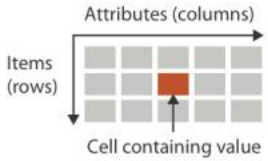
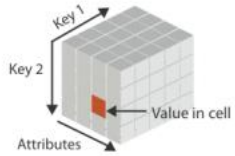
TABLES

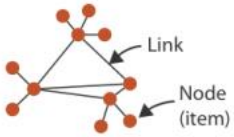
DATASET TYPES

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

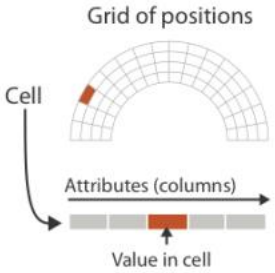


→ *Multidimensional Table*





→ *Trees*



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” visualization

~1000 – need analytical methods

How many records?

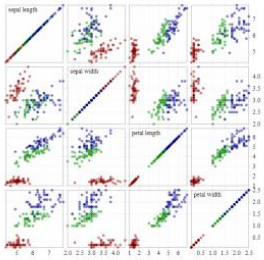
~ 1000 – “just” vis is fine

>> 10 000 – need analytical methods

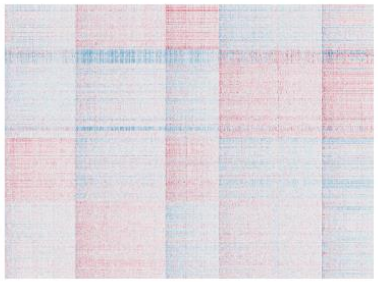
Homogeneity

- Same data type?
- Same scales?

ANALYTIC COMPONENT

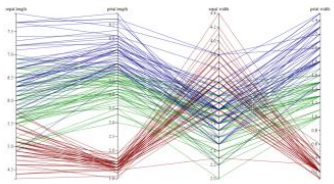


Scatterplot Matrices
[Bostock]

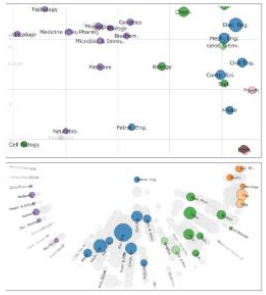


Pixel-based visualizations / heat maps

Multidimensional Scaling
[Doerk 2011]



Parallel Coordinates
[Bostock]



[Chuang 2012]

