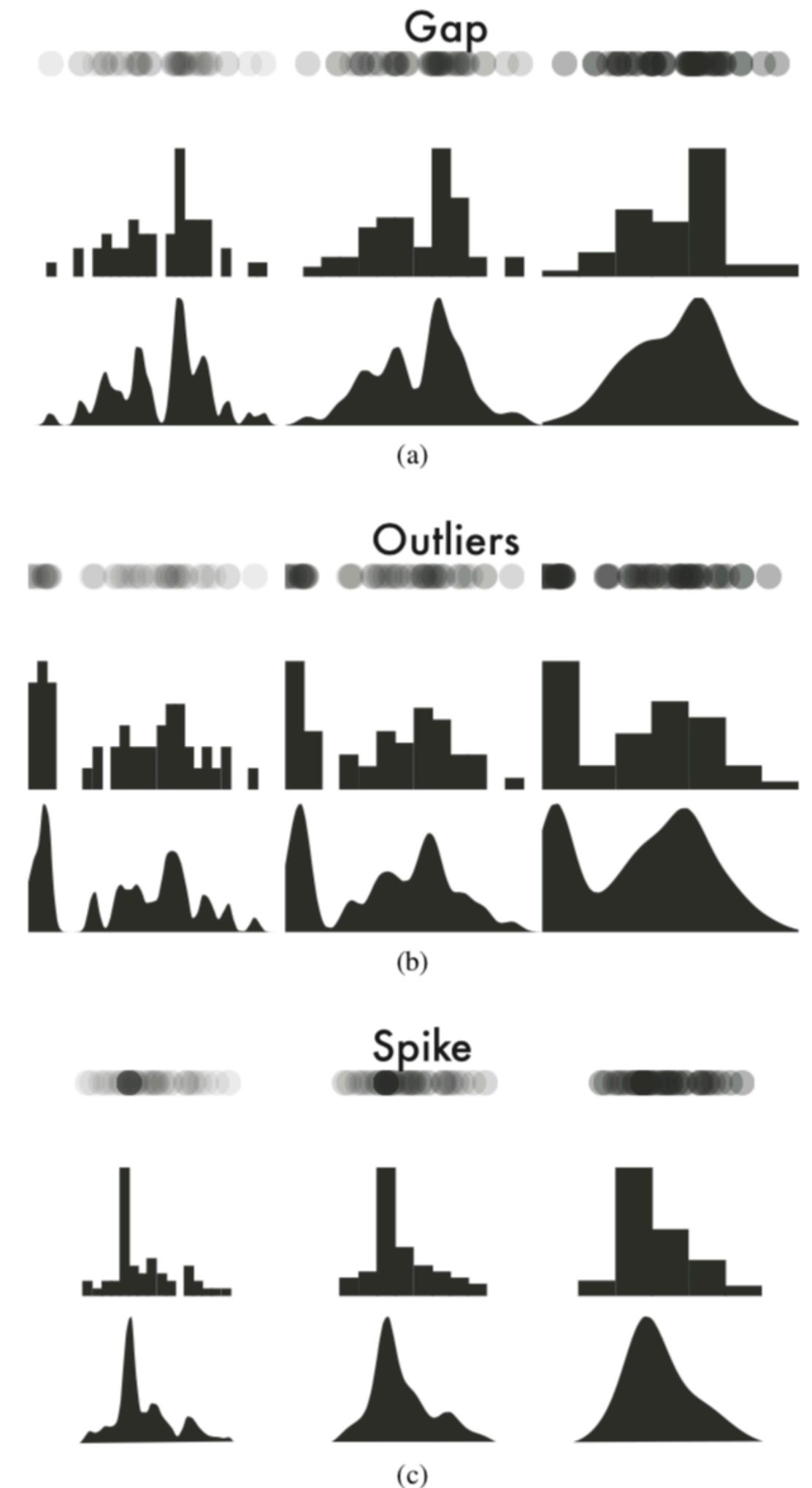
19 charts are random samples from a gaussian.
1 chart has 20% of samples with identical value



Detecting Data Flaws

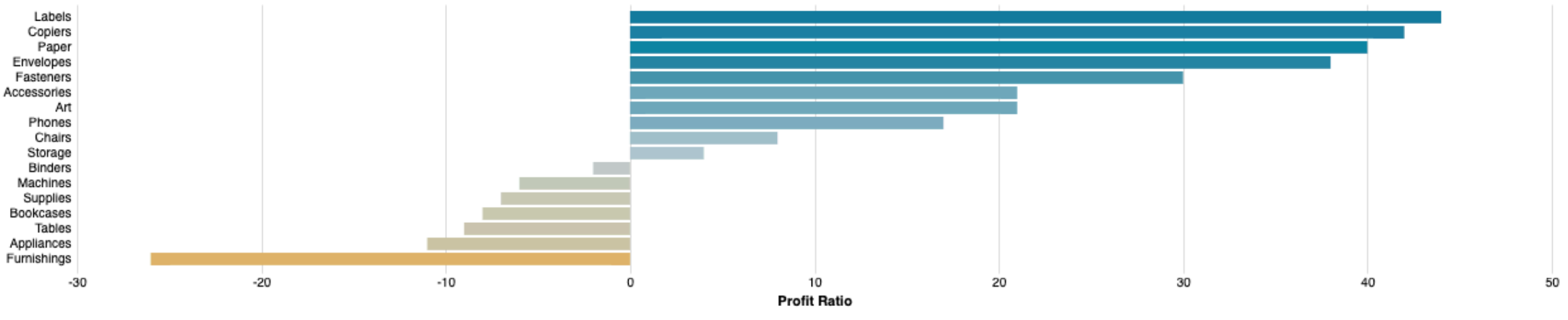
Tricky with aggregate
visualization

Bin size / kernel type /
bandwidth / visualization
choice all affect different
situations

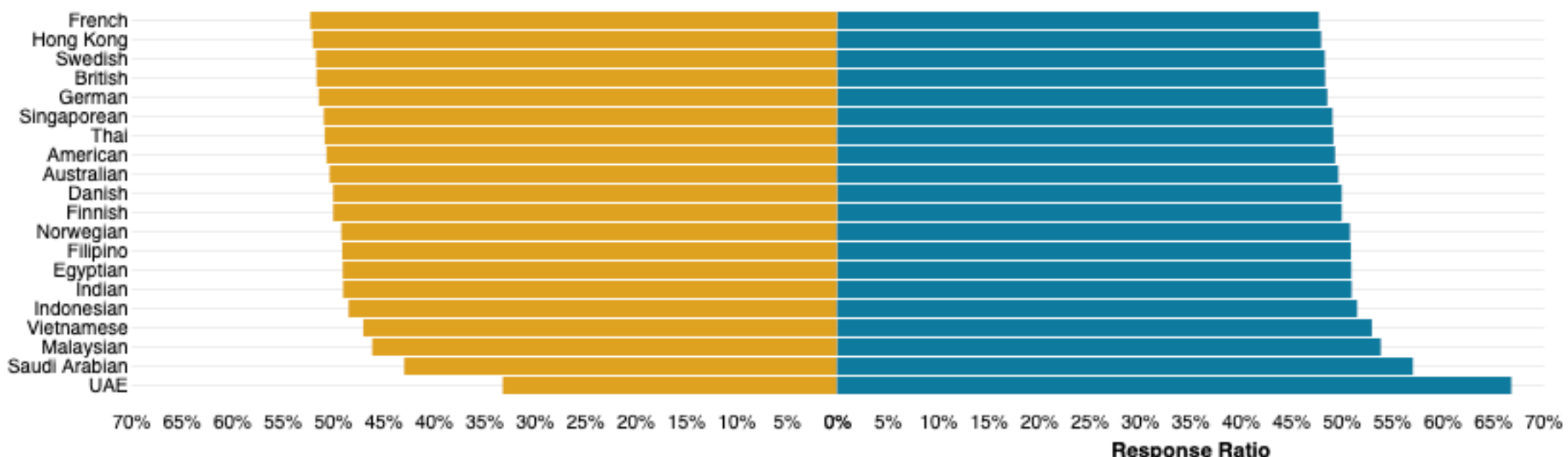


Deviation

Comparison to Reference Point



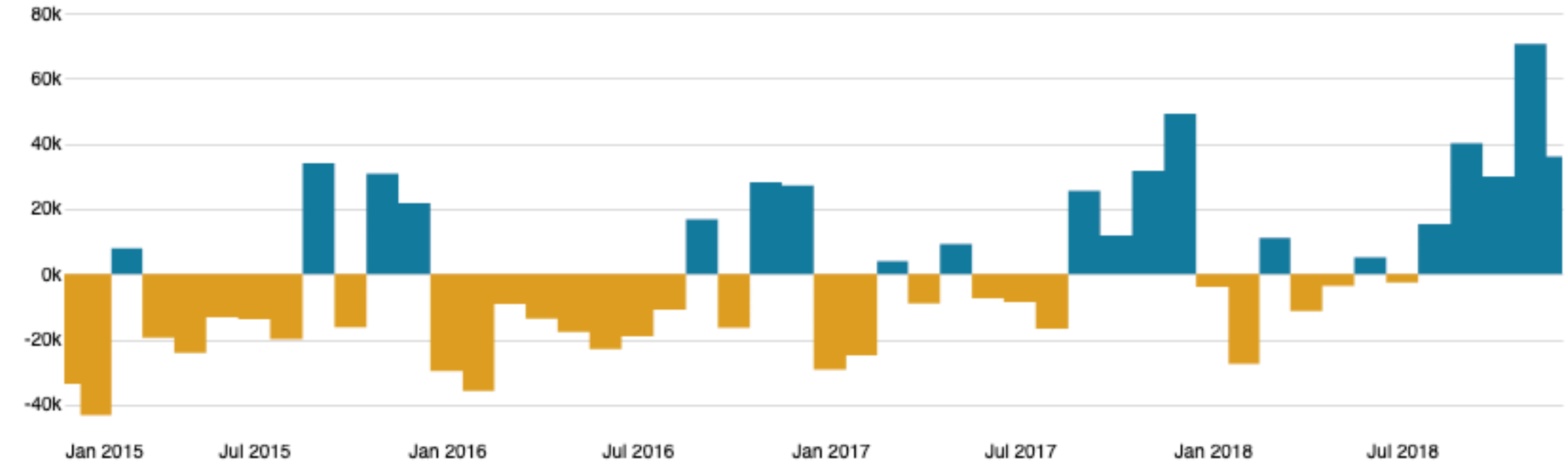
Diverging Bar Chart



Juxtaposing Two Variables (male/female)

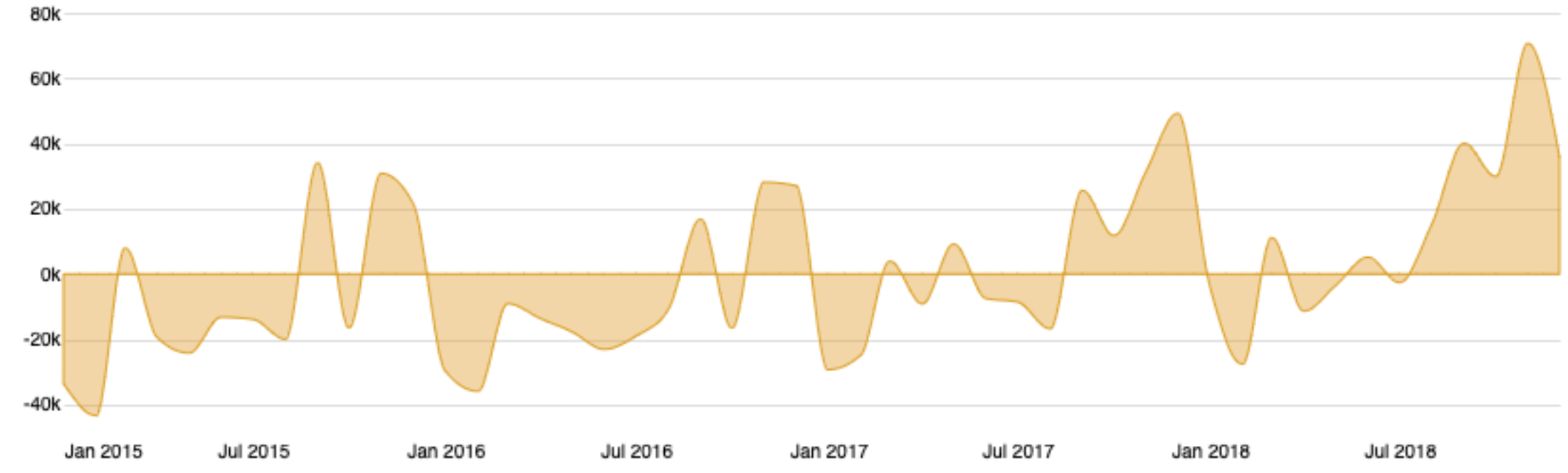
Surplus/deficit filled line

The shaded area of these charts allows a balance to be shown; either against a baseline or between two serie



Surplus/deficit filled area

Same as before.



Change over Time

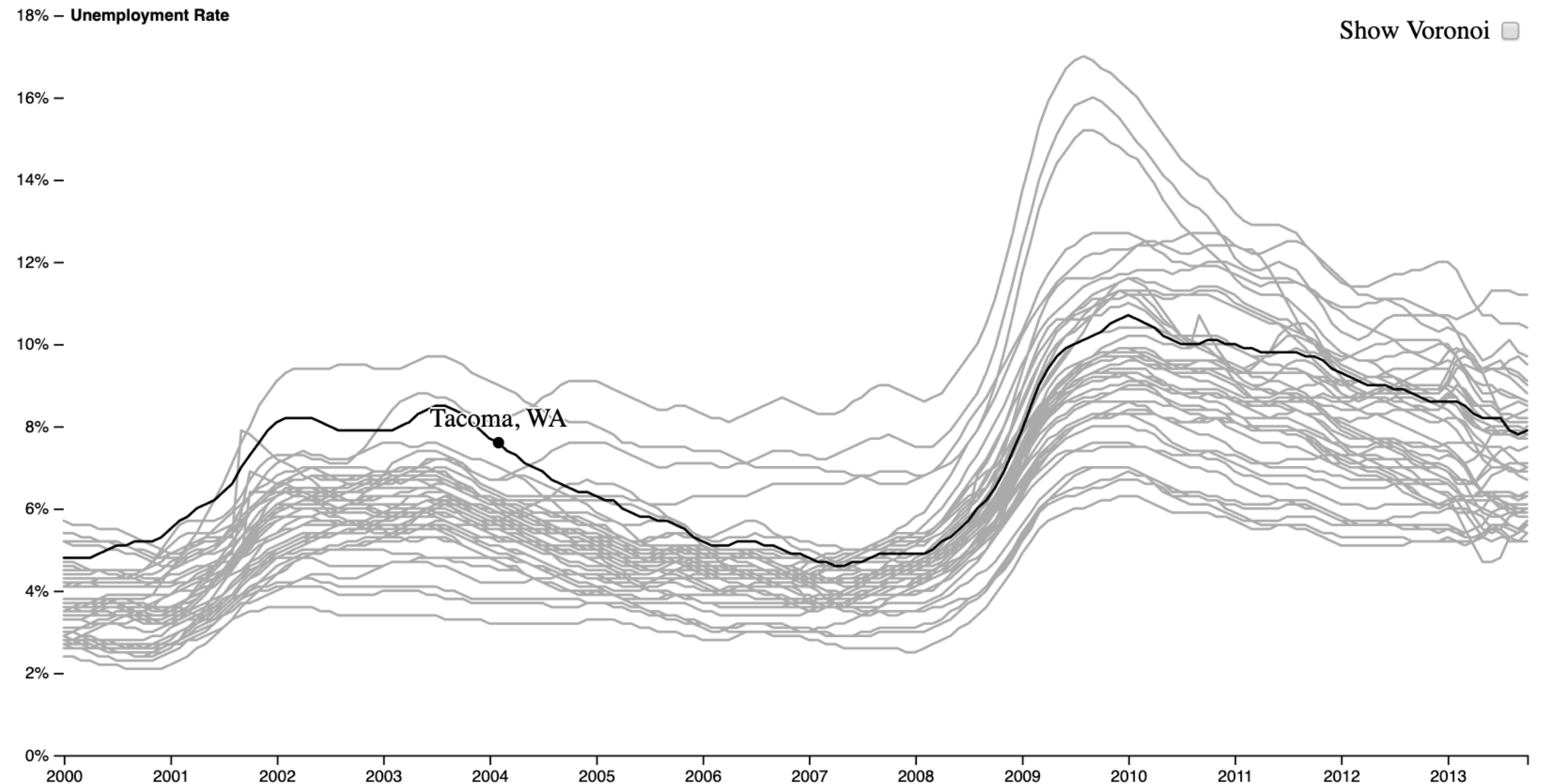
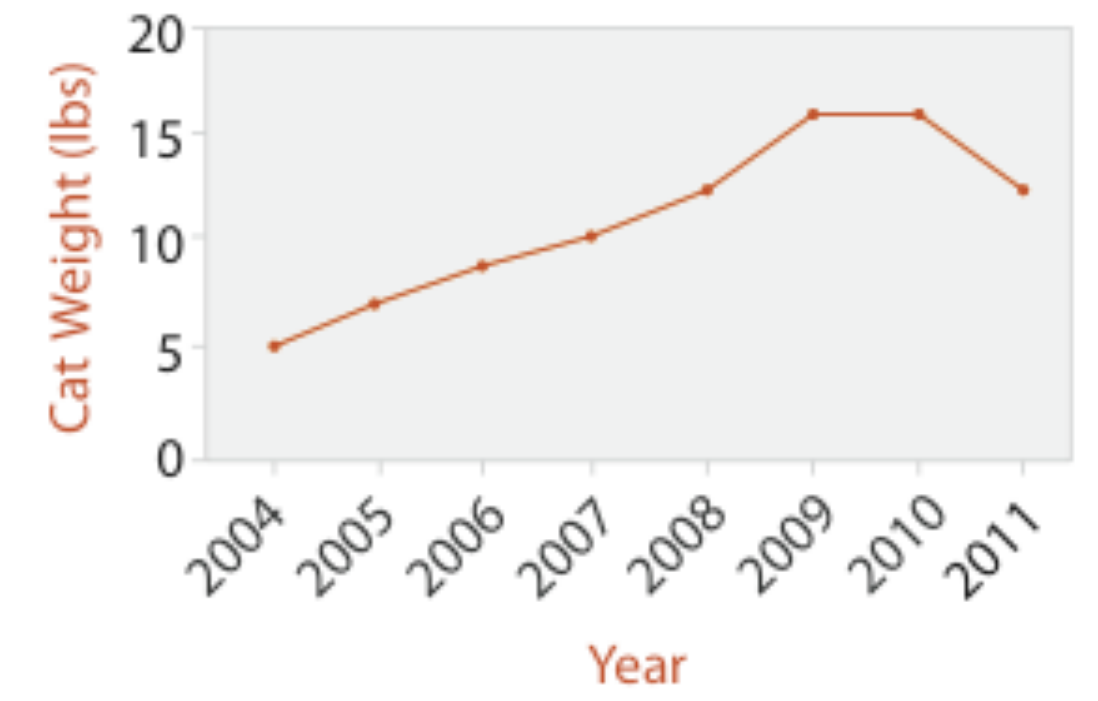
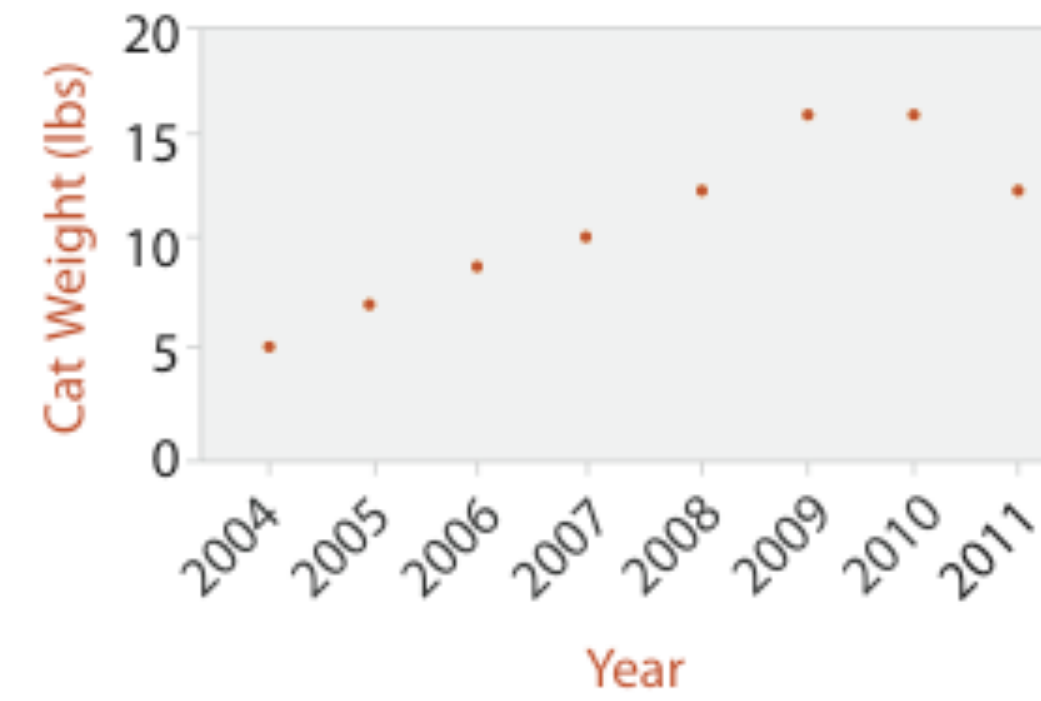
Line Chart

Simple

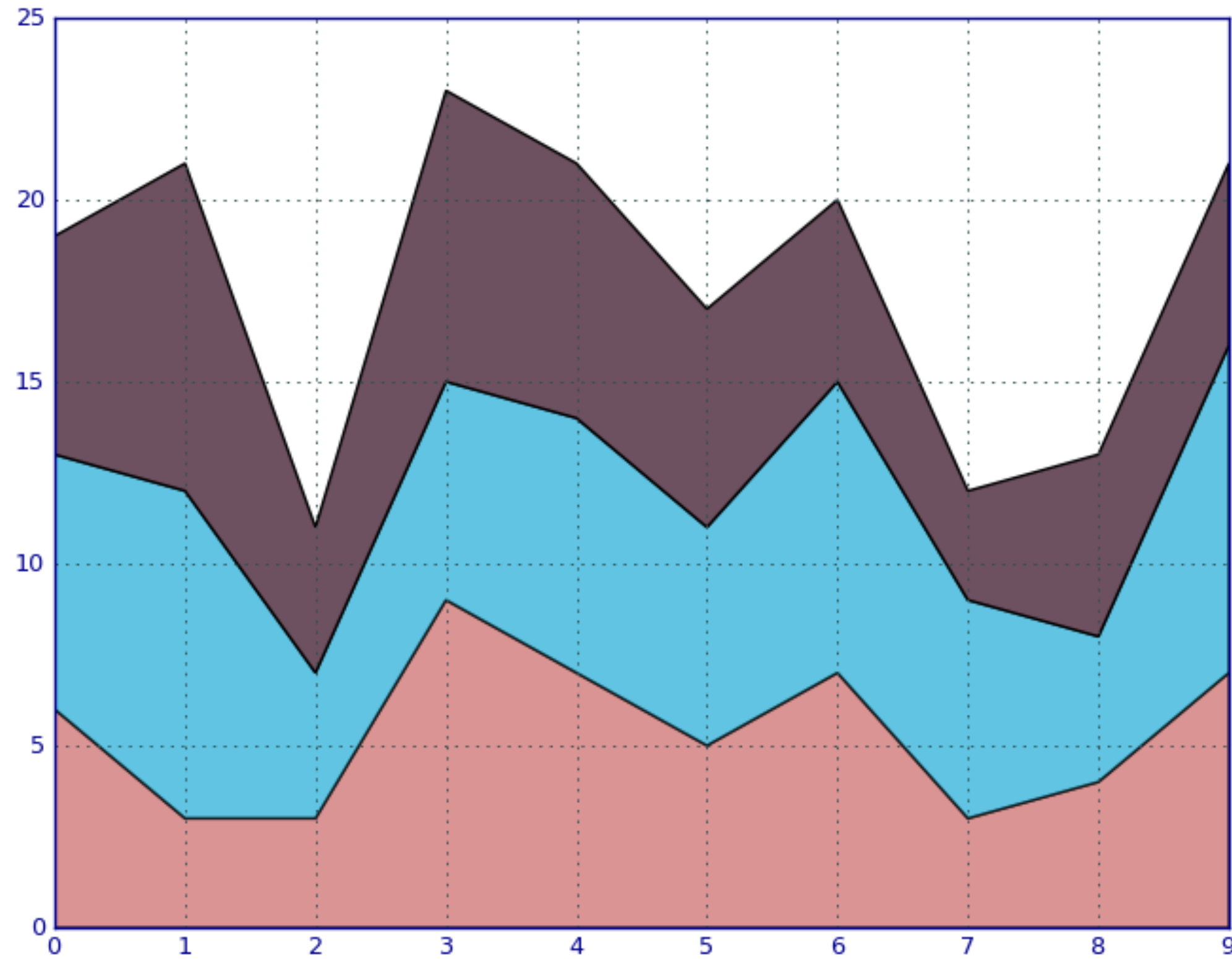
Familiar

Accurate

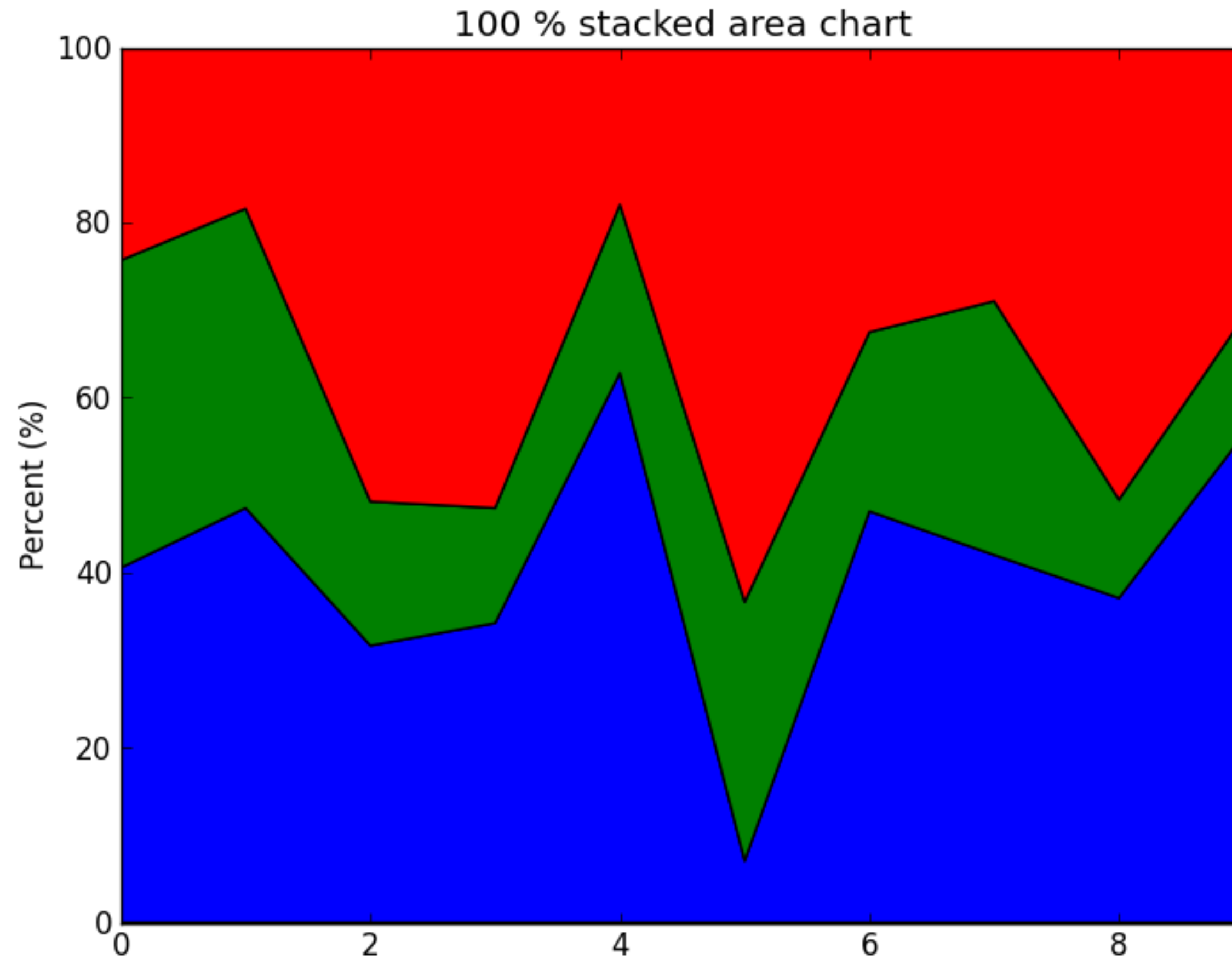
Fairly Scalable



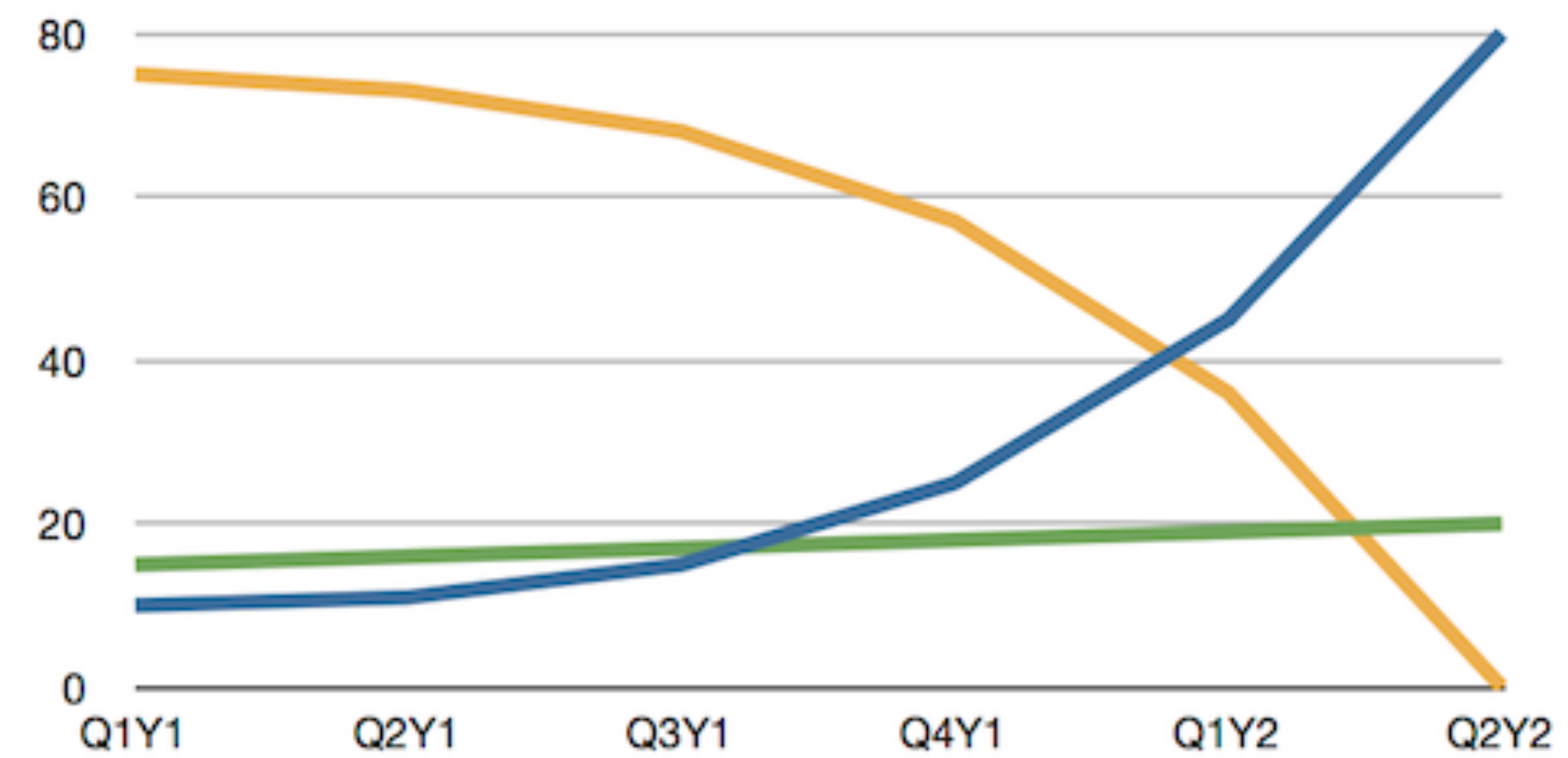
Stacked Area Chart



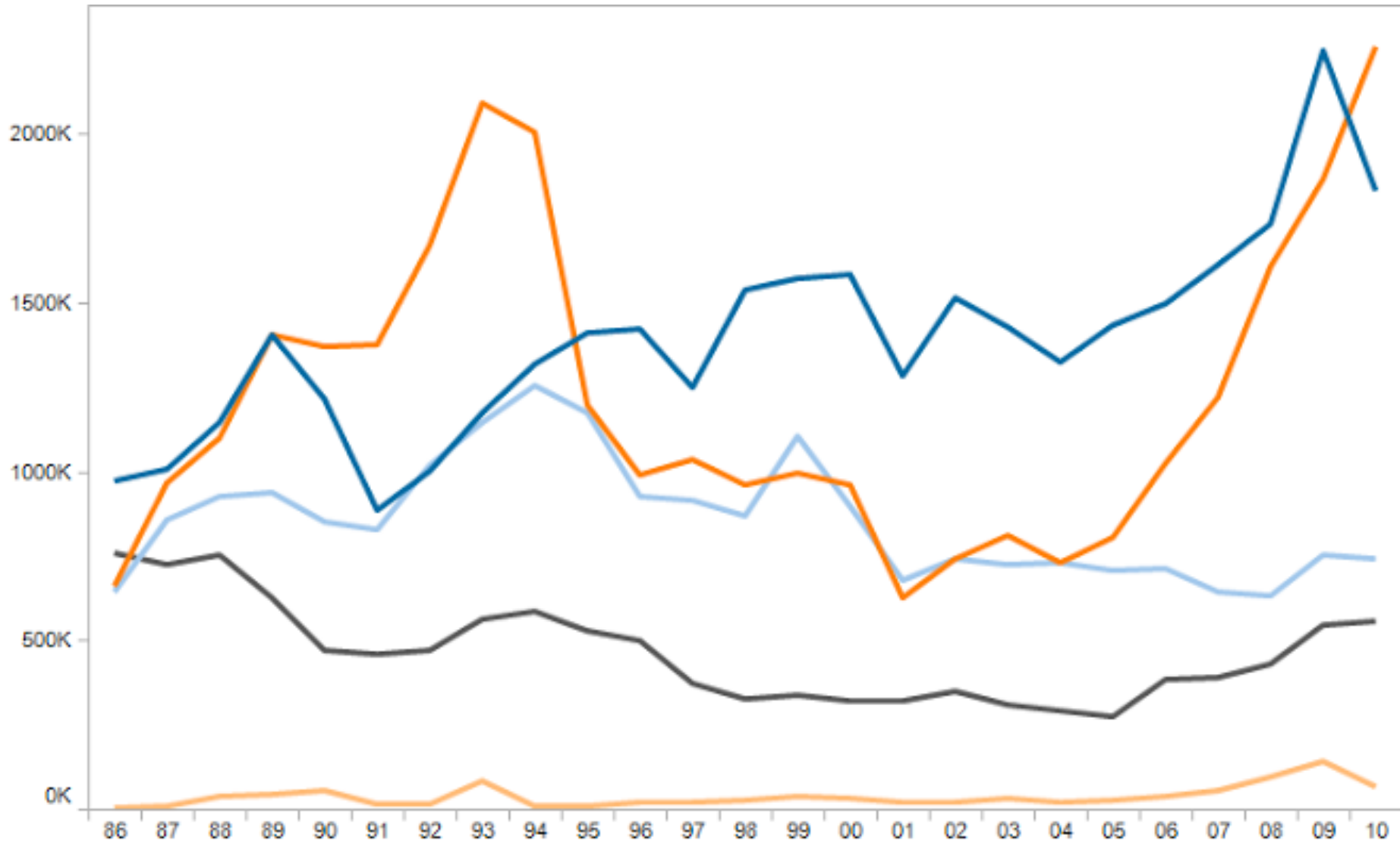
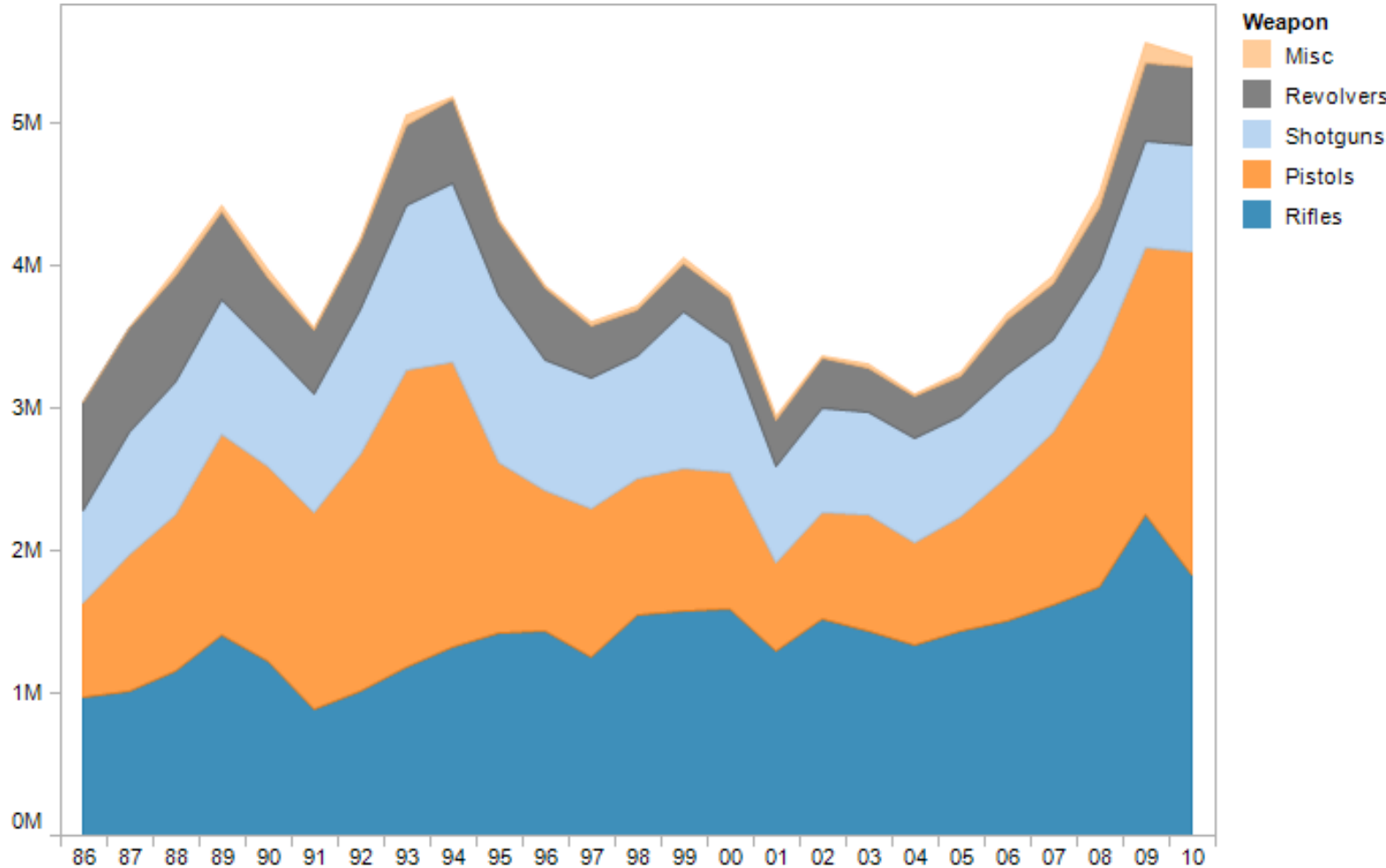
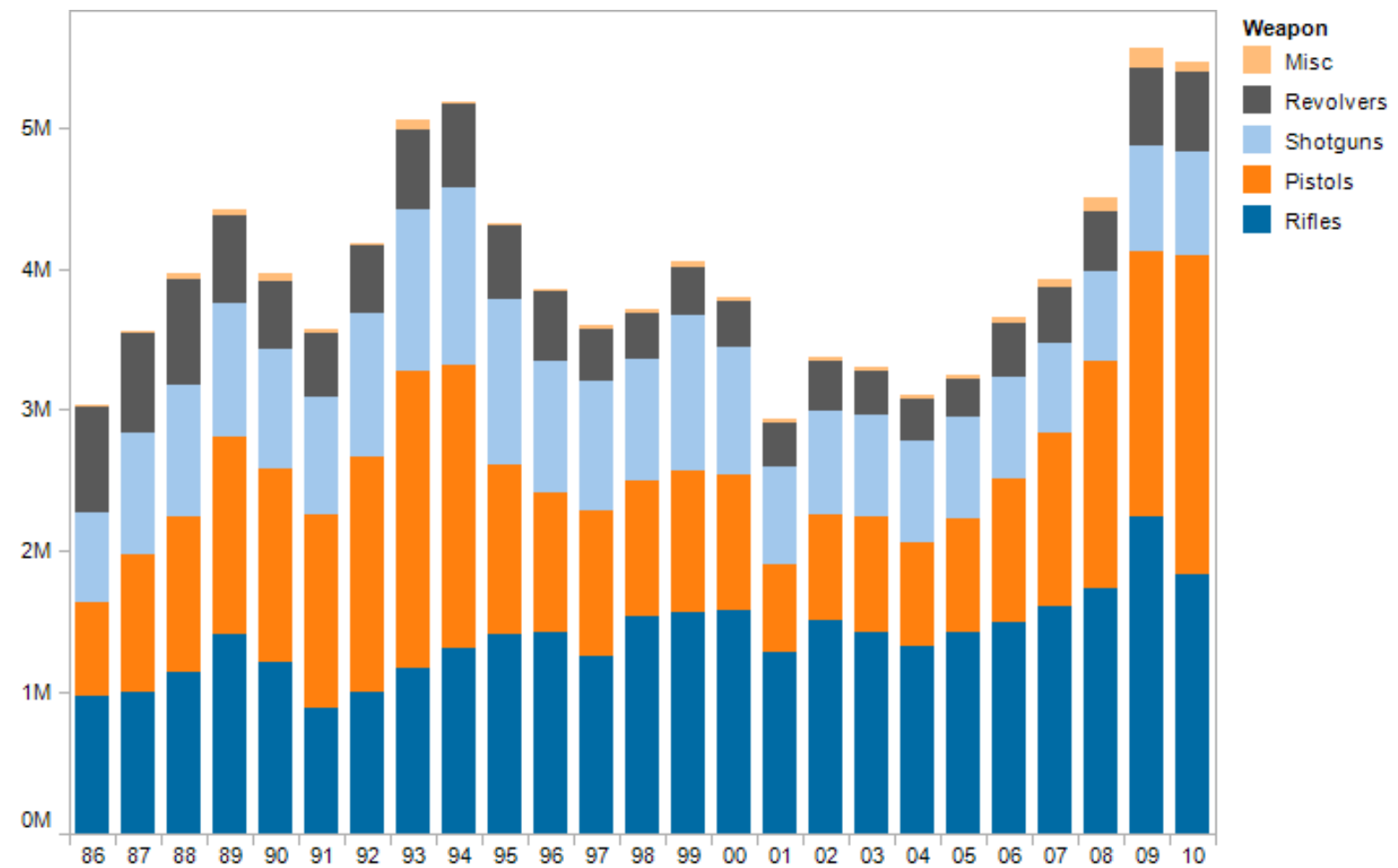
100% Stacked Area Chart



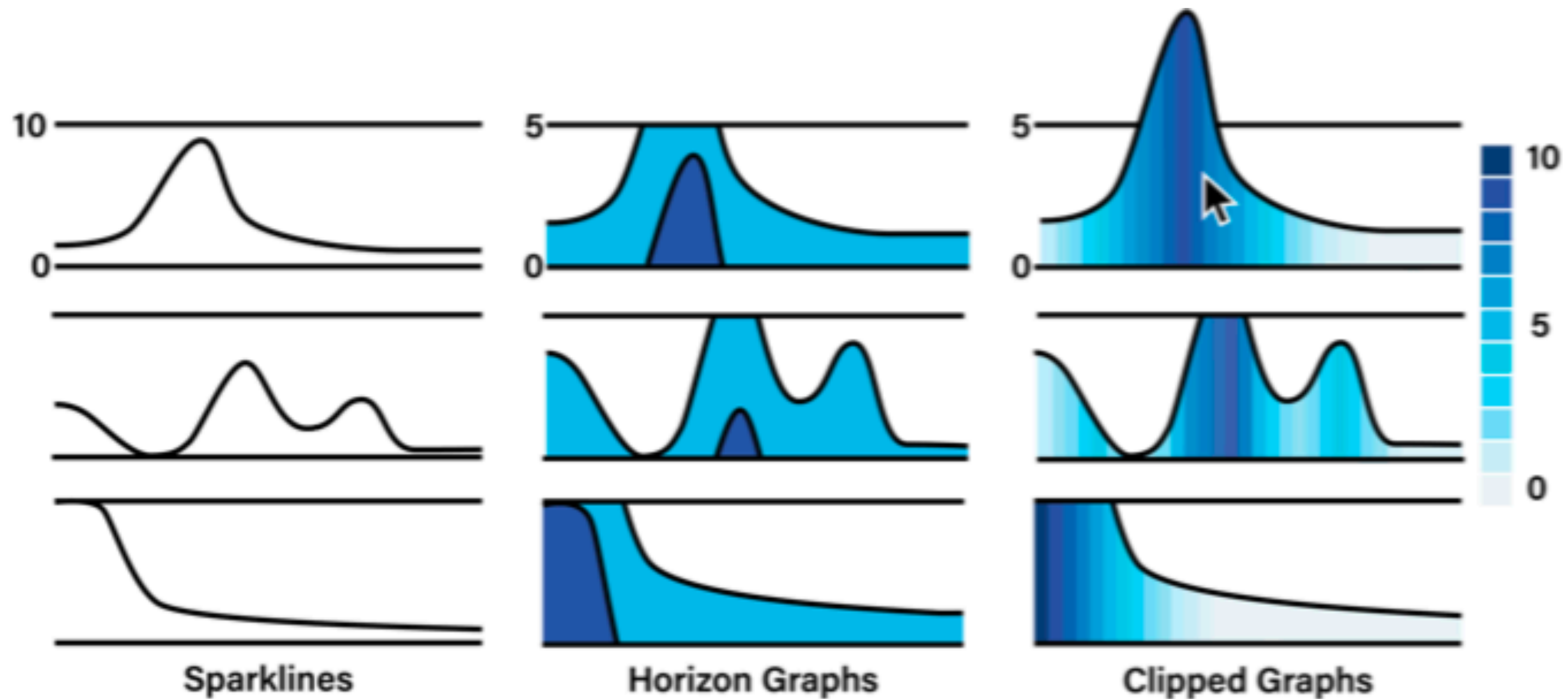
Stacked Area vs. Line Graphs



Can you spot the trends?