no / little analytics ⟷ **strong analytics component**

TECHNIQUES AND TASKS

Magnitude

Distribution

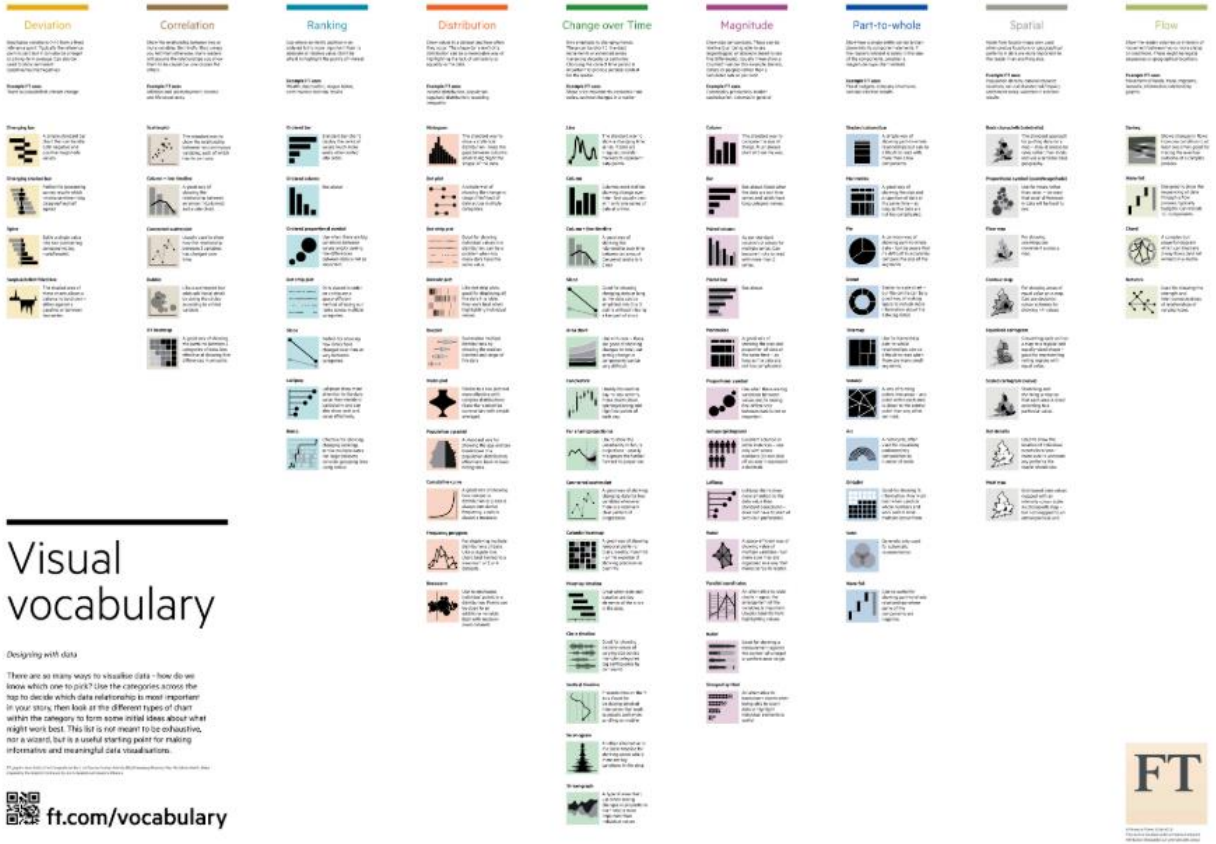
Ranking

Part to whole

Deviation

Correlation

Change over time



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

MAGNITUDE

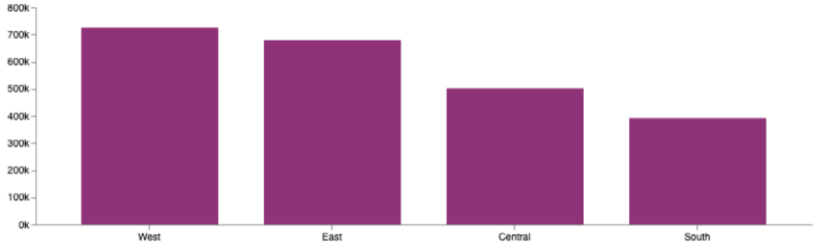
Show size comparison

- Relative: to see bigger/larger
- Absolute: to see differences

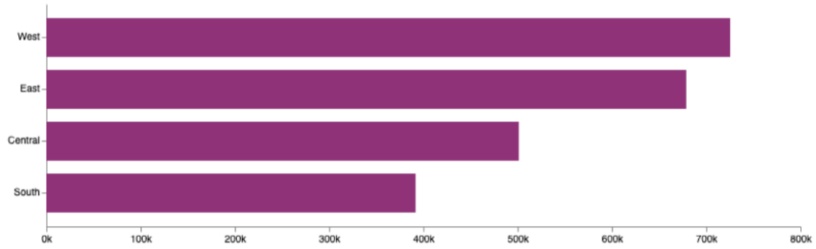
Usually for “counted” numbers

Rather than a calculated rate or per cent

BAR CHARTS

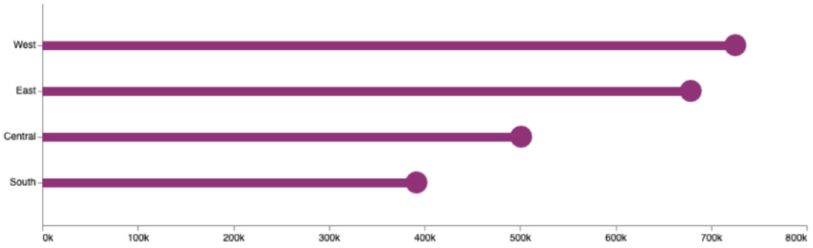


Vertical Bar Chart / Column Chart



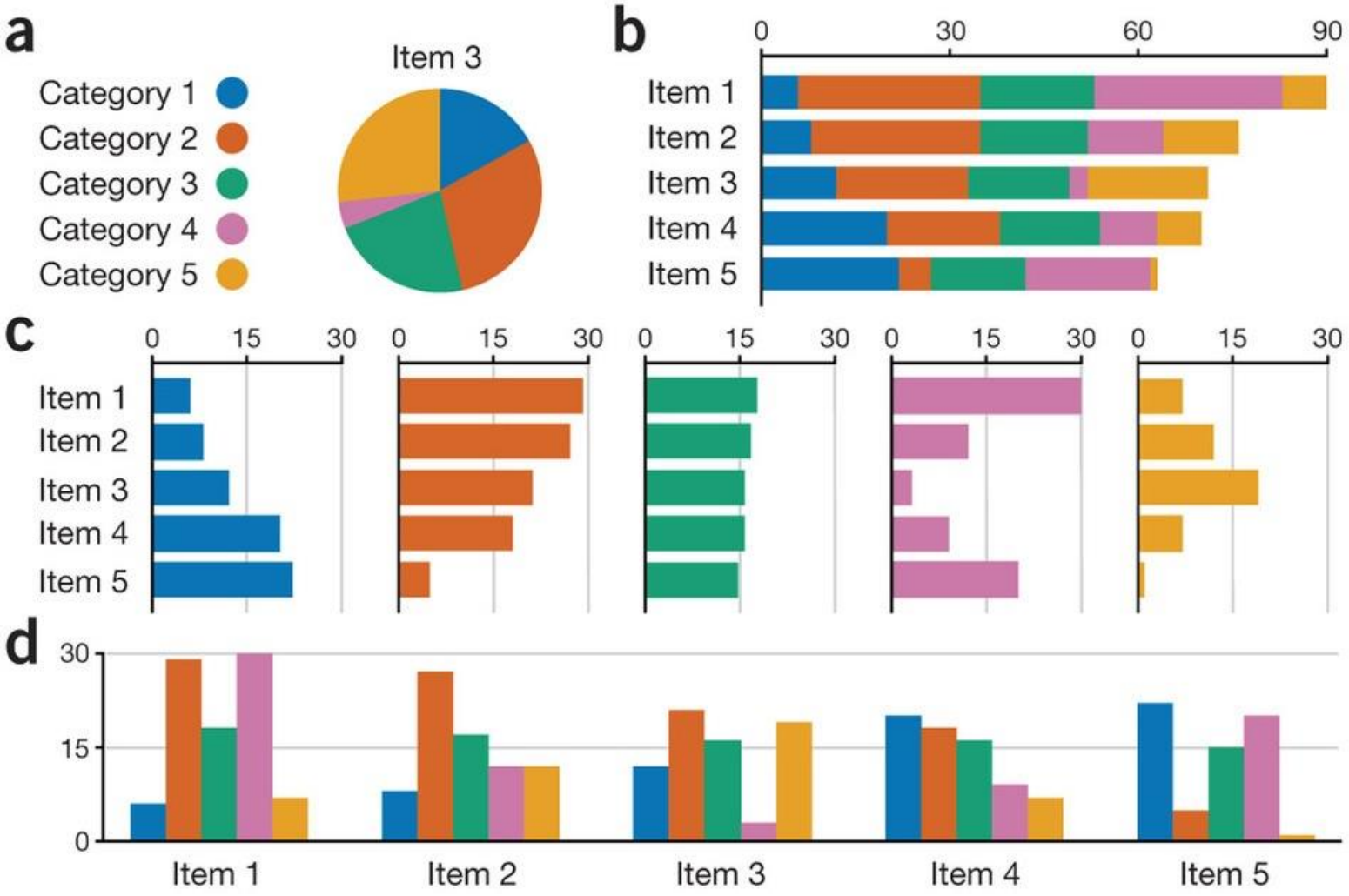
Horizontal Bar Chart

Grouped Bar Chart



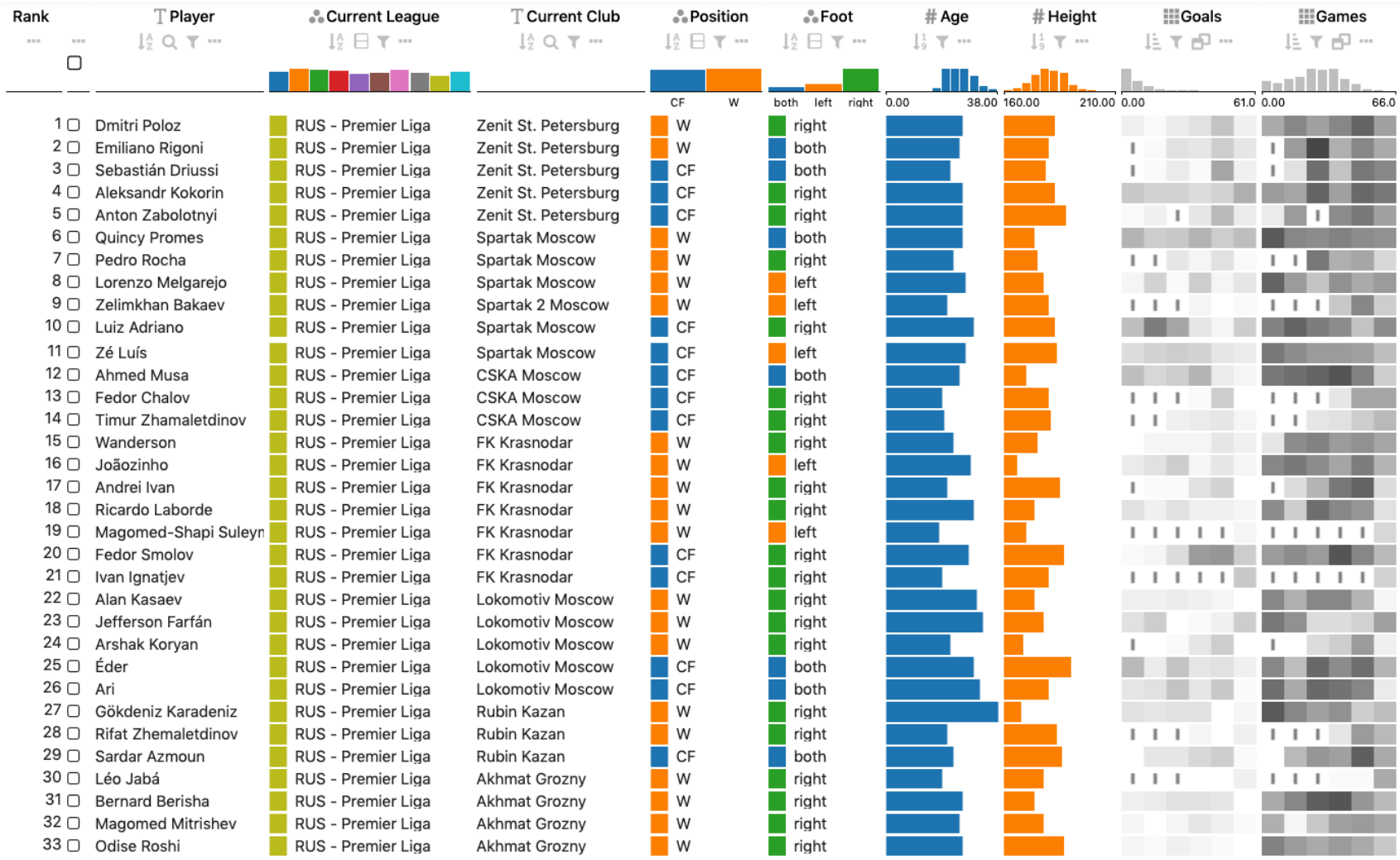
Lollipop Chart

BAR CHARTS TYPES



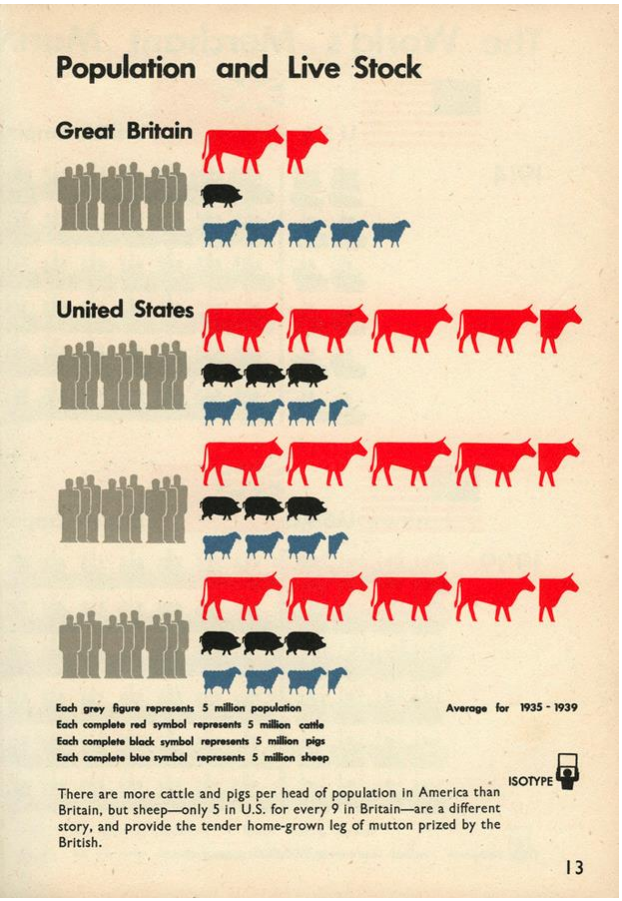
Streit & Gehlenborg, PoV, Nature Methods, 2014

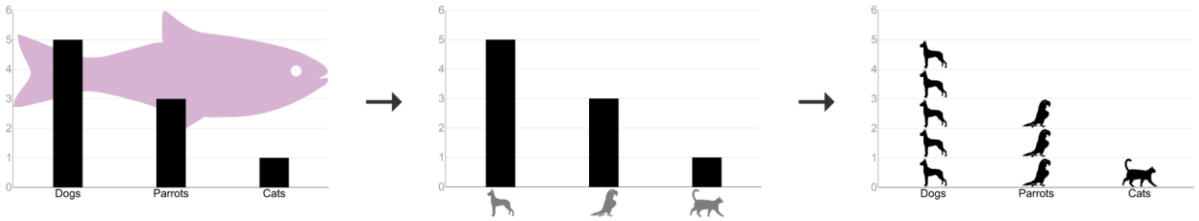
BAR CHARTS IN COMPLEX VIS



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

ISO TYPE VISUALIZATION





Otto and Marie Neurath

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

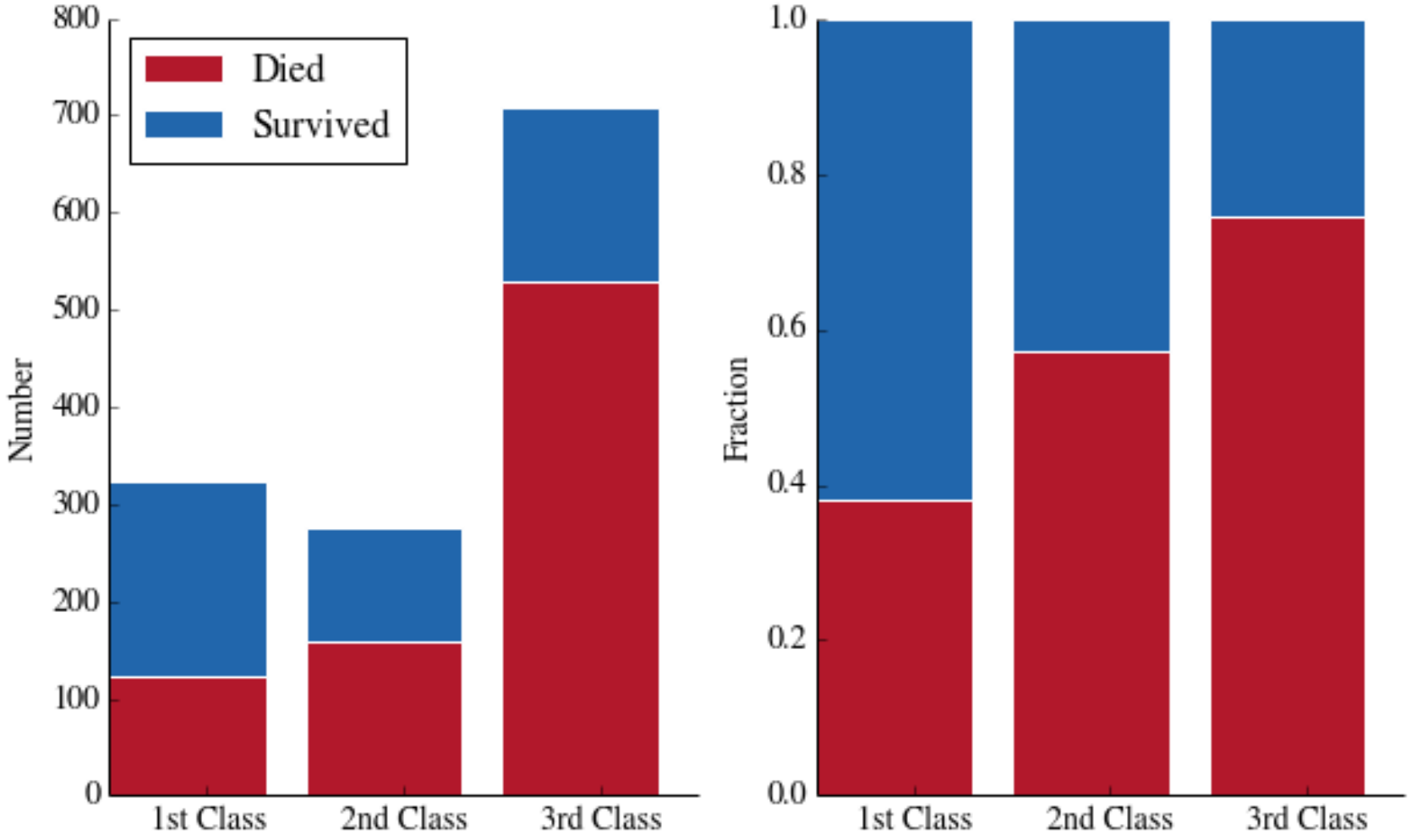
PART OF WHOLE

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

If the interest is solely in the size of the components

Consider magnitude-type chart

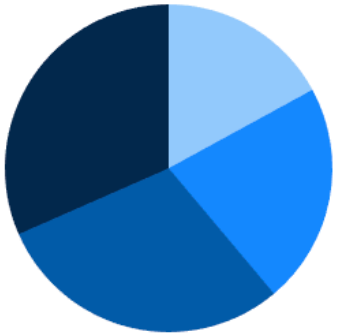
STACKED BAR CHART



PIE AND DONUT CHARTS

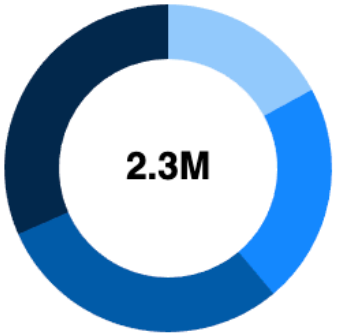
Pie Edit

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



Donut

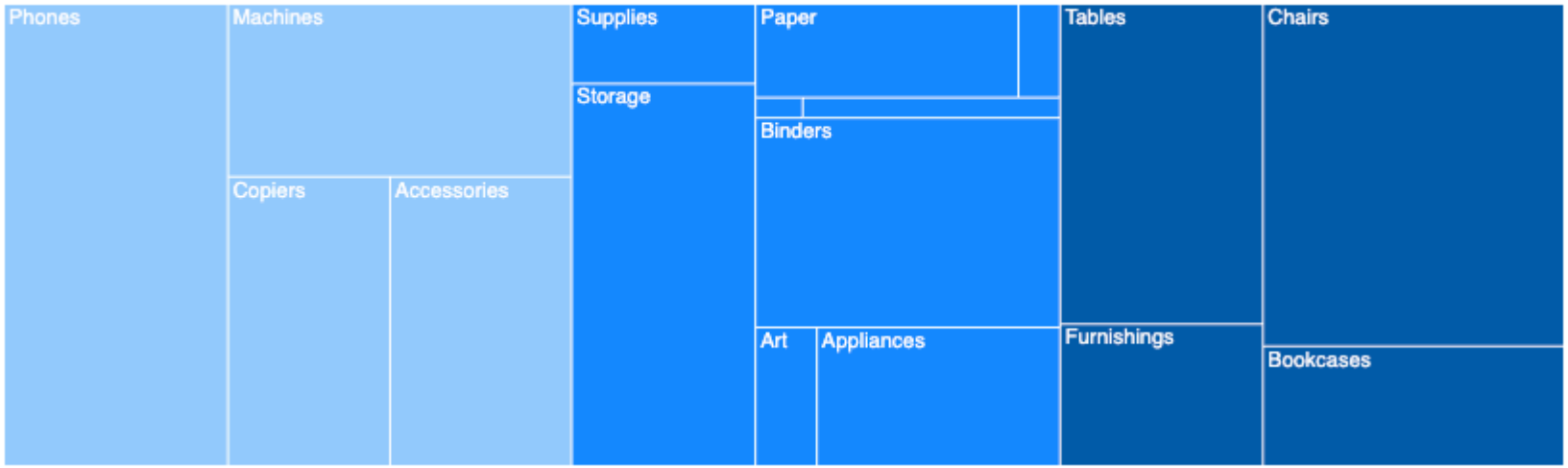
Similar to a pie chart - but the centre can be a good way of making space to include more information about the data (eg. total)



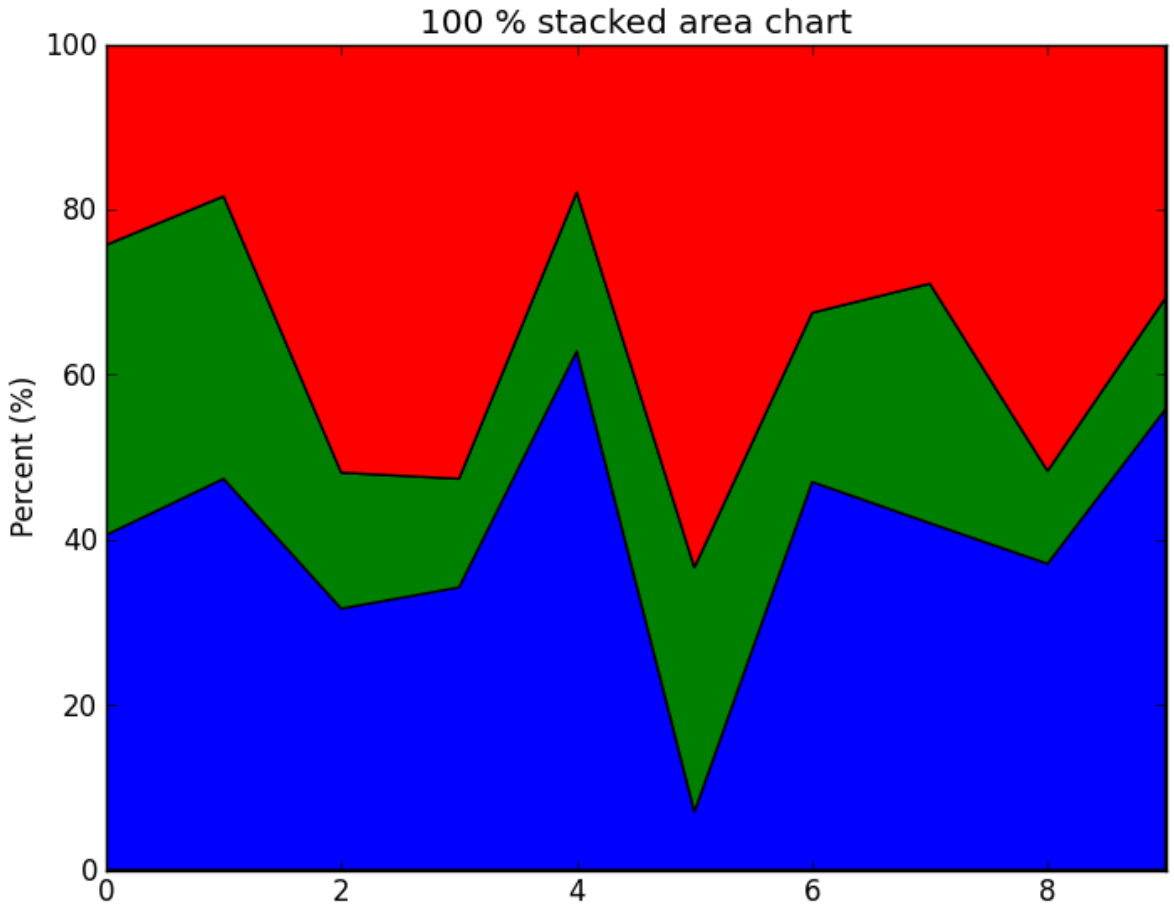
TREEMAP

Treemap

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



STACKED AREA



DISTRIBUTION

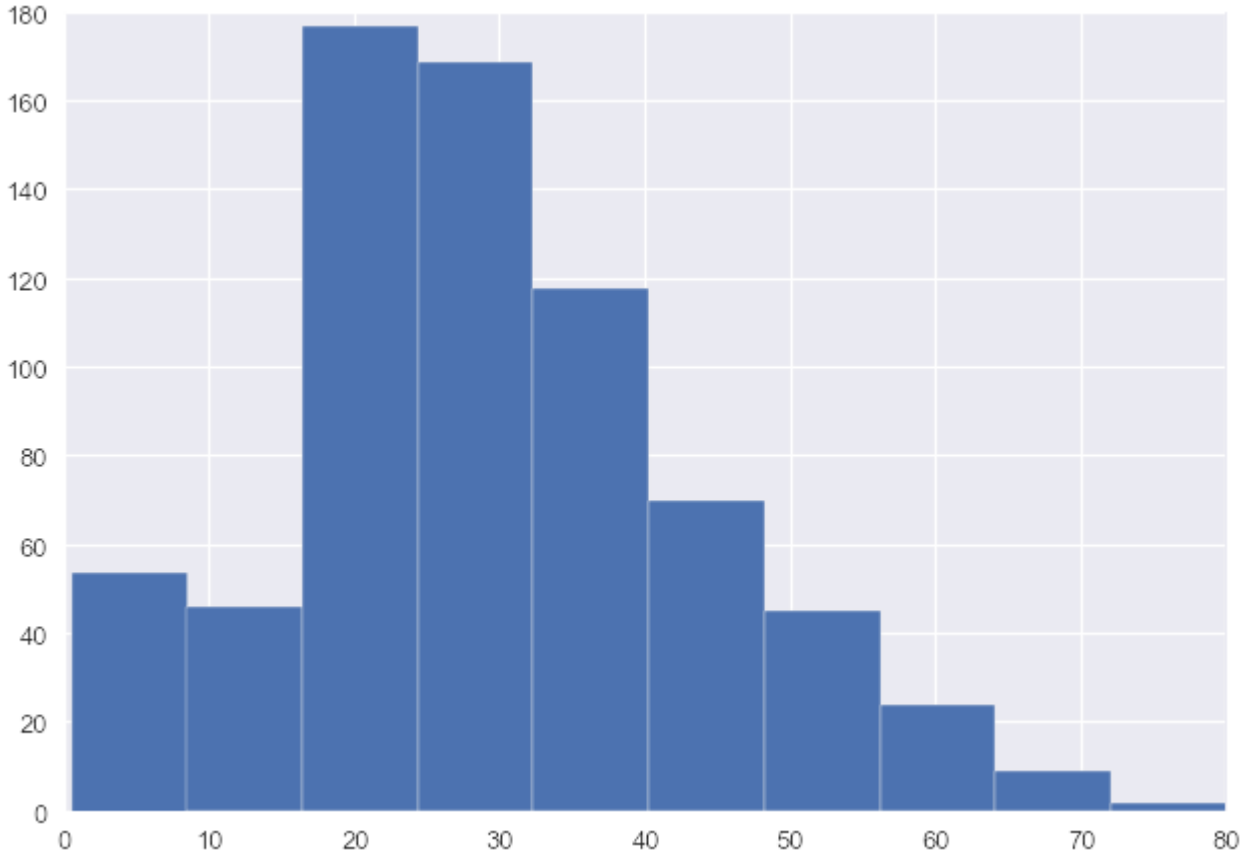
Aggregating large data vectors

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

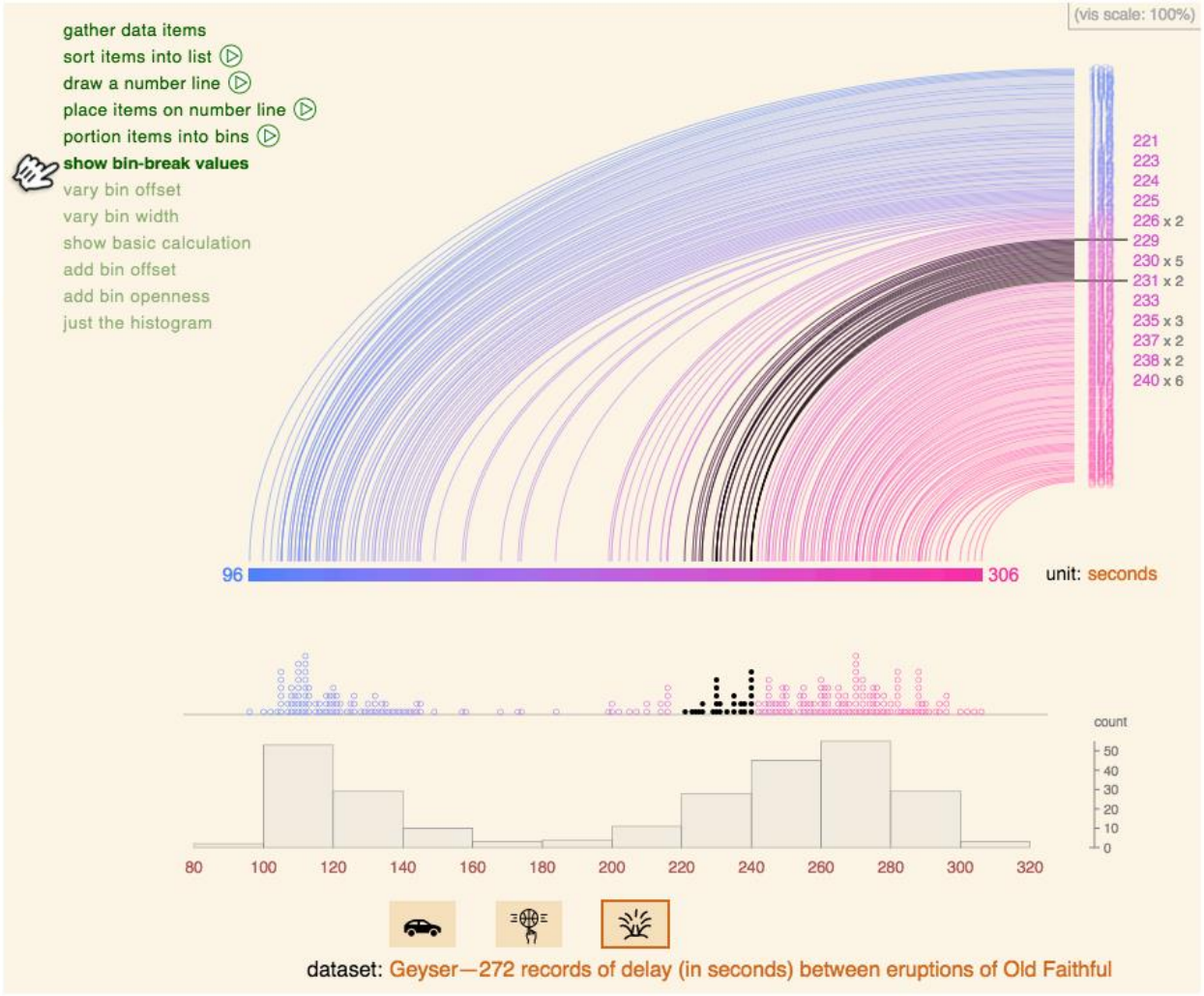
Pro: Compact representation

Con: Works only if data is "well behaved" for the type of distribution visualization

WHAT IS A HISTOGRAM



WHY HISTOGRAM IS HARD



http://tinlizzie.org/histograms/

HISTOGRAM

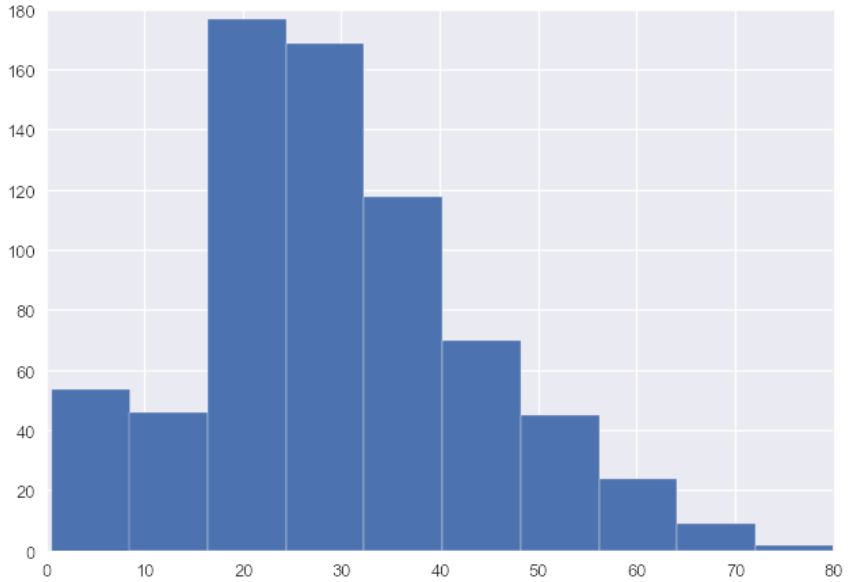
Good #bins hard to predict

Make interactive

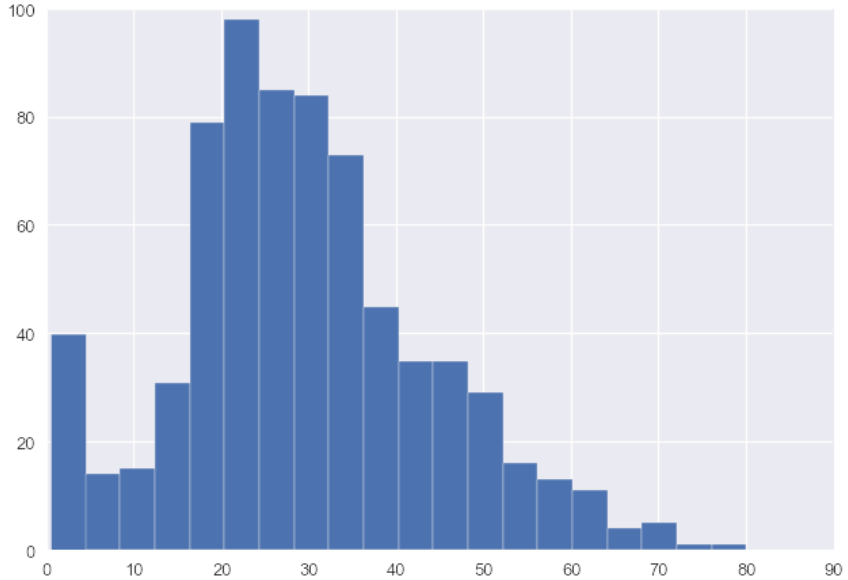
Rules of thumb:

#bins = sqrt(n)

#bins = $\log_2(2) + 1$

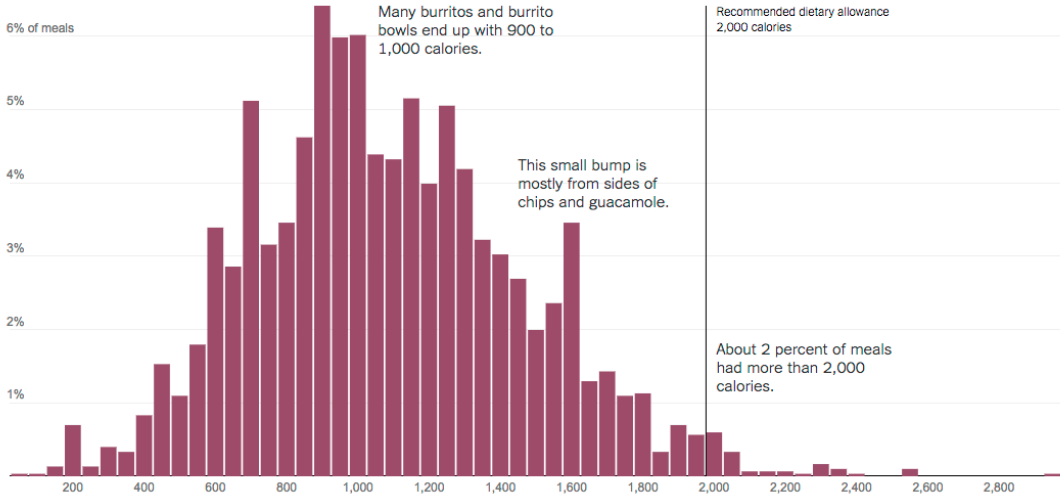


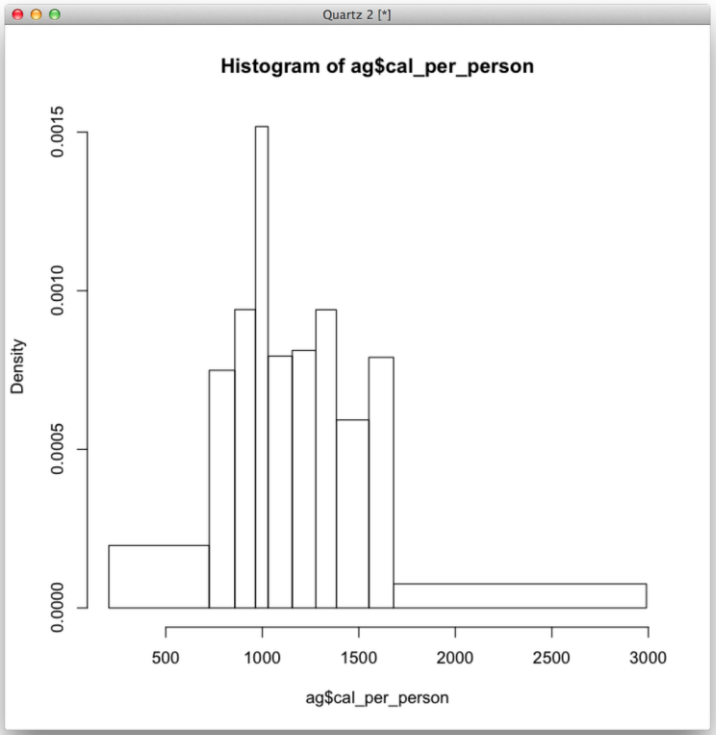
10 bins



20 bins

UNEQUAL BIN WIDTH

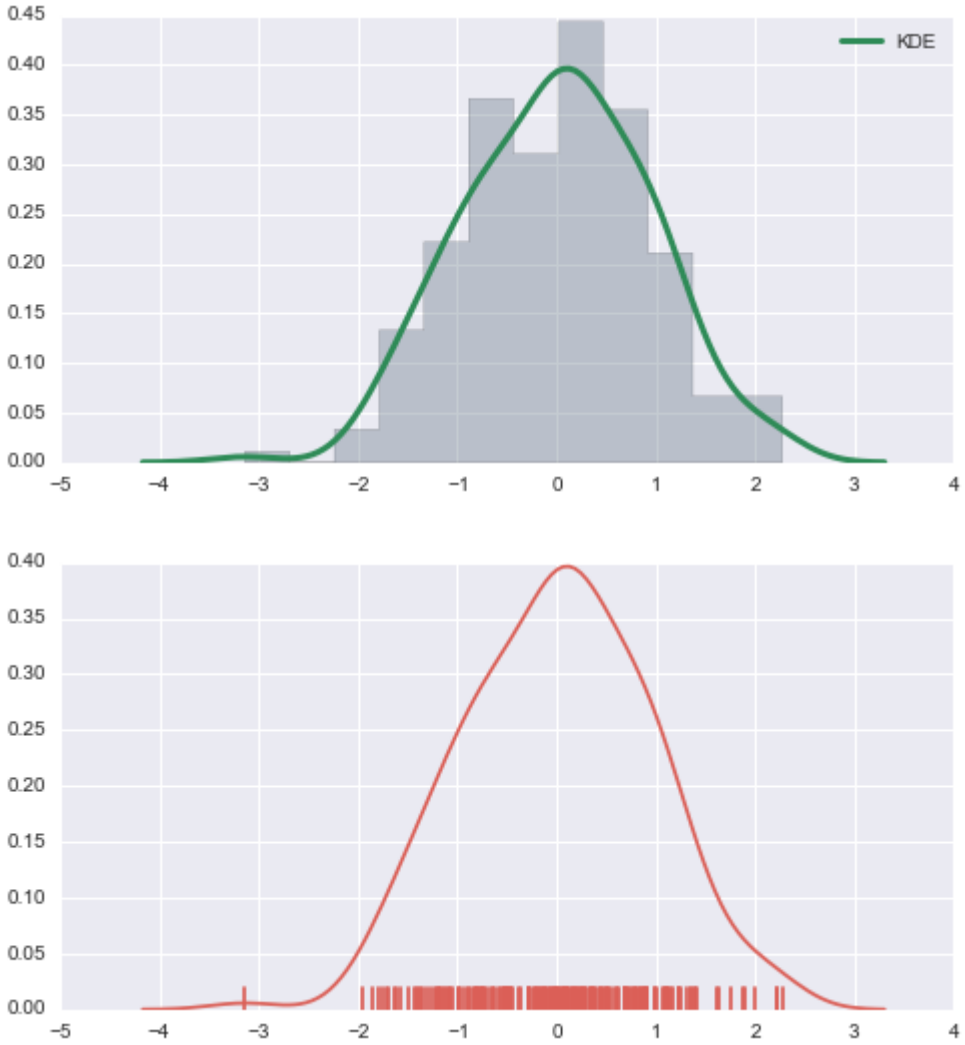


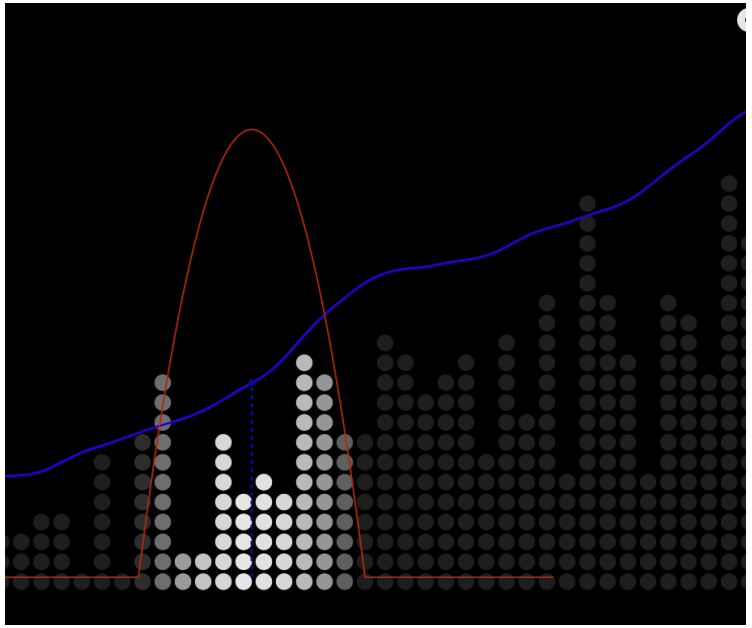


Can be useful if data is much sparser in some areas than others
Show density as area, not high

https://www.nytimes.com/interactive/2015/02/17/upshot/what-do-people-actually-order-at-chipotle.html?_r=1

DENSITY PLOTS (KERNEL DENSITY ESTIMATION)





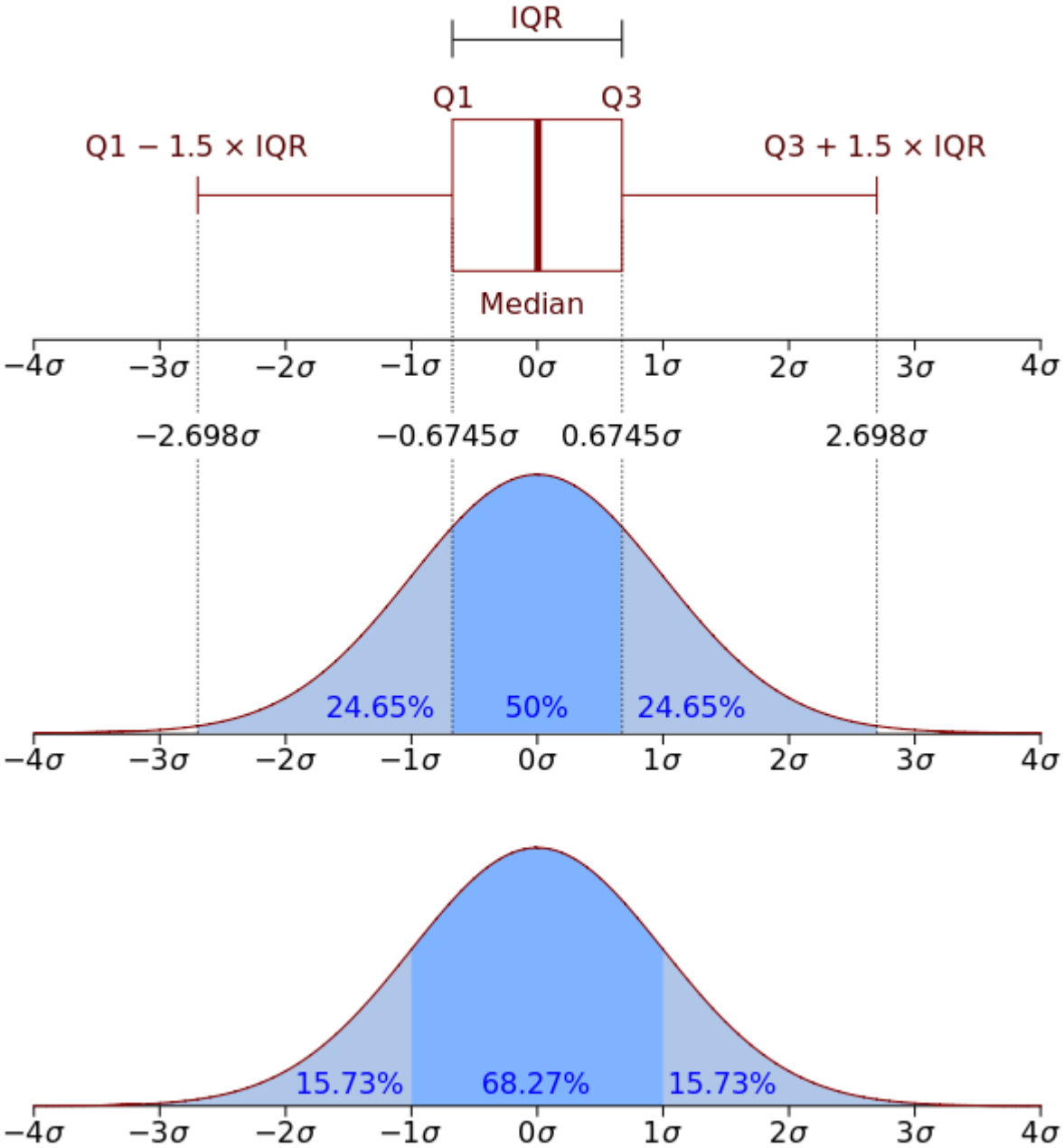
https://mathisonian.github.io/kde/

BOX PLOTS

aka Box-and-Whisker Plot

Show outliers as points!