Multiple Line Charts



Sparklines
















Small line charts

can be embedded in text
or part of a table

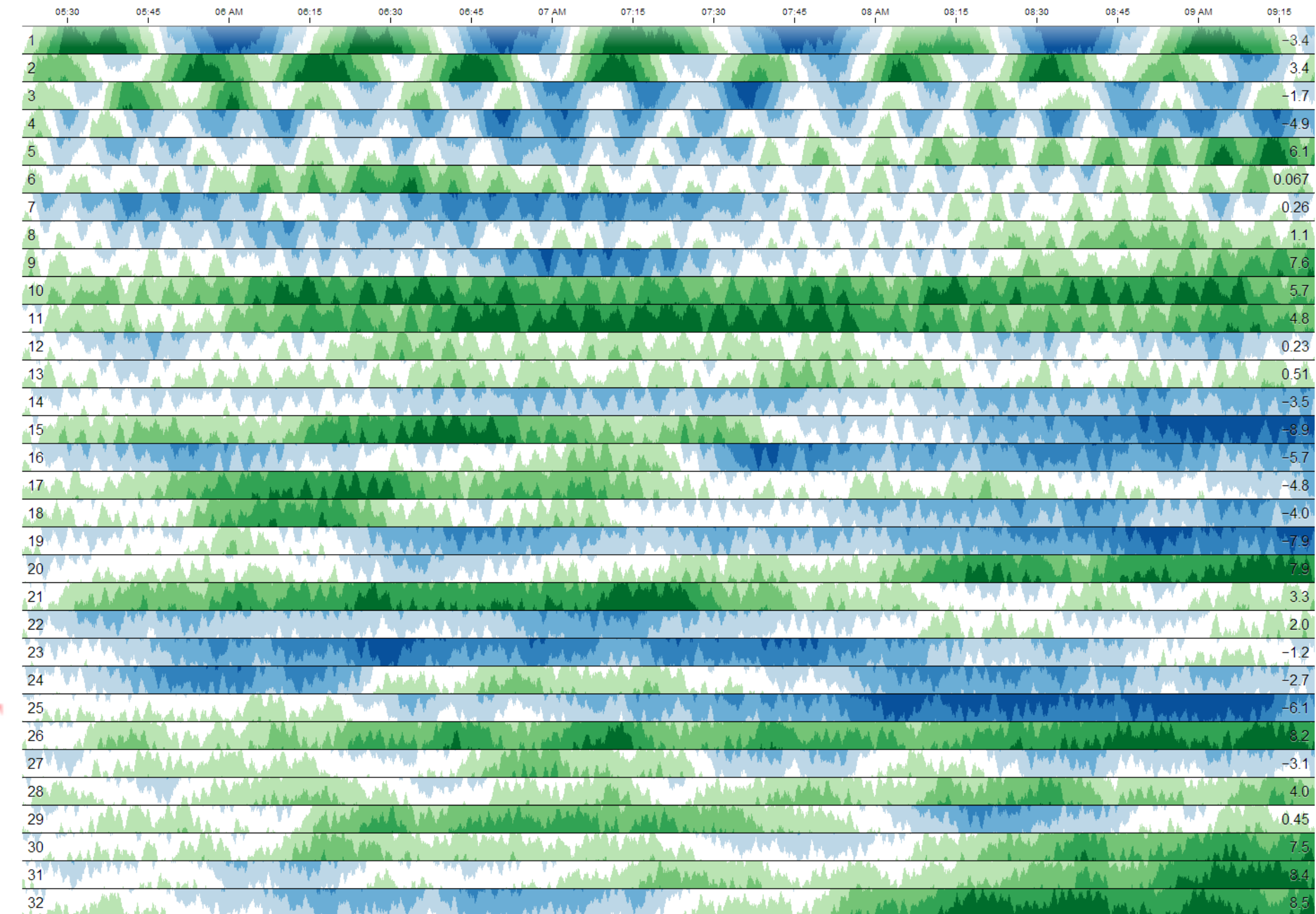
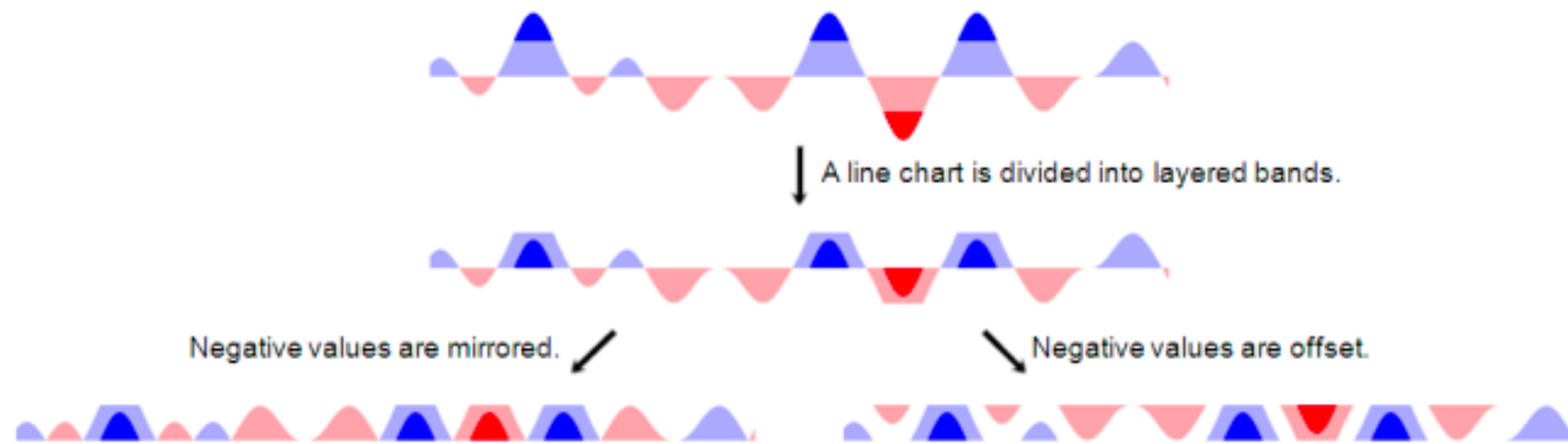
Mauricio Pochettino has lead Spurs on their best run **8TH**  **2ND** in 24 years of the Premier League

Alibaba stock is at 5 yr high **93.89**  **152.11** as of July 2017

The FTSE100 Brexit bounce **5562**  **7501** continues one year on from the vote last summer

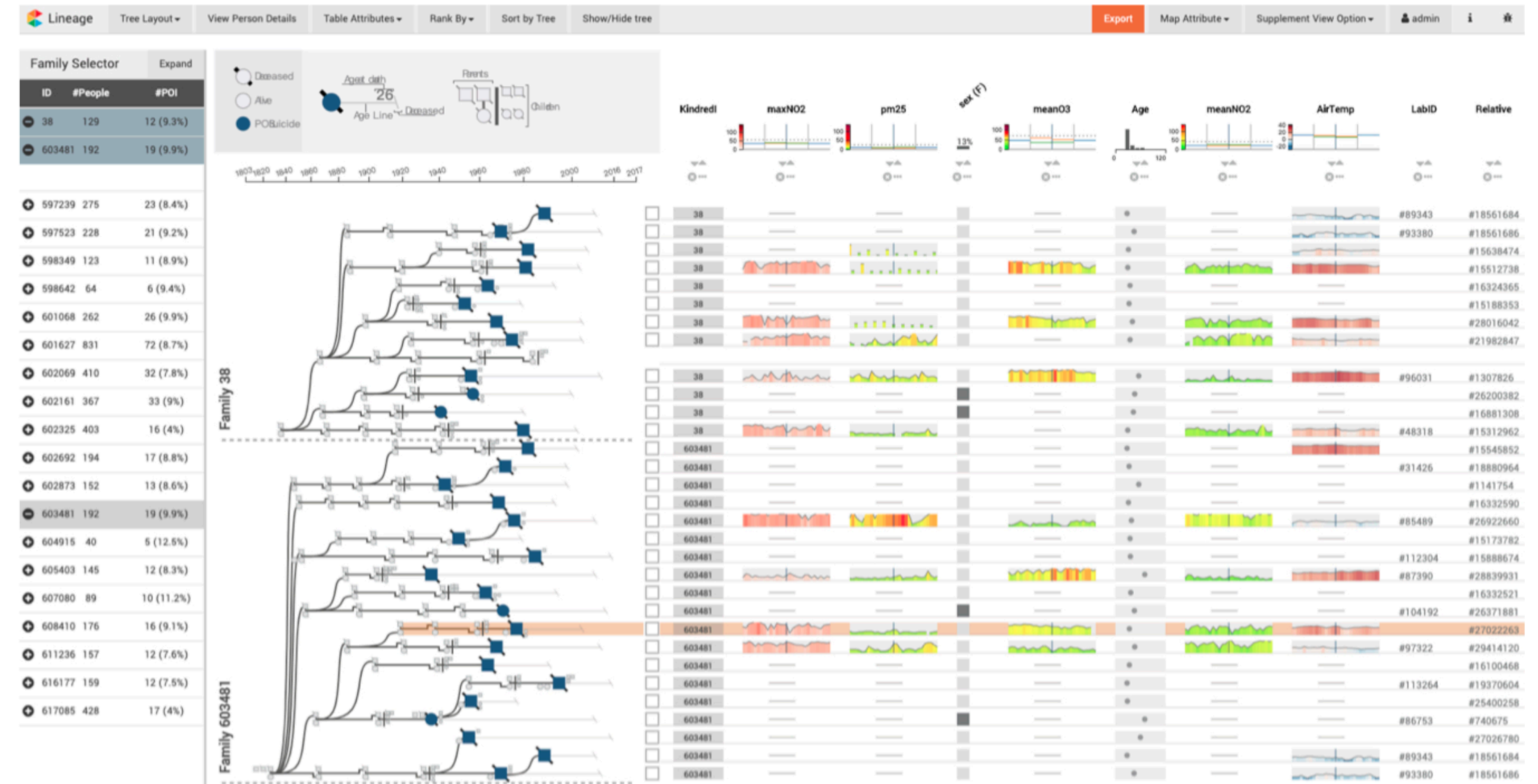
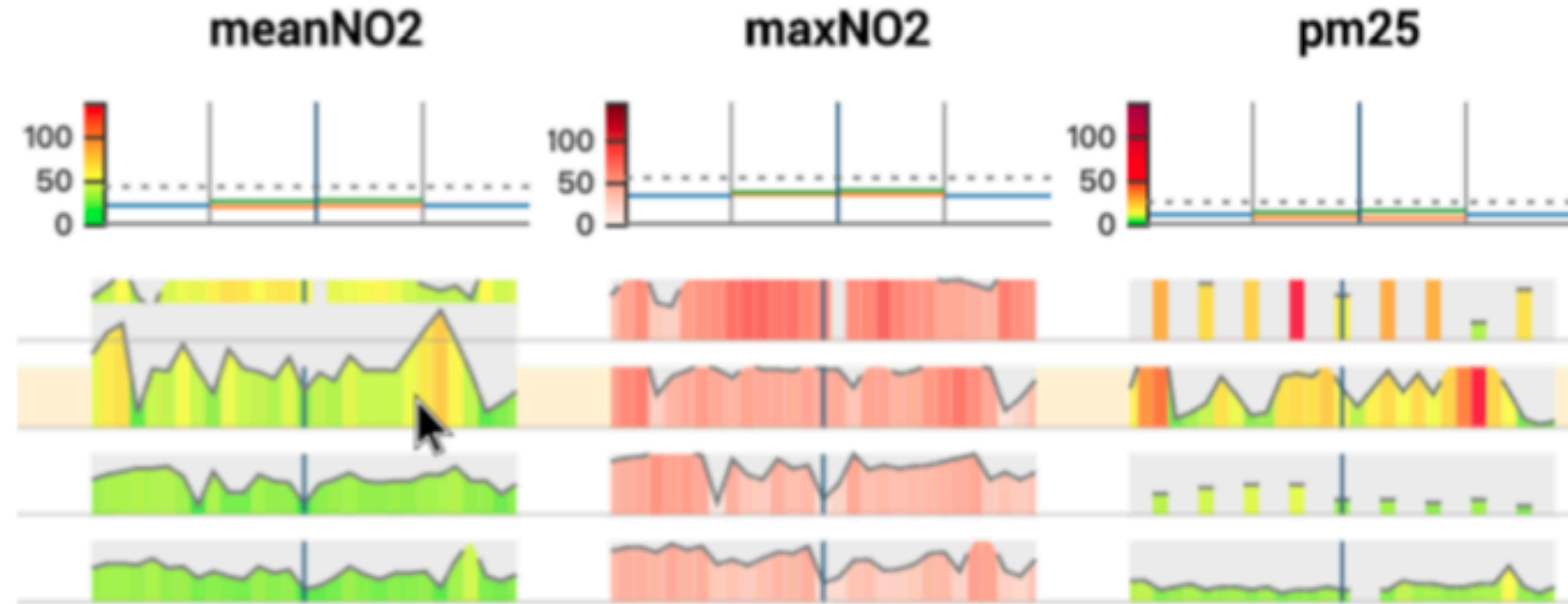
Symbol	Bid	Ask	Last	Change	T	Chart	Volume	High	Low	Value Change	Value	Gain
DELL	89 3/4	89 13/16	89 3/4	+ 1 1/4	↑		10,310,100	90 1/8	88 1/2	+1.41% 250	17,950	+273.72% 13,147
CPQ	48 7/16	48 9/16	48 7/16	- 13/16	↓		25,628,700	51 1/4	1/4	-1.65% -81	4,844	+60.79% 1,831
SDTI	26 1/4	26 3/8	26 3/8	+ 1/2	↓		504,600	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
RNVK	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 Graphs



<http://square.github.io/cubism/>

Clipped Graphs

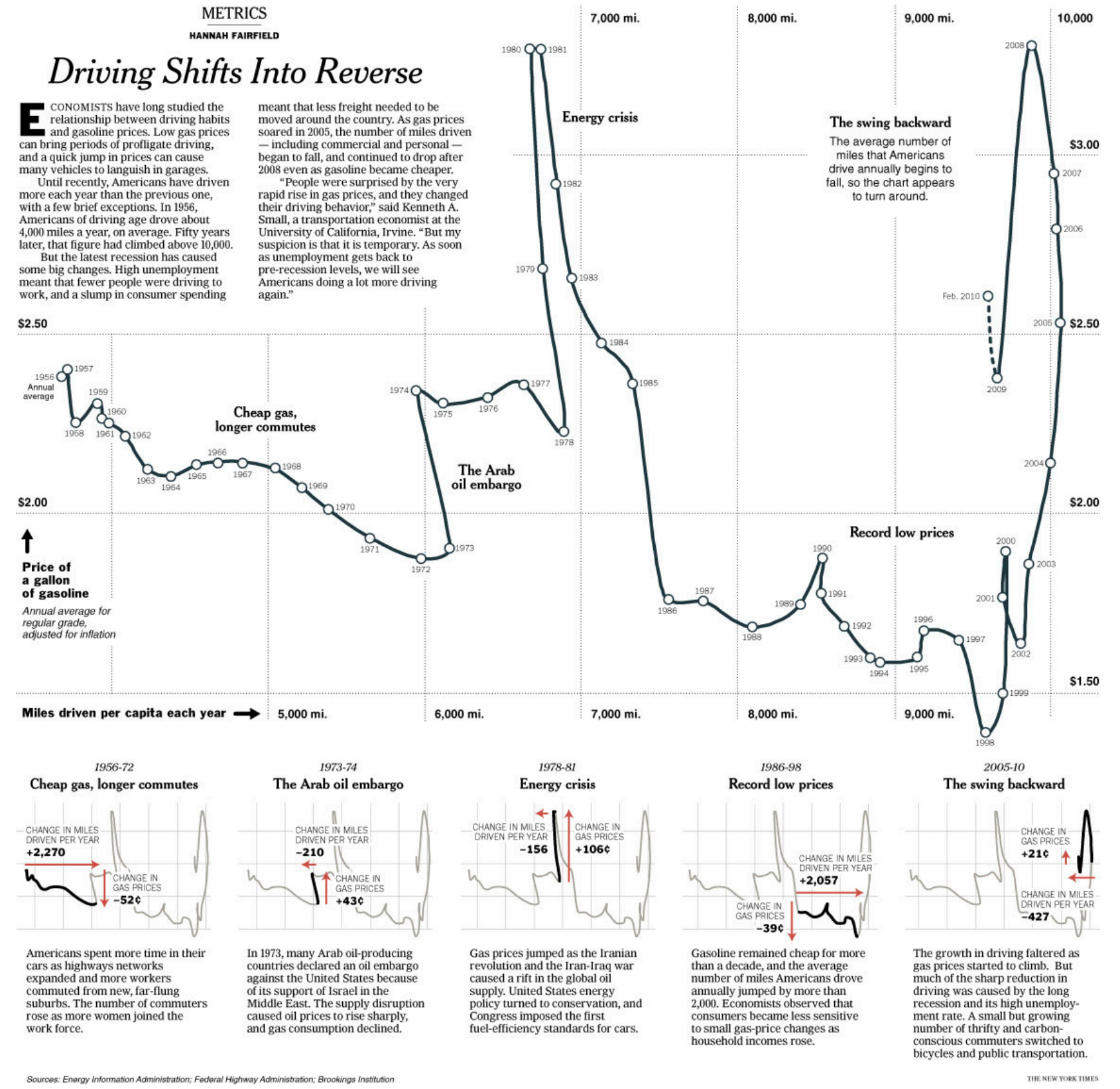


Connected Scatterplot

Two Variables + Time
 Only one per Chart!
 Labels important

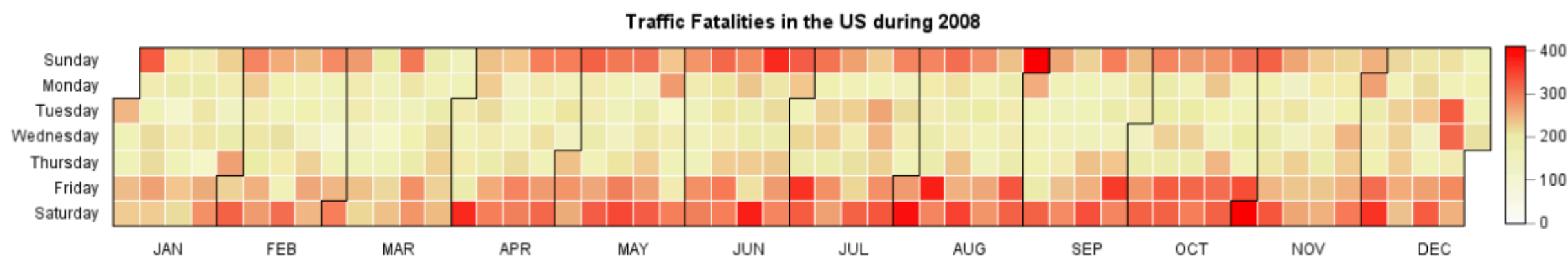
Connected scatterplot

A good way of showing changing data for two variables whenever there is a relatively clear pattern of progression.

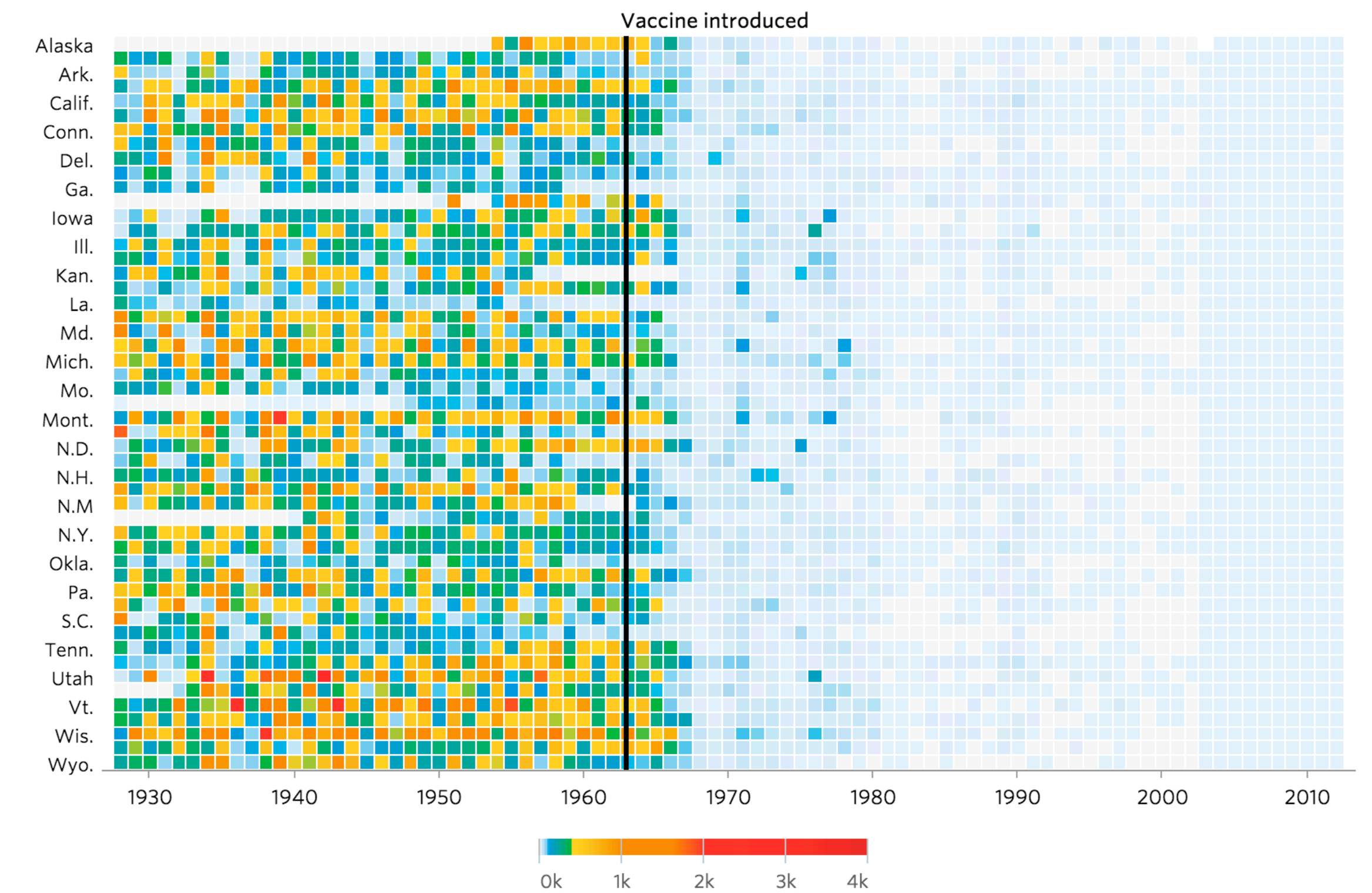


Heat Map and Calendar Heat Map

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

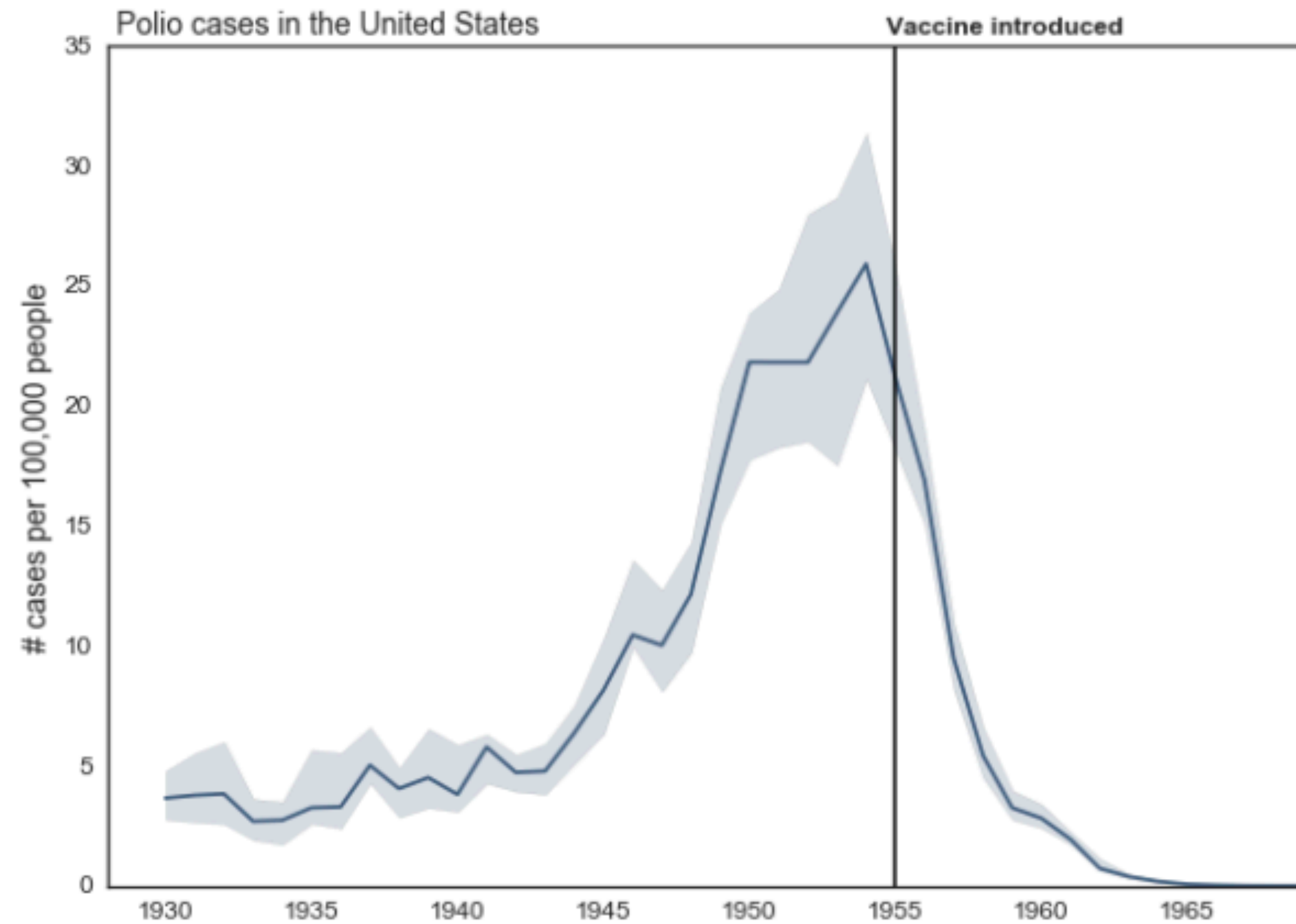


Measles



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.

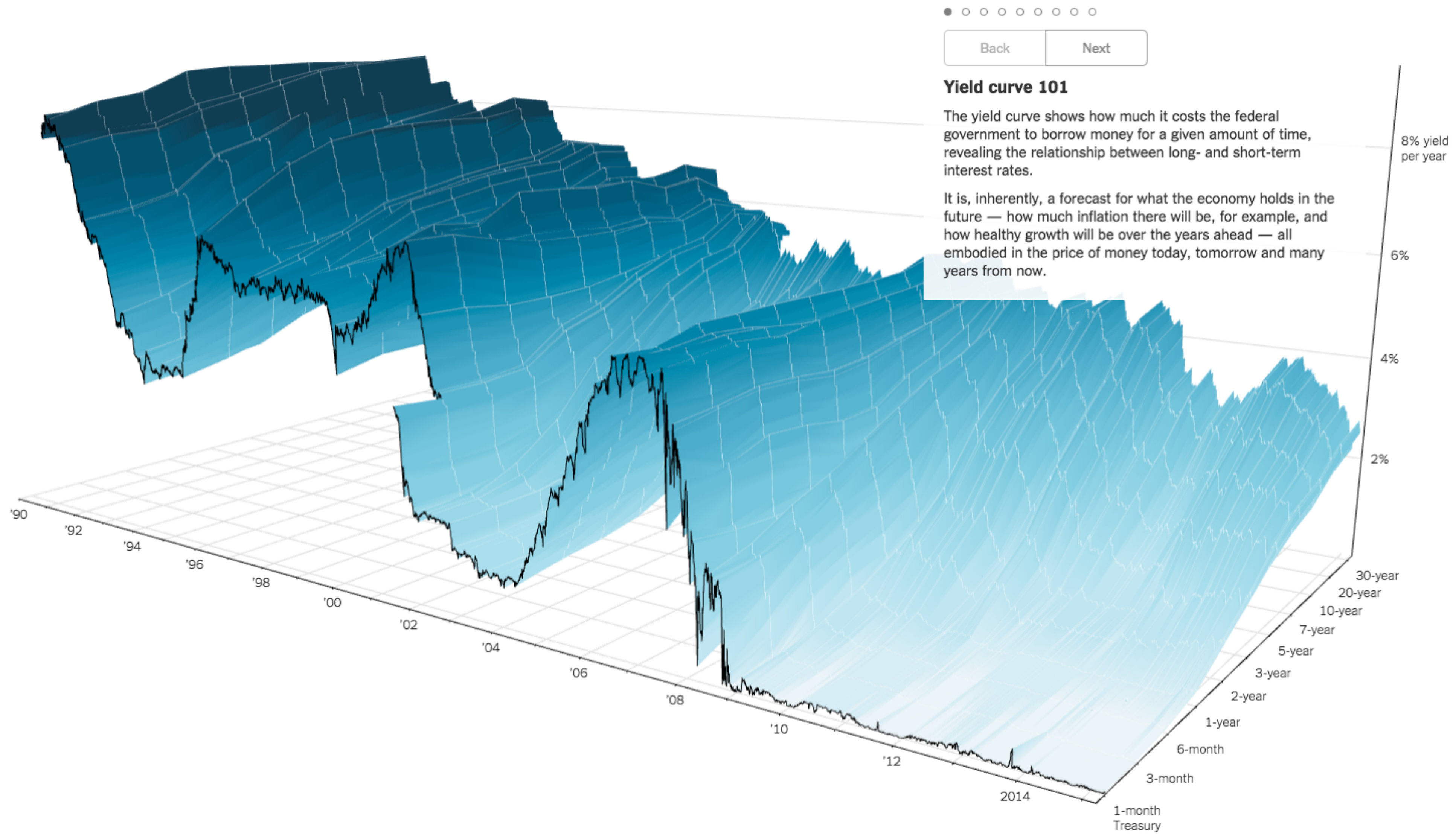
Sometimes you can Show Too Much Data



Data source: Project TYCHO (tycho.pitt.edu) | Author: Randy Olson (randalolson.com / @randal_olson)

<http://www.randalolson.com/2016/03/04/revisiting-the-vaccine-visualizations/>

Design Critique



Document: <https://goo.gl/W6w0il>
 Website: <http://goo.gl/D3mlsy>

Context / Critiques

<https://vimeo.com/127205447>

<https://community.jmp.com/t5/JMP-Blog/Graph-makeover-3-D-yield-curve-surface/ba-p/30573>

<http://www.visualisingdata.com/2015/03/when-3d-works/>

Ranking

Ranking Exercise

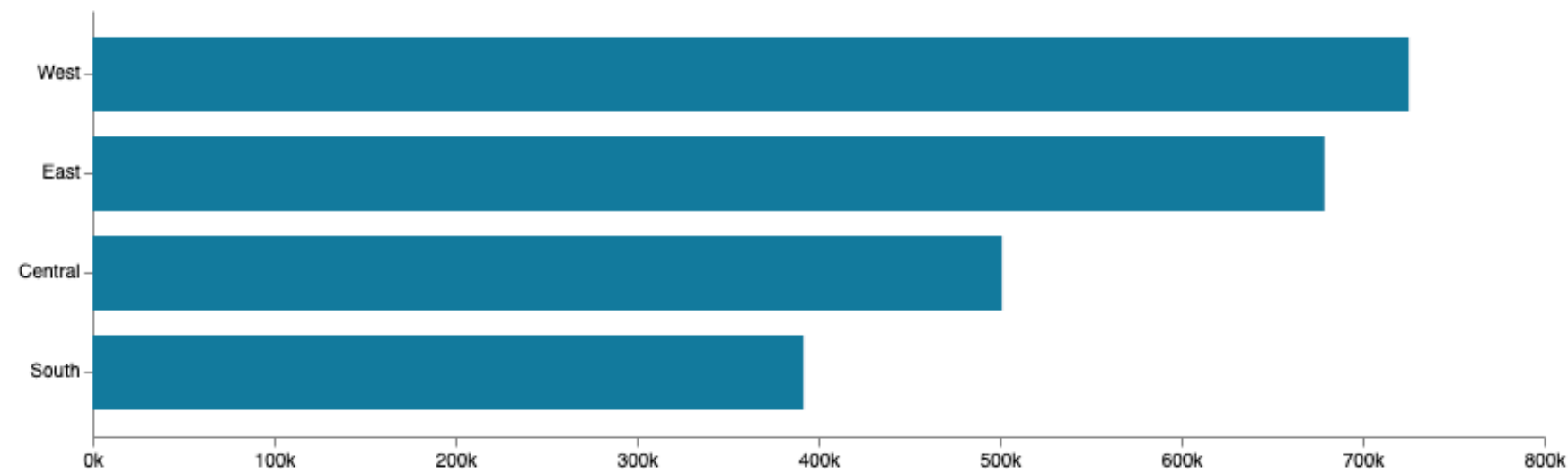
	1	2	3	4	5	6	7
Bavaria	8	6	2	4	2	1	3
Dortmund	1	1	5	2	3	8	8
Leipzig	2	2	1	1	1	2	4
Leverkusen	5	5	4	8	7	6	7
Moenchengladbach	10	7	8	7	6	5	1
Wolfsburg	6	4	3	5	8	7	2

Design a visualization showing the ranking of these football clubs over time.

Ranking

Ordered bar

Standard bar charts display the ranks of values much more easily when sorted into order



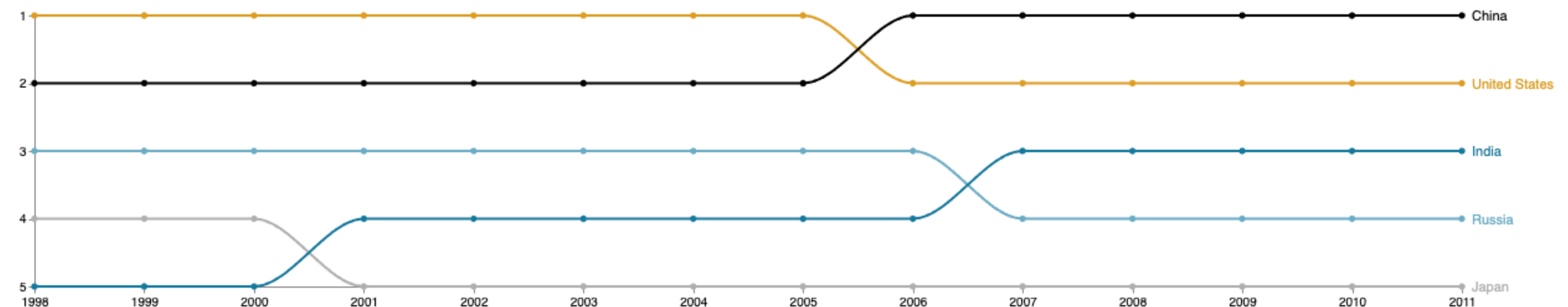
Edi

Magnitude Visualization + Sorting

Bump Charts for Rankings over Time

Bump

Effective for showing changing rankings across multiple dates. For large datasets, consider grouping lines using colour.



Edi

Temporal Rankings

		Pts ▾	W	D	L	Ho	Aw	GF	GA	GD
1	■ Paris SG	38	11	5	3	20	18	36	12	24
2	□ Lyon	13	11	5	3	25	13	33	17	16
3	■ Marseille	38	12	2	5	17	21	24	20	4
4	□ Rennes	32	10	2	7	16	16	29	24	5
5	■ Lorient	31	8	7	4	19	12	32	29	3
6	■ Valenciennes	29	8	5	6	21	8	31	24	7
7	□ Bordeaux	29	6	11	2	15	14	21	14	7
8	□ Lille	29	7	8	4	18	11	24	18	6
9	□ Nice	29	7	8	4	21	8	26	26	0
19	□ Troyes	13	2	7	10	11	2	20	37	-17
20	□ Nancy	11	1	8	10	7	4	15	33	-18

(b)

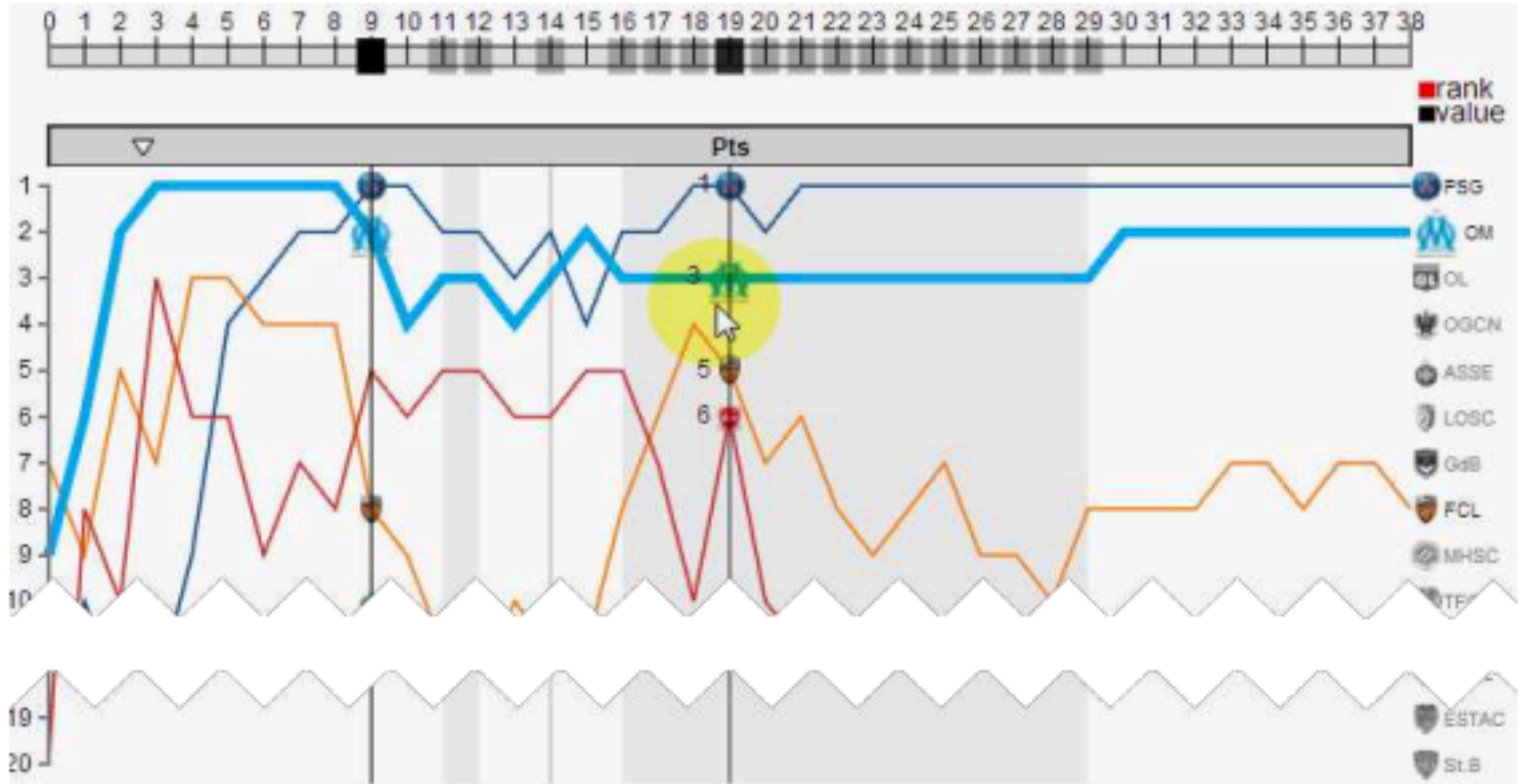
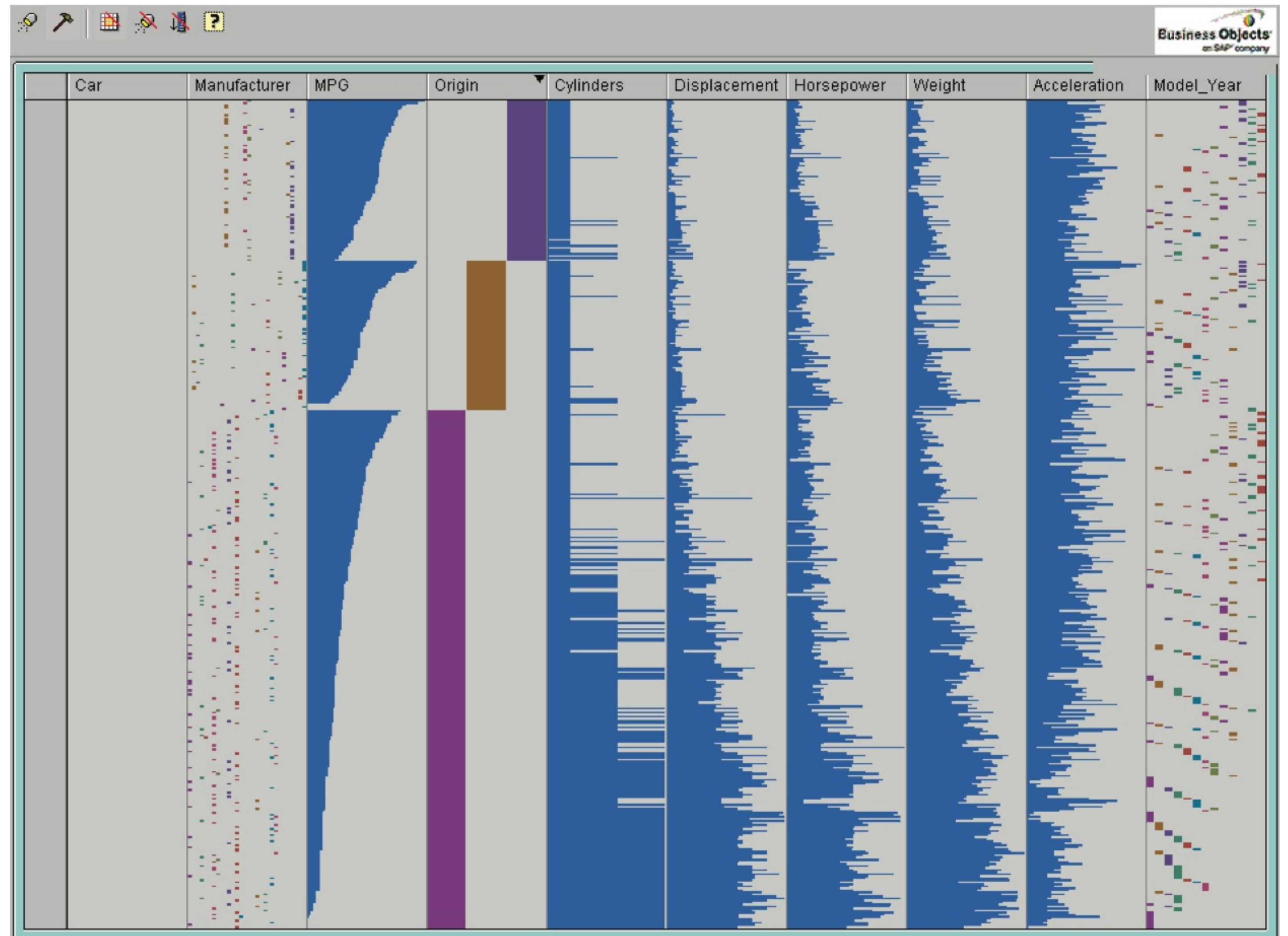
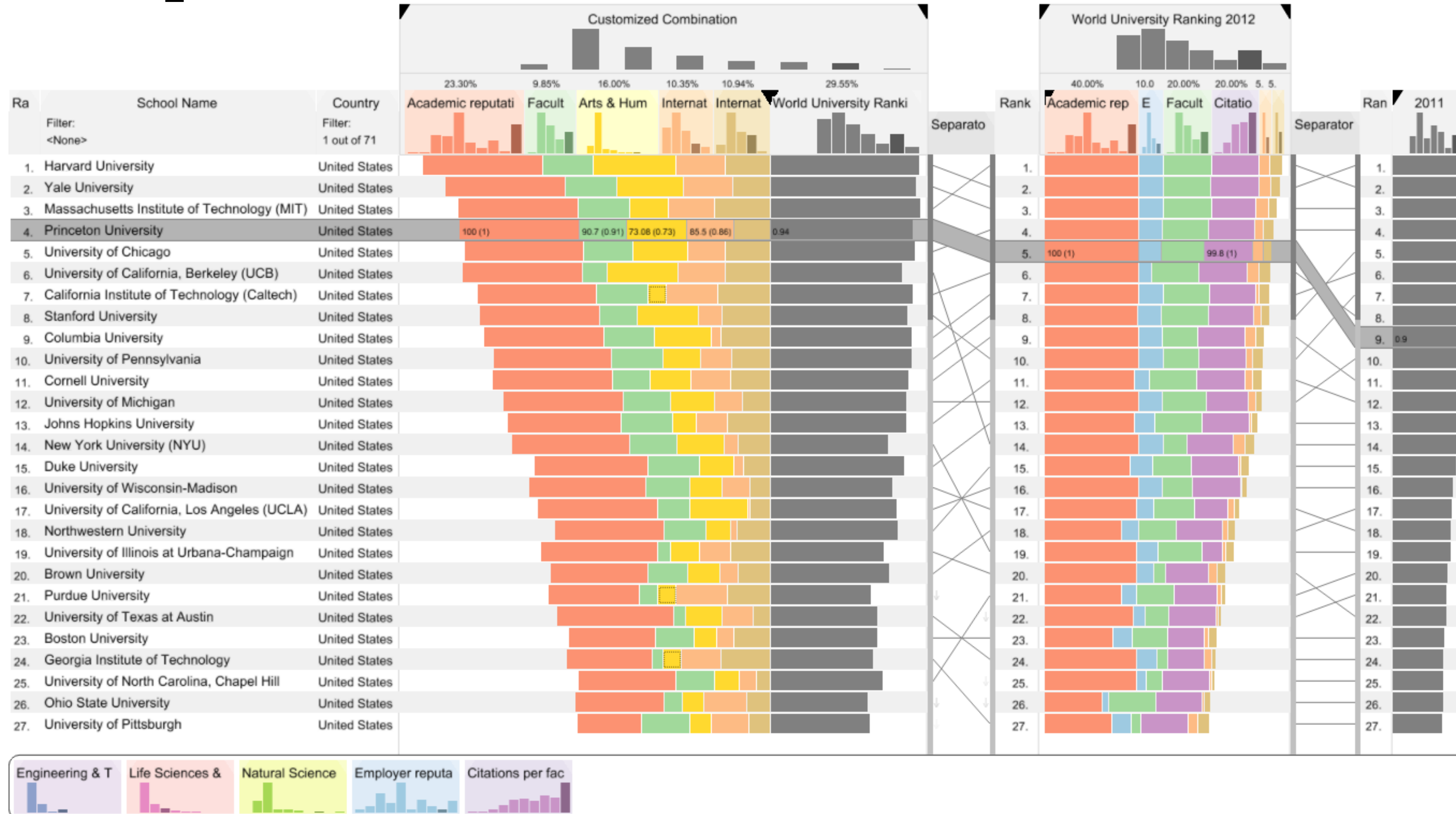


Table Lens

Interactive table-based representation



LineUp



Rankings are Popular

U.S. and Canada Box Office
Weekend of Aug. 30, 2013 - Sep. 01, 2013 | All-Time U.S. and Canada Box Office

This Wk	Last Wk	Title	Distributor	Weekend Gross	Cumulative Gross	Wks Out	# of Theaters
		EST MODUS IN REBUS	The Weinstein Company	\$20,201,300	\$79,466,400	3	3330
				\$18,472,900	\$18,472,900	1	2735

Journal Rankings
Subject Area: Computer Science, Year: 2012.

Title	SJR	H Index	Total Docs. (2012)	Total Docs. (3 years)	Total Refs.	Total Cites (3 years)	Citable Docs. (3 years)	Cites / Doc. (2 years)	Ref. / Doc.	Country
1 IEEE Transactions on Pattern Analysis and Machine Intelligence	8,094	200	196	579	8,249	5,724	543	9.43	42.09	USA
2 ACM Computing Surveys	6,751	81	12	65	4,962	894	10	8.72	130.58	USA
3 Foundations and Trends in Information Retrieval	6,536	12	38	10	664	151	64	7.14	221.33	USA
4 International Journal of Computer Vision	6,167	121	3	301	5,111	151	10	6.12	44.83	USA
5 Foundations and Trends in Networking	5,997	11	114	9	1,746	71	91	0.25	0.00	USA
6 Journal of the ACM	5,949	81	0	98	968	482	9	4.91	0.00	USA
7 Foundations and Trends in Computer Graphics and Information Theory	5,949	6	27	11	334	175	11	13.86	35.85	USA
8 ACM Transactions on Intelligent Systems and Vision	5,576	8	2	59	1,000	11	65	15.38	167.00	USA
9 Foundations and Trends in Communications and Information Theory	5,078	12	59	11	1,746	71	10	8.72	42.09	USA
10 Proceedings of the Annual ACM Symposium on Theory of Computing	4,919	130	2	11	2,329	175	9	7.14	130.58	USA
11 IEEE Communications Magazine	4,844	35	281	5	1,000	11	91	6.12	221.33	USA
12 IEEE Transactions on Evolutionary Computation	4,844	104	91	779	1,746	71	65	4.91	0.00	USA
13 IEEE Journal on Selected Areas in Communications	4,602	104	54	249	2,743	4,470	5	4.43	39.47	USA
14 IEEE Transactions on Information Theory	4,452	52	17	747	1,675	242	7,13	86.00	0.00	USA
15 IEEE Transactions on Fuzzy Systems	4,430	15	222	184	2,743	4,470	5	4.43	39.47	USA
16 IEEE Transactions on Database Systems	4,223	204	17	747	1,675	242	7,13	86.00	0.00	USA
17 IEEE Transactions on Signal Processing	4,131	176	235	504	2,251	667	326	17.874	14.933	USA
18 IEEE Transactions on Wireless Communications	4,128	106	504	307	14,944	2,536	38	2.160	2.160	USA
19 IEEE Transactions on Software Engineering	3,851	52	77	471	17,874	326	631	7.99	30.14	USA
20 IEEE Transactions on Computational Mathematics and Bioinformatics	3,715	151	22	307	14,944	2,536	38	2.160	2.160	USA
21 IEEE Transactions on Information Theory	3,648	104	601	81	3,315	6,037	1,471	302	78	USA
22 IEEE Transactions on Database Systems	3,426	25	454	29	1,612	18,136	266	1,486	1,583	USA
23 Foundations of Computational Mathematics	3,393	100	85	164	11,615	7,233	354	1,486	1,583	USA
24 IEEE Transactions on Software Engineering					1,046	7,327	866	75	5,83	USA

QS World University Rankings - 2012/2013

Rank	School Name	Country	QS Stars Rating	Overall	Academic Reputation	Employment
1	Massachusetts Institute of Technology (MIT)	United States	5 stars	100.00	100	100
2	University of Cambridge	United Kingdom	5 stars	99.78	100	100
3	Harvard University	United States	5 stars	99.15	100	100
4	UCL (University College London)	United Kingdom	5 stars	99.15	100	100
5	University of Oxford	United Kingdom	5 stars	99.15	100	100
6	Imperial College London	United Kingdom	5 stars	99.15	100	100
7	Yale University	United States	5 stars	99.15	100	100
8	University of Chicago	United States	5 stars	99.15	100	100
9	Princeton University	United States	5 stars	99.15	100	100
10	California Institute of Technology	United States	5 stars	99.15	100	100



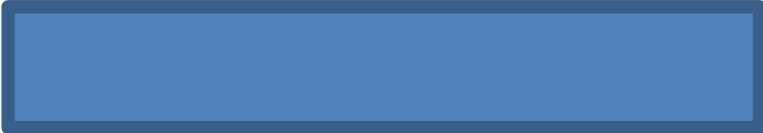


Ranking.com

Web Rank	Domain	TrustGauge	Company Name
1	google.com	10	Google
2	youtube.com	7	YouTube, Inc.
3	facebook.com	10	Facebook
4	yahoo.com	8	Yahoo! Inc.
5	wikipedia.org	10	Wikimedia Foundation
6	amazon.com	7	Amazon.com
7	live.com	7	N/A
8	msn.com	7	Microsoft Corp
9	bing.com	10	N/A
10	blogspot.com	6	PBPATEL
11	microsoft.com	7	Microsoft Corporation
12	ebay.com	7	eBay, Inc.
13	ask.com	10	Ask Jeeves, Inc.
14	twitter.com	7	Twitter, Inc.
15	wordpress.com	8	JDC Global Ventures LLC
16	linkedin.com	7	N/A
17	craigslist.org	7	CRAIGSLIST, INC
18	tumblr.com	6	N/A
19	aol.com	7	N/A
20	adobe.com	7	Adobe Systems Incorporated
21	huffingtonpost.com	7	N/A
22	go.com	7	N/A
23	cnn.com	7	N/A
24	about.com	7	N/A
25	ehow.com	7	N/A

Things to do in Atlanta

Atlanta Botanical Garden
Ranked #2 of 177 attractions in Atlanta
"Sweet Garden" 09/03/2013
"One of the coolest things in Atlanta" 09/03/2013
Category: Gardens
Owner description: Step into a world of magic and serenity at the Atlanta Botanical Garden... more »

Harvard University (United States) 93.7
Massachusetts Institute of Technology (United States) 93.7
Princeton University (United States) 93.1
University of Cambridge (United Kingdom) 92.7
Imperial College London (United Kingdom) 90.6
University of California, Berkeley (United States) 90.5
Stanford University (United States) 90.4

Rank	University	Score
1.	MIT	
2.	Harvard	
3.	Princeton	
4.	Cambridge	
5.	Oxford	

Support Multiple
Attributes

Combiner functions: $f(A, B, C)$

(Weighted) sum

$$\text{Score} = w_a A + w_b B + w_c C$$

→ Serial

Maximum

$$\text{Score} = \max(A, B, C)$$

→ Parallel

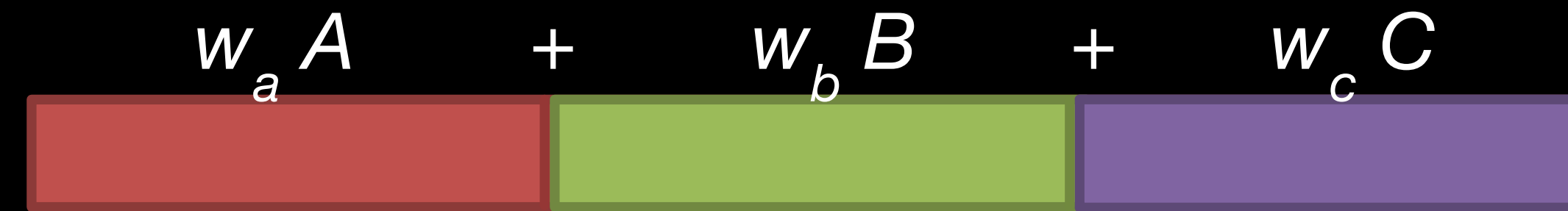
Product

Nesting

...

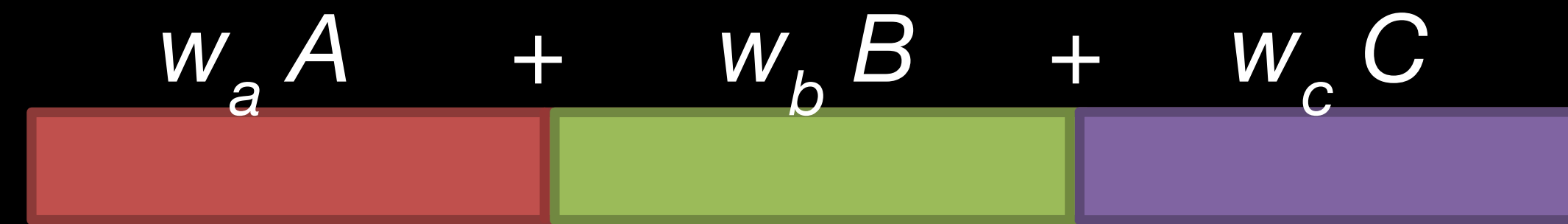
→ Complex
Combiners

Serial Combiner (as Stacked Bar)



Rank	University	A	B	C
1.	MIT			
2.	Harvard			
3.	Princeton			
4.	Cambridge			
5.	Oxford			

Serial Combiner (as Stacked Bar)



Rank	University	A	B	C
1.	MIT			
2.	Harvard			
3.	Princeton			
4.	Cambridge			
5.	Oxford			

Serial Combiner (as Stacked Bar)



Rank	University	A	B	C
1.	MIT	Large red segment	Medium green segment	Medium purple segment
2.	Harvard	Large red segment	Medium green segment	Medium purple segment
3.	Princeton	Medium red segment	Medium green segment	Medium purple segment
4.	Cambridge	Large red segment	Small green segment	Small purple segment
5.	Oxford	Large red segment	Very small green segment	Small purple segment

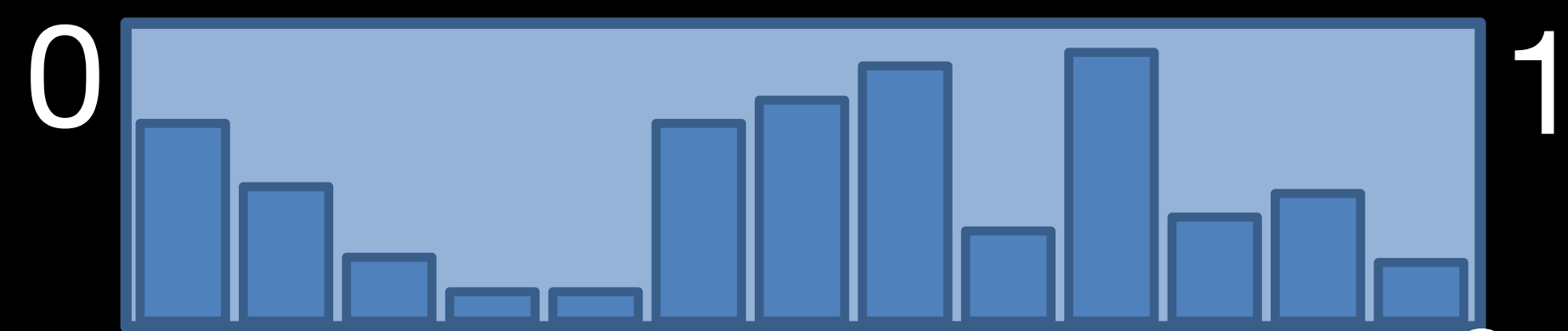
Rank	School Name	Country	Faculty/student ratio	Employer reputation	Citations per faculty
1.	American University	United States			
2.	Arizona State University	United States			
3.	Aston University	United Kingdom			
4.	Birkbeck College, University of L	United Kingdom			
5.	Boston College	United States			
6.	Boston University	United States			
7.	Brandeis University	United States			
8.	Brown University	United States			
9.	Brunel University	United Kingdom			
10.	California Institute of Technology	United States			
11.	Cardiff University	United Kingdom			
12.	Case Western Reserve University	United States			
13.	City University London	United Kingdom			
14.	College of William & Mary	United States			
15.	Colorado State University	United States			
16.	Columbia University	United States			
17.	Cornell University	United States			
18.	Cranfield University	United Kingdom			
19.	Dartmouth College	United States			
20.	Drexel University	United States			
21.	Duke University	United States			
22.	Durham University	United Kingdom			

Filter:
<None>

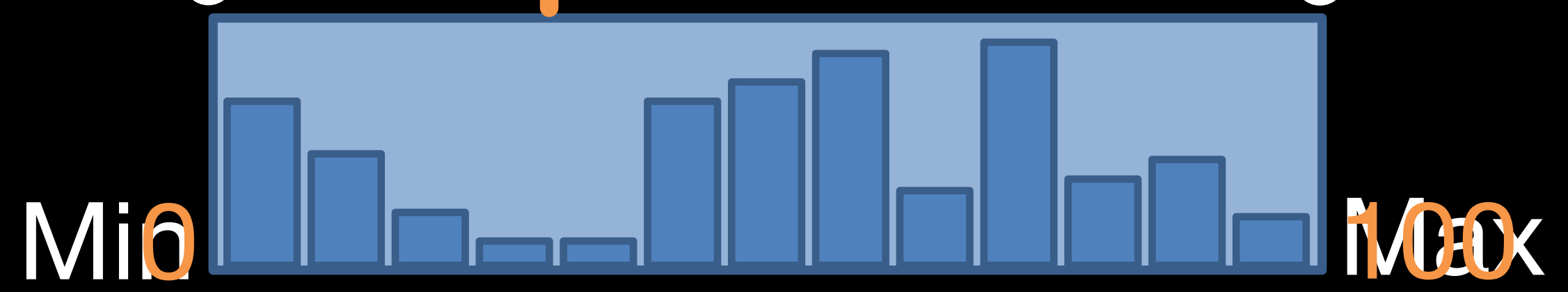
Filter:
2 out of 72

Flexible Mapping of Attributes to Scores

Transformed



Input

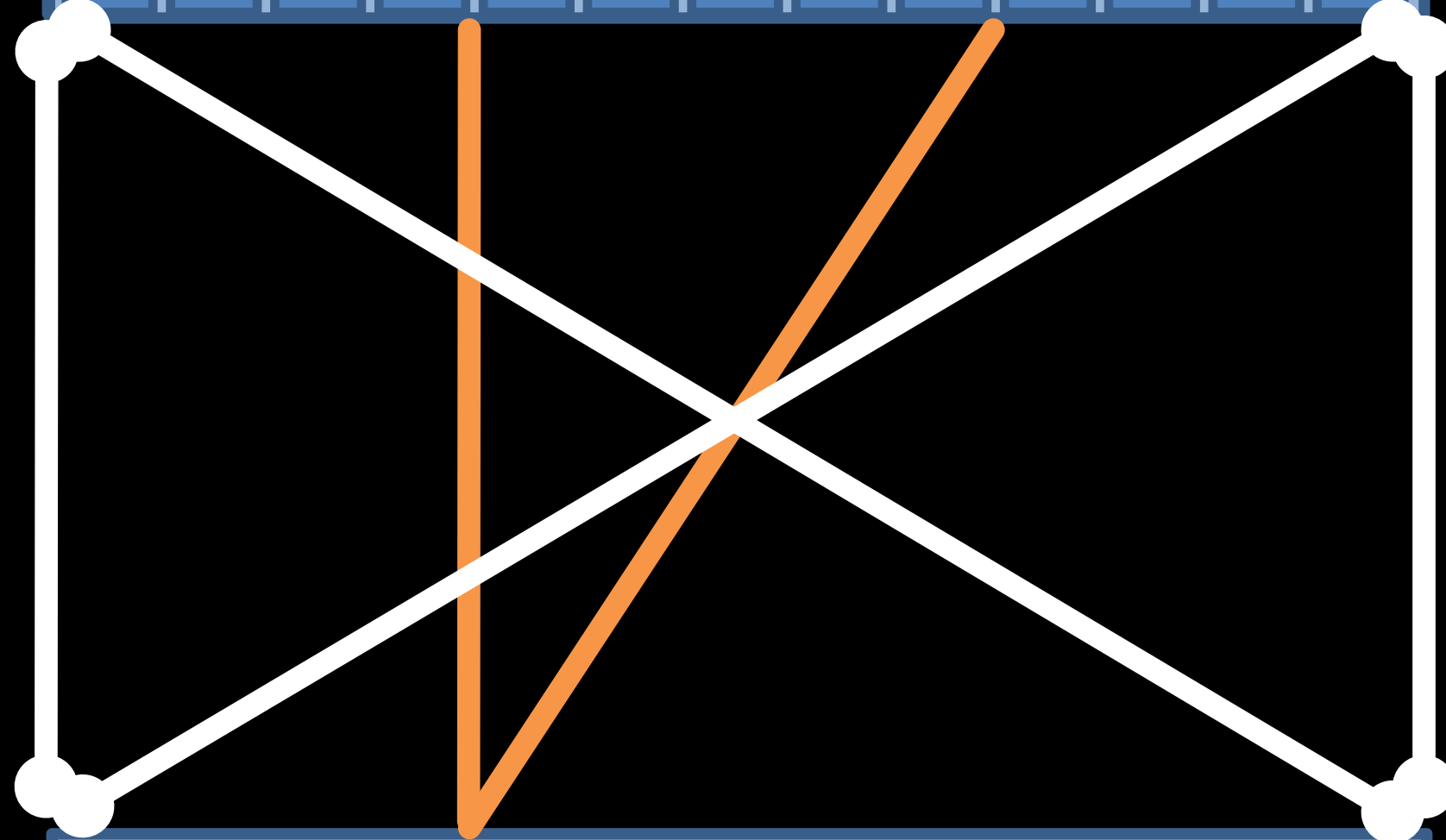
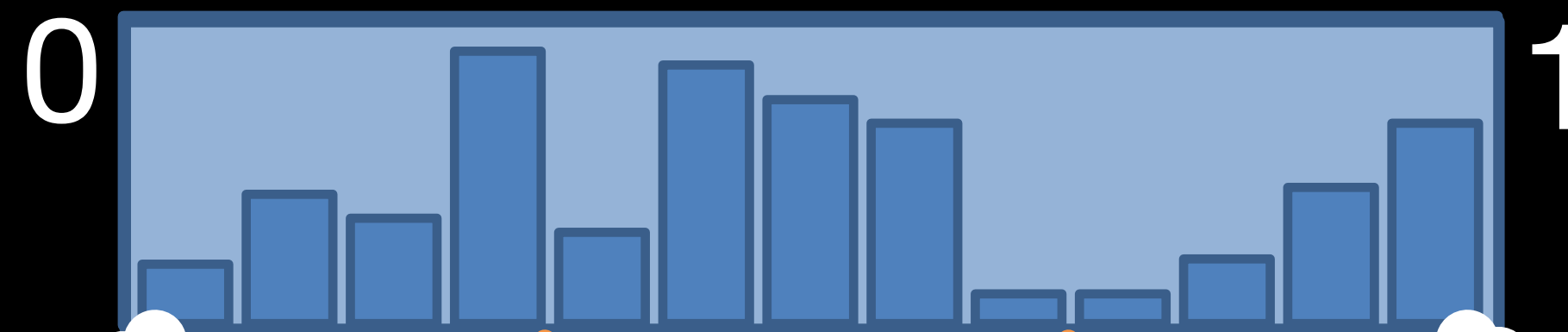


Min

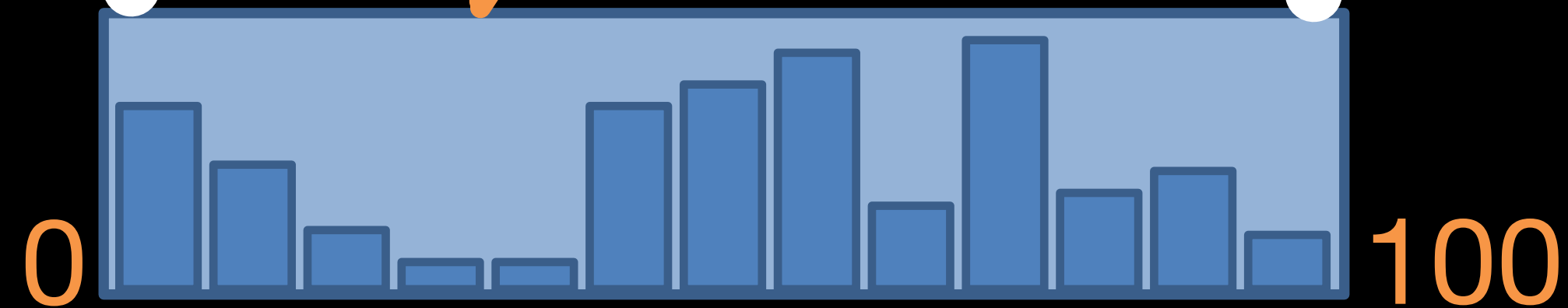
Max

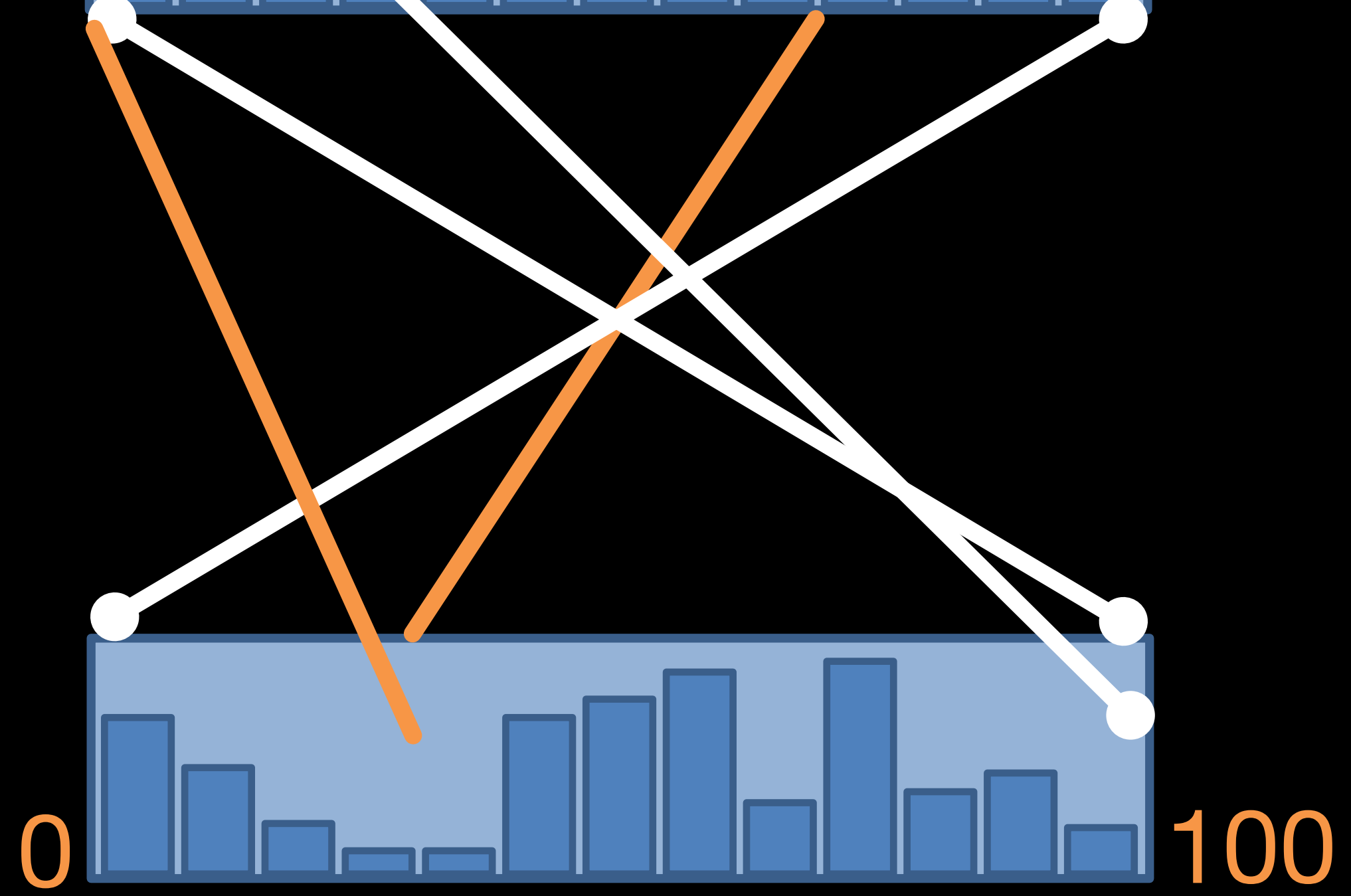
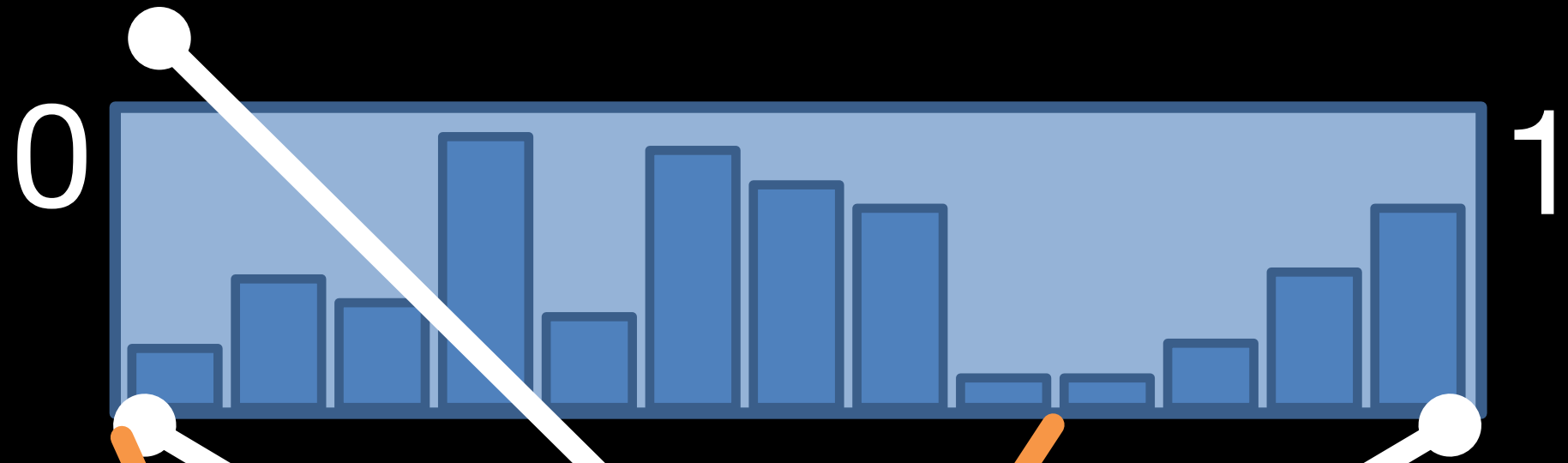


Transformed



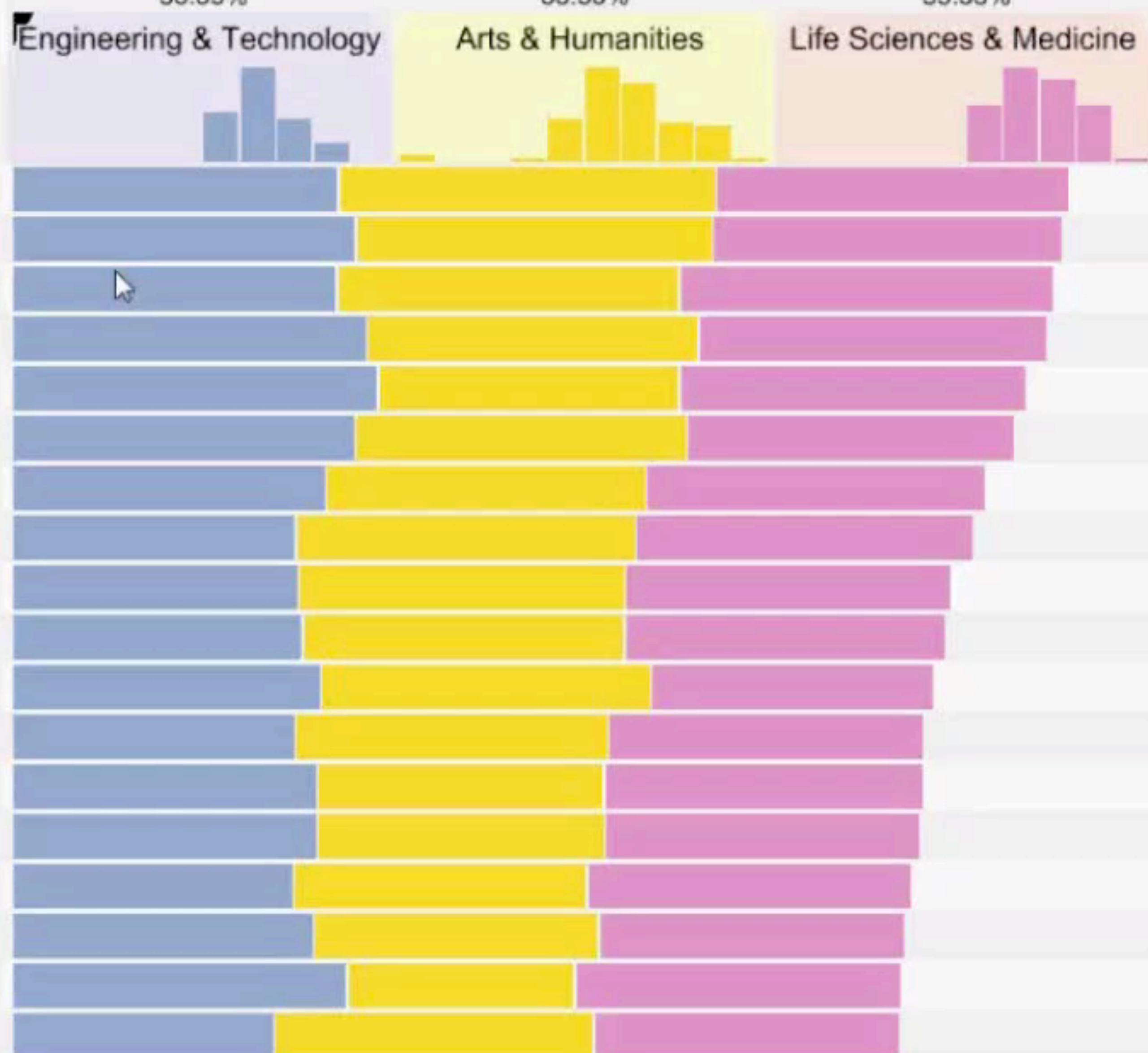
Input





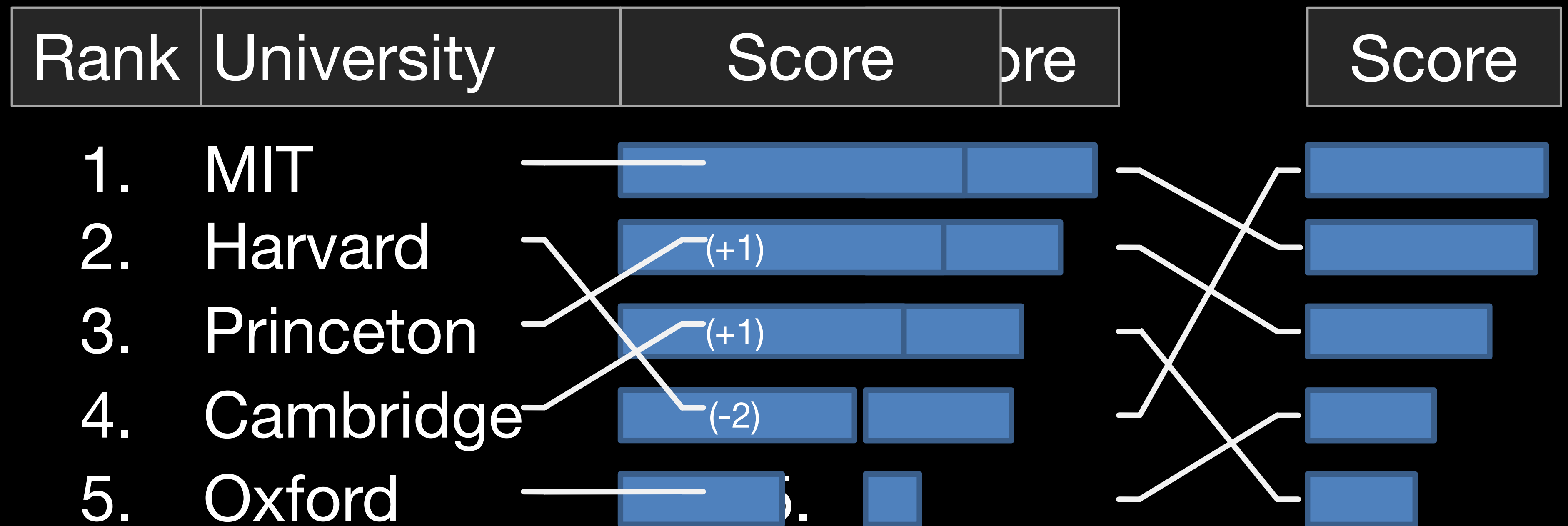
SUM (Engineering & Technology, Arts & Humanities, Life Sciences & Medicine)

Ran	School Name	Country
Filter:		Filter:
<None>		2 out of 43
1.	University of Oxford	United Kingdom
2.	University of Cambridge	United Kingdom
3.	Harvard University	United States
4.	Stanford University	United States
5.	Massachusetts Institute of Technology (MIT)	United States
6.	University of California, Berkeley (UCB)	United States
7.	University of California, Los Angeles (UCLA)	United States
8.	Yale University	United States
9.	UCL (University College London)	United Kingdom
10.	Columbia University	United States
11.	Princeton University	United States
12.	University of Edinburgh	United Kingdom
13.	University of Michigan	United States
14.	Cornell University	United States
15.	University of Pennsylvania	United States
16.	The University of Manchester	United Kingdom
17.	Imperial College London	United Kingdom
18.	University of Chicago	United States