Bad for non-normal distributed data
Especially bad for bi- or multimodal distributions



ONE BOXPLOT, FOUR DISTRIBUTIONS

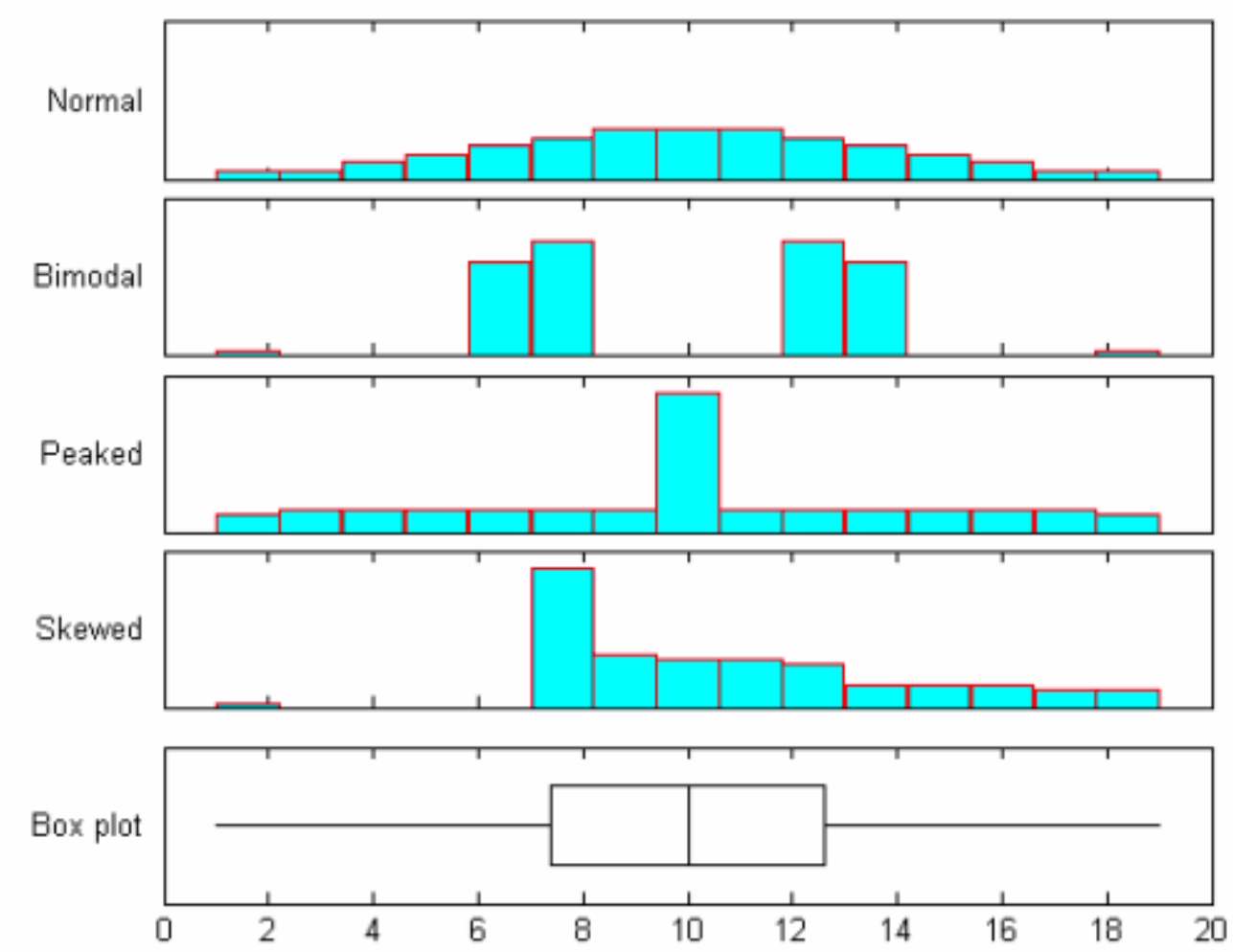


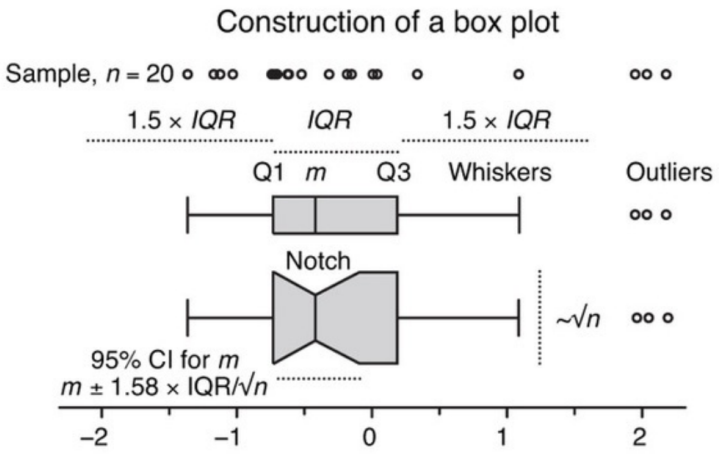
Figure 1: Histograms and box plot: four samples each of size 100

NOTCHED BOX PLOTS

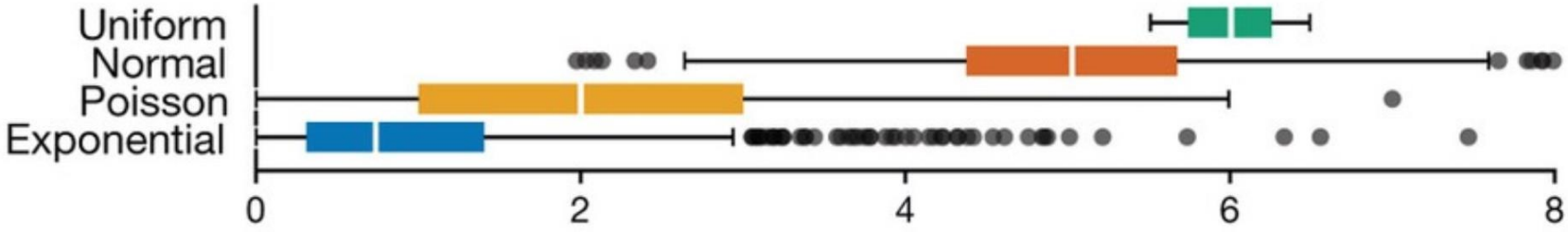
Notch shows
m +/- 1.5i x IQR/sqrt(n)

-> 95% Confidence Intervall

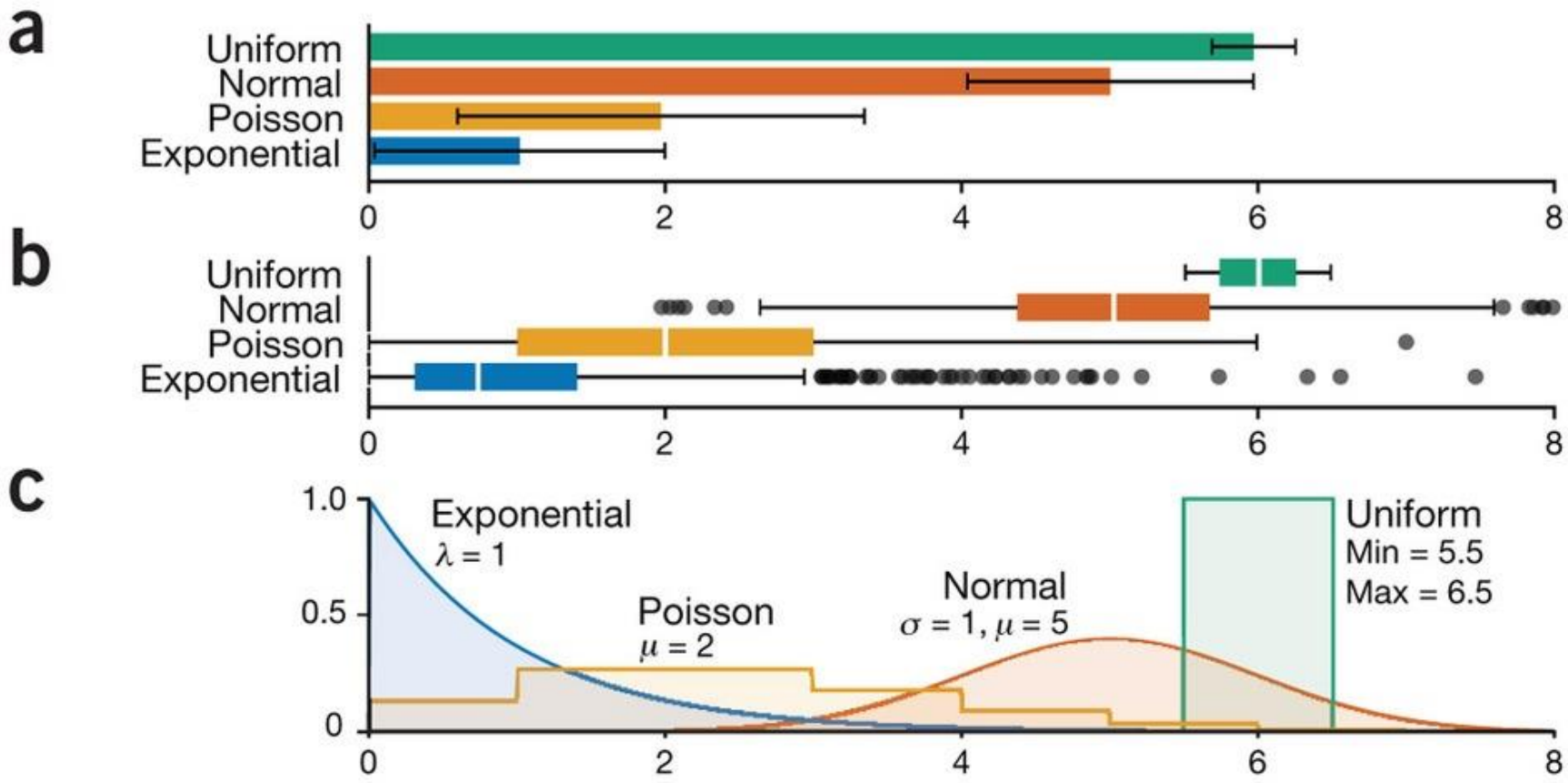
A guide to statistical significance.