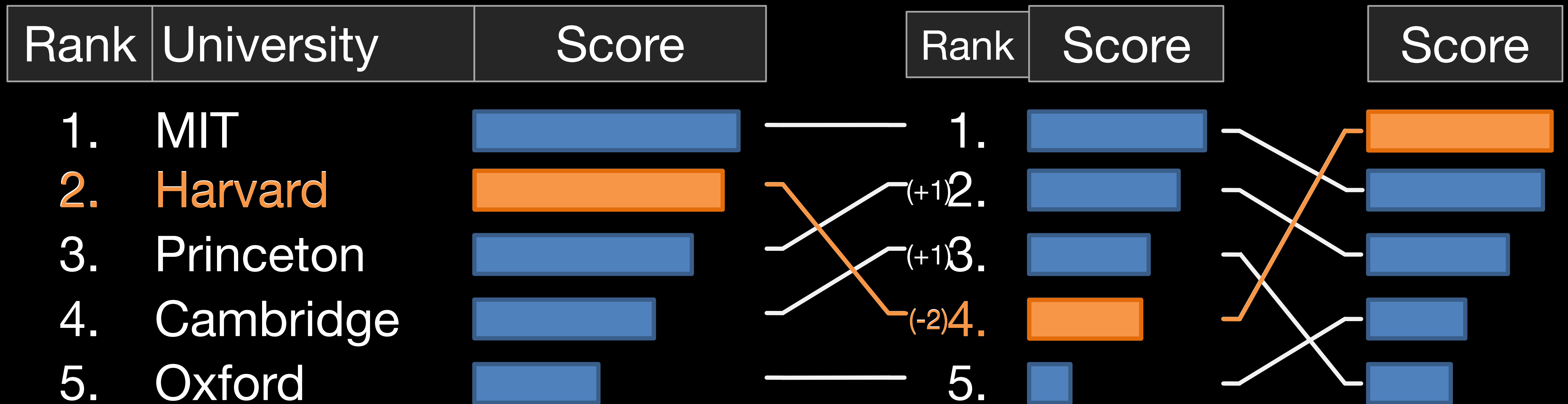


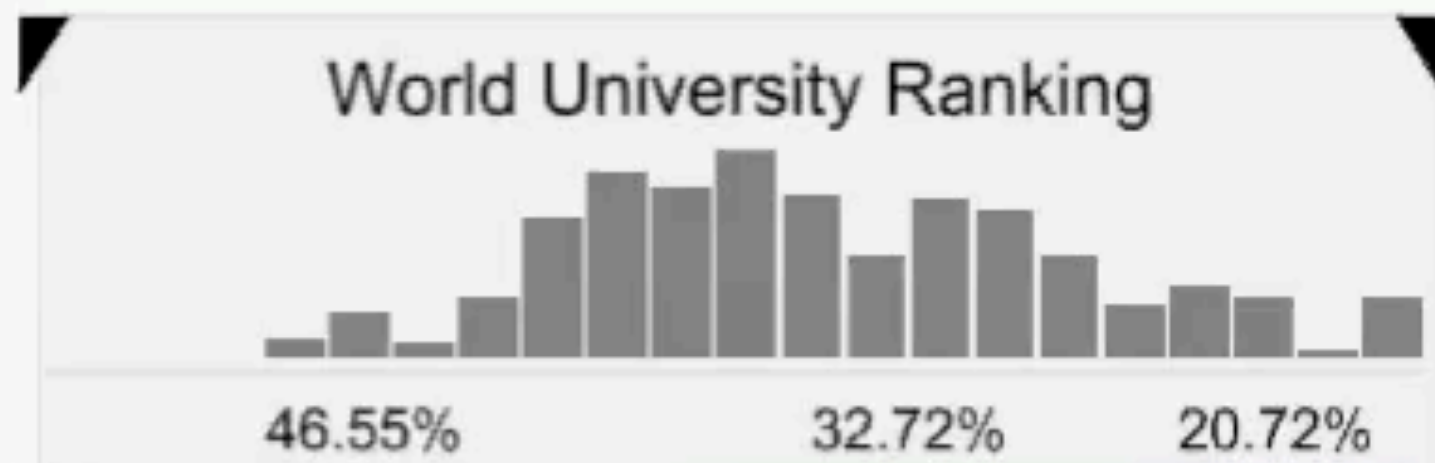
Compare Rankings

Bump Charts

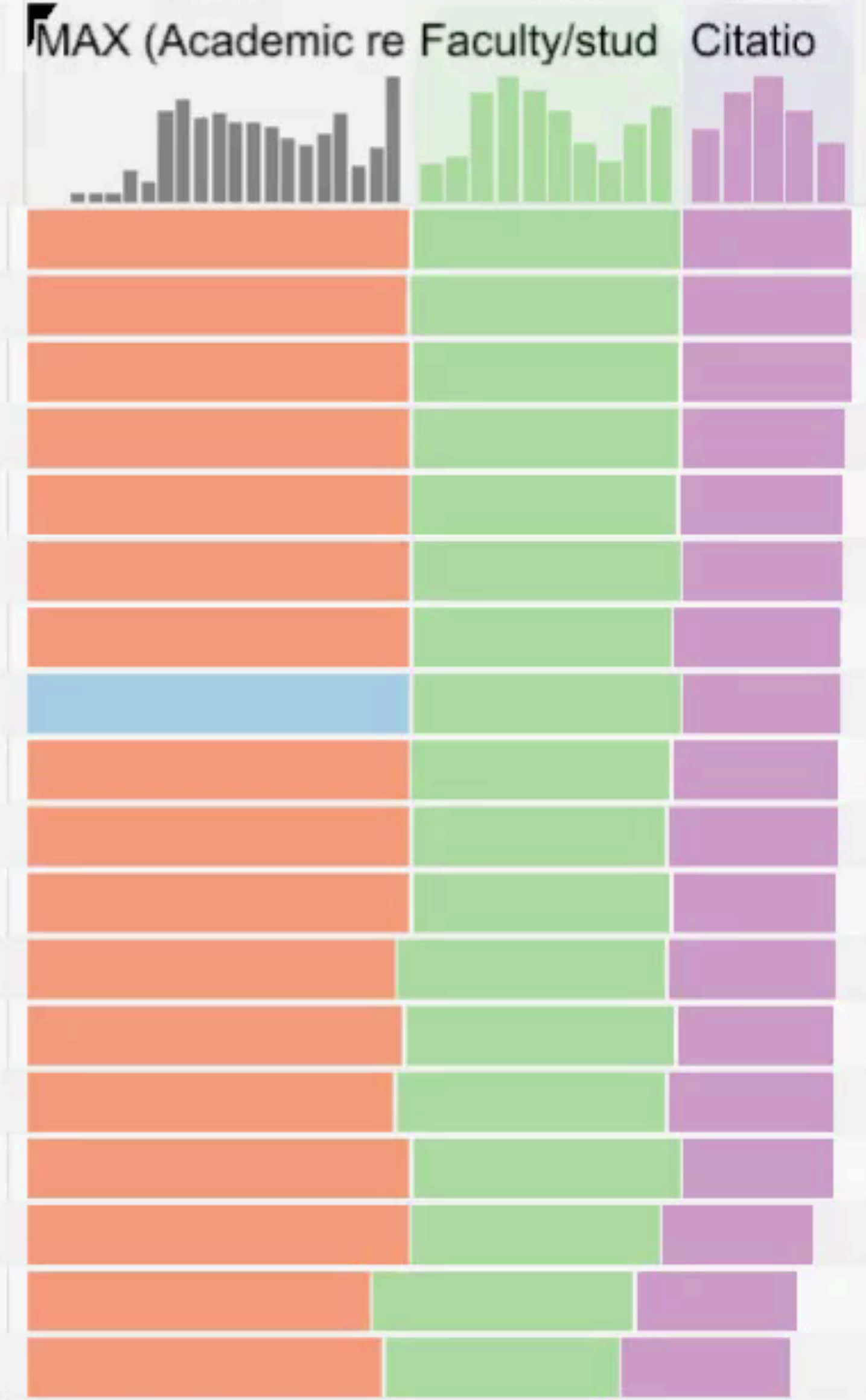


Bump Charts





Rank	School Name
Filter:	<None>
1.	Massachusetts Institute of Te
2.	California Institute of Technol
3.	Harvard University
4.	University of Cambridge
5.	UCL (University College Lond
6.	University of Oxford
7.	Princeton University
8.	Imperial College London
9.	University of Chicago
10.	Stanford University
11.	Columbia University
12.	Duke University
13.	University of Pennsylvania
14.	Johns Hopkins University
15.	Yale University
16.	University of Michigan
17.	Ecole normale supérieure, Pa
18.	Northwestern University

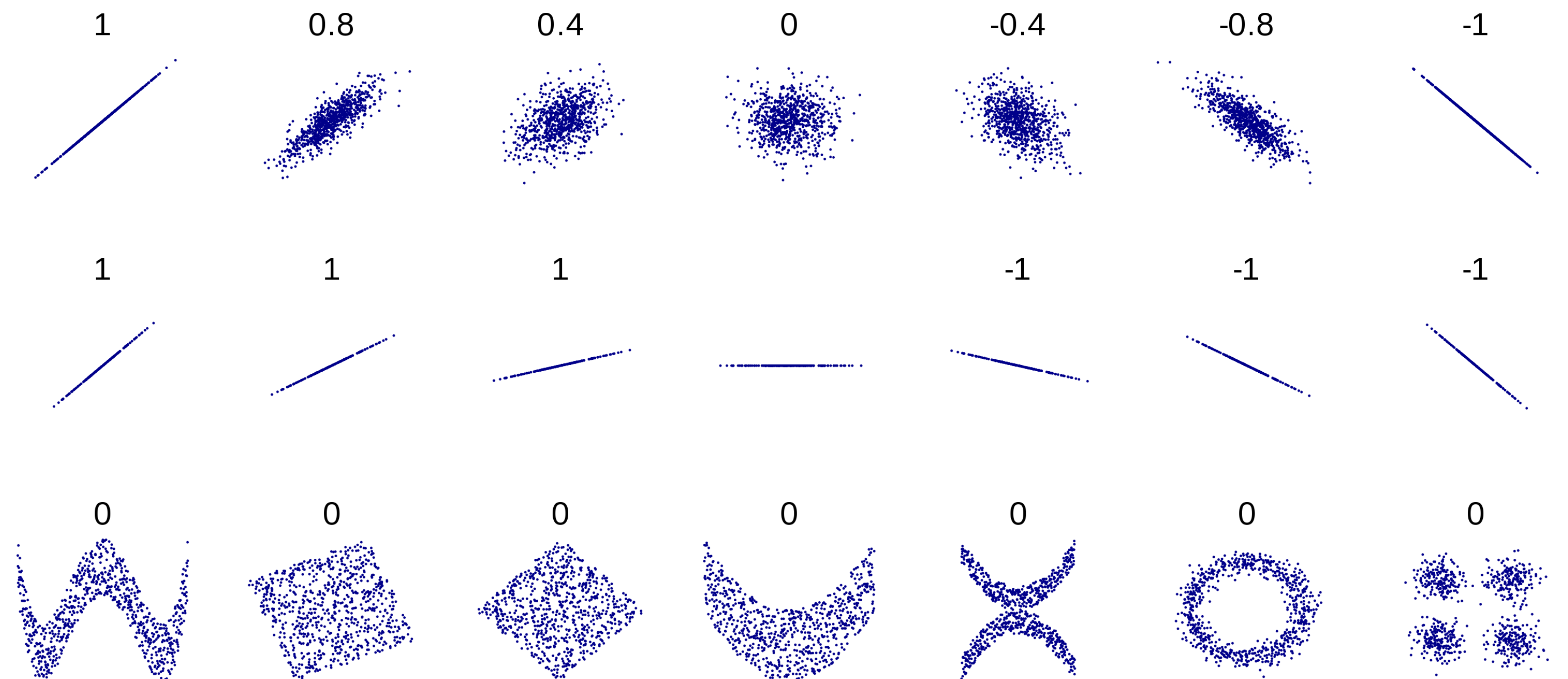


<https://lineup.js.org/>

Correlation

What is Correlation

How do two or more variables behave relative to each other?

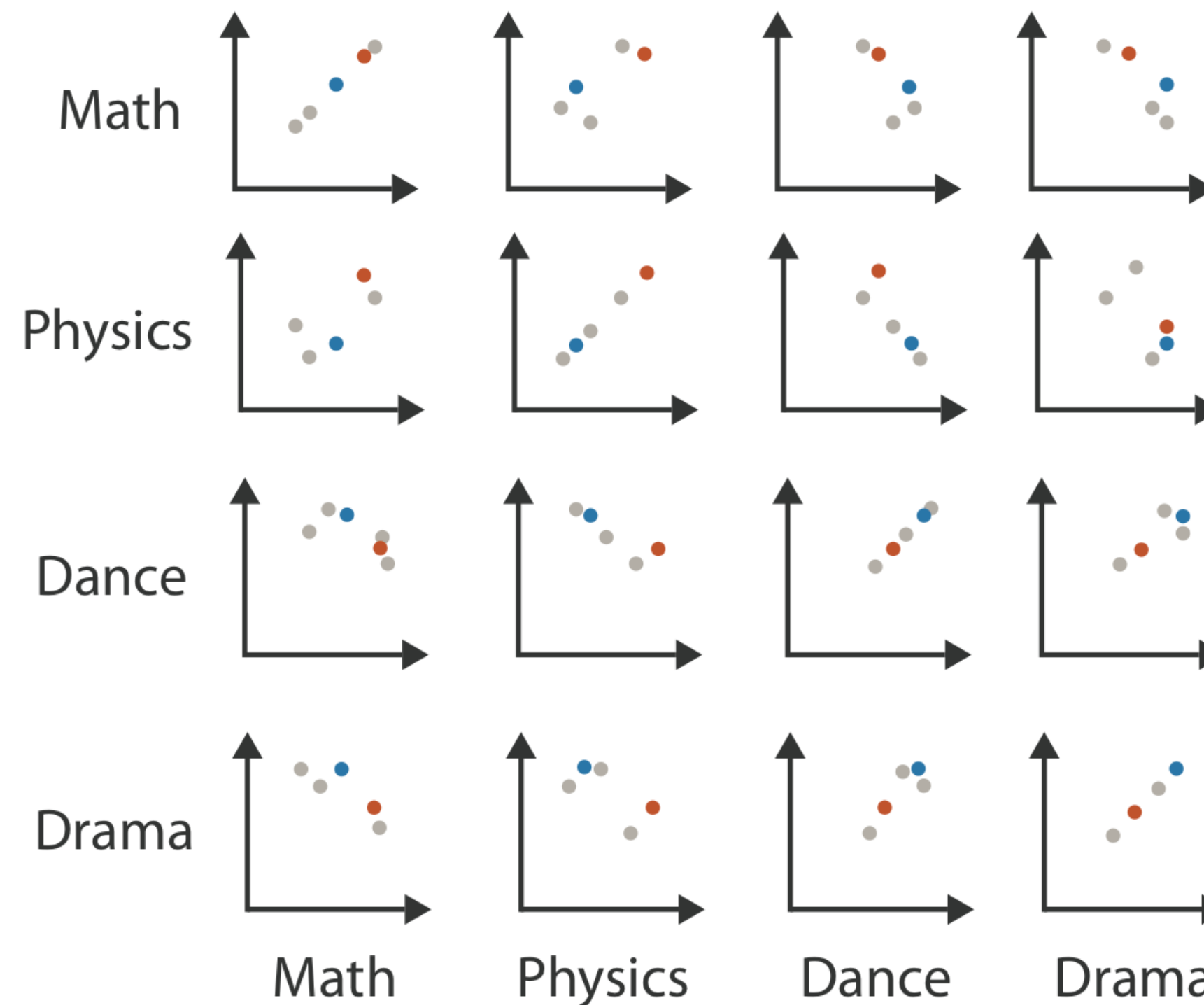


Axis-Based Techniques

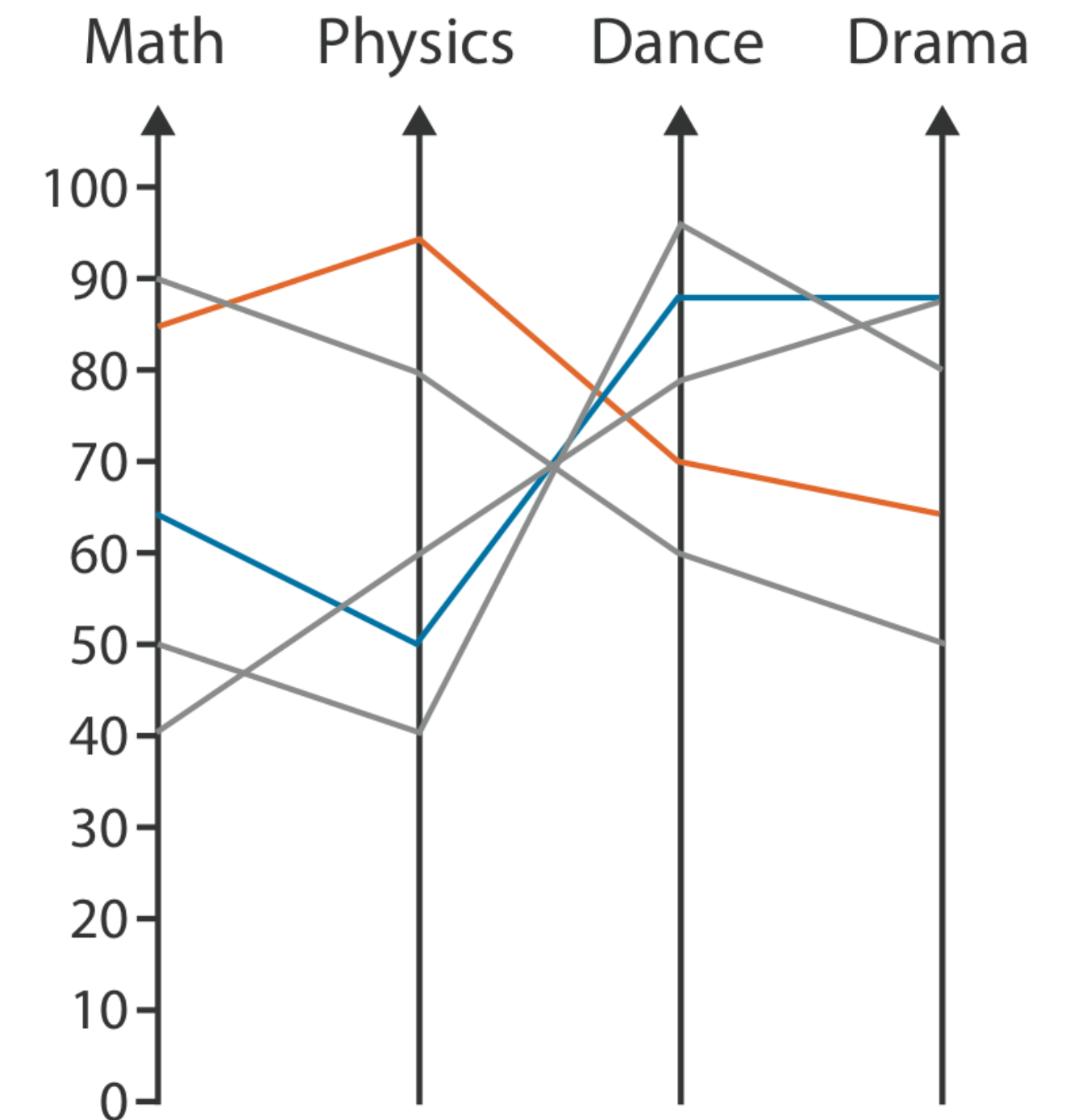
Table

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



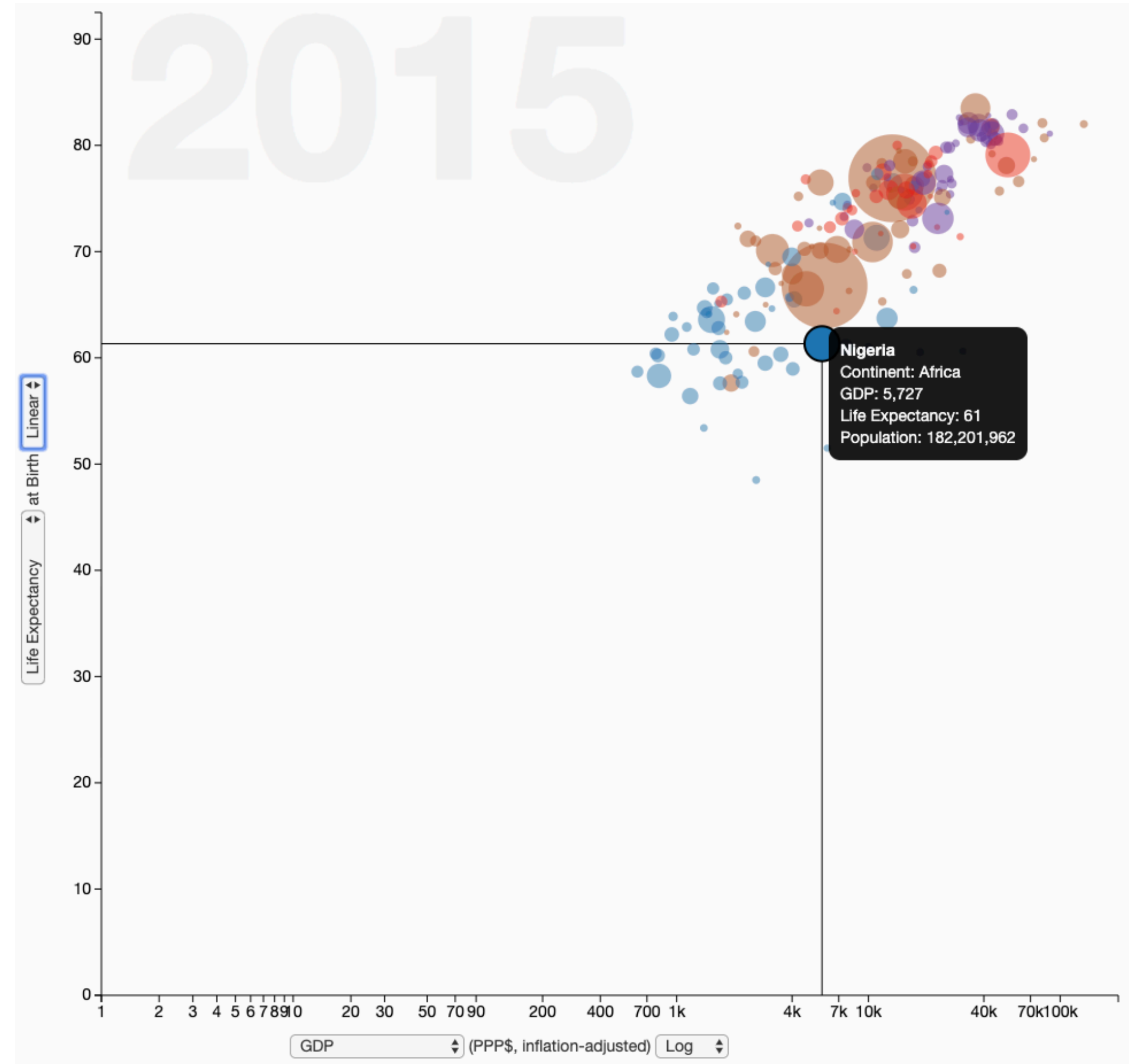
Scatterplots

Scatterplots

Two orthogonal axis
visualizing one
dimension each.

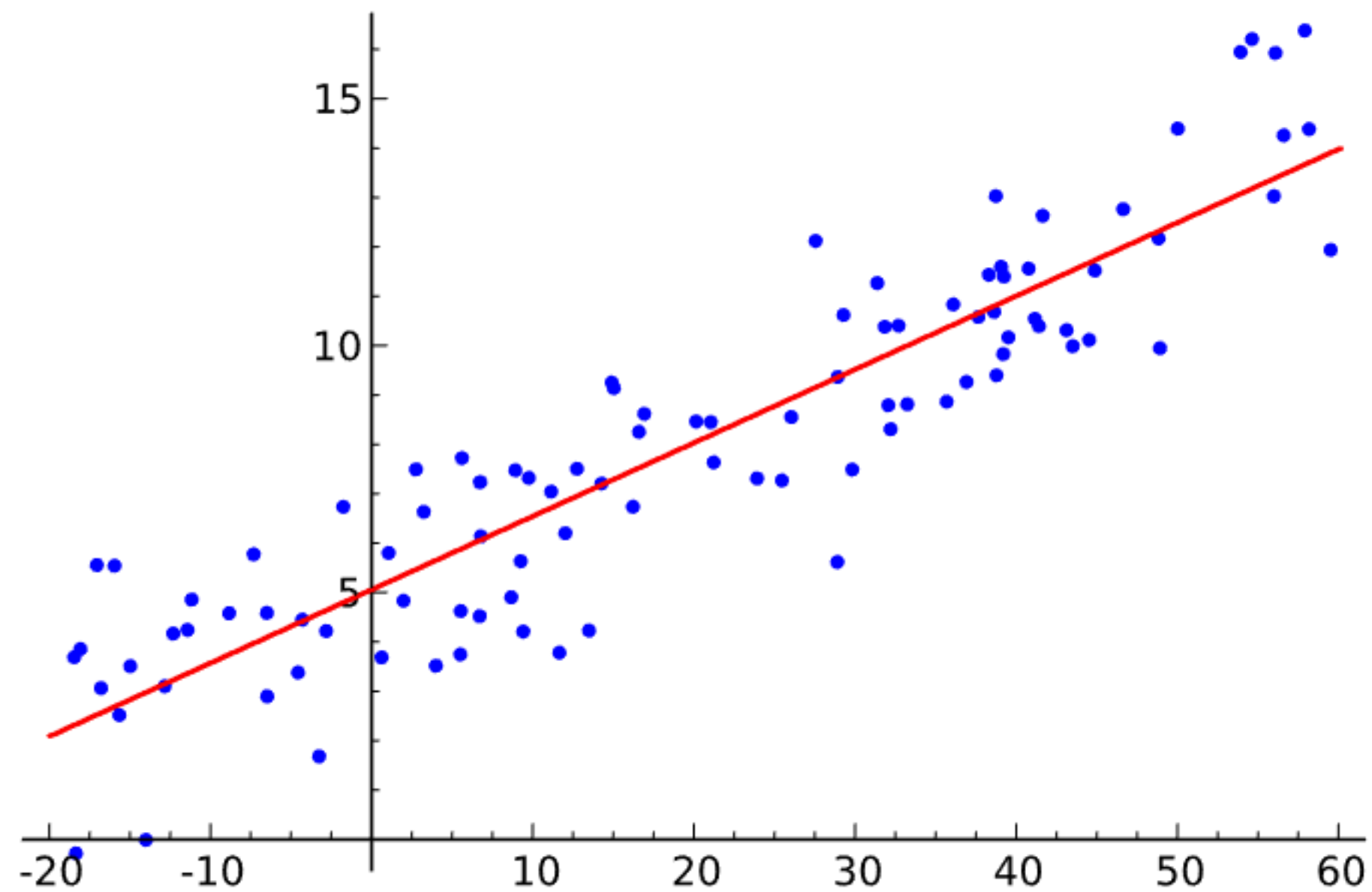
How to encode the
mark?

How to deal with many
points?



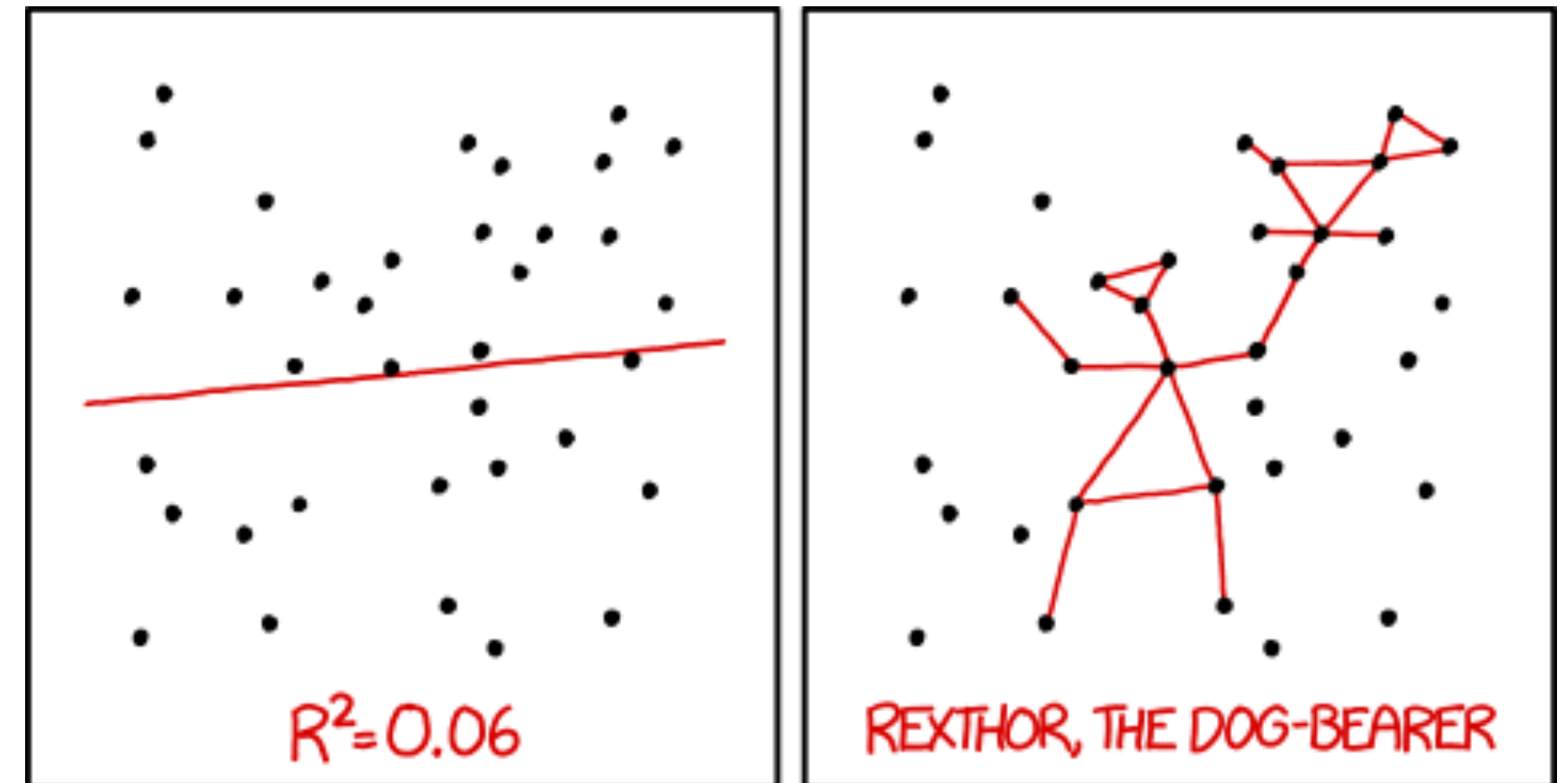
Regression Lines

$$y \sim \beta_0 + \beta_1 x$$



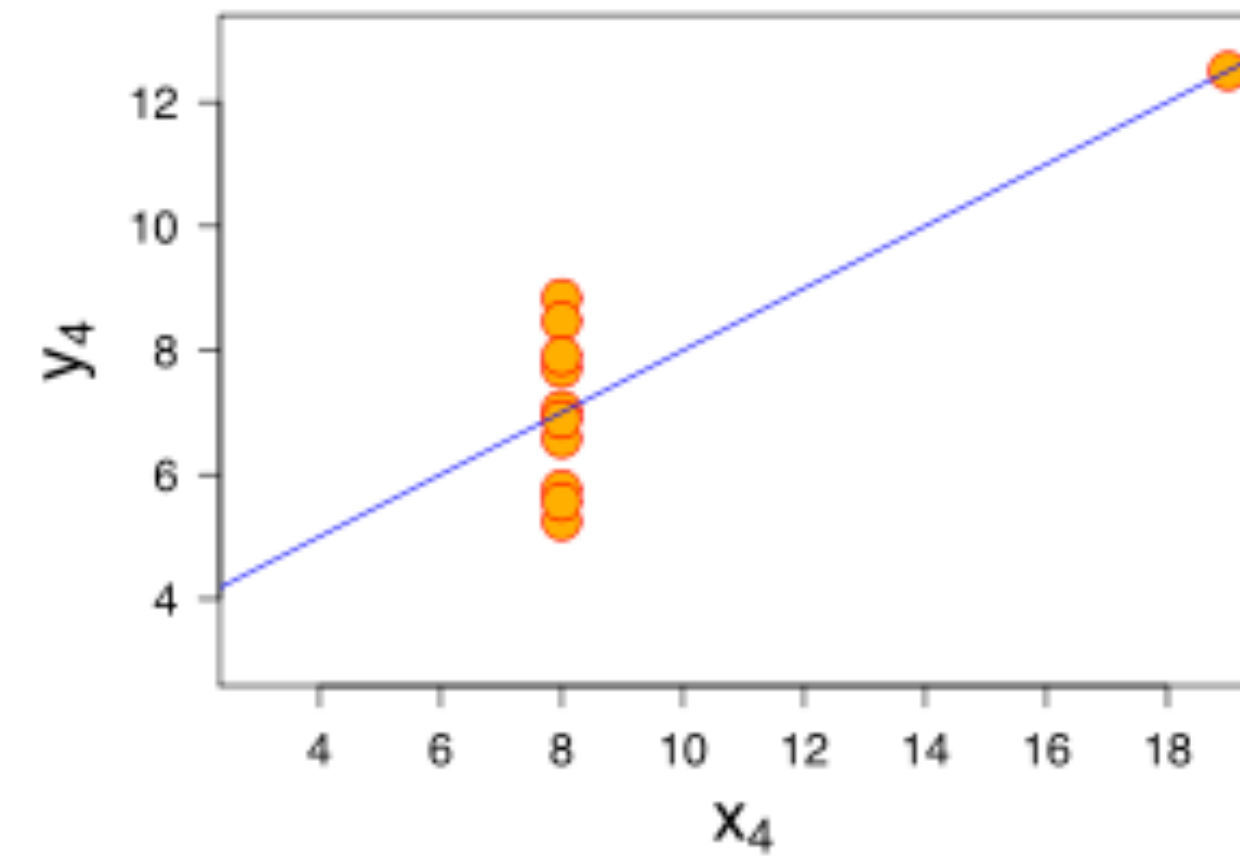
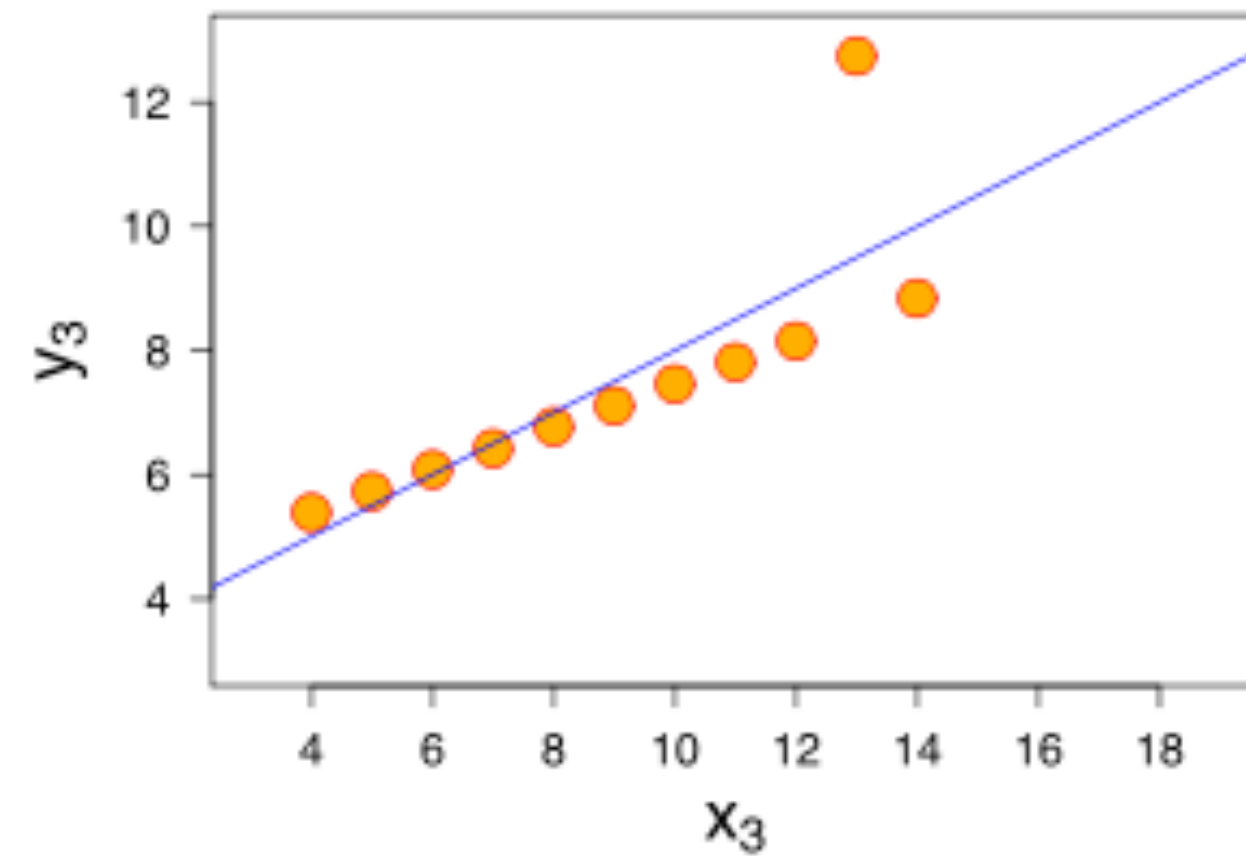
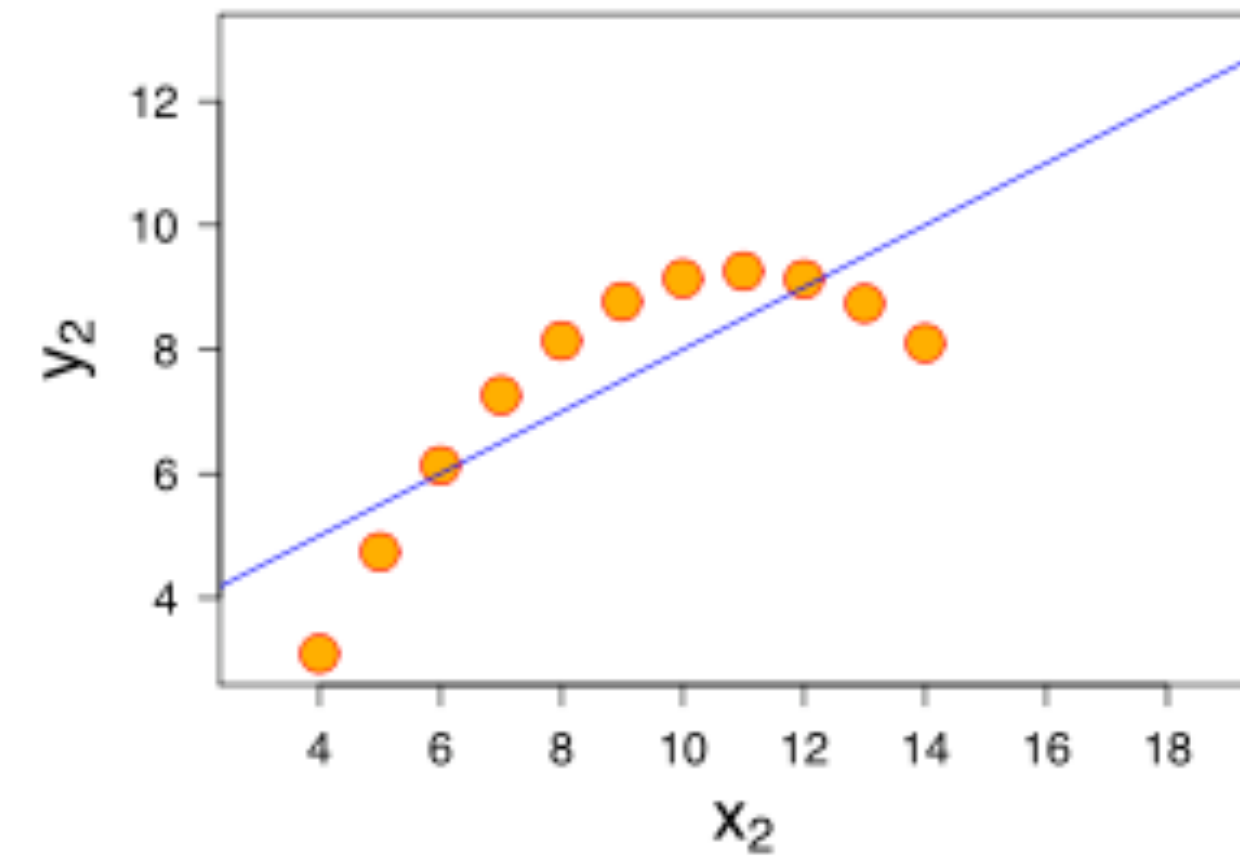
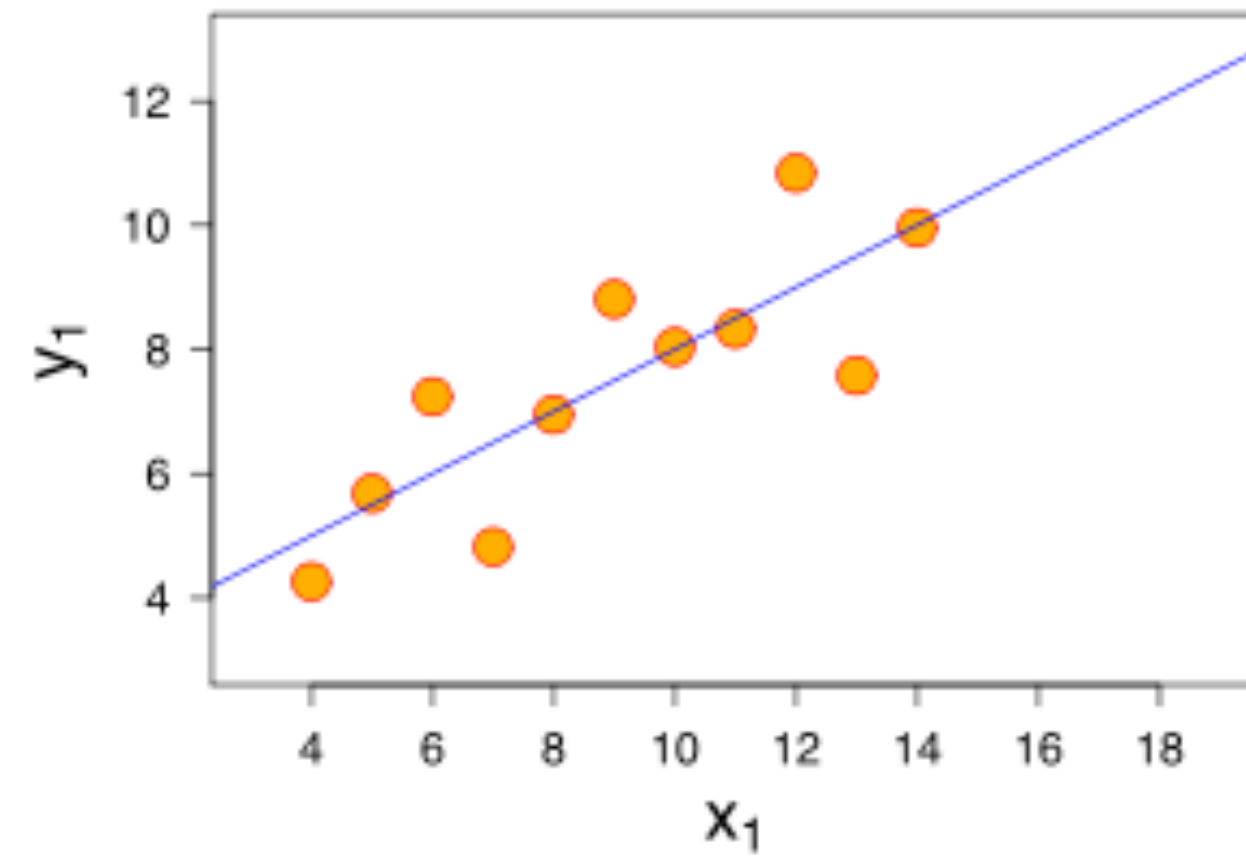
Goal: Find the best values of β_0 and β_1 , denoted $\hat{\beta}_0$ and $\hat{\beta}_1$, so that the prediction $y = \hat{\beta}_0 + \hat{\beta}_1 x$ “best fits” the data.