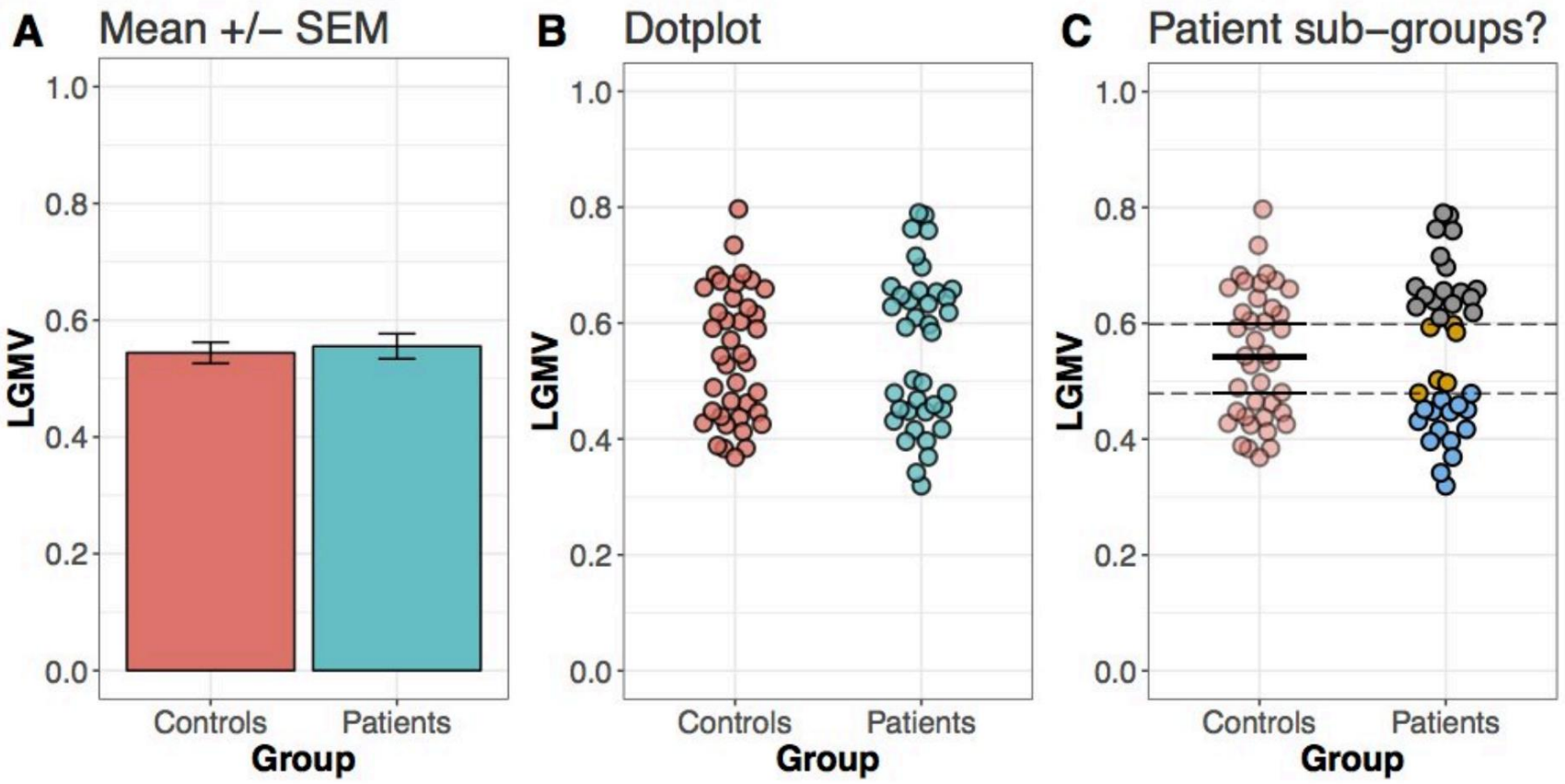
BOX PLOTS



COMPARISON

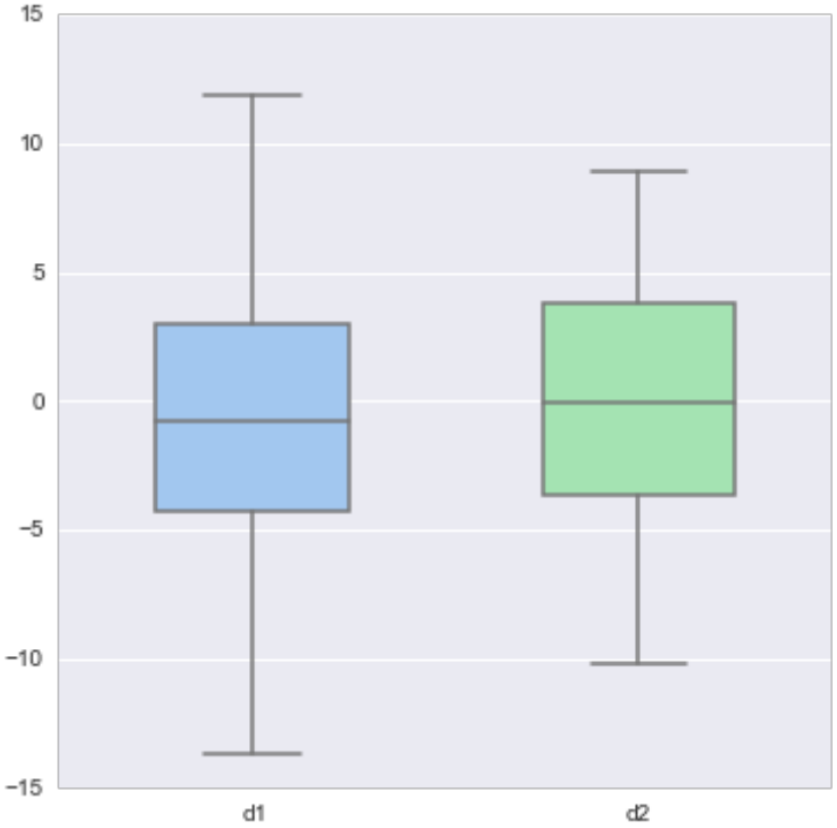

a
Uniform
Normal
Poisson
Exponential
0
2
4
6
8
b
Uniform
Normal
Poisson
Exponential
0
2
4
6
8
c
1.0
0.5
0
Exponential
λ = 1
Poisson
μ = 2
Normal
σ = 1, μ = 5
Uniform
Min = 5.5
Max = 6.5
0
2
4
6
8

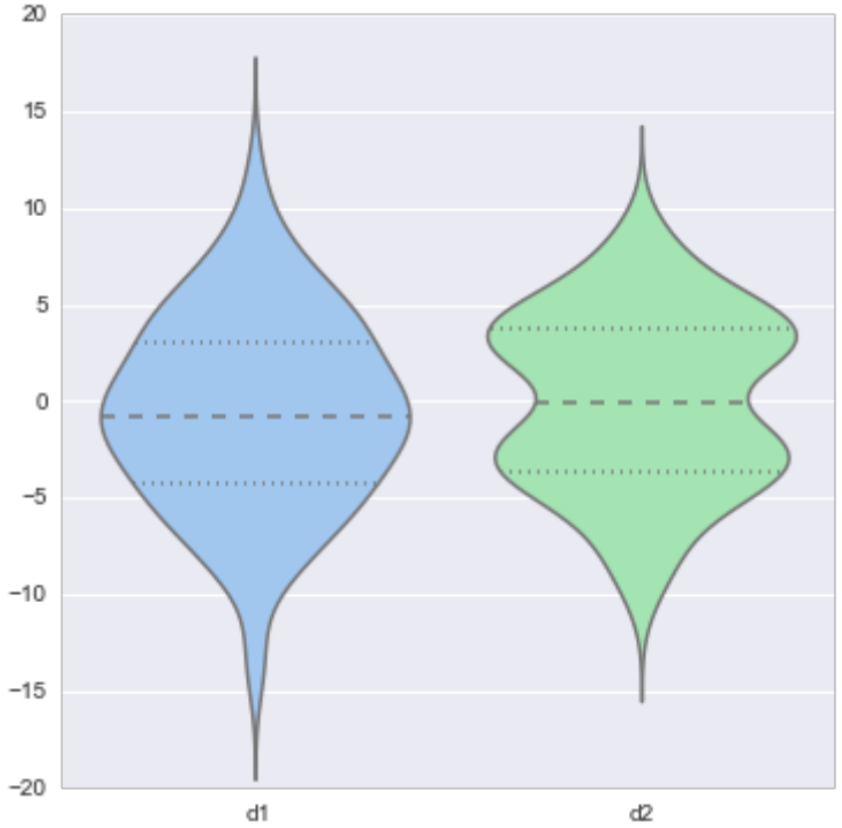
BAR CHARTS VS DOTPLOTS



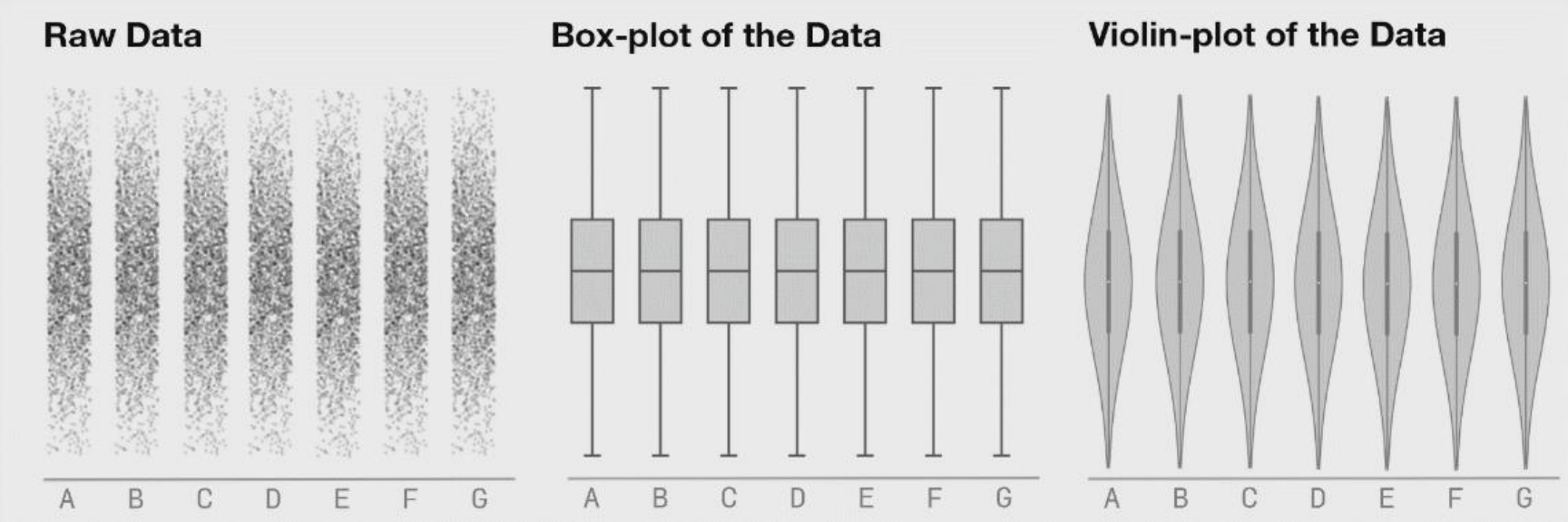
VIOLIN PLOT

Violin plot = box plot + probability density function





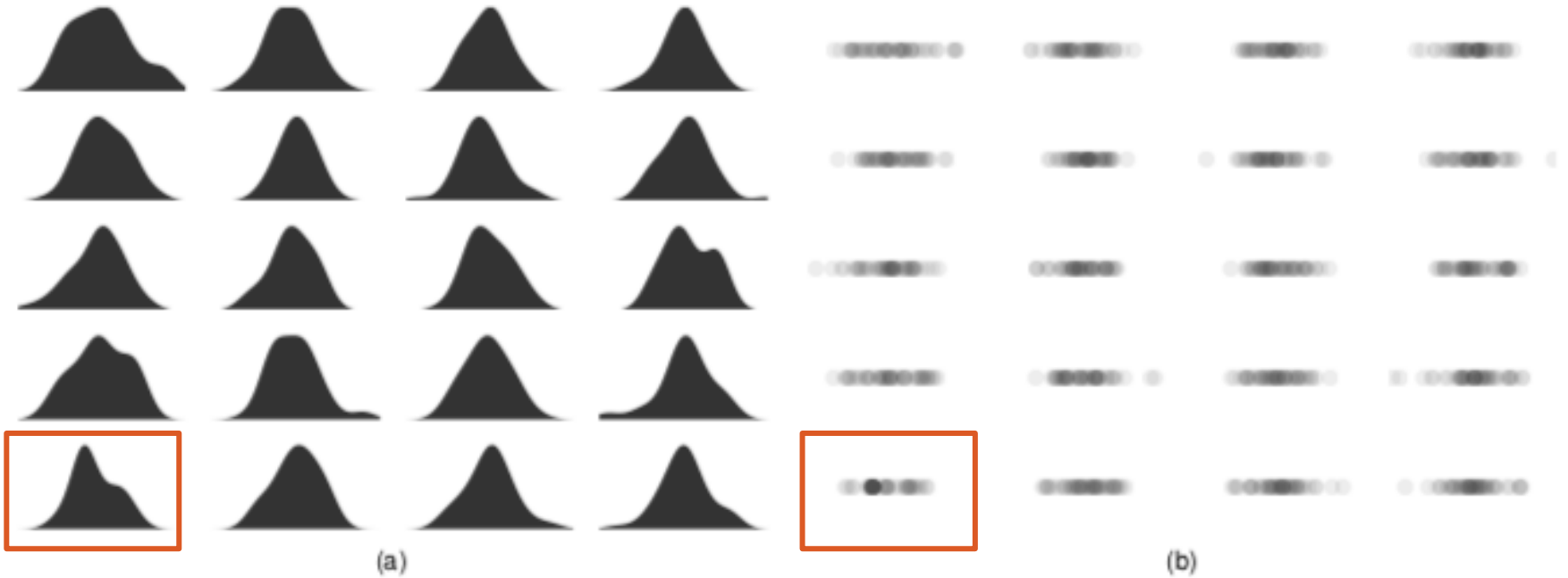
DIFFERENT DISTRIBUTIONS



https://blog.bioturing.com/2018/05/16/5-reasons-you-should-use-a-violin-graph/

DATA FLAWS

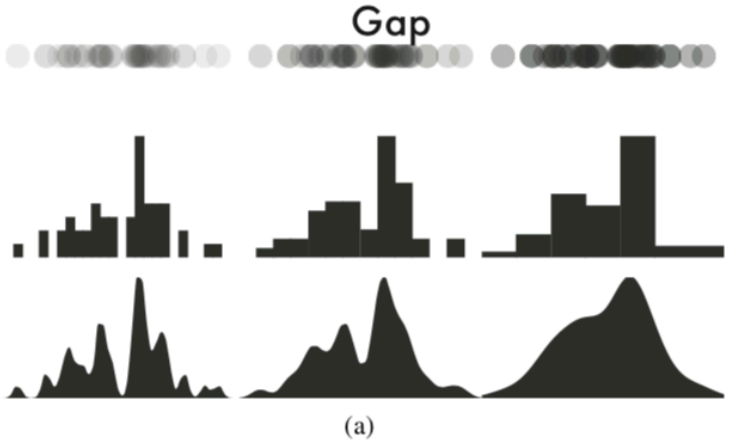
19 charts are random samples from a gaussian.
1 chart has 20% of samples with identical value



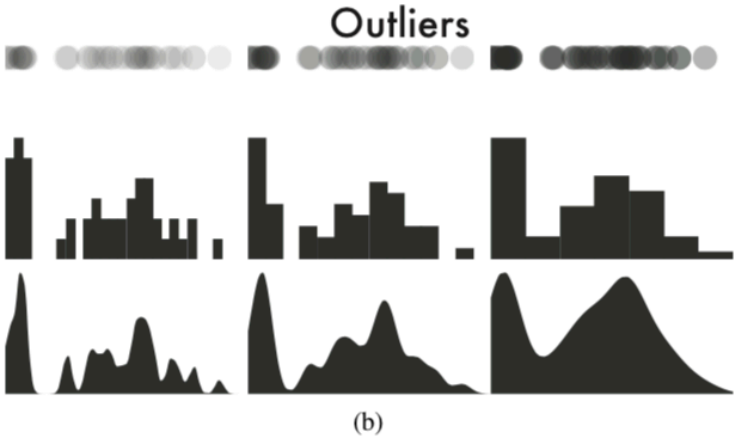
DETECTING DATA FLAWS

Tricky with aggregate visualization

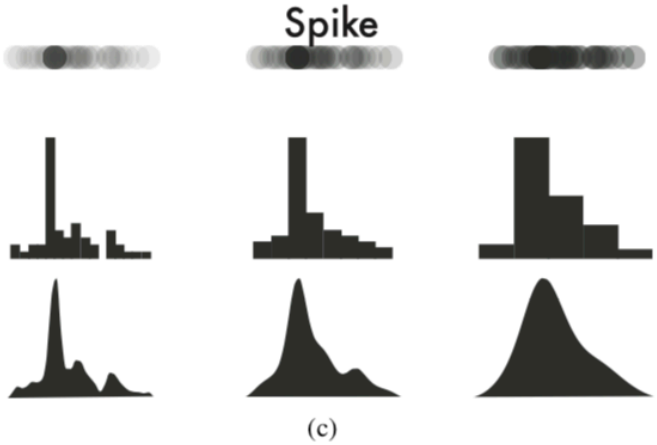
Bin size / kernel type / bandwidth / visualization choice all affect different situations