Approach: use least squares to minimize the sum of the squares of the errors



I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

Anscombe's Quartet

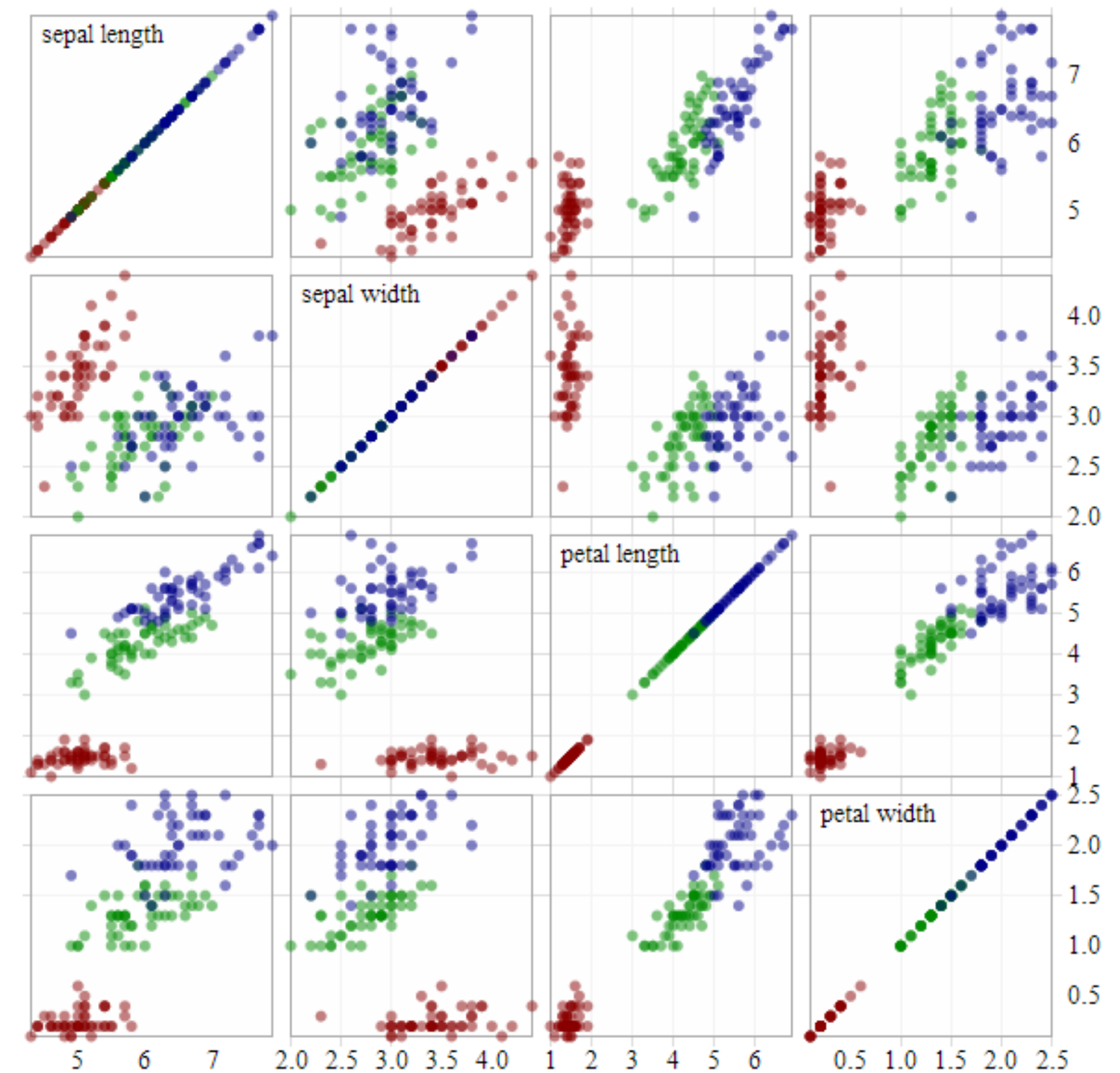


Scatterplot Matrices (SPLOM)

Matrix of size $d \times d$

Each row/column is one dimension

Each cell plots a scatterplot of two dimensions



Scatterplot Matrices

Limited scalability (~20 dimensions, ~500-1k records)

Brushing is important

Often combined with “Focus Scatterplot” as F+C technique

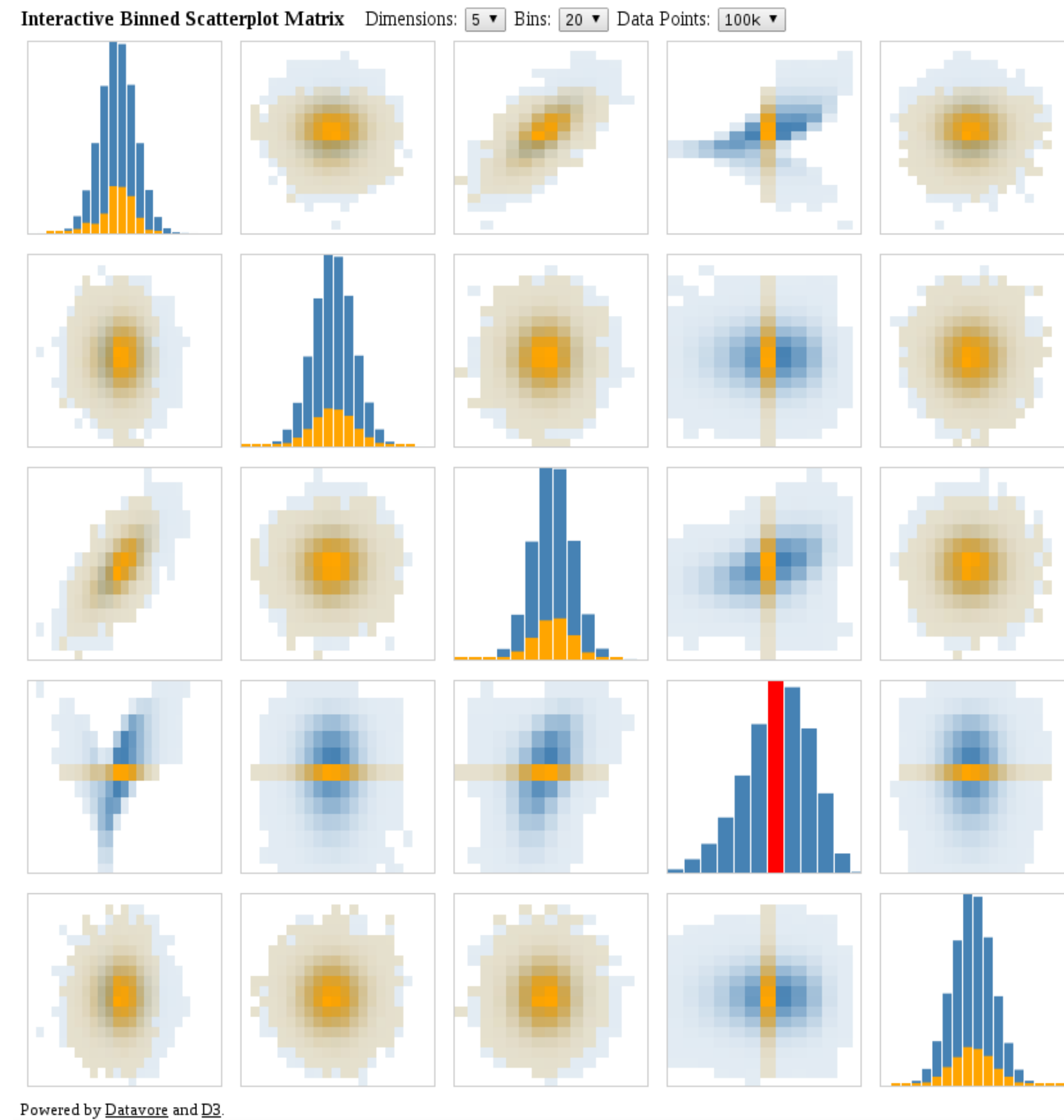
Algorithmic approaches:

Clustering & aggregating records

Choosing dimensions

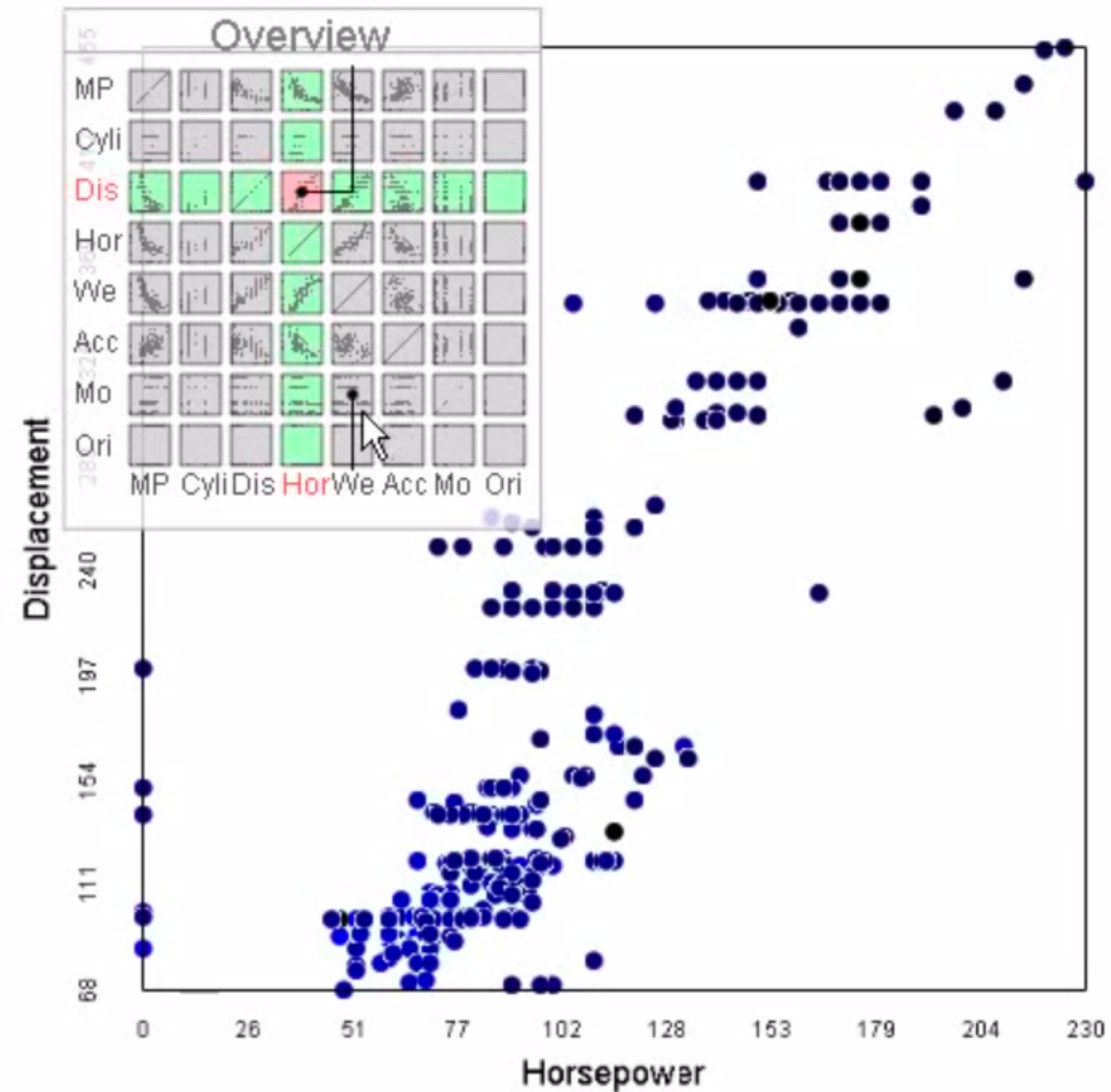
Choosing order

SPLOM Aggregation - Heat Map



Datavore: <http://vis.stanford.edu/projects/datavore/splom/>

SPLOM F+C, Navigation



[Elmqvist]

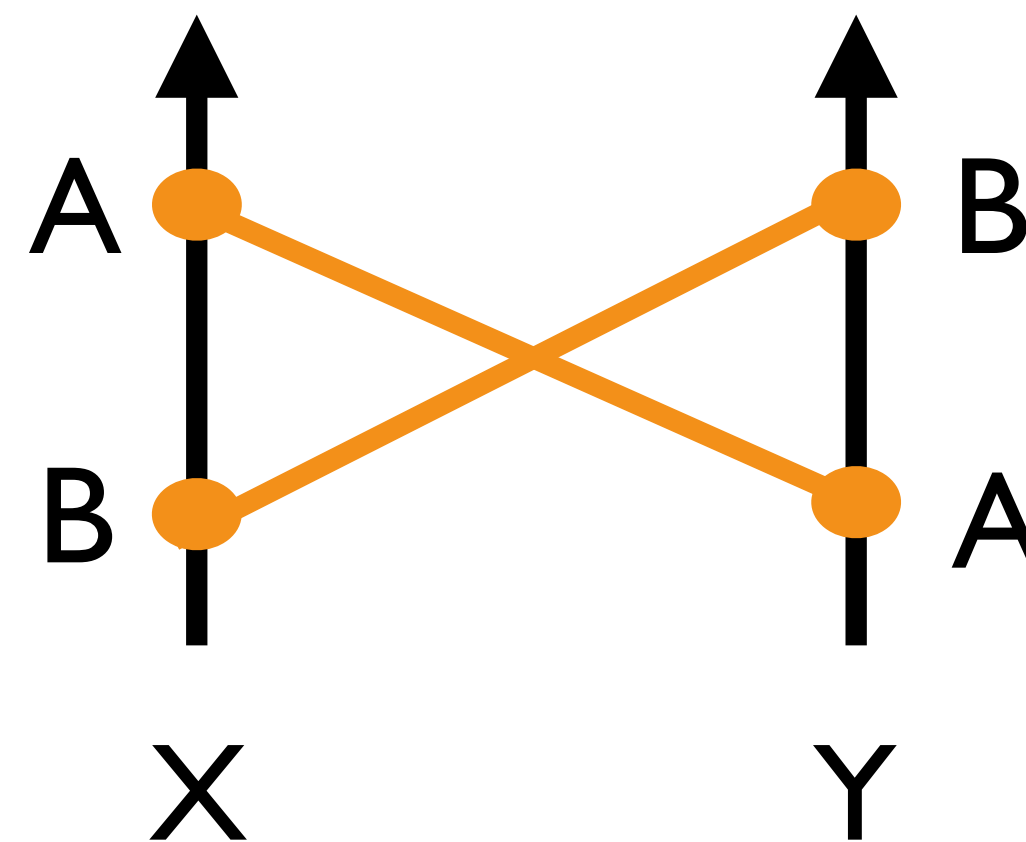
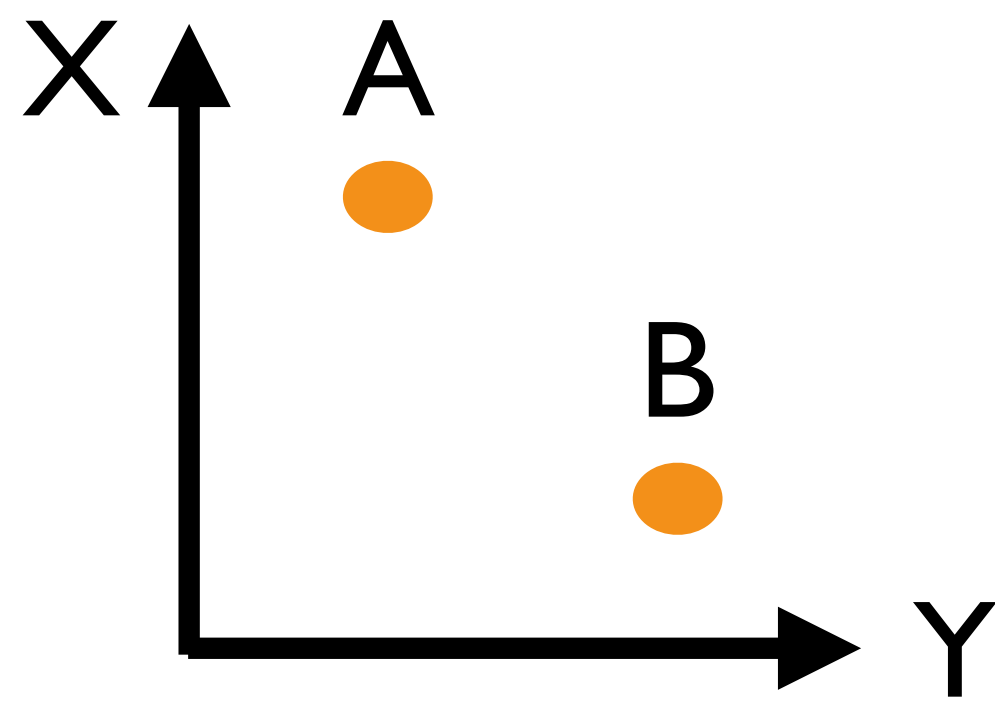
Parallel Coordinates

Parallel Coordinates (PC)

Inselberg 1985

Axes represent attributes

Lines connecting axes represent items



Parallel Coordinates

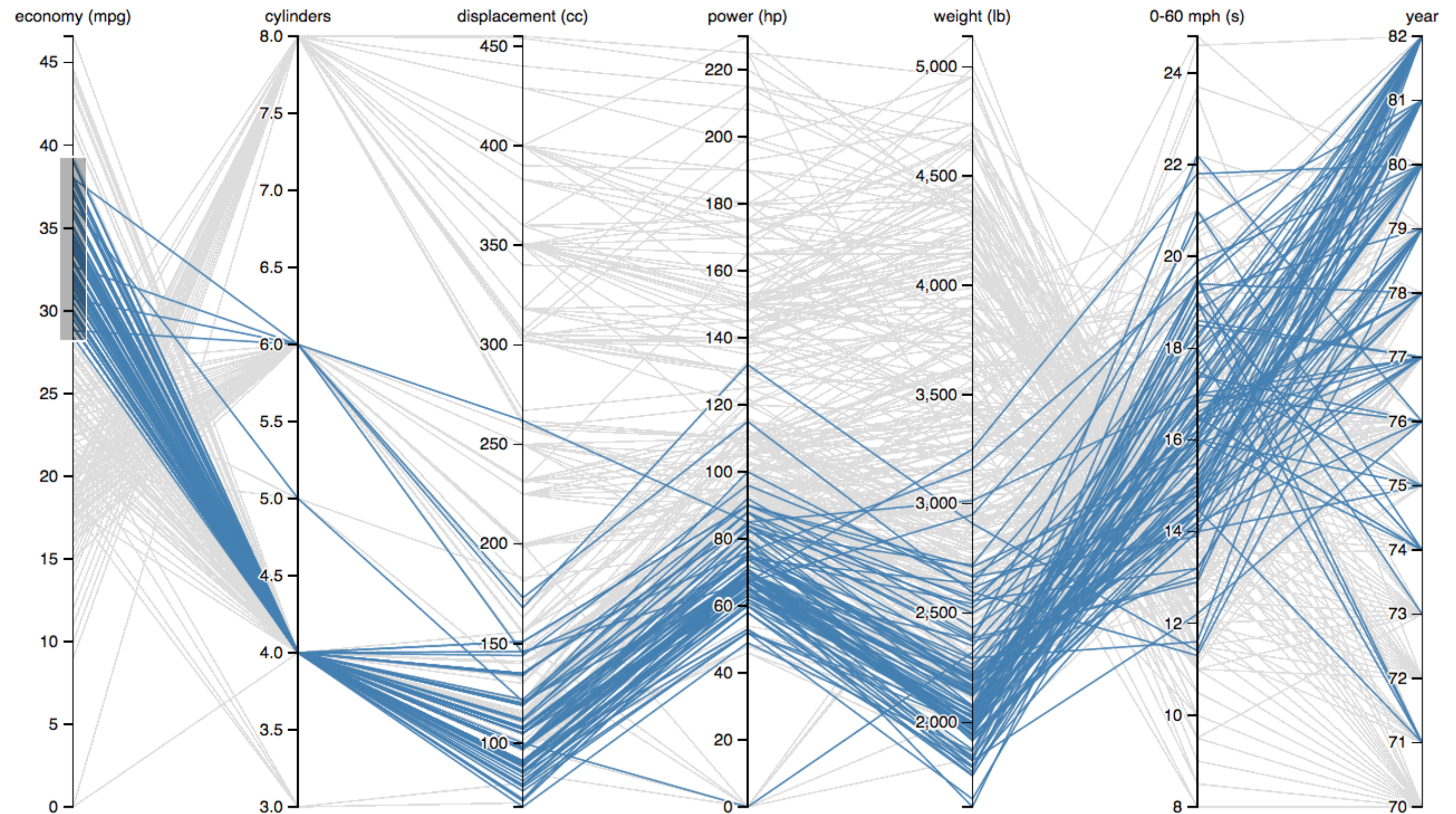
Each axis represents dimension

Lines connecting axis represent records

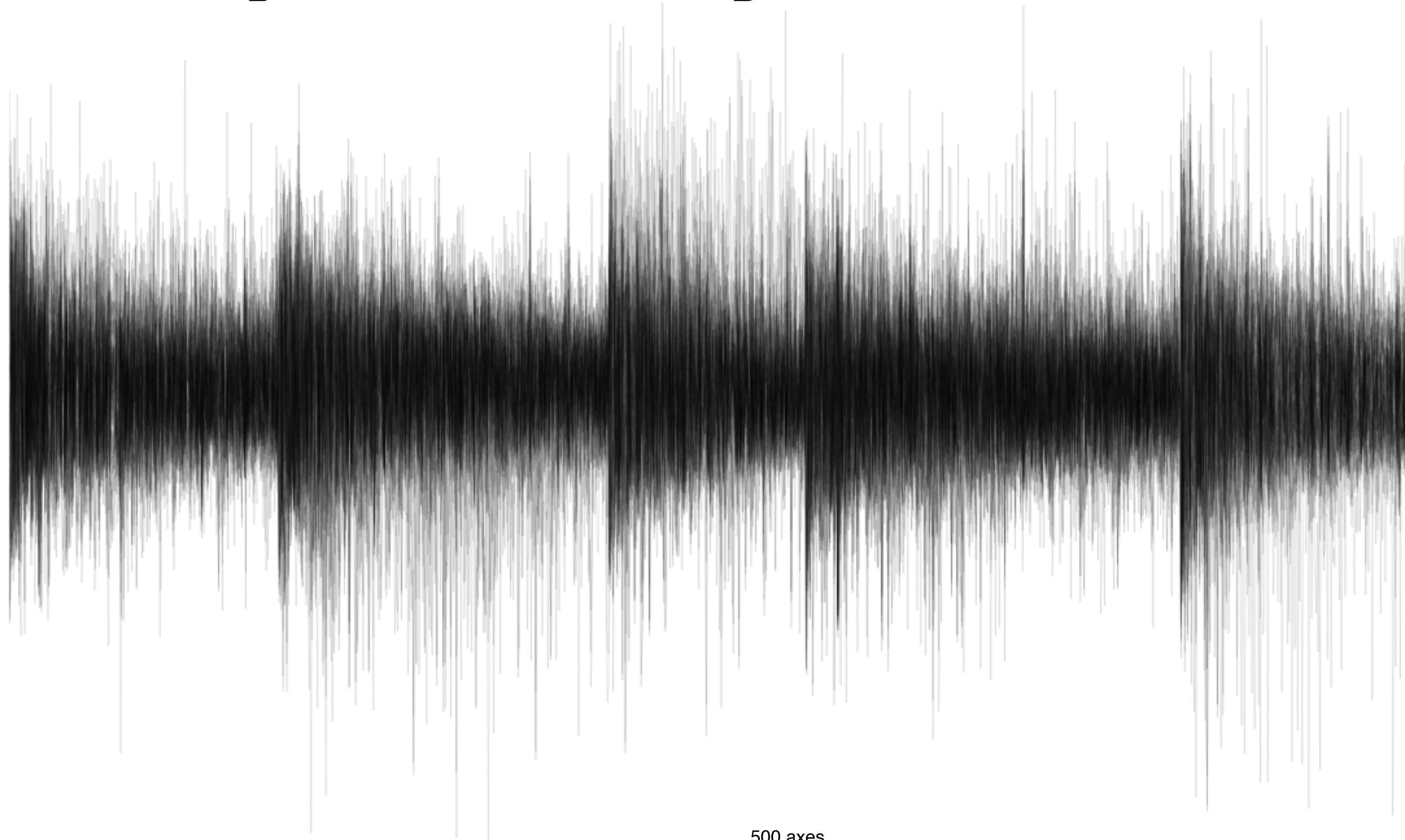
Suitable for

all tabular data types

heterogeneous data



PC Limitation: Scalability to Many Dimensions



500 axes

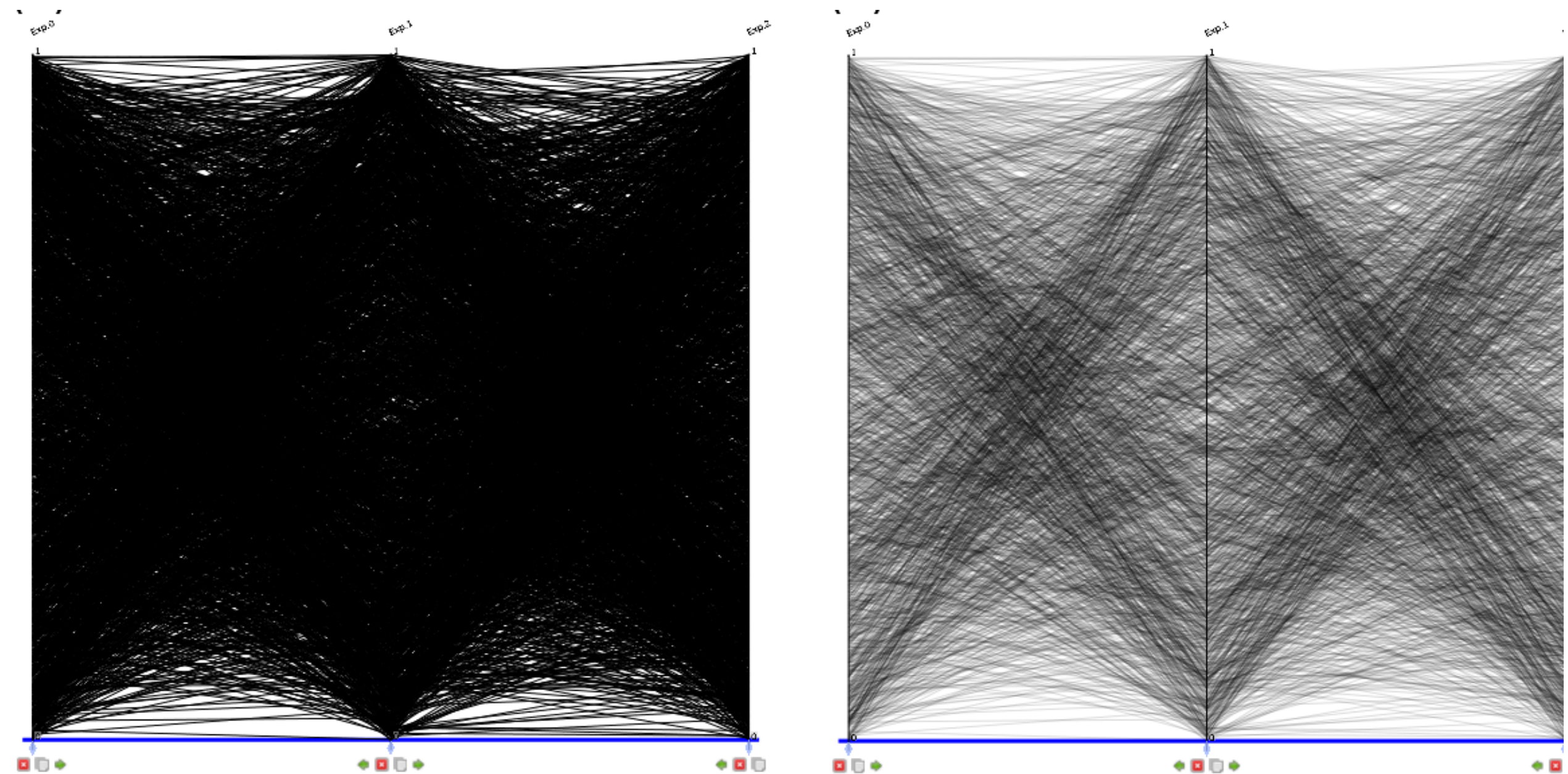
PC Limitation: Scalability to Many Items

Solutions:

Transparency

Bundling, Clustering

Sampling



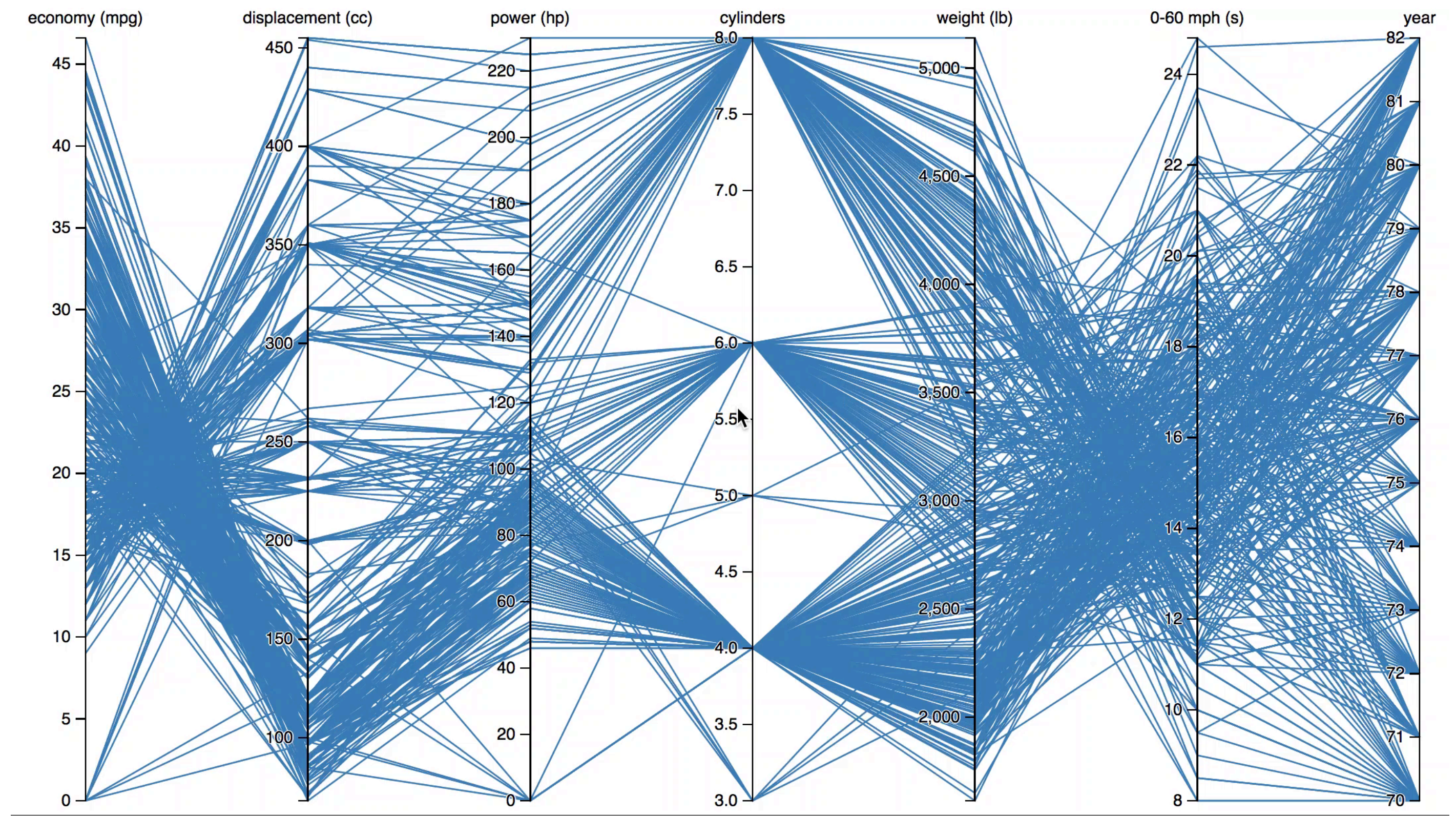
PC Limitations

Correlations only between adjacent axes

Solution: Interaction

Brushing

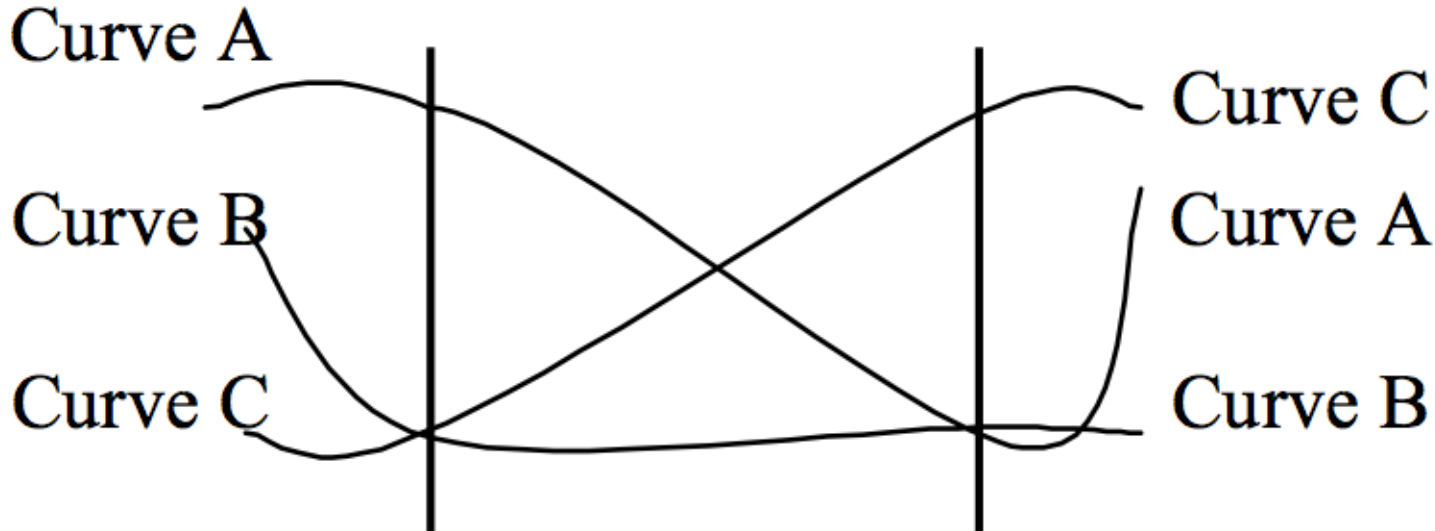
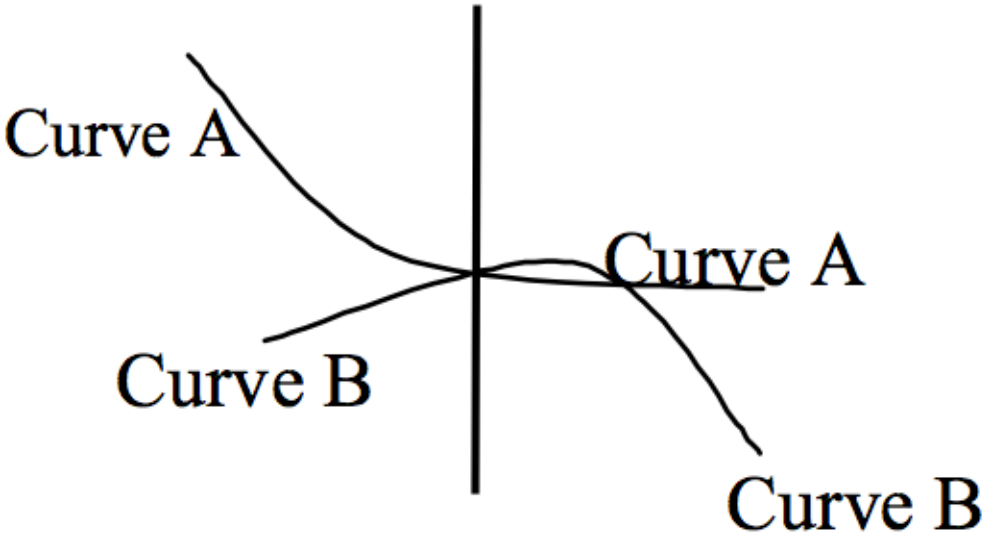
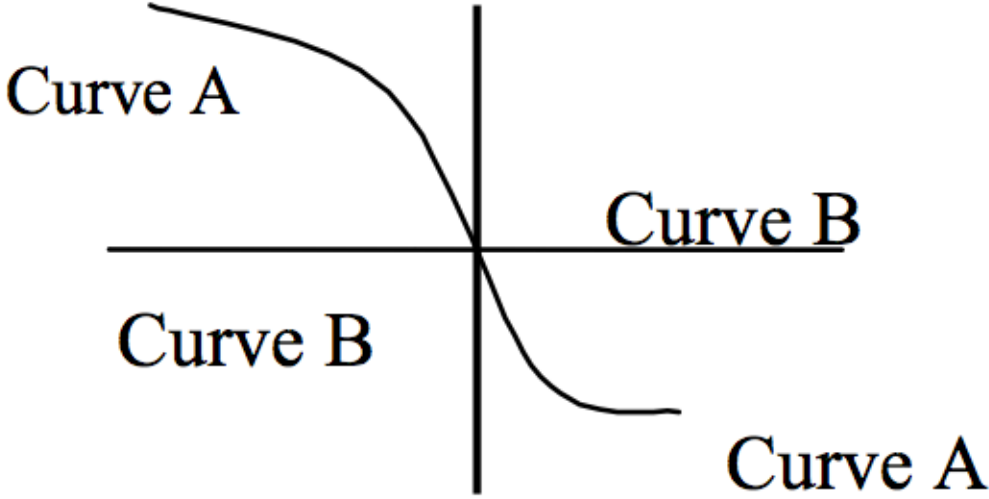
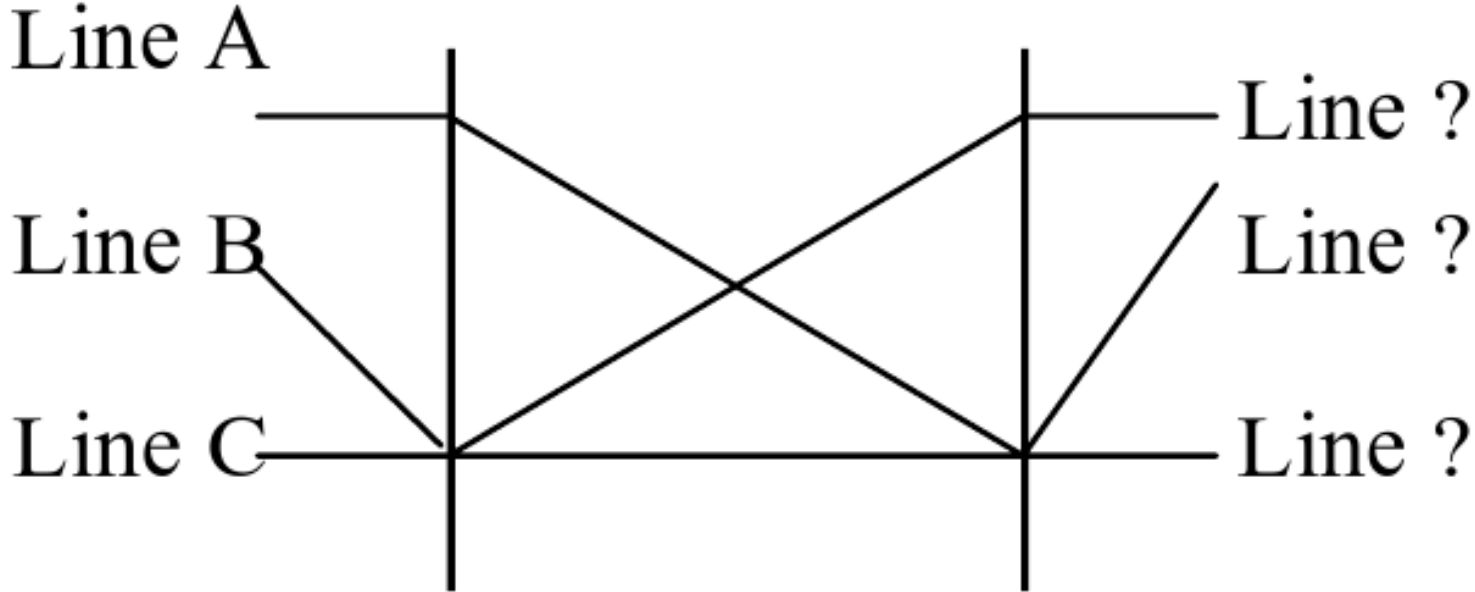
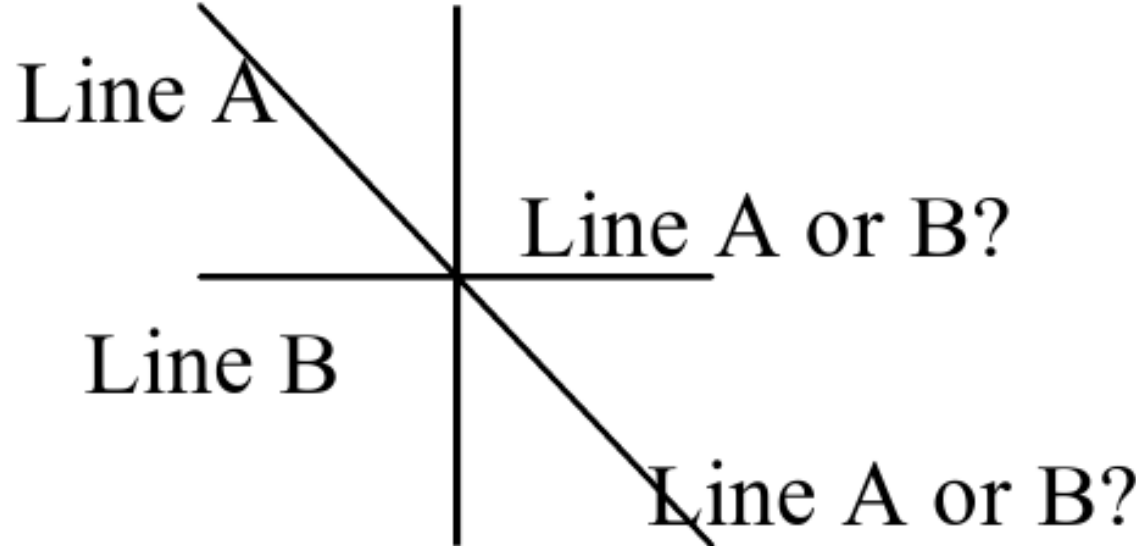
Let user change order



PC Limitation: Ambiguity

Solutions:

Brushing
Curves



Parallel Coordinates

Shows primarily relationships between adjacent axis

Limited scalability (~50 dimensions, ~1-5k records)

Transparency of lines

Interaction is crucial

Axis reordering

Brushing

Filtering

Algorithmic support:

Choosing dimensions

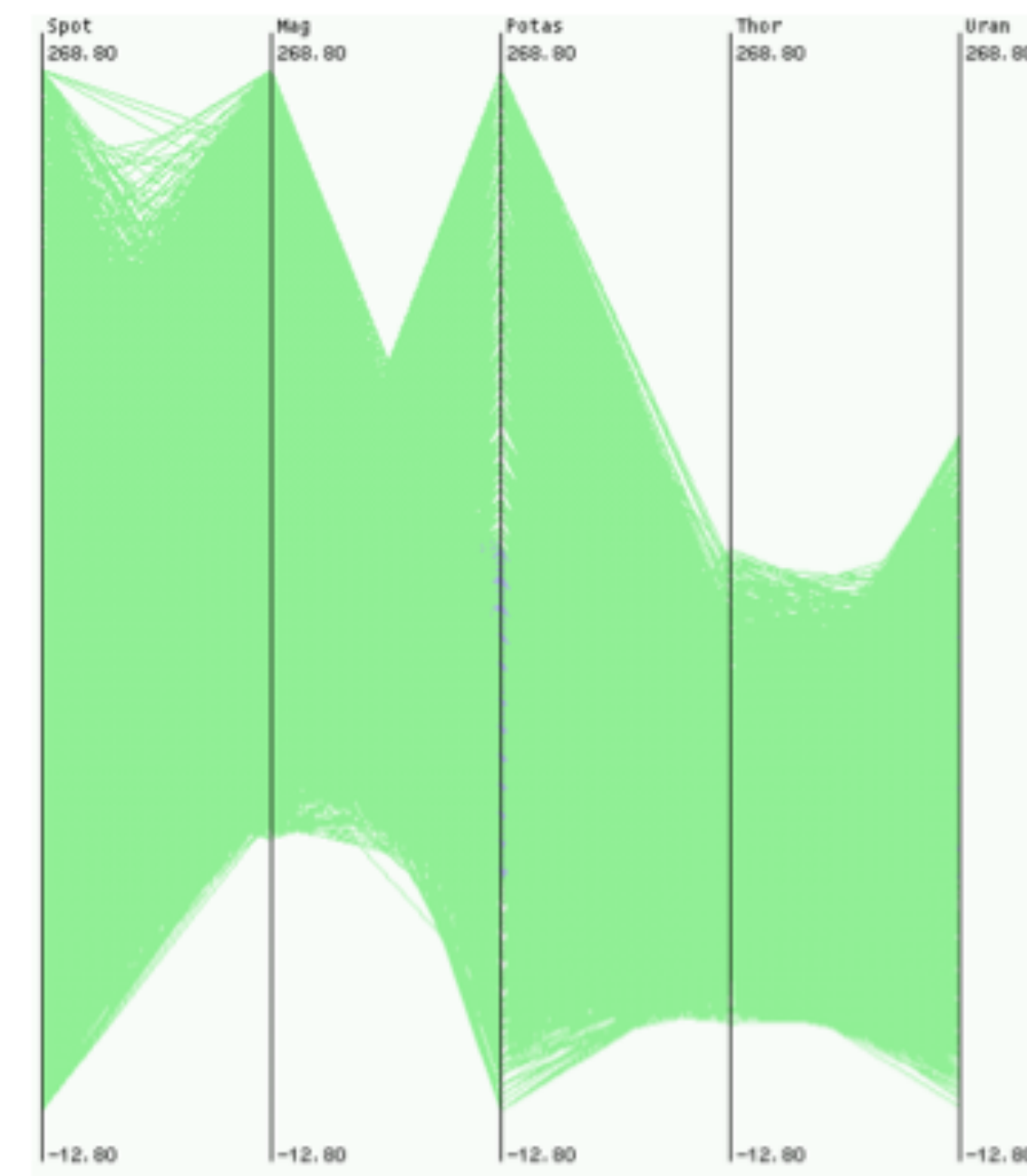
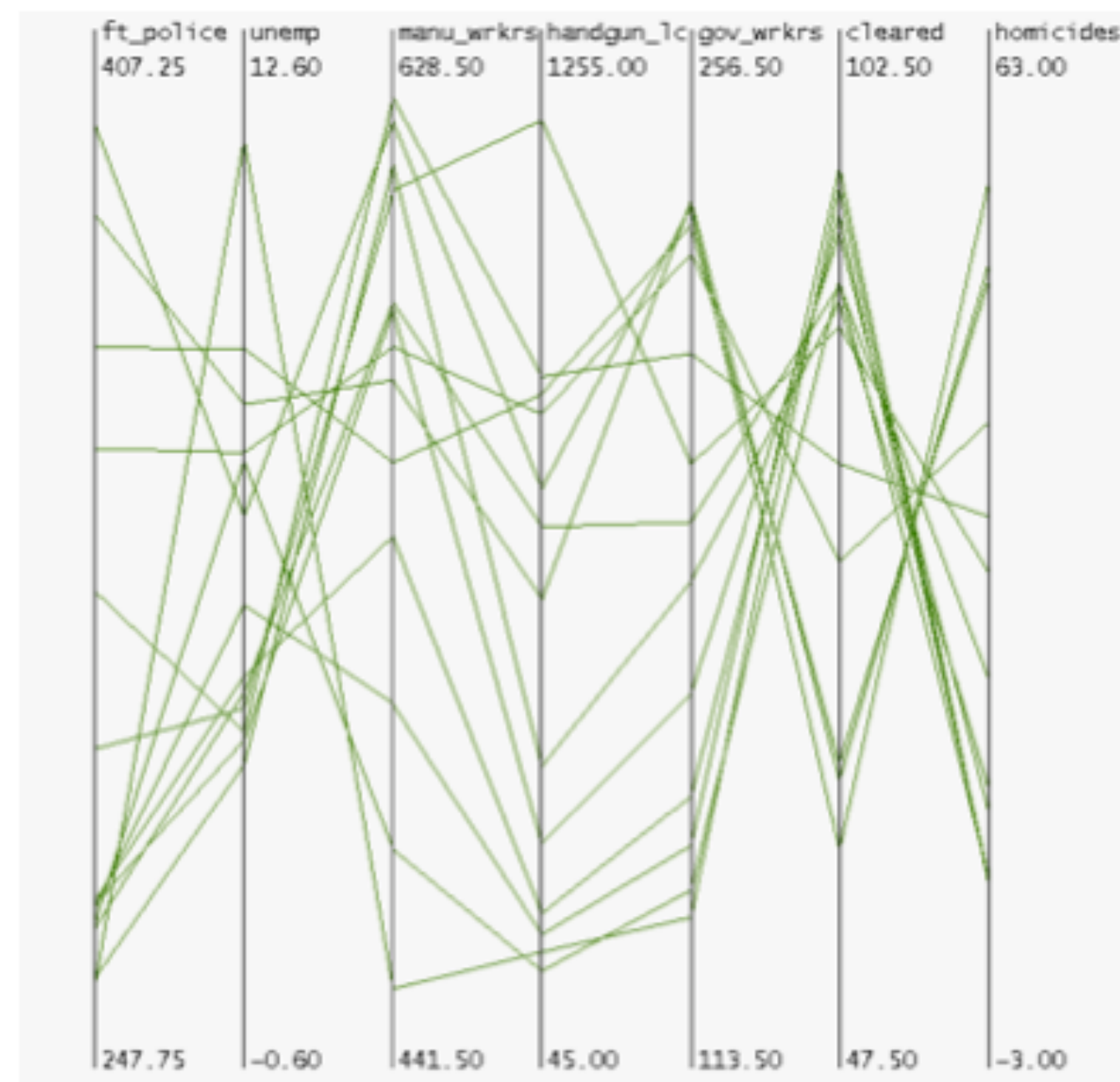
Choosing order

Clustering & aggregating records

HIERARCHICAL PARALLEL COORDINATES

goal: scale up parallel coordinates to large datasets

challenge: overplotting/occlusion



HPC: ENCODING DERIVED DATA

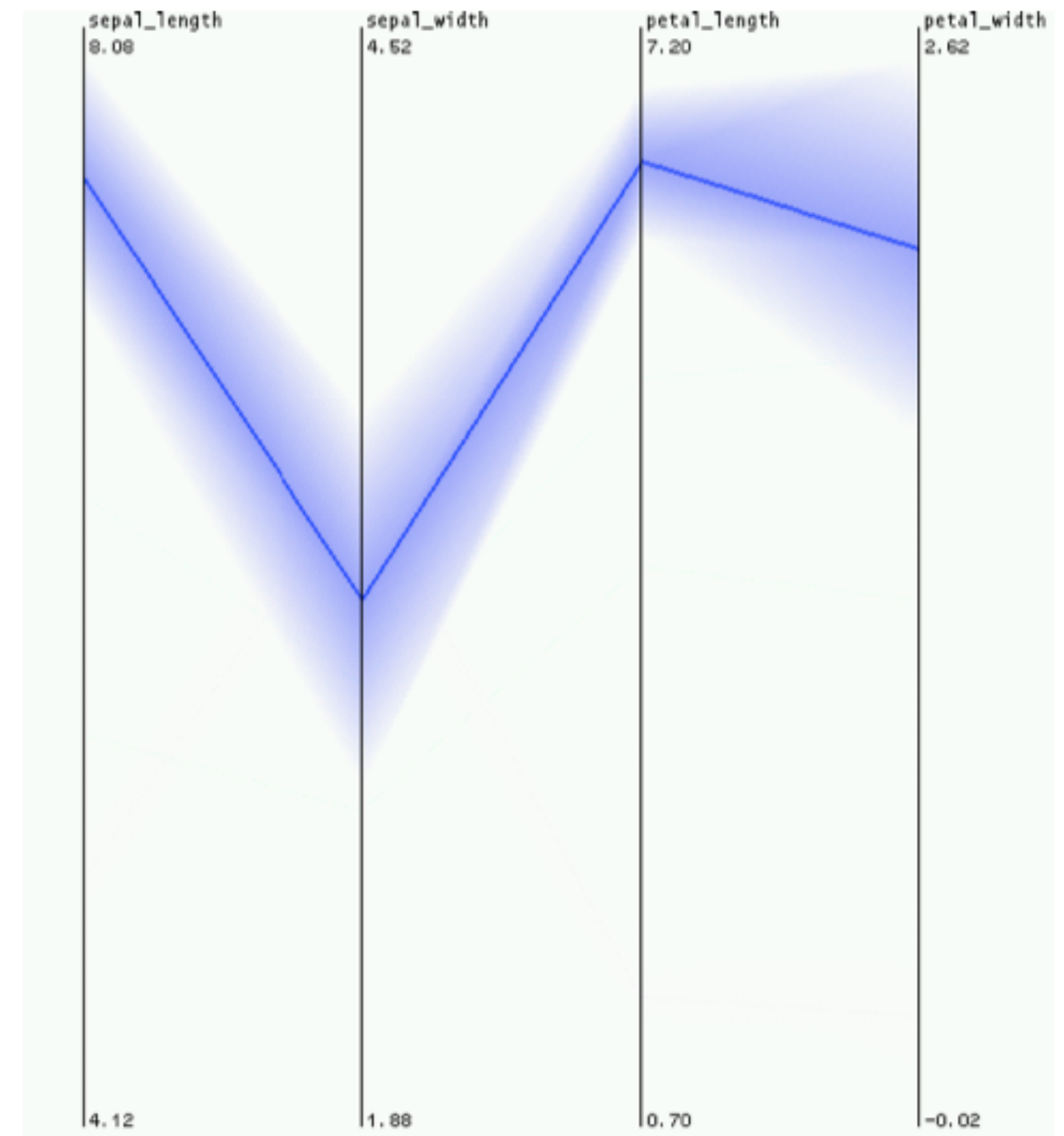
visual representation: variable-width opacity bands

show whole cluster, not just single item

min / max: spatial position

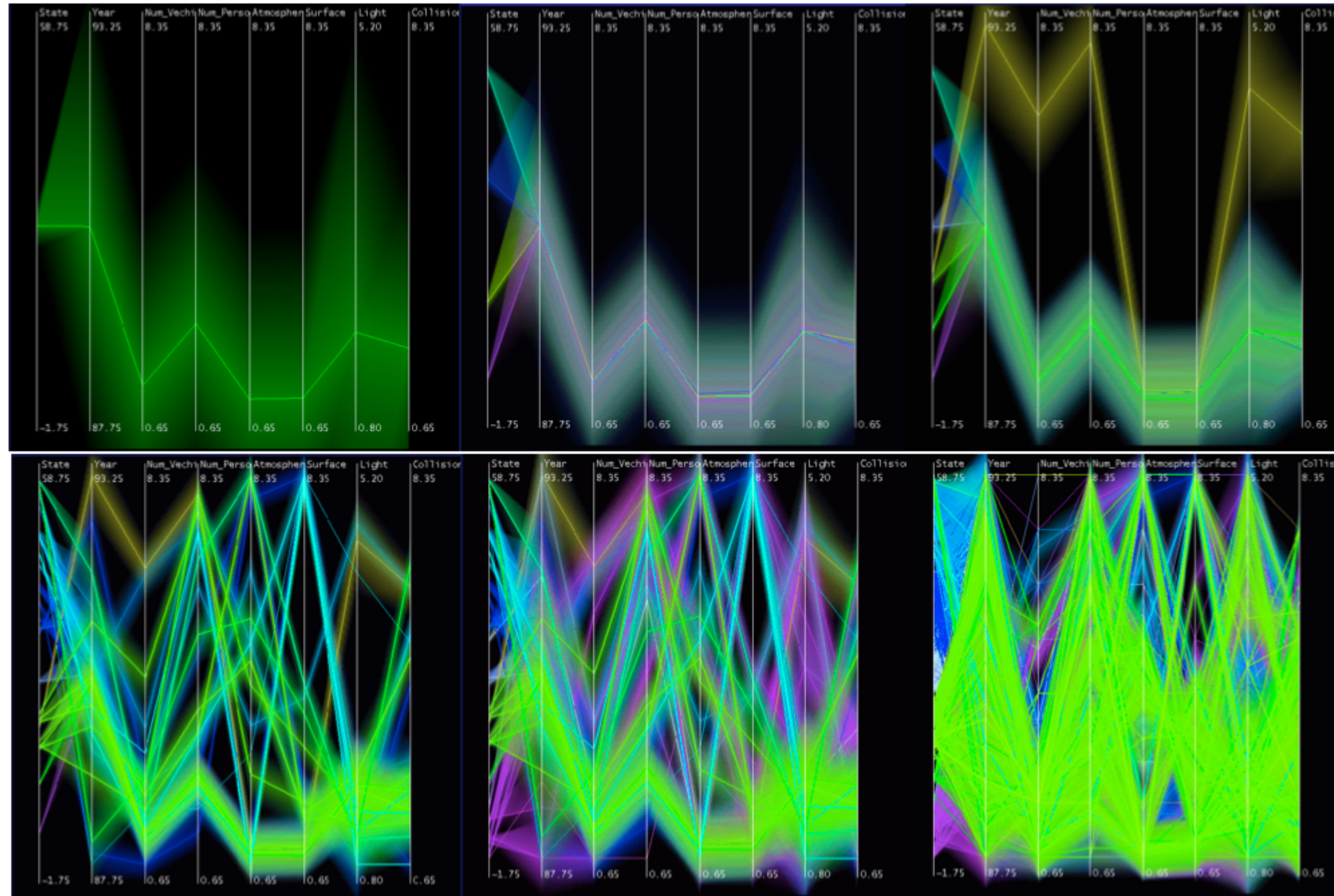
cluster density: transparency

mean: opaque



HPC: INTERACTING WITH DERIVED DATA

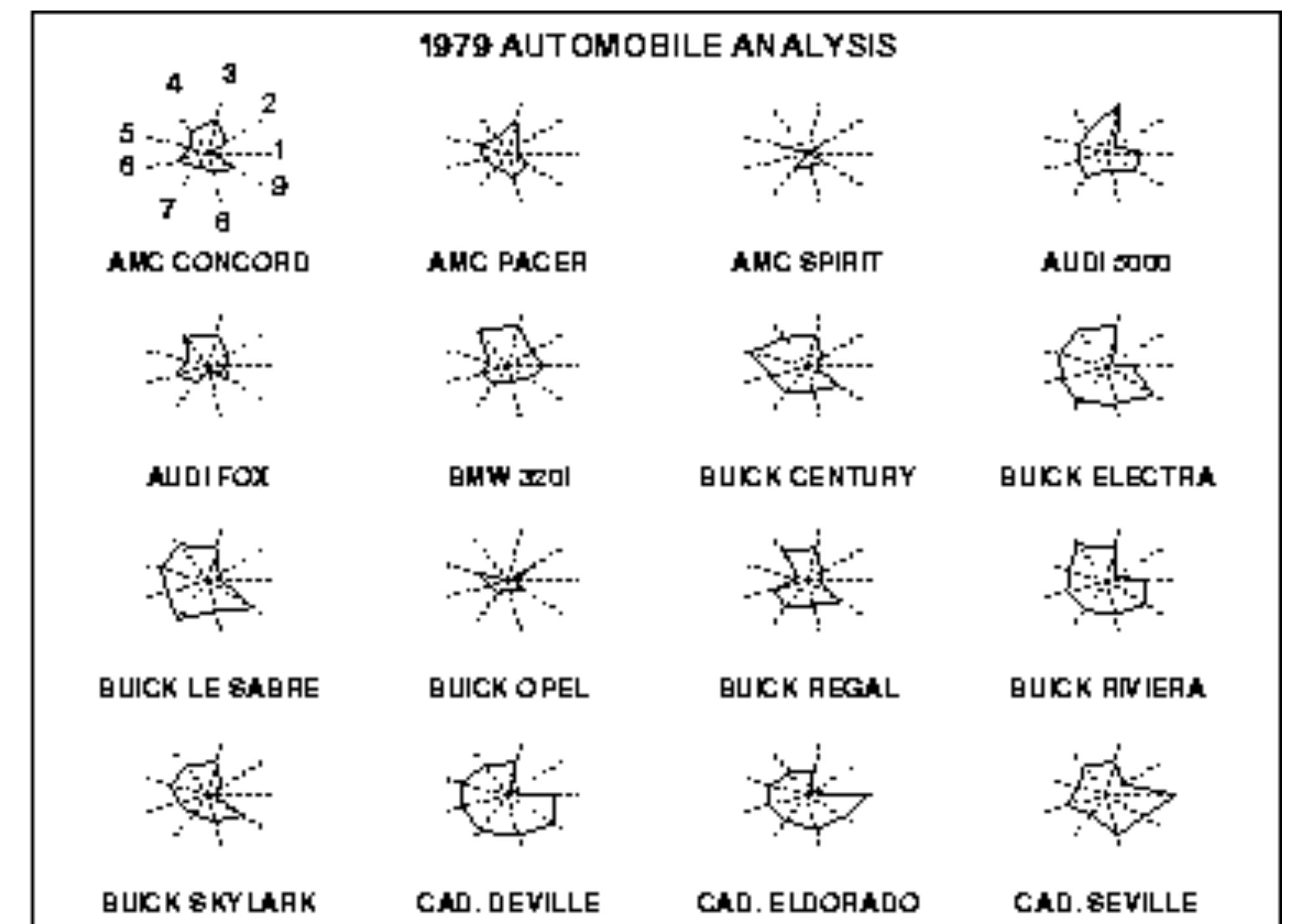
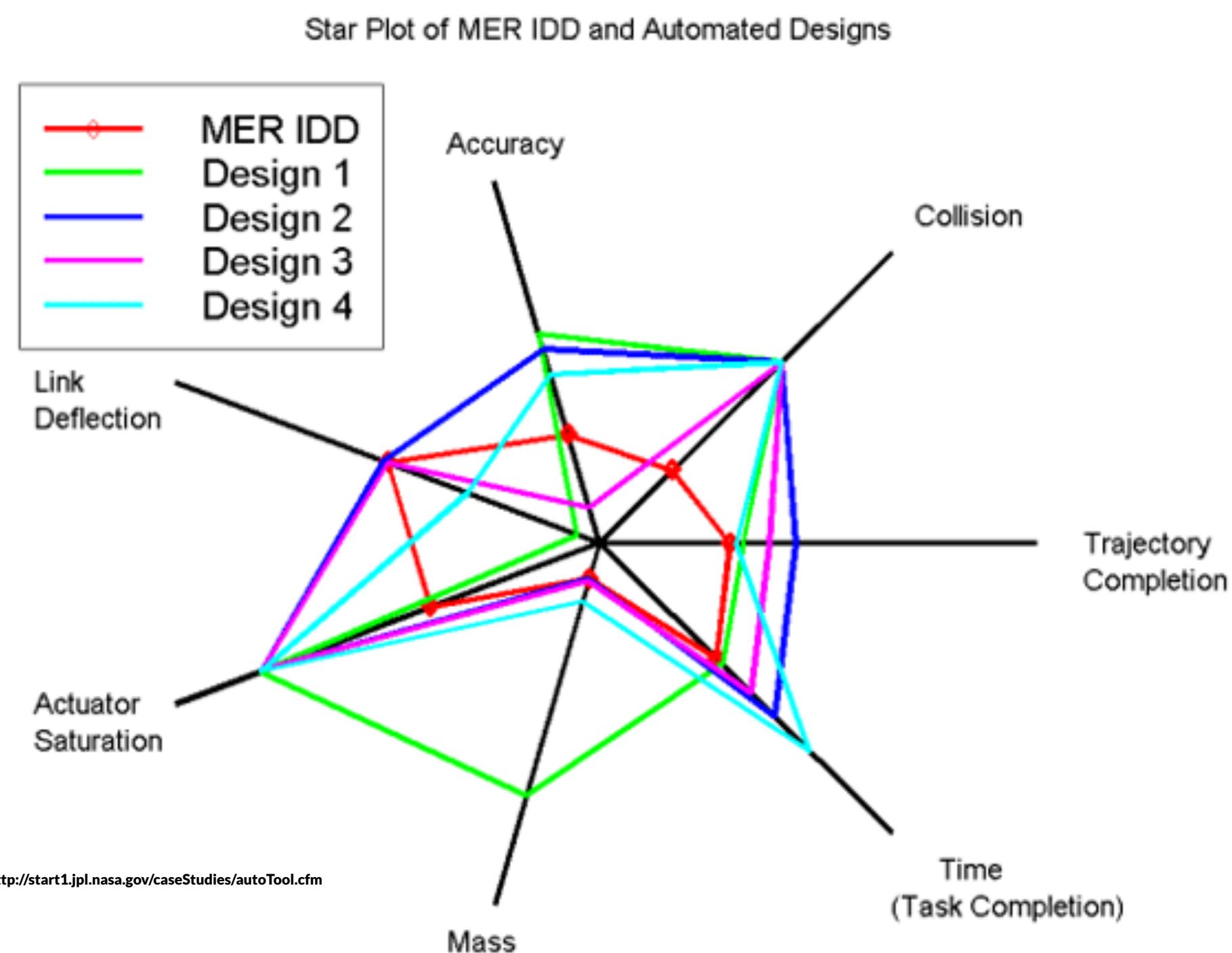
interactively change level of detail to navigate cluster hierarchy



Star Plot

[Coekin1969]

Similar to parallel coordinates
Radiate from a common origin



<http://www.itl.nist.gov/div898/handbook/eda/section3/starplot.htm>

<http://blocks.org/kevinschau/raw/8833989/>

Data Reduction

Sampling

Don't show every element, show a (random) subset

Efficient for large dataset

Apply only for display purposes

Outlier-preserving approaches

Filtering

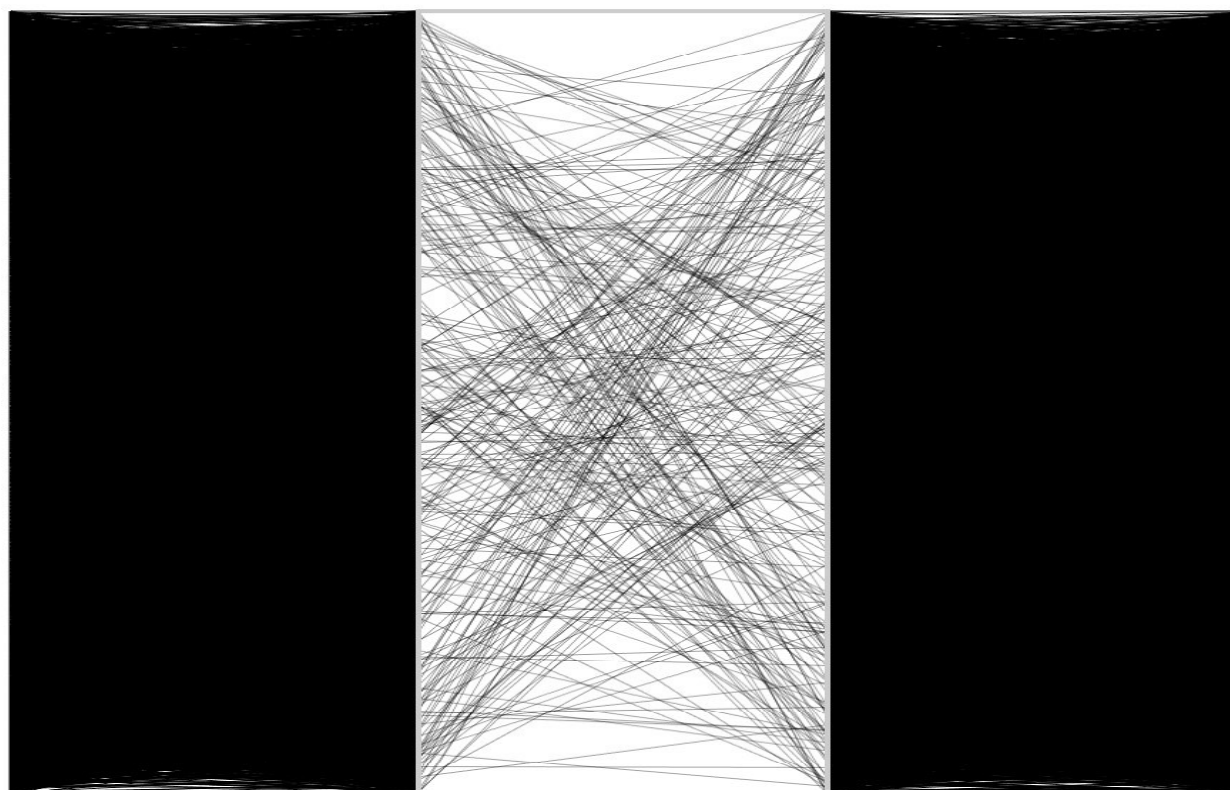
Define criteria to remove data, e.g.,

minimum variability

> / < / = specific value for one dimension

consistency in replicates, ...

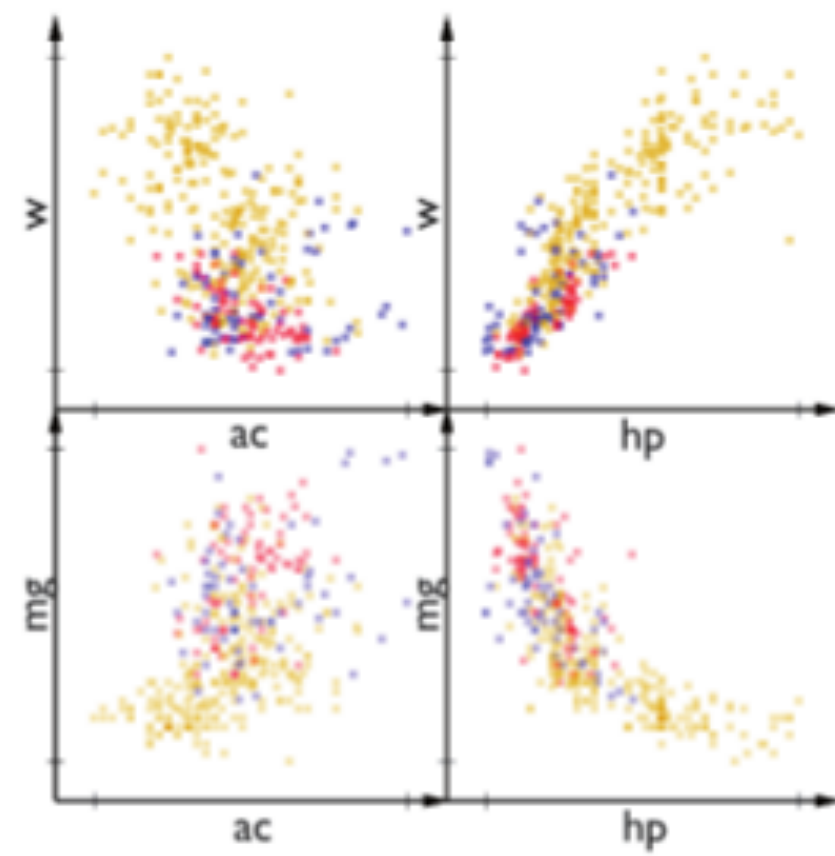
Can be interactive, combined with sampling



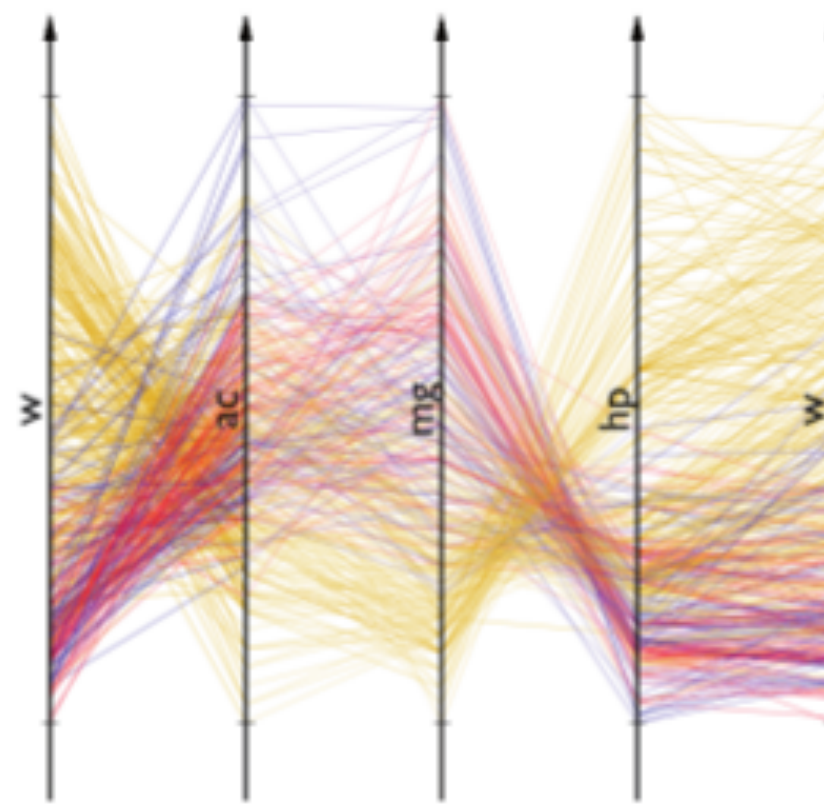
[Ellis & Dix, 2006]

Hybrids with Axis

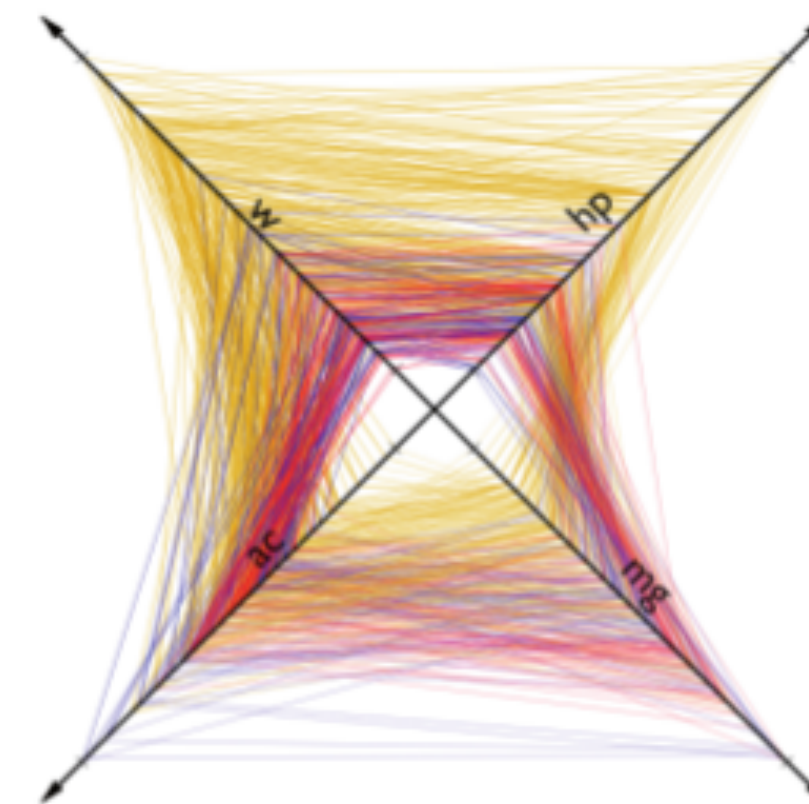
Flexible Linked Axes (FLINA)



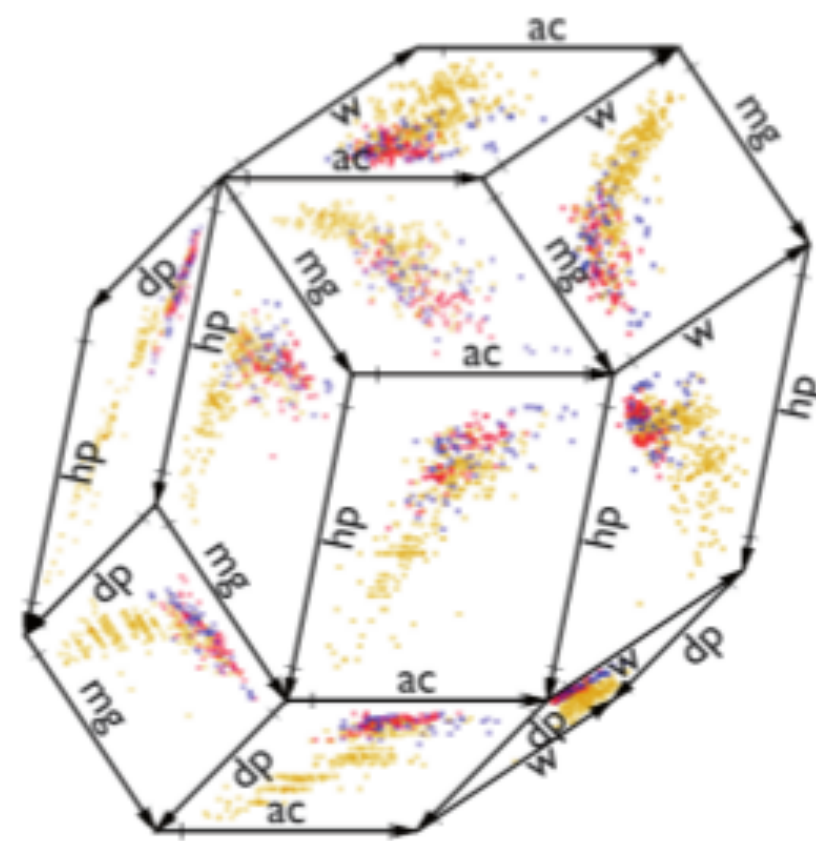
(a) scatterplots



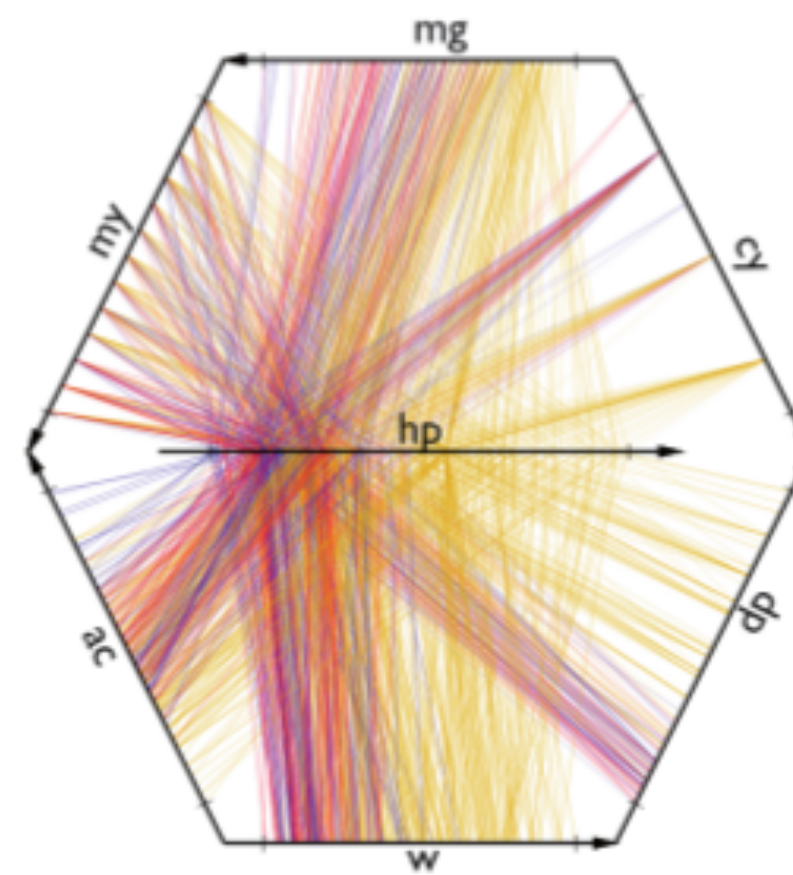
(b) Parallel Coordinates Plot



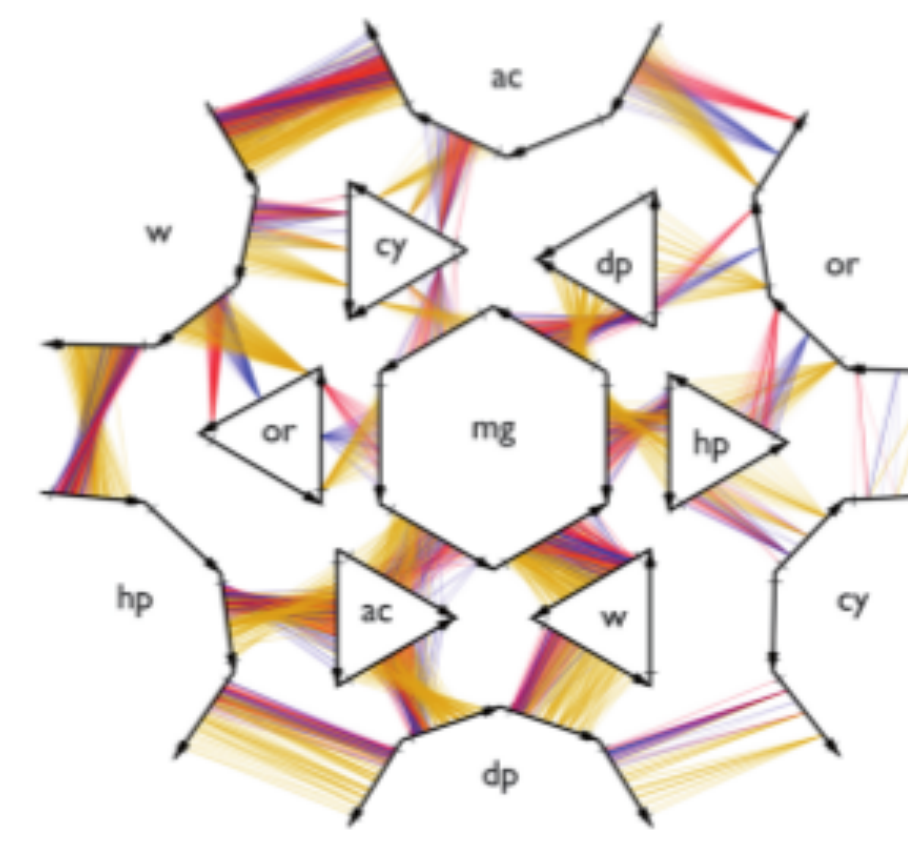
(c) radar chart



(d) Hyperbox

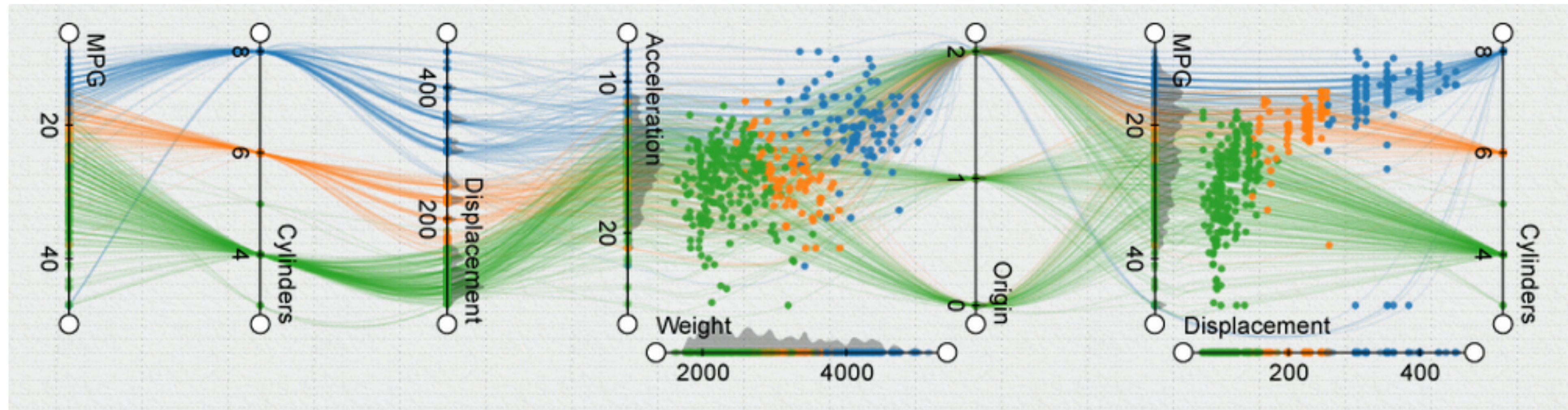


(e) Time Wheel



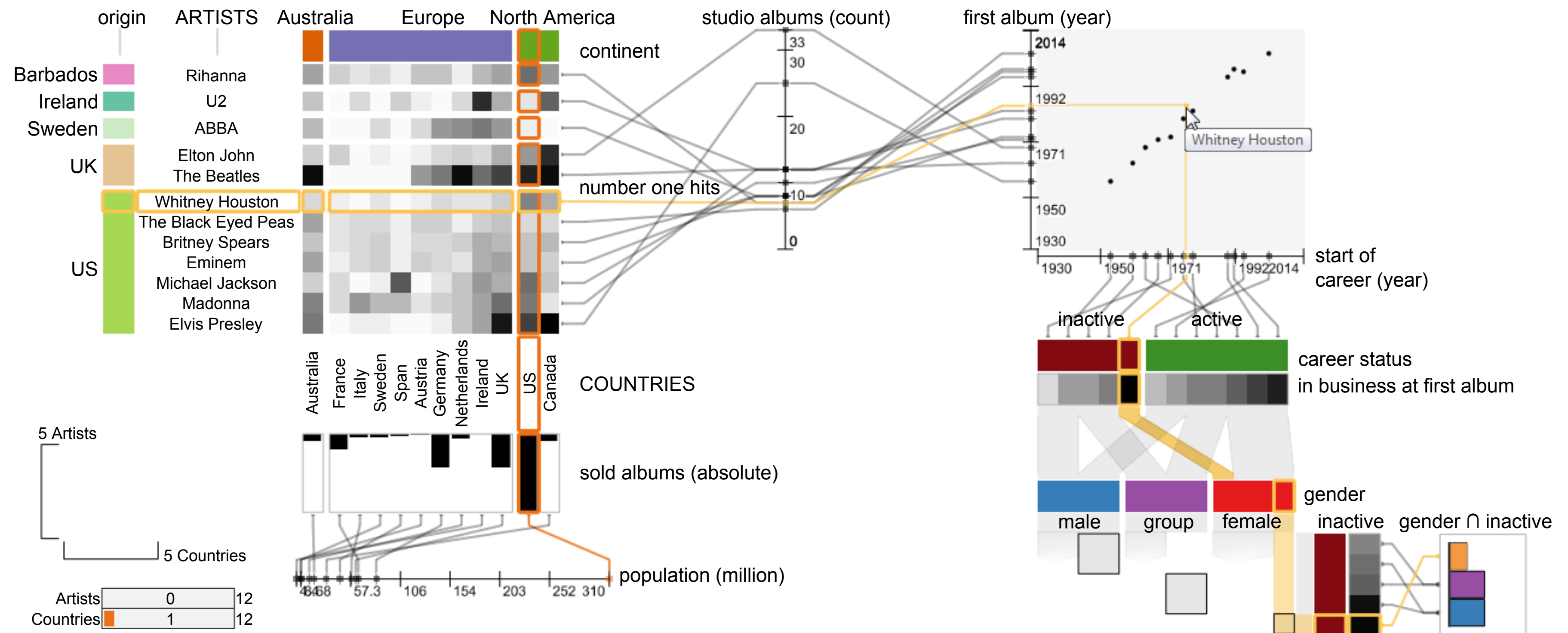
(f) Many-to-many PCP

Web-based implementation of FLINA concept



<http://vis.pku.edu.cn/mddv/val/>

Domino



Parallel Sets

Parallel Sets

builds on PC to better handle categorical data

discrete

small number of values

no implied ordering between attributes

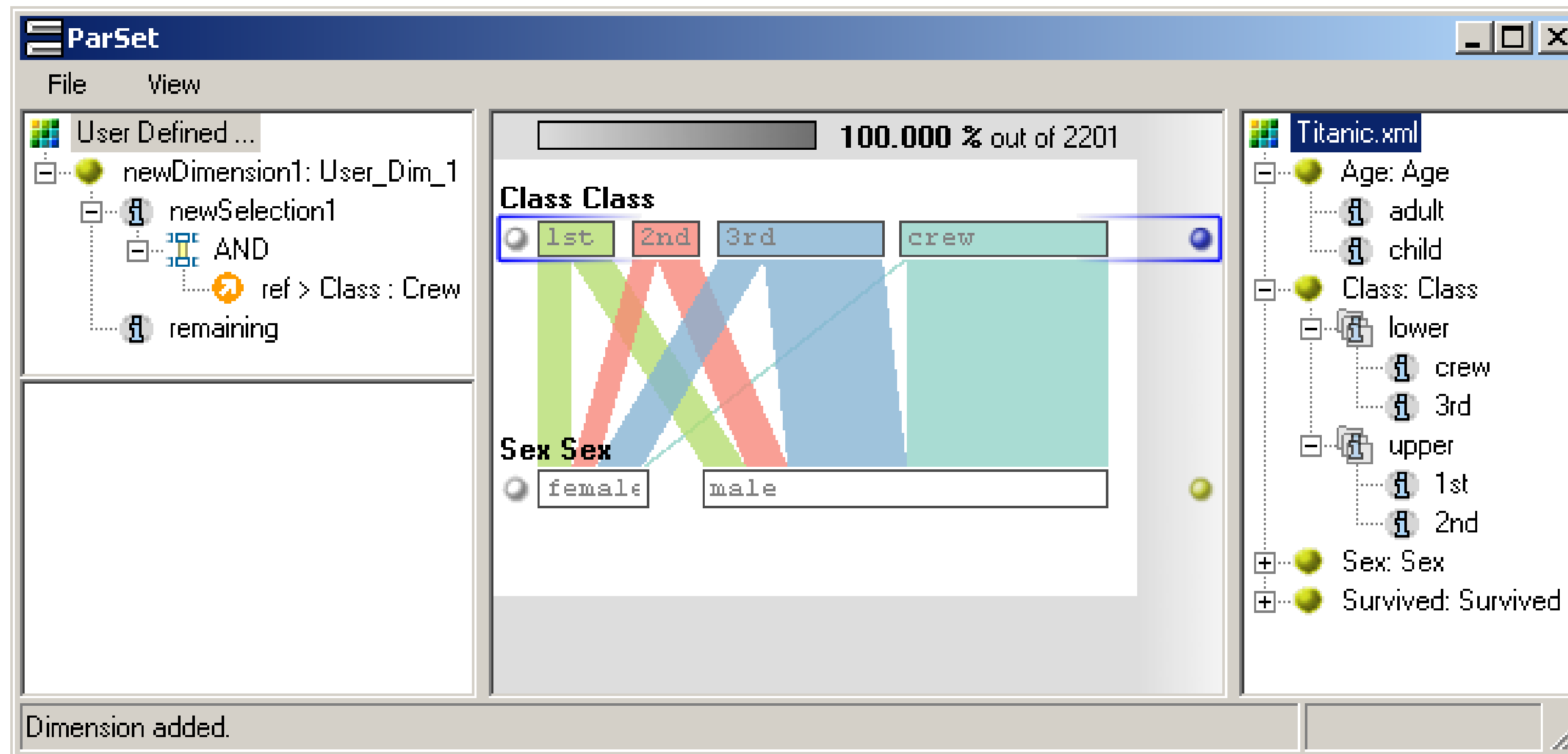
task: find relationship between attributes