(a)



(b)

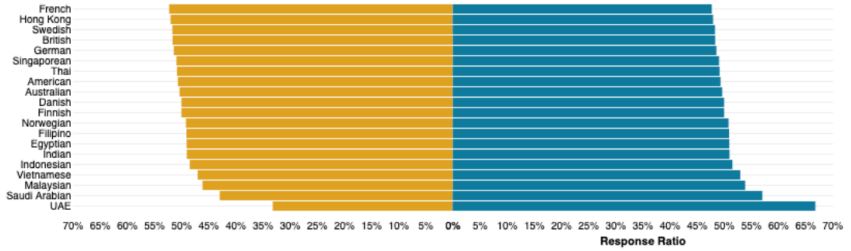


(c)

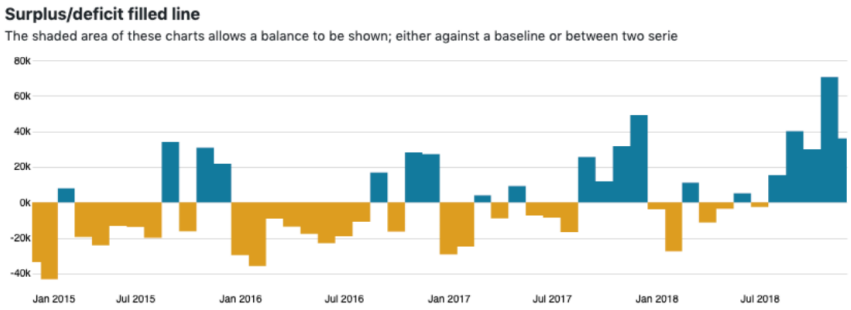
DEVIATION

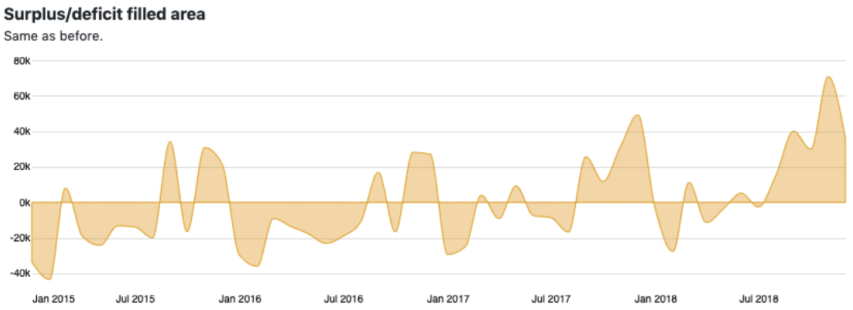
COMPARISON TO REFERENCE POINT

Diverging Bar Chart



Juxtaposing Two Variables (male/female)