interaction driven technique

Visual Encoding

boxes scaled by frequency

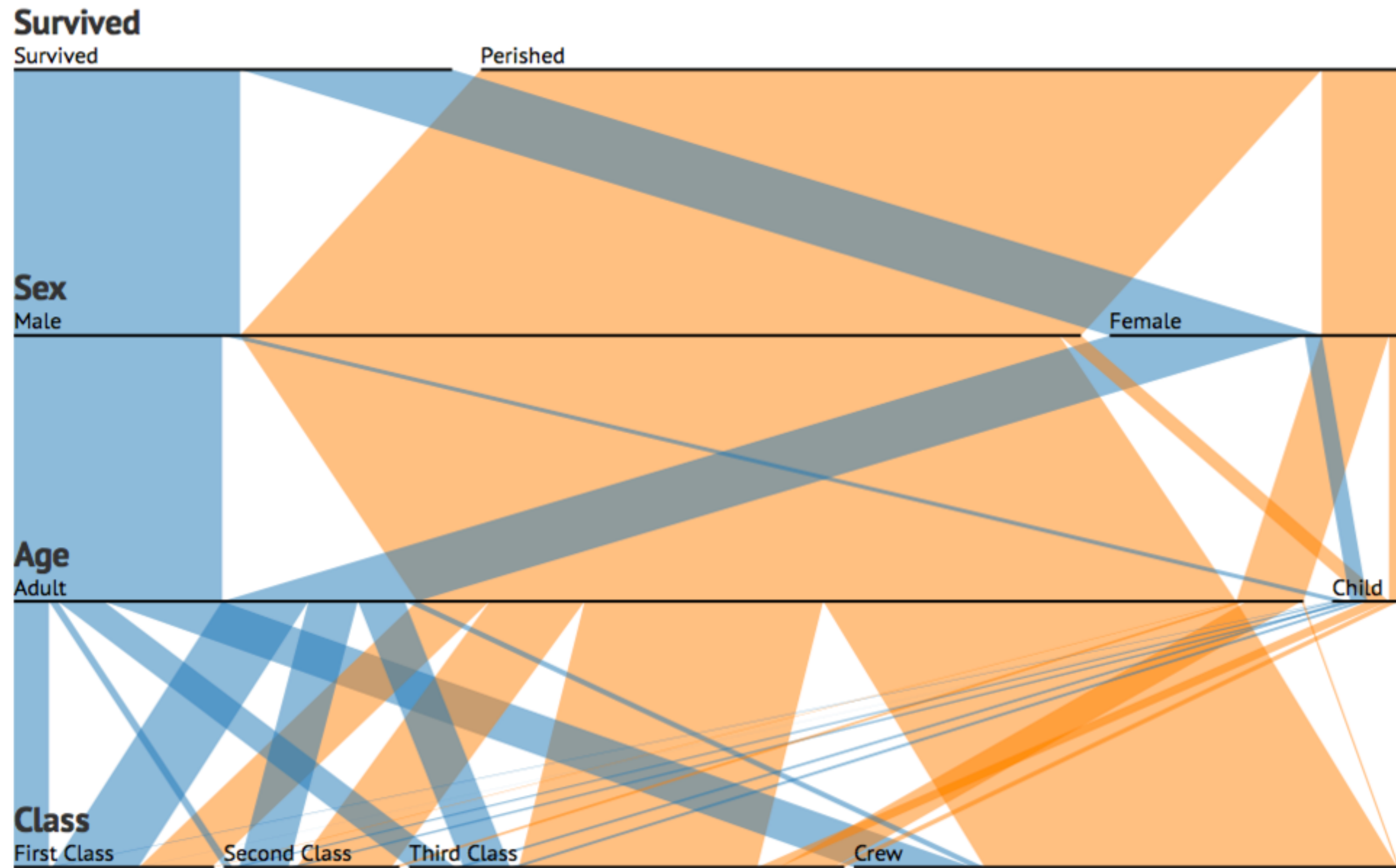
color coded by values for current active dimension



Parallel Sets

A visualisation technique for multidimensional categorical data.

Titanic Survivors

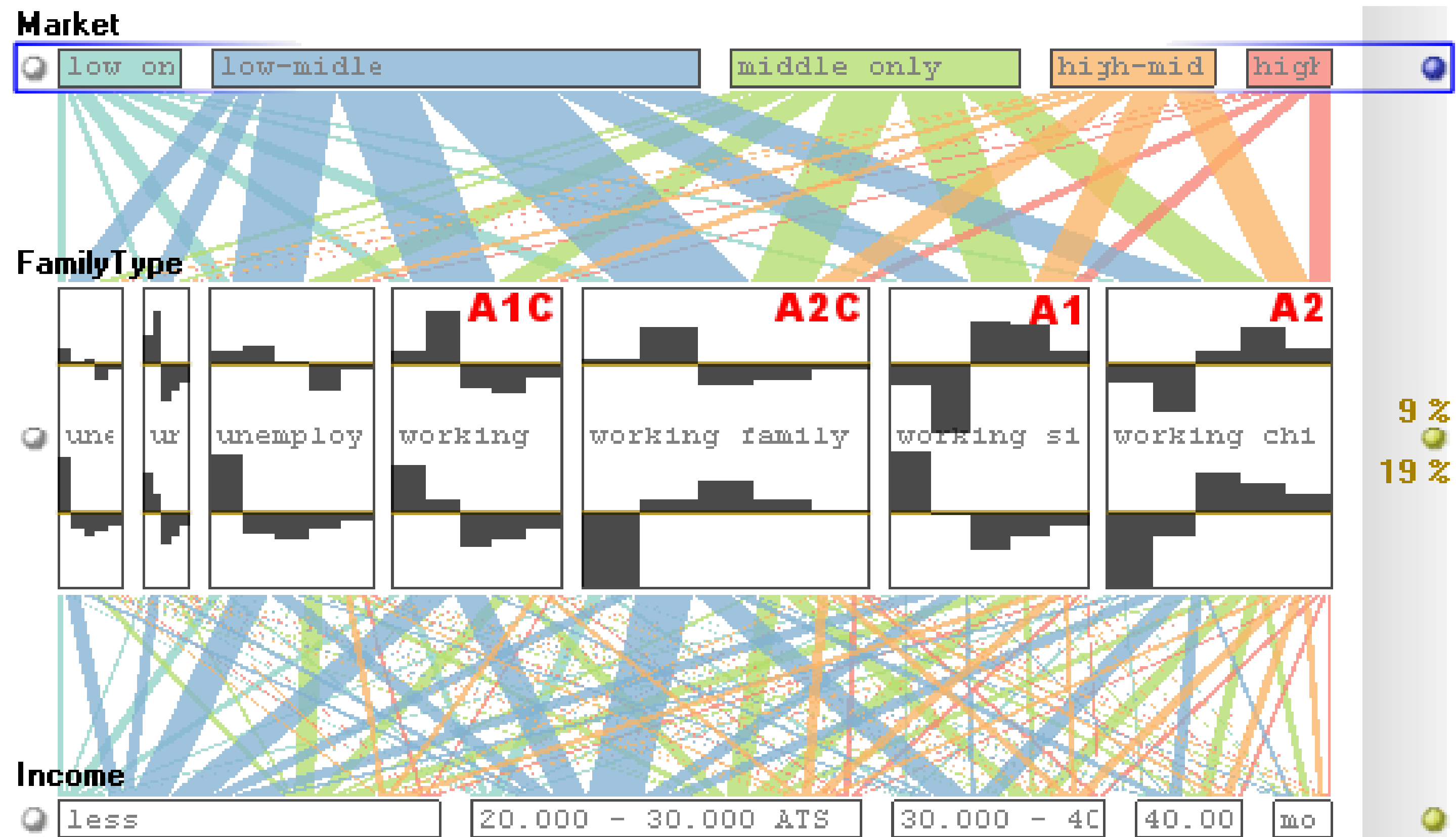


Curves?

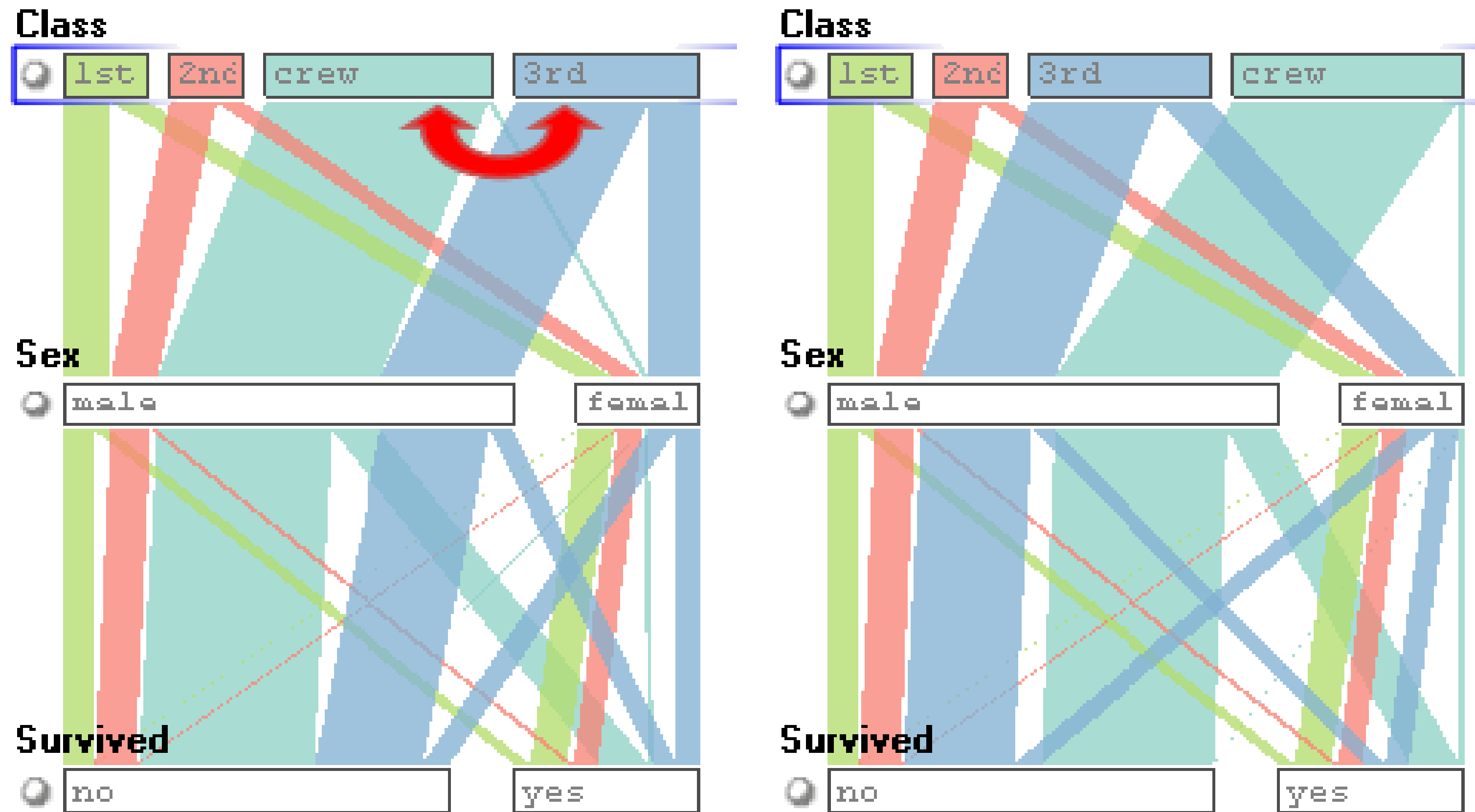
Data: [Robert J. MacG. Dawson.](#)

Visual Encoding

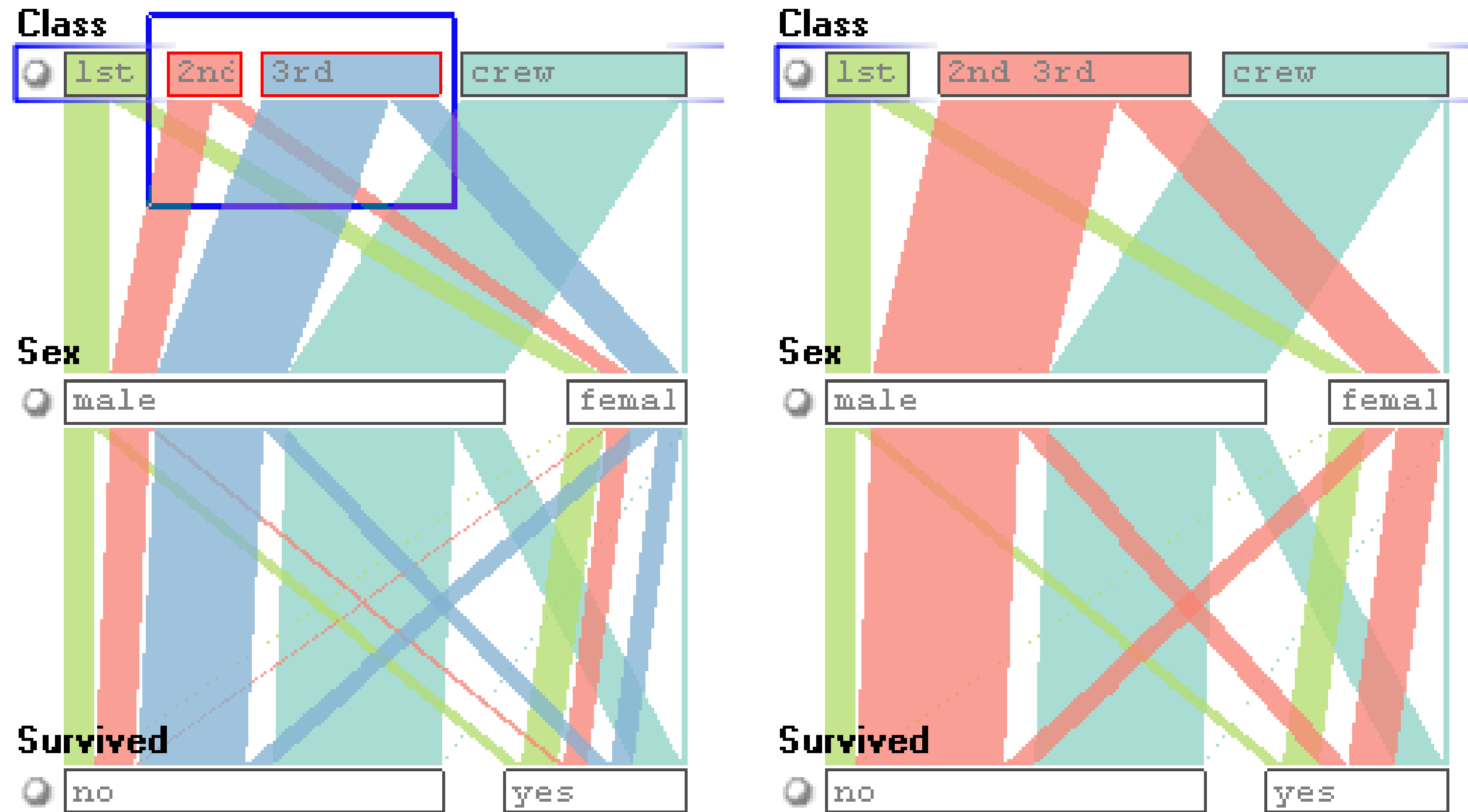
Boxes expand to show histogram



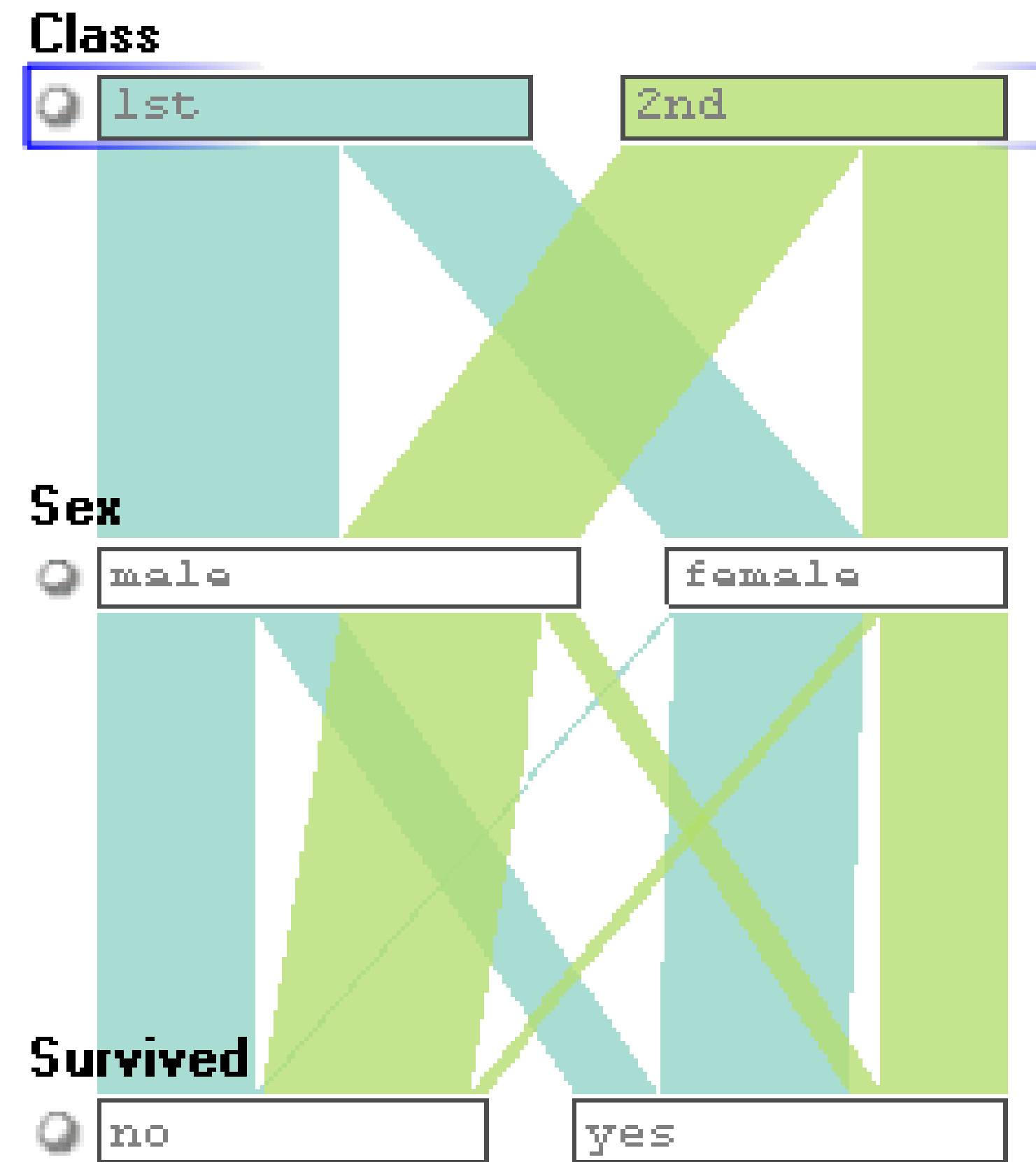
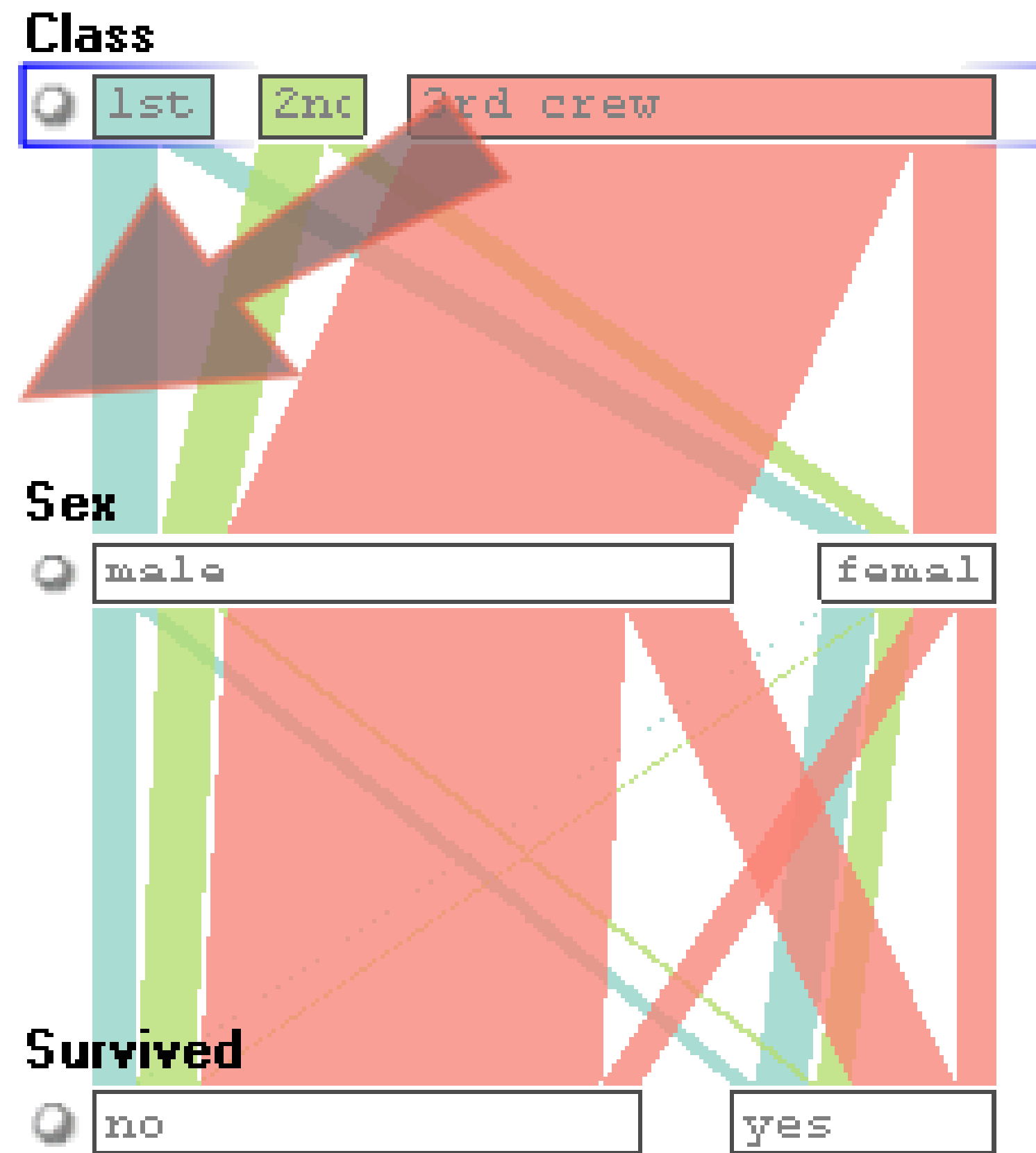
Interaction: Reorder



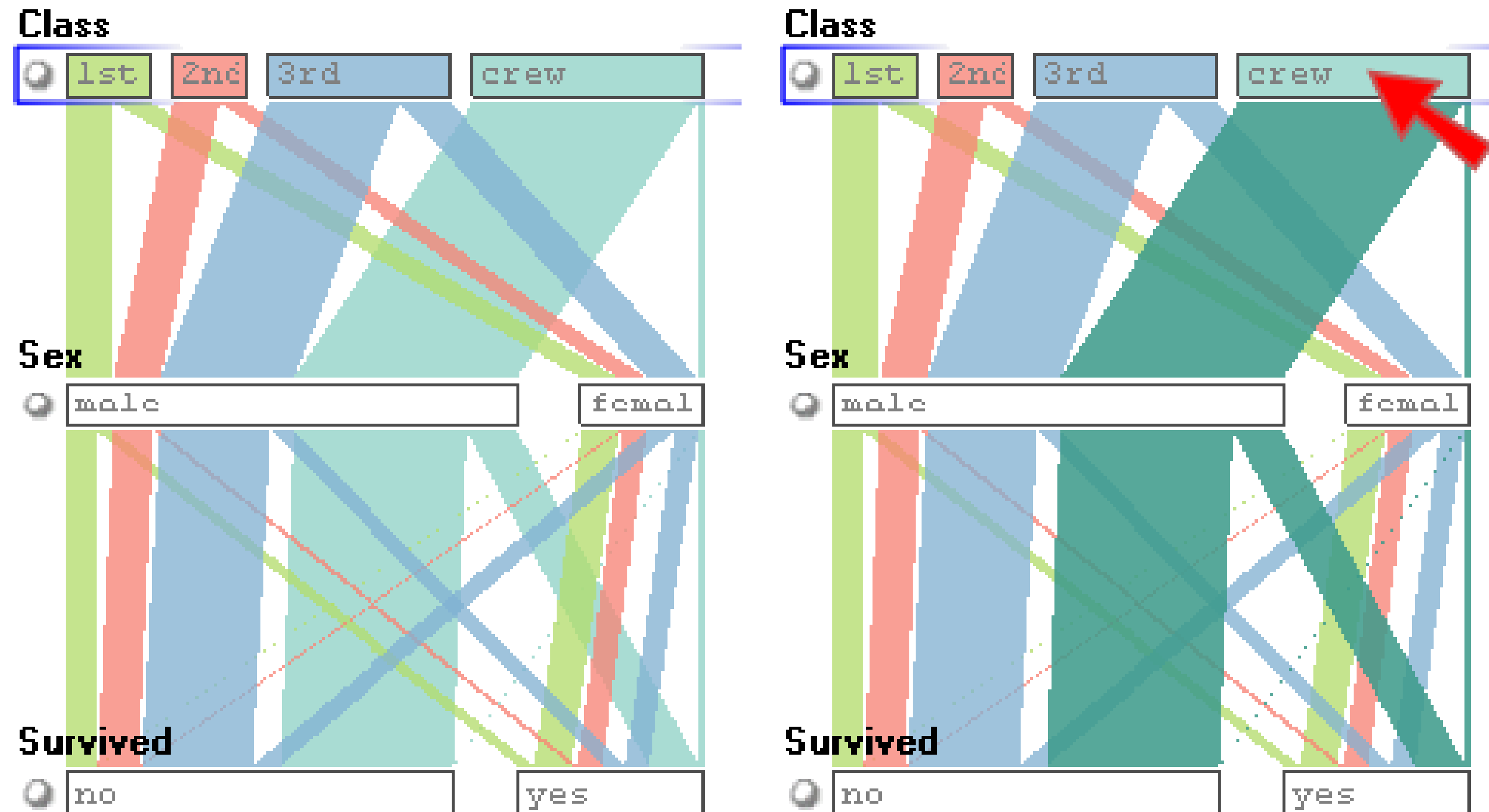
Interaction: Aggregate



Interaction: Filter



Interaction: Highlight



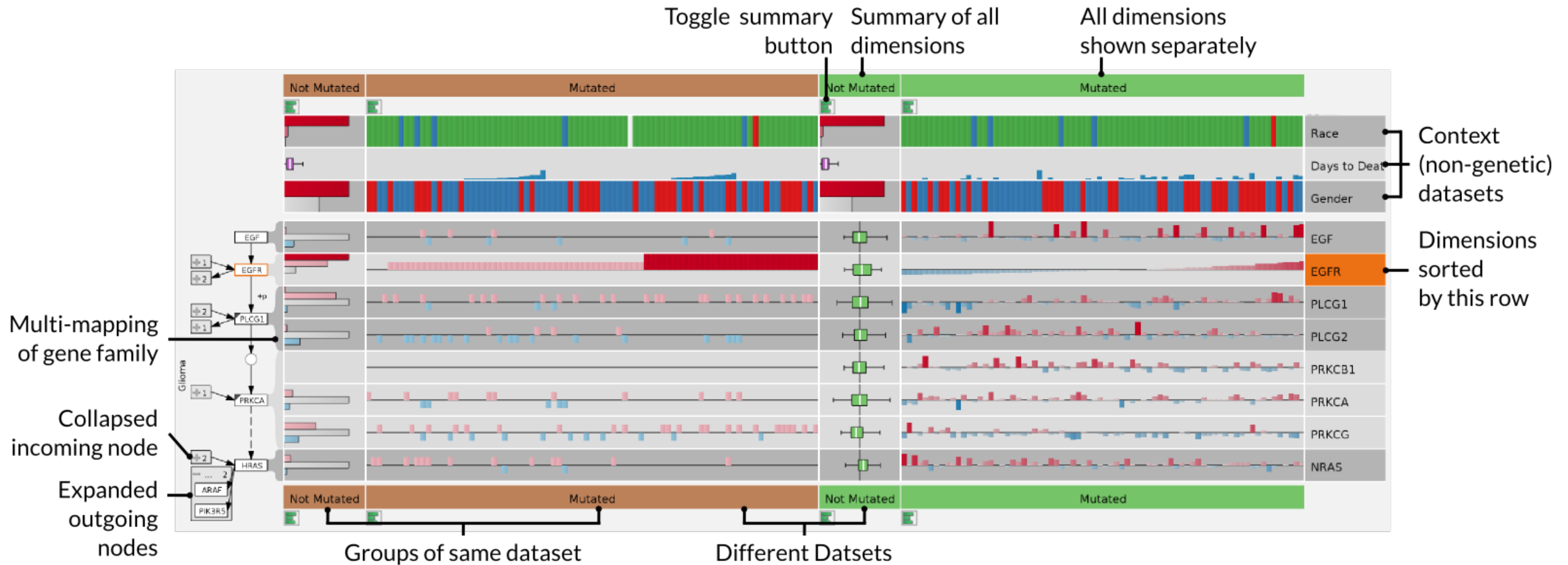
Tabular / Grid / Matrix - Based Representations

Tabular Representation

Like spreadsheet: each variable in it's own column

Visual encodings to make it scalable

Combining Various Charts



Taggle

A	Rank	S	AIDS_Countries	Continent	Human deve...	Ppl knowing...	Discriminat...	Urban Pop (...)	Discriminat...	Density (P/...
		<input type="checkbox"/>								
▼	1	<input type="checkbox"/>	Singapore	Asia	4 Very hig...	--	Unknown	--	Unknown	
	2	<input type="checkbox"/>	Malta	Europe	4 Very hig...	--	Unknown		Unknown	
	3	<input type="checkbox"/>	Bangladesh	Asia	2 Medium ...		Unknown		Unknown	
	4	<input type="checkbox"/>	Maldives	Asia	3 High hu...	--	Unknown		Unknown	
	5	<input type="checkbox"/>	Barbados	North Ame...	3 High hu...		Unknown		Unknown	
	6	<input type="checkbox"/>	Mauritius	Africa	3 High hu...		Unknown		Unknown	
	7	<input type="checkbox"/>	Lebanon	Asia	3 High hu...		Unknown		Unknown	
	8	<input type="checkbox"/>	Republic of Korea	Asia	4 Very hig...	--	Unknown		Unknown	
	9	<input type="checkbox"/>	Netherlands	Europe	4 Very hig...	--	Unknown		Unknown	
	10	<input type="checkbox"/>	Rwanda	Africa	1 Low hu...		Low		Low	
	11	<input type="checkbox"/>	Burundi	Africa	1 Low hu...		Slight		Slight	
	12	<input type="checkbox"/>	India	Asia	2 Medium ...	--	Unknown		Unknown	
	13	<input type="checkbox"/>	Haiti	North Ame...	1 Low hu...		Medium		Medium	
	14	<input type="checkbox"/>	Israel	Asia	4 Very hig...	--	Unknown		Unknown	
	15	<input type="checkbox"/>	Belgium	Europe	4 Very hig...	--	Unknown		Unknown	
	16	<input type="checkbox"/>	Philippines	Asia	2 Medium ...		Unknown		Unknown	
	17	<input type="checkbox"/>	Japan	Asia	4 Very hig...	--	Unknown		Unknown	
	18	<input type="checkbox"/>	Sri Lanka	Asia	3 High hu...		Unknown		Unknown	
	19	<input type="checkbox"/>	Viet Nam	Asia	2 Medium ...		Unknown		Unknown	
	20	<input type="checkbox"/>	El Salvador	North Ame...	2 Medium ...		Unknown		Unknown	
	21	<input type="checkbox"/>	United Kingdom of Great Britain	Europe	4 Very hig...	--	Unknown		Unknown	
	22	<input type="checkbox"/>	Trinidad and Tobago	North Ame...	3 High hu...		Unknown		Unknown	
	23	<input type="checkbox"/>	Jamaica	North Ame...	3 High hu...		Predomina...		Predomina...	
	24	<input type="checkbox"/>	Pakistan	Asia	2 Medium ...		Medium		Medium	
	25	<input type="checkbox"/>	Germany	Europe	4 Very hig...	--	Unknown		Unknown	
	26	<input type="checkbox"/>	Luxembourg	Europe	4 Very hig...	--	Unknown		Unknown	
	27	<input type="checkbox"/>	Dominican Republic	North Ame...	3 High hu...		Medium		Medium	
	28	<input type="checkbox"/>	Switzerland	Europe	4 Very hig...	--	Unknown		Unknown	
	29	<input type="checkbox"/>	Nigeria	Africa	1 Low hu...		Medium		Medium	
	30	<input type="checkbox"/>	Democratic People's Republic	Asia	2 Medium ...	--	Unknown		Unknown	

Pixel Based Displays

Each cell is a “pixel”, value encoded in color / value

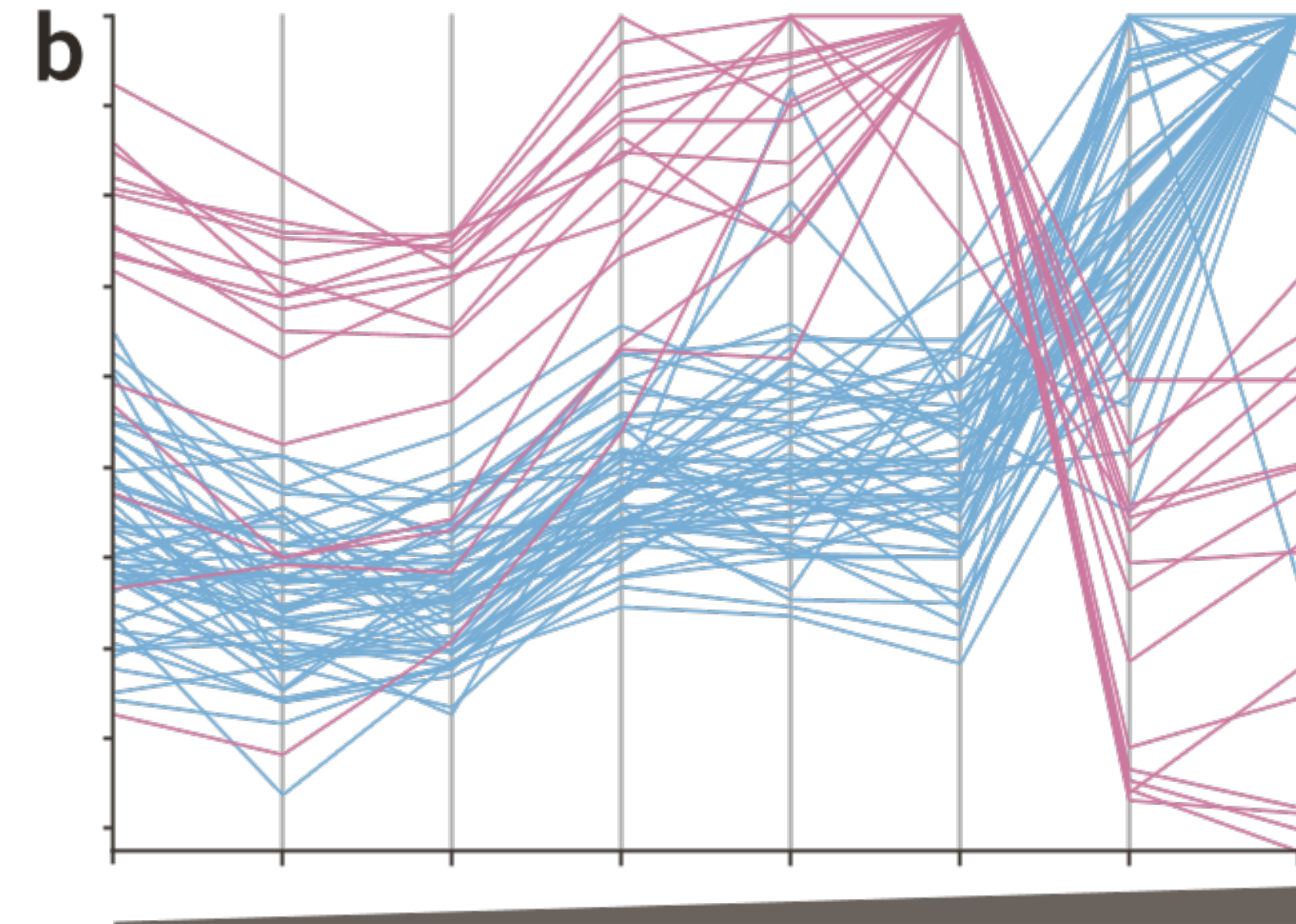
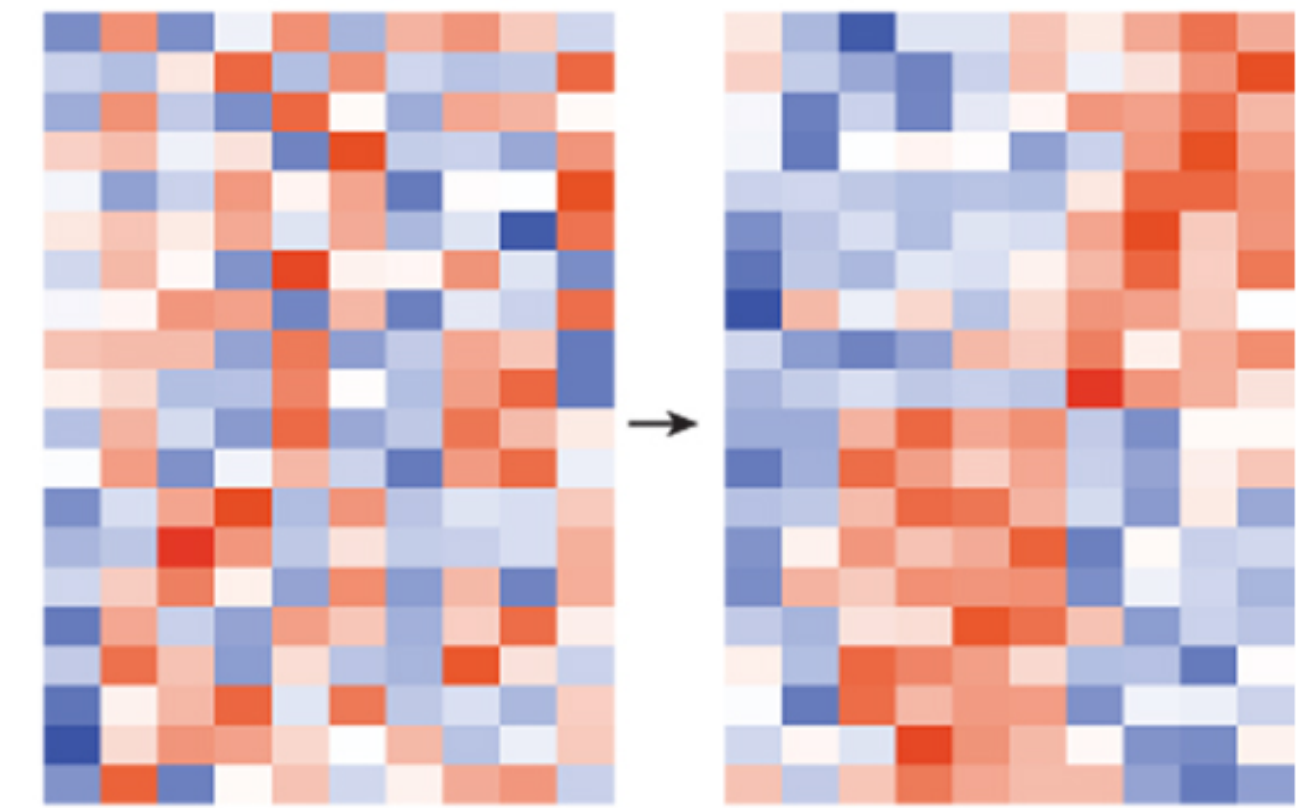
Ordering critical for interpretation

If no ordering inherent, clustering is used

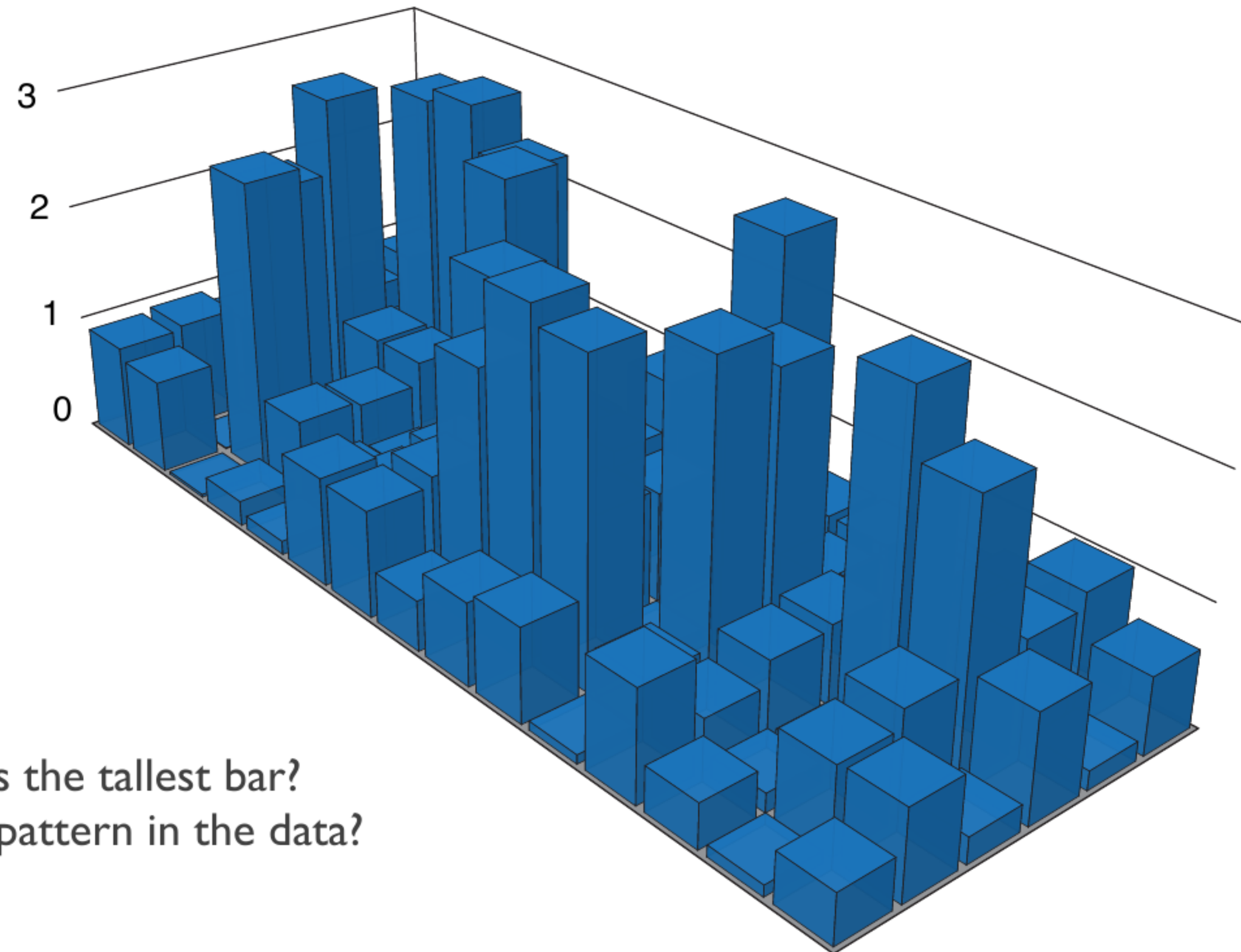
Scalable – 1 px per item

Good for homogeneous data

same scale & type

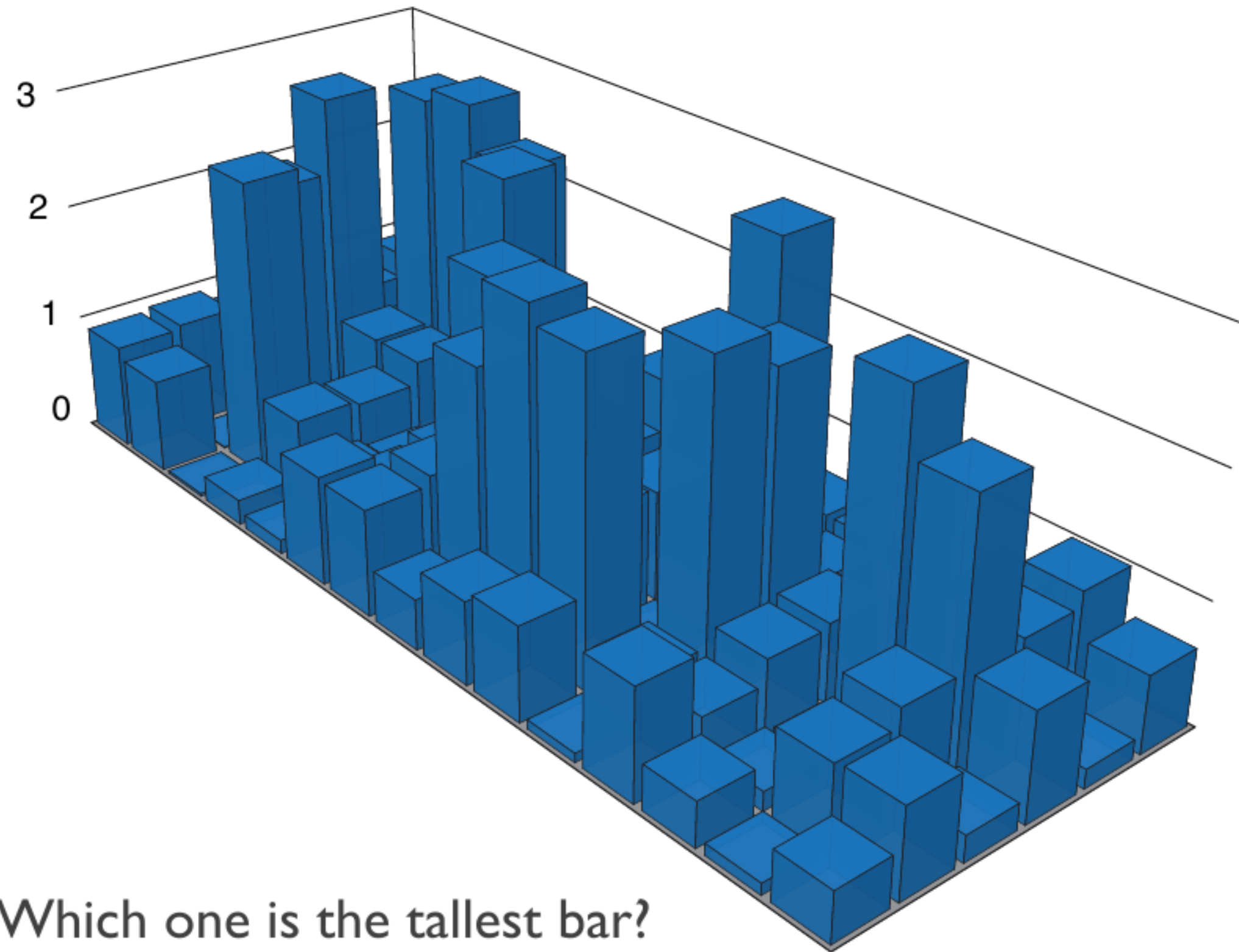


3D Pitfall: Occlusion & Perspective

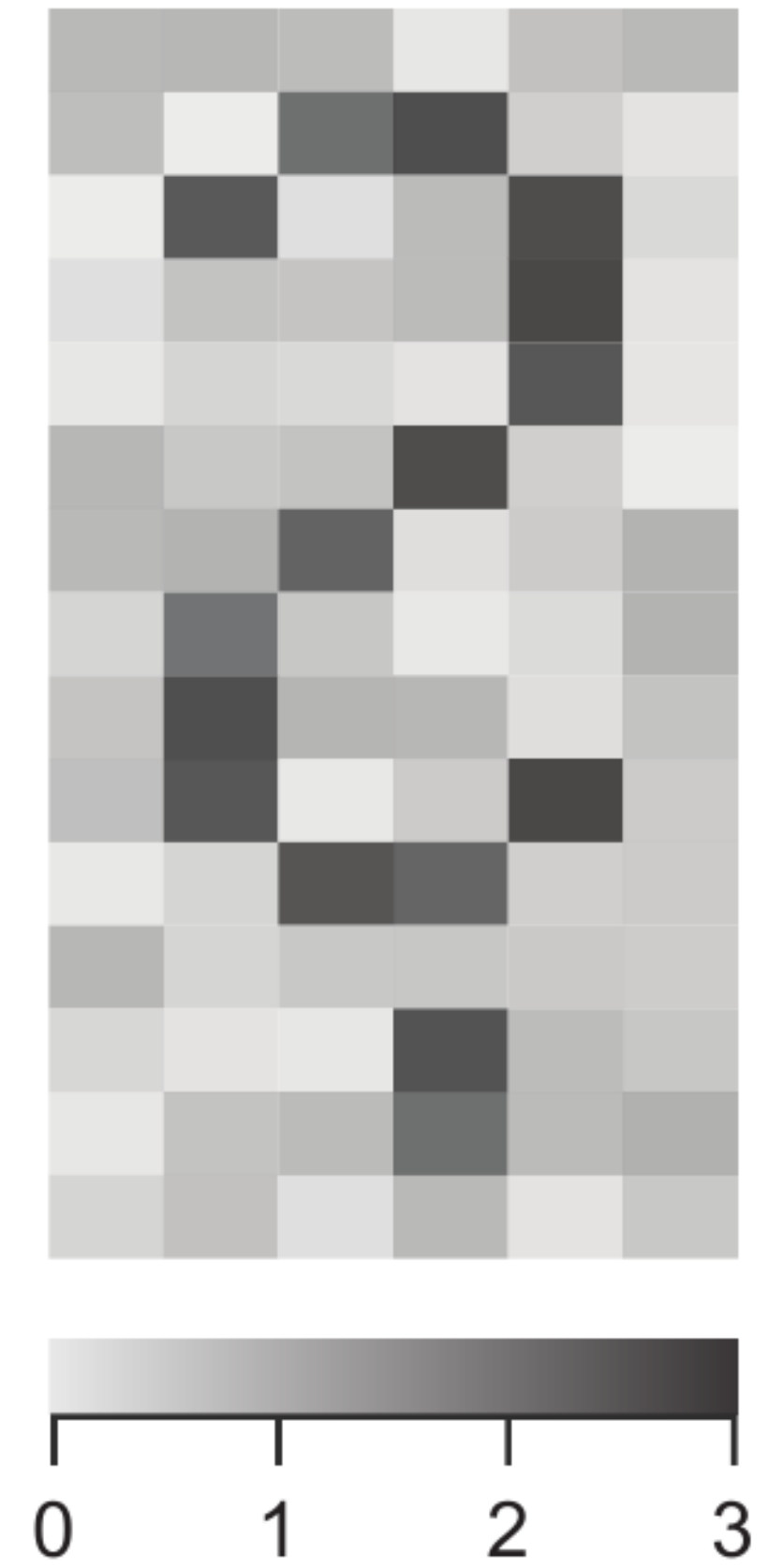


Which one is the tallest bar?
What is the pattern in the data?

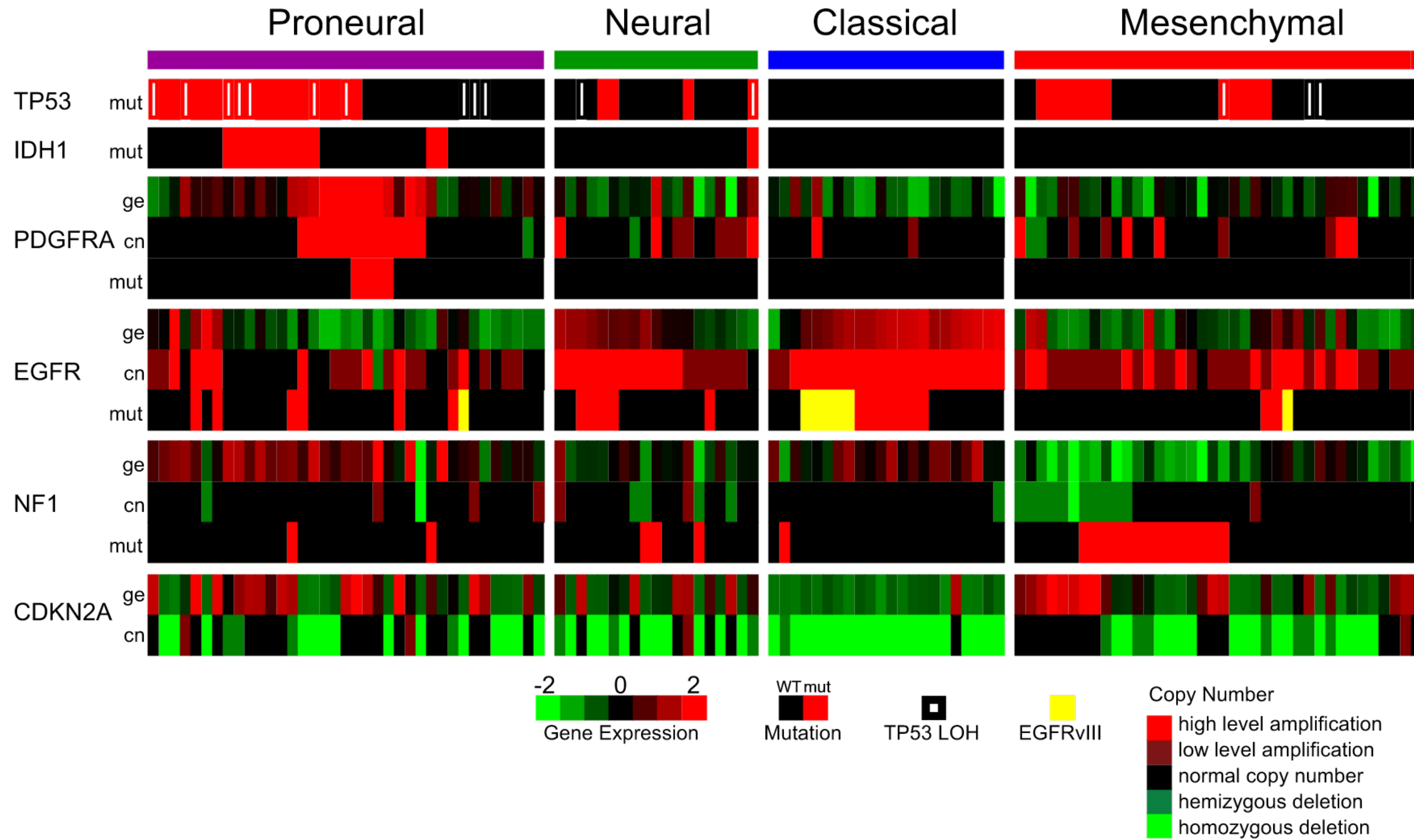
3D Pitfall: Occlusion & Perspective



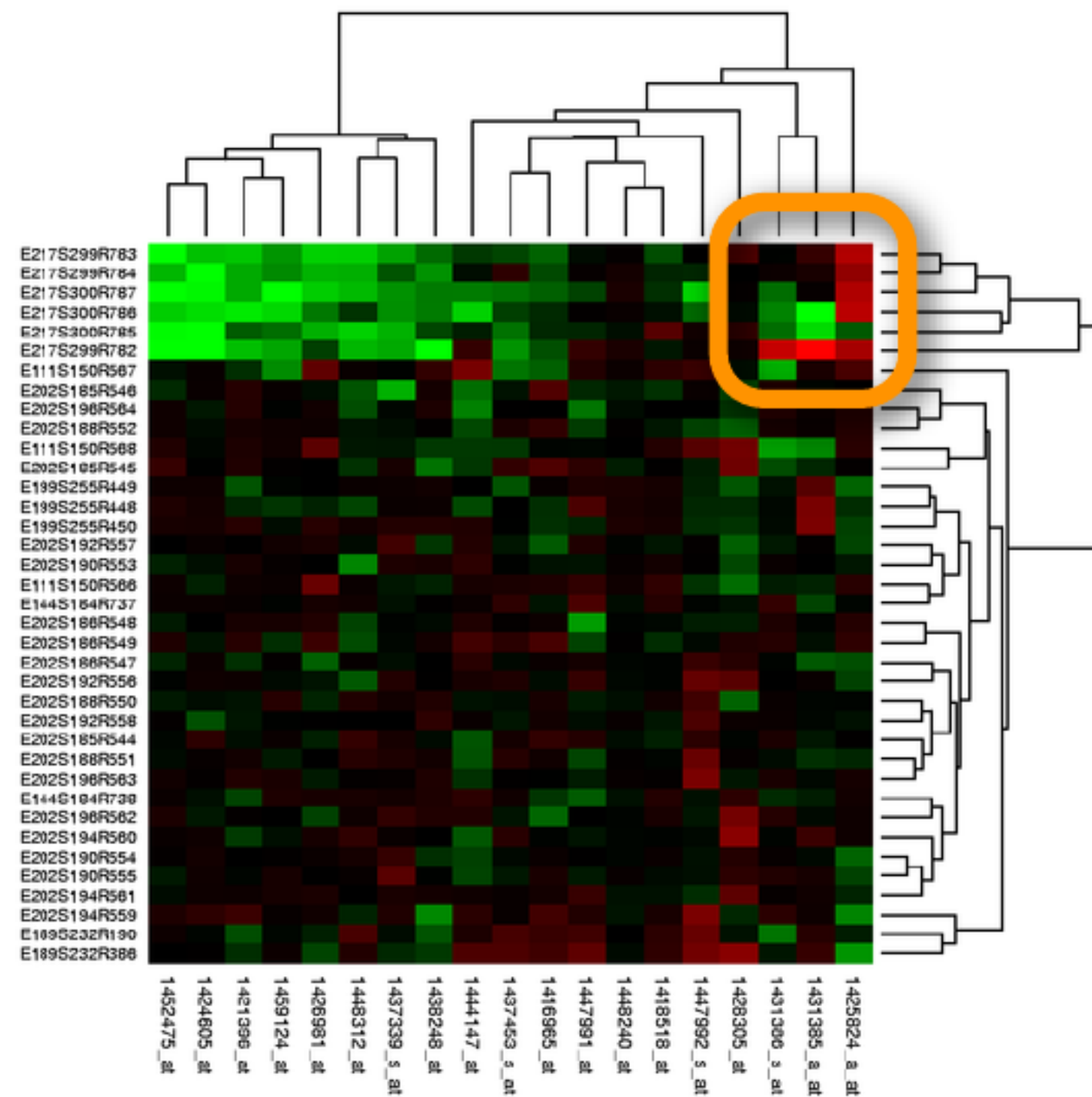
Which one is the tallest bar?
What is the pattern in the data?



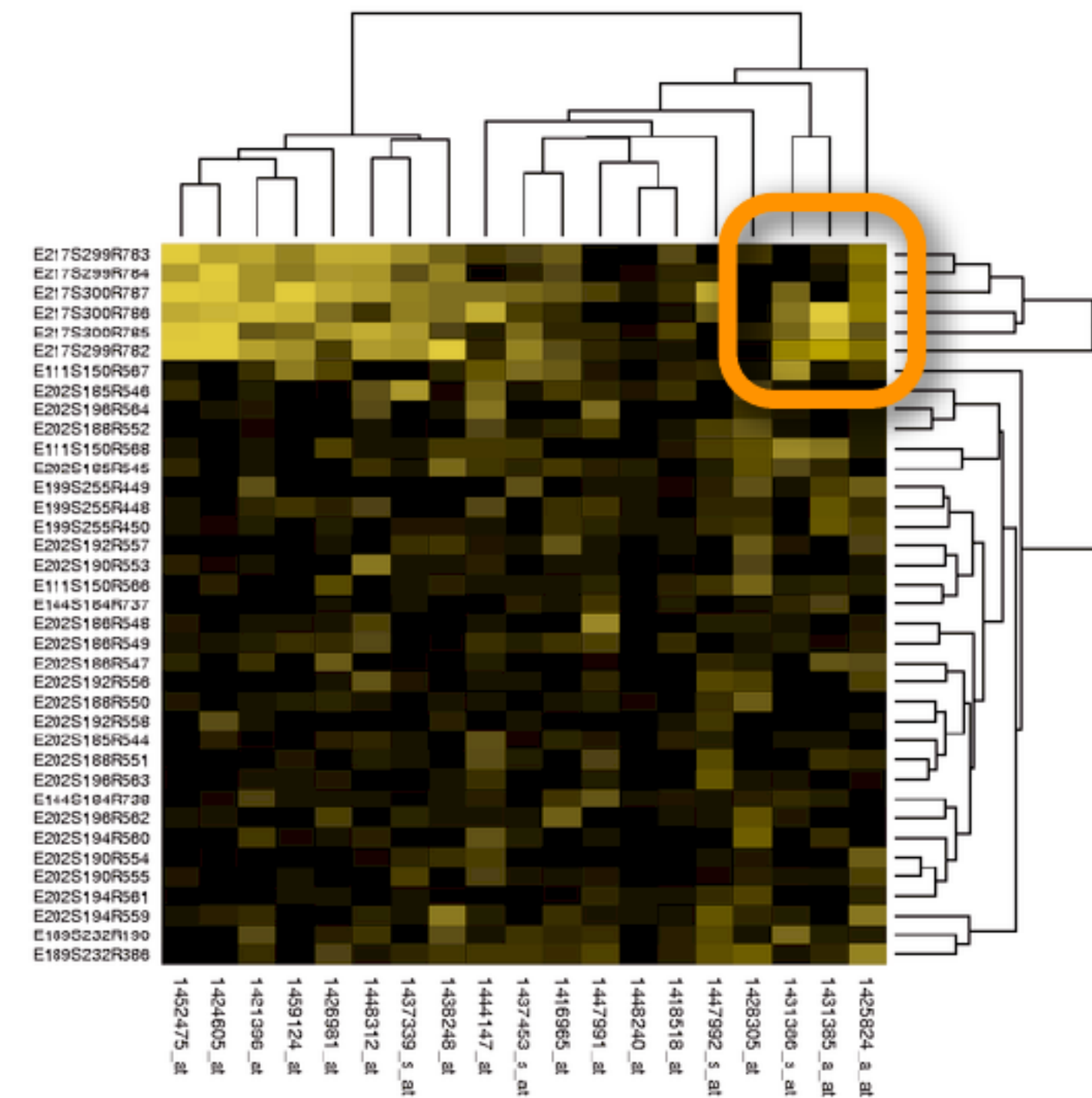
Heterogeneous Data?



Bad Color Mapping

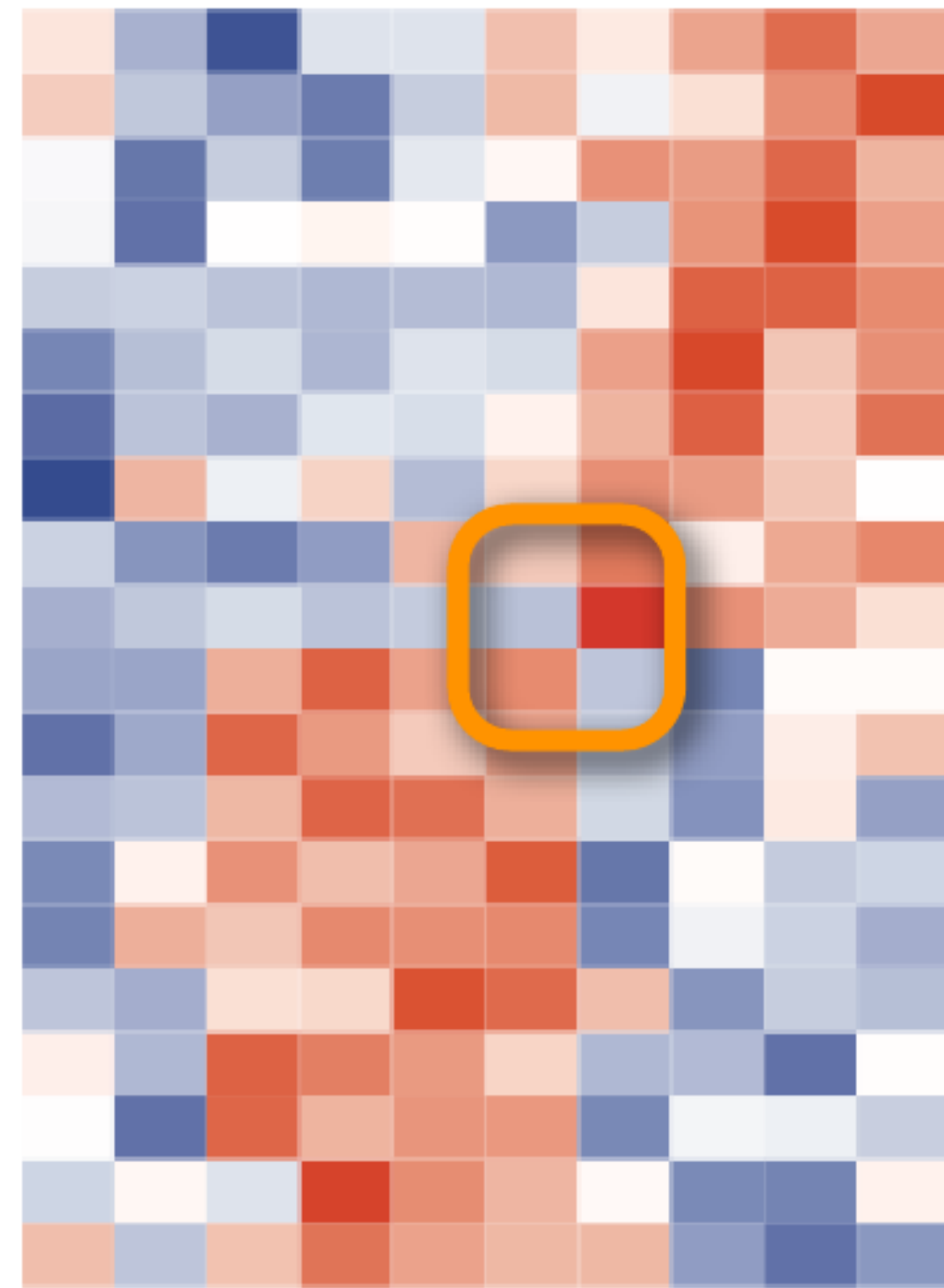


Normal Vision

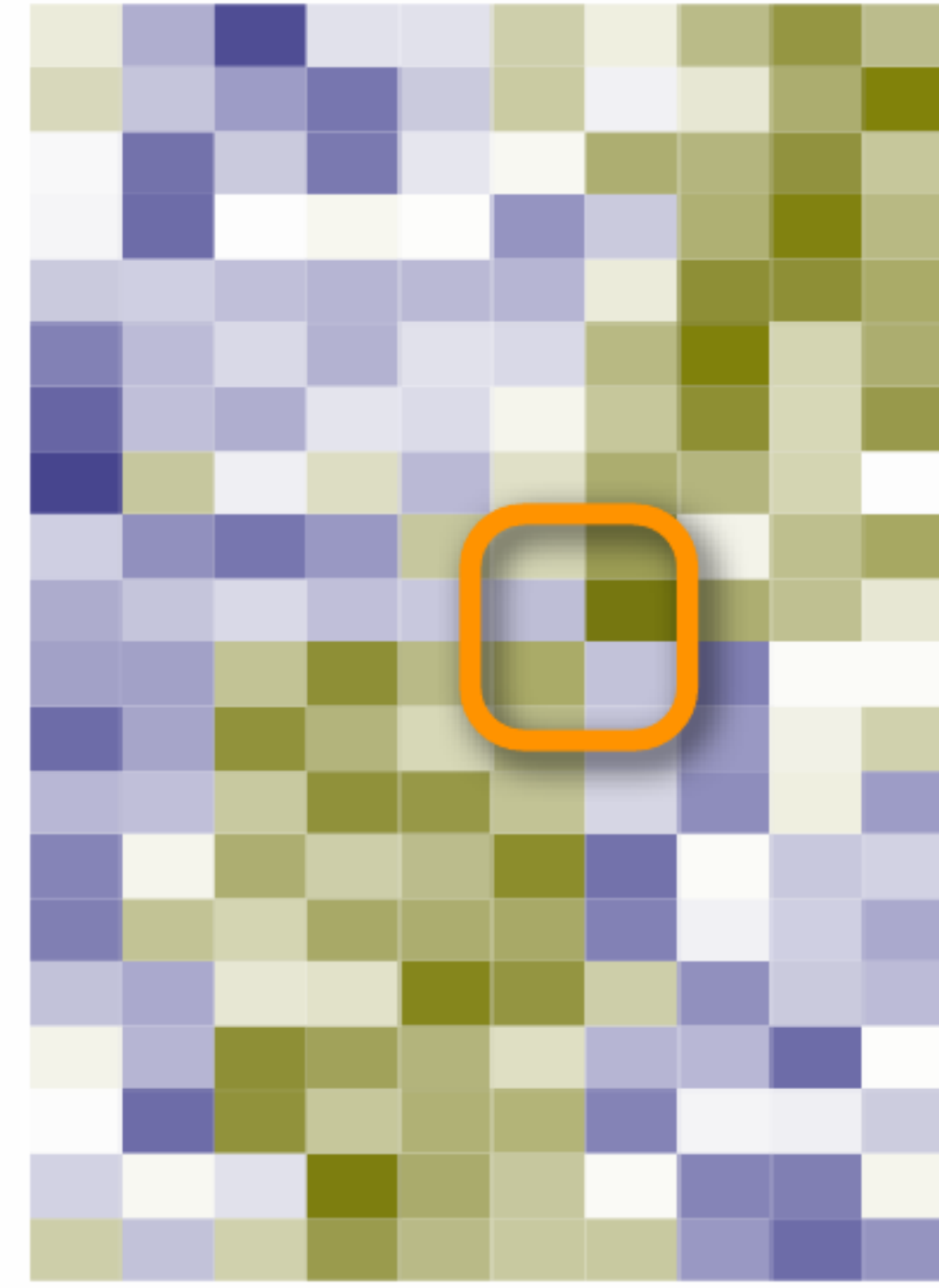


Deuteranope Vision
("Red-Green Blindness")

Good Color Mapping

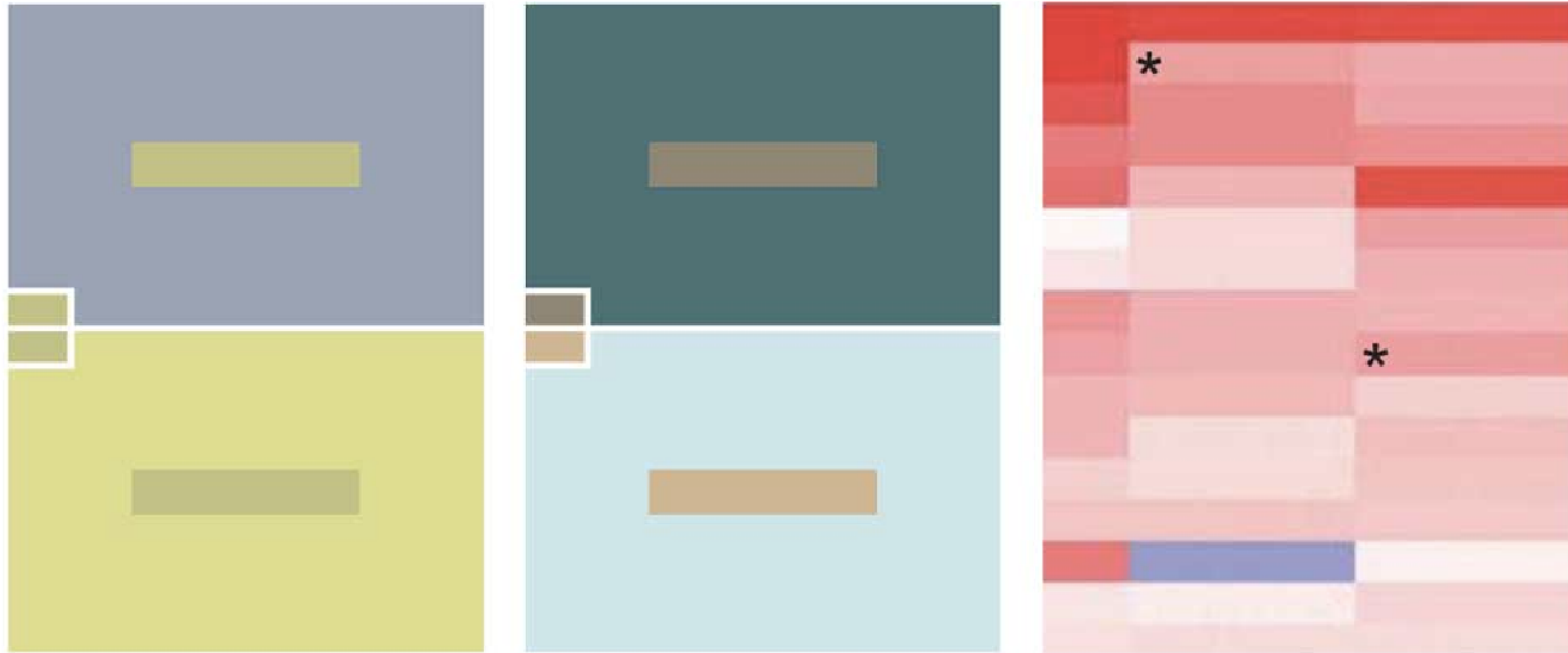


Normal Vision

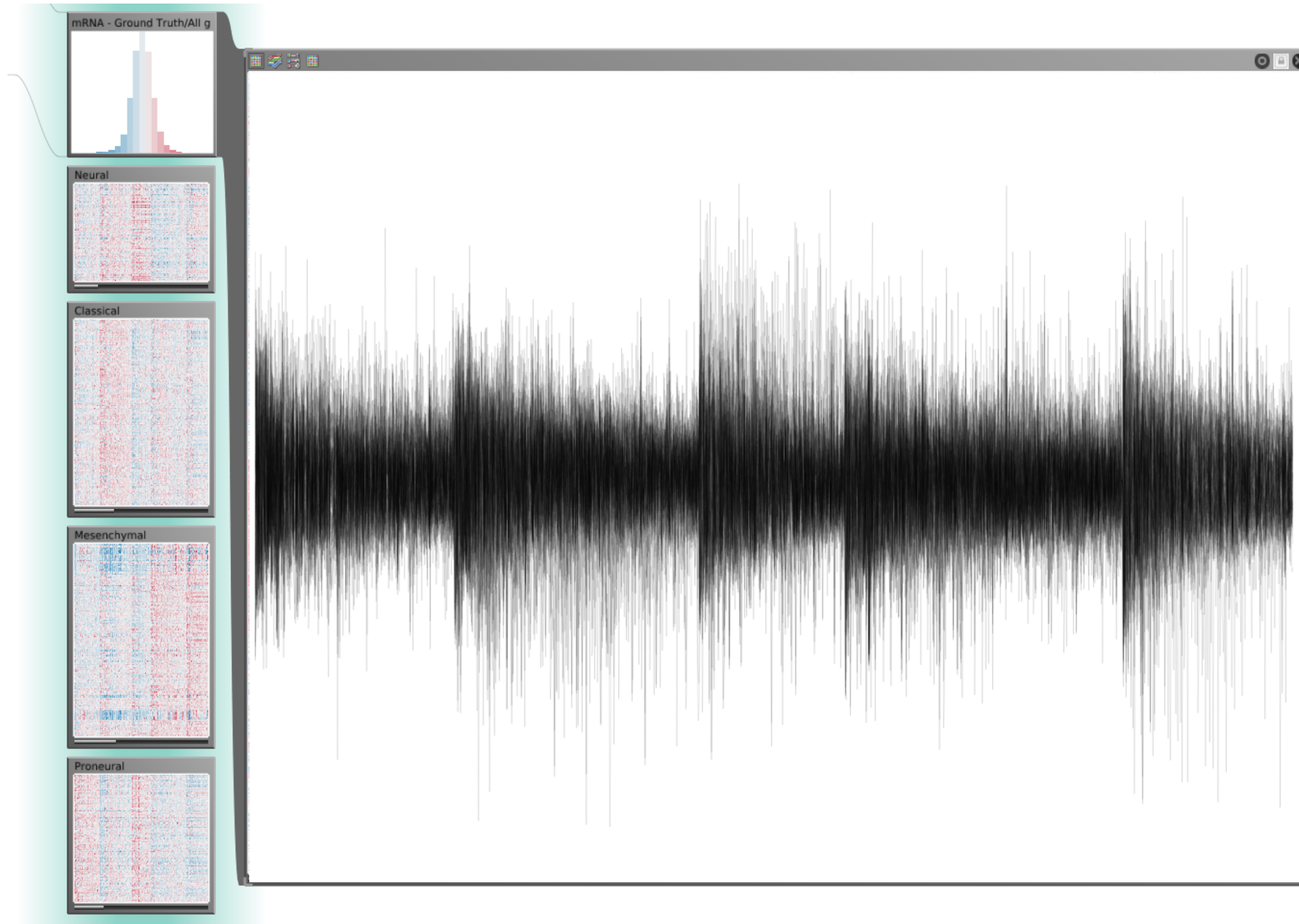


Deuteranope Vision
("Red-Green Blindness")

Color is relative!



Clustered Heat Map



Filling Space

Non-Tabular Space Filling Layouts

HiVE example: London property

partitioning attributes

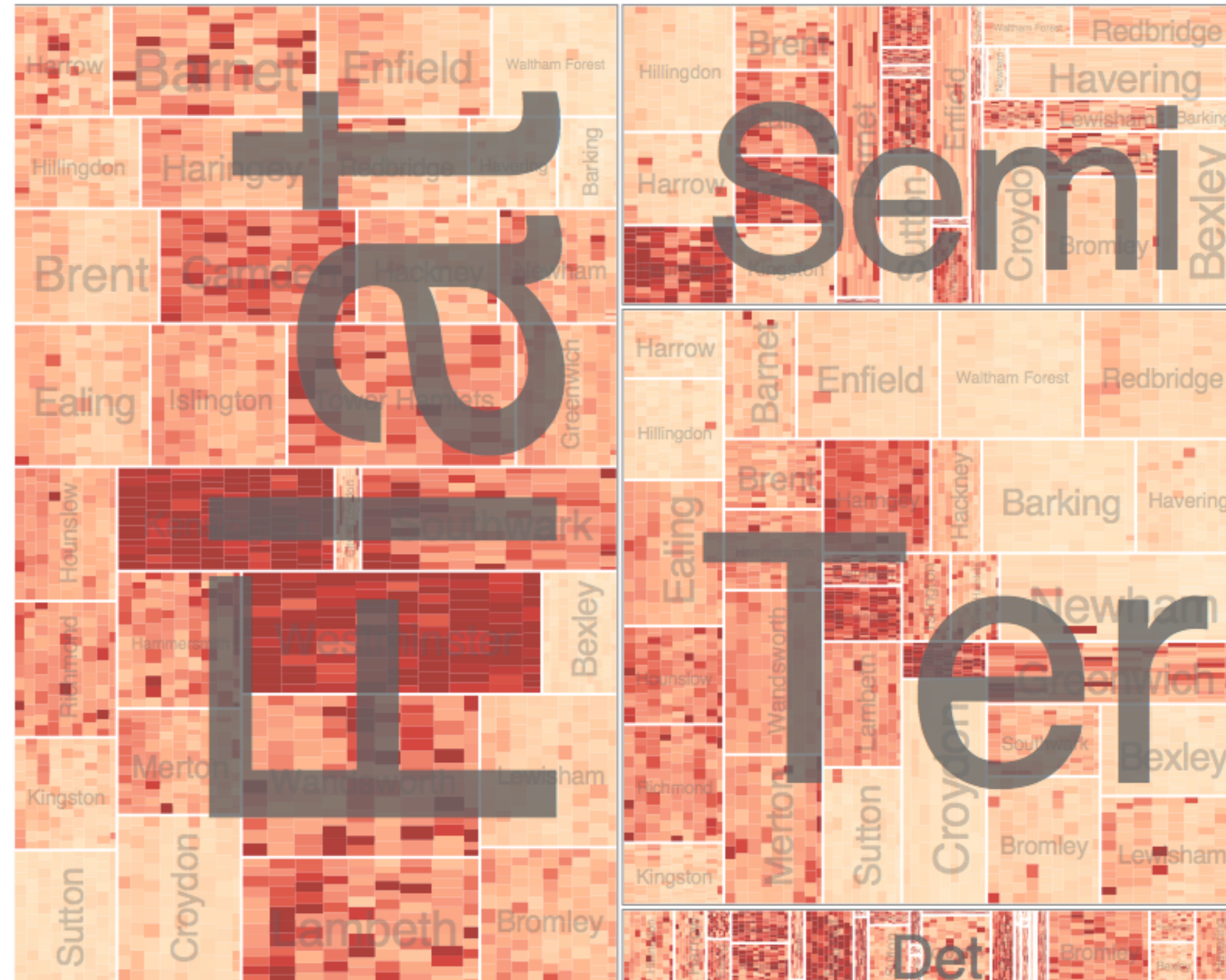
house type
neighborhood
sale time

encoding attributes

average price (color)
number of sales (size)

results

between neighborhoods,
different housing distributions
within neighborhoods,
similar prices



Dense pixel display: VisDB

represent each data item, or each attribute in an item as a single pixel

can fit as many items on the screen as there are pixels, on the order of millions

relies heavily on color coding

challenge: what's the layout?

The data...

large database where each item has multiple attributes (on the order of 10)

goal: visualize the relevance of set of items which satisfy a query

plot out data items in a spiral pattern, ordered by relevance