CHANGE OVERTIME

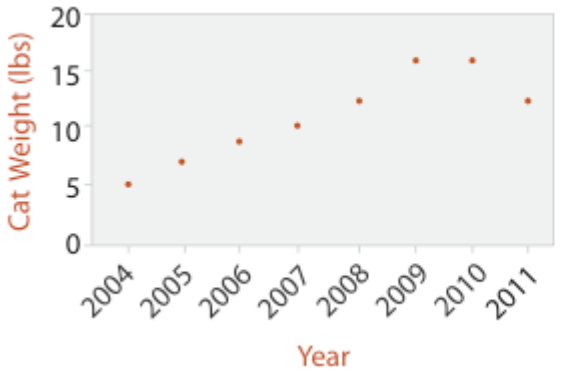
LINE CHART

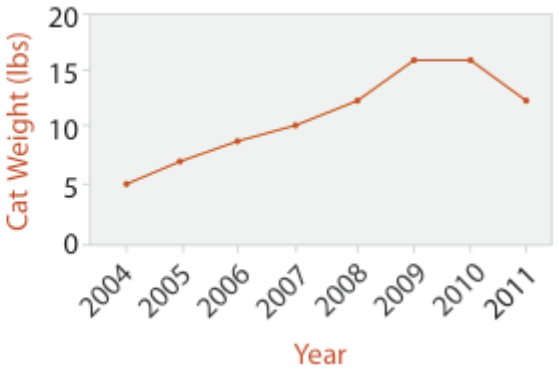
Simple

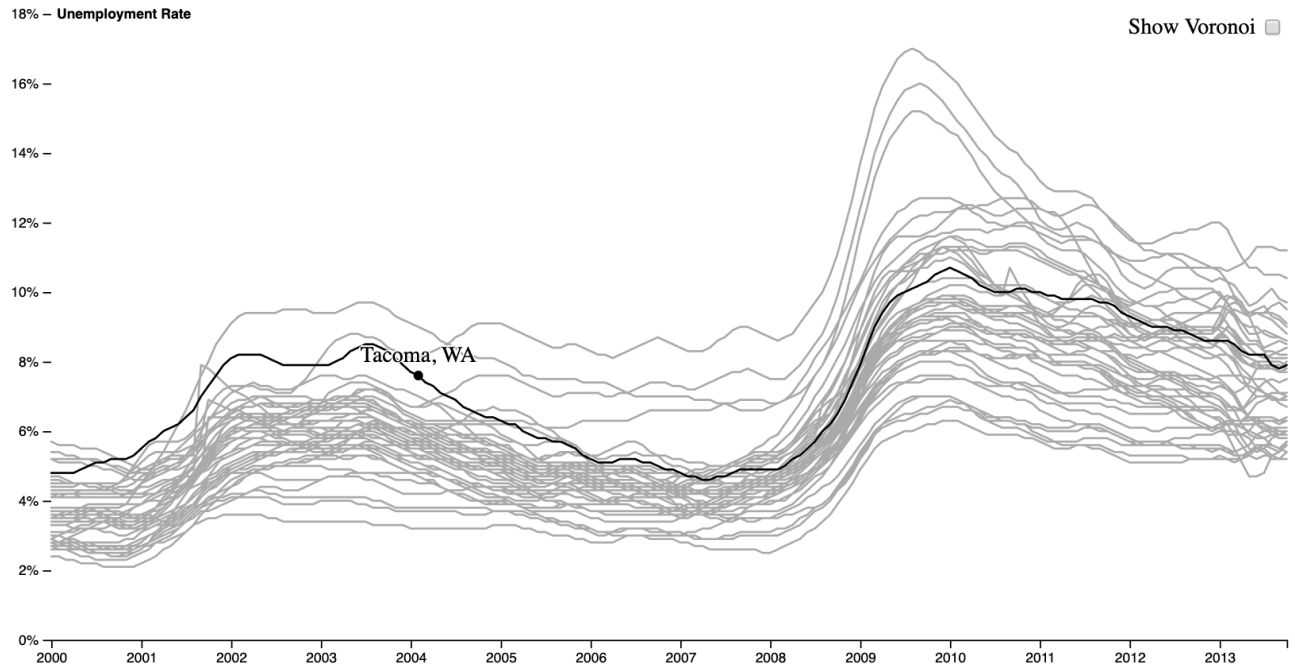
Familiar

Accurate

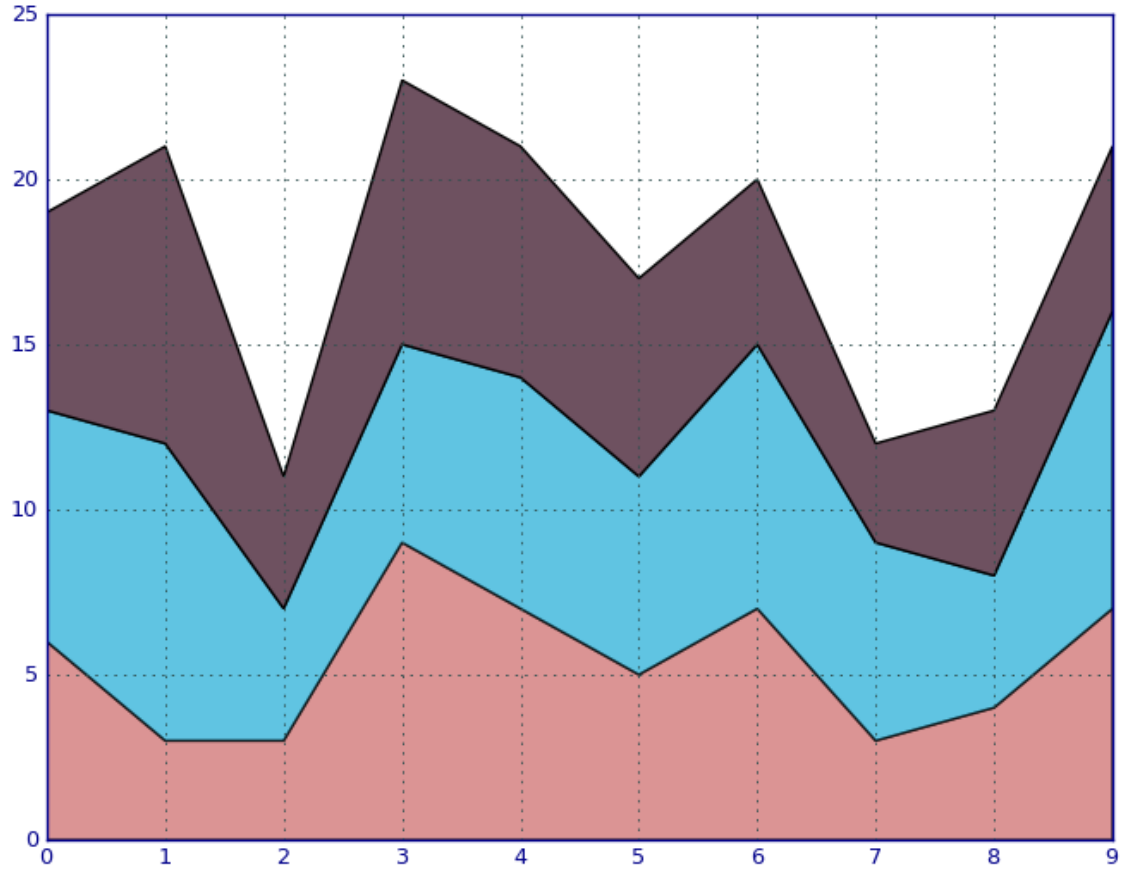
Fairly Scalable

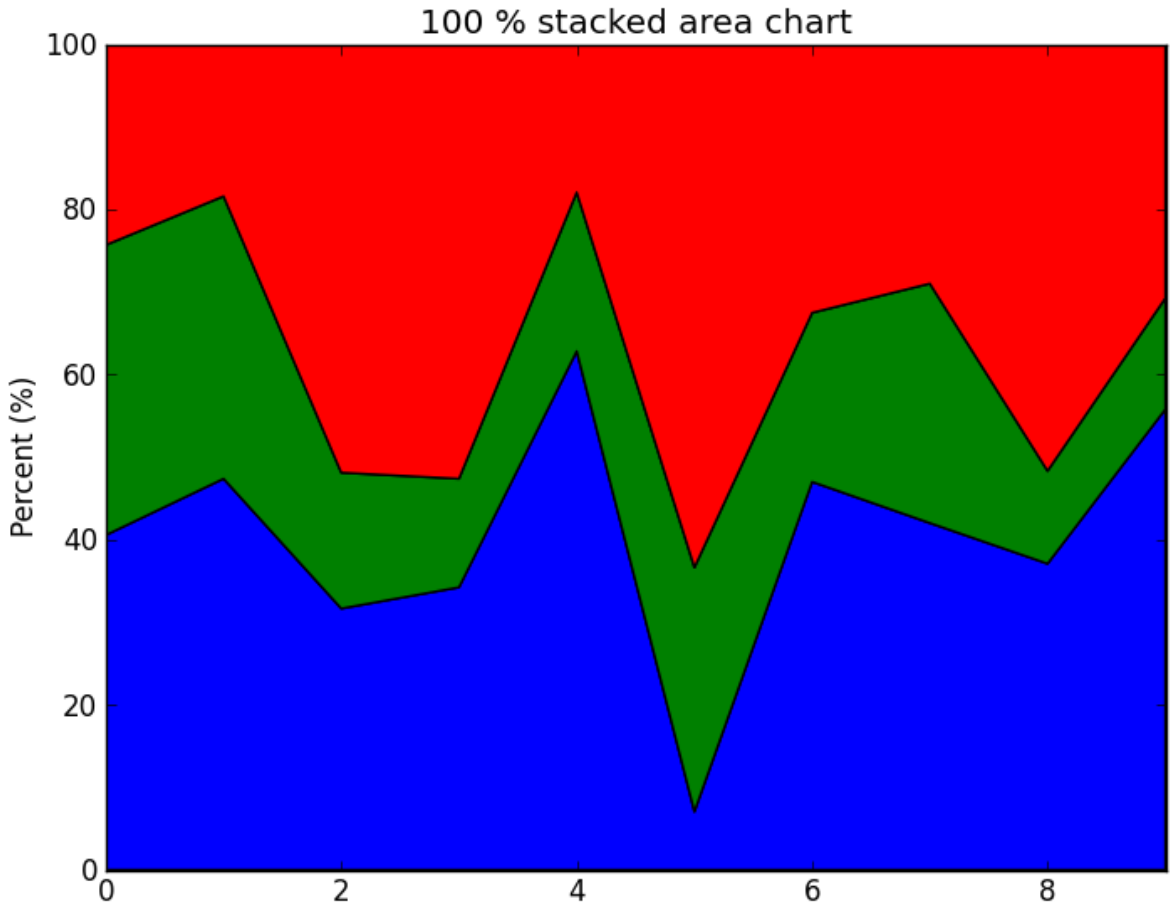




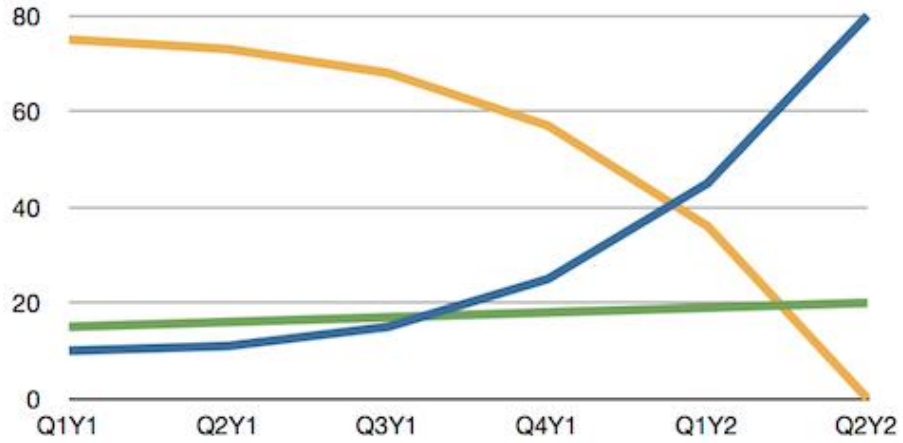


STACKED AREA

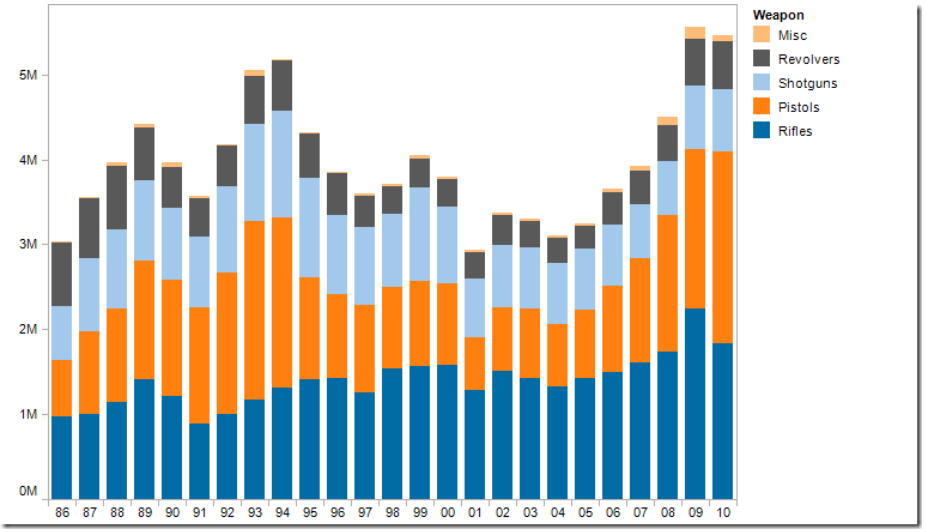


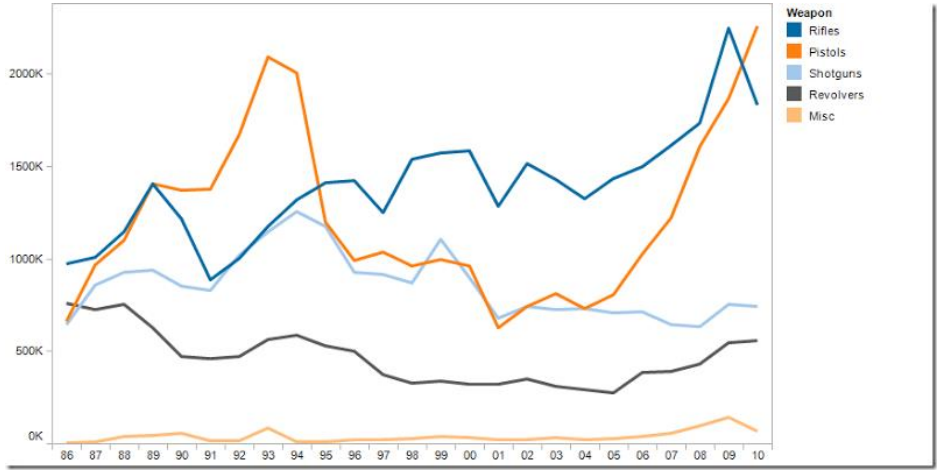

100 % stacked area chart
Percent (%)
100
80
60
40
20
0
0
2
4
6
8

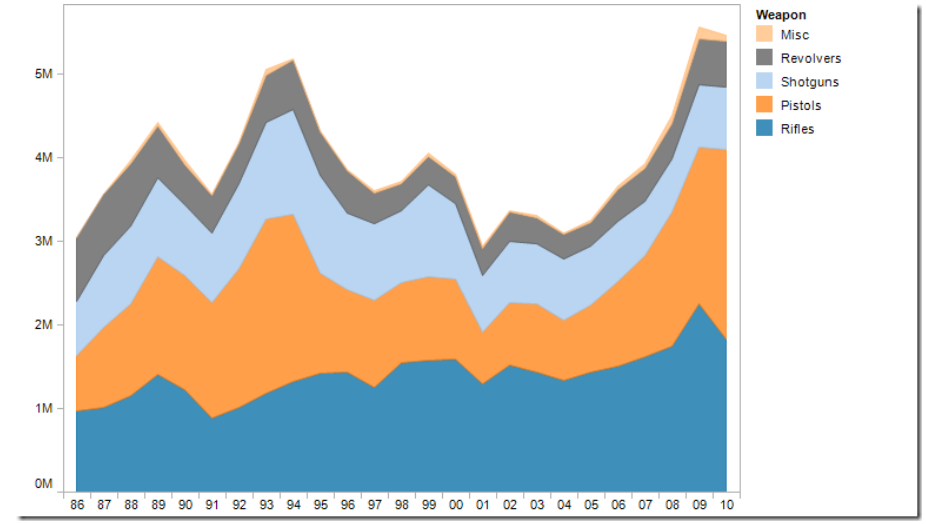
CAN YOU SPOT THE TREND?



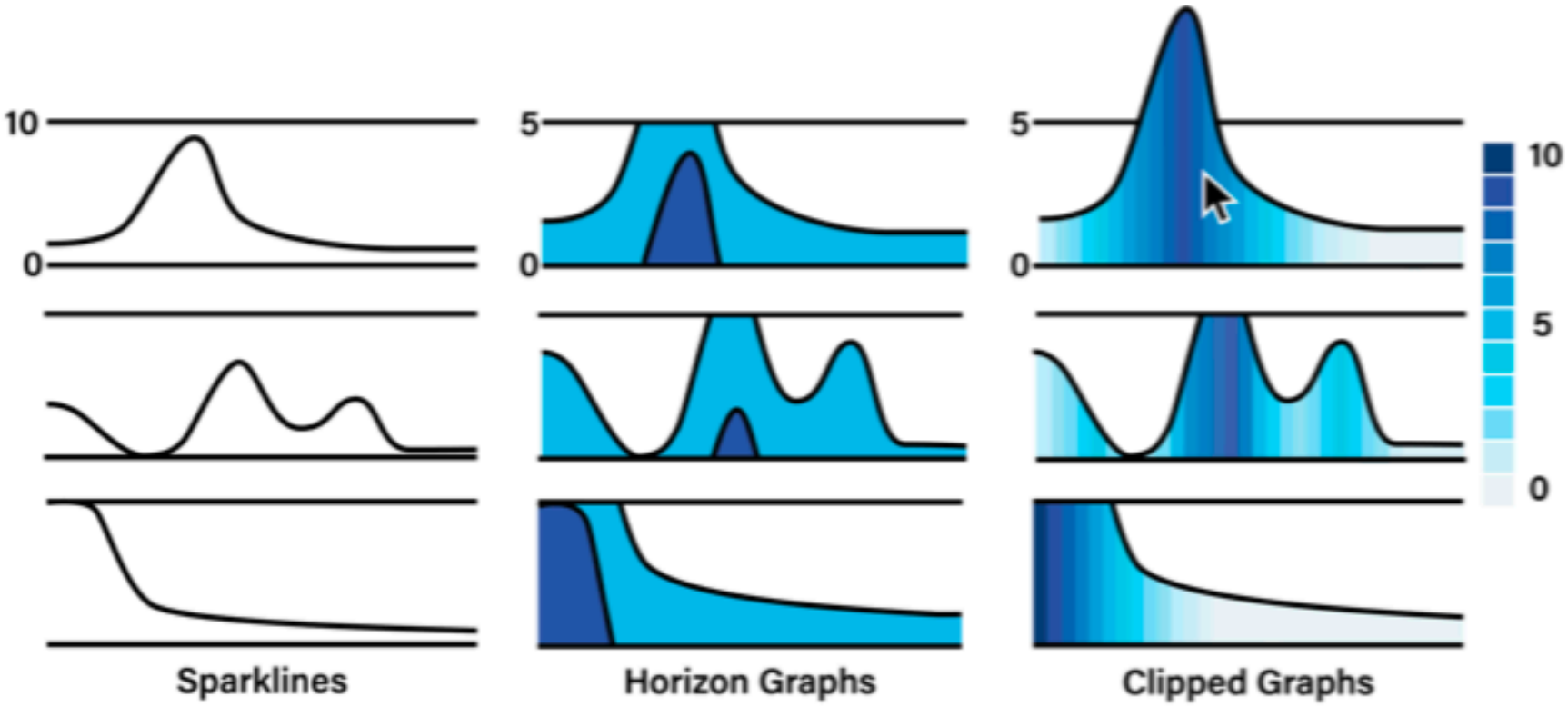
CAN YOU SPOT THE TREND?







MULTIPLE LINE CHARTS



SPARKLINES

Small line charts. can be embedded in text or part of a table

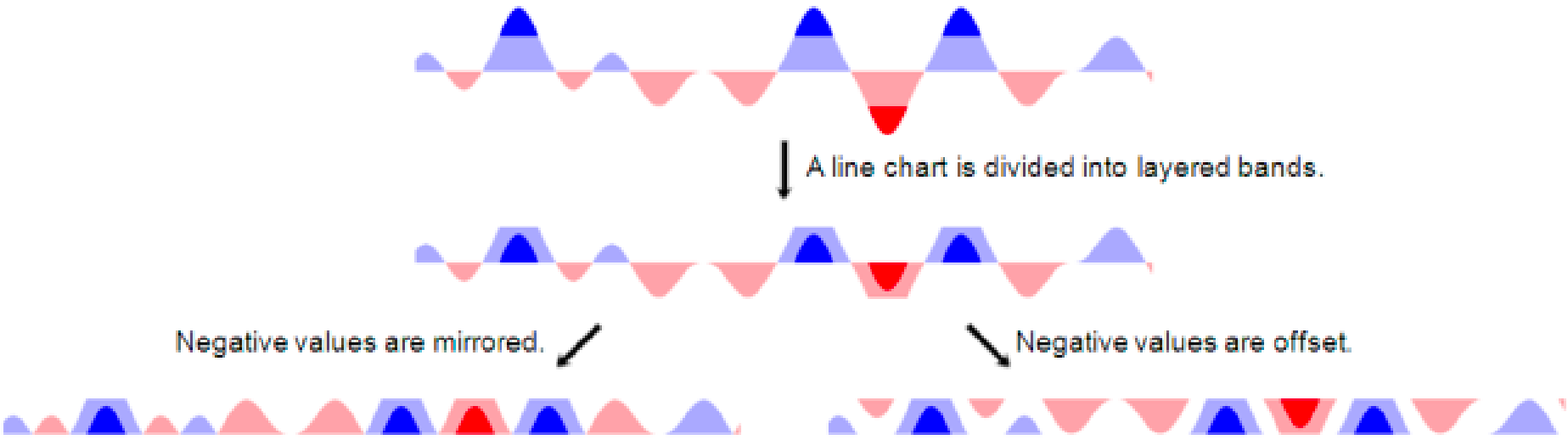
Symbol	Bid	Ask	Last	Change	T	Chart	Volume	High	Low	Value Change		Value	Gain	
DELL	89 3/4	89 13/16	89 3/4	+ 1 1/4			10,310,100	90 1/8	88 1/2	+1.41%	250	17,950	+273.72%	13,147
CPQ	48 7/16	48 9/16	48 7/16	- 13/16			25,628,700	51 1/4	1/4	-1.65%	-81	4,844	+60.79%	1,831
SDTI	26 1/4	26 3/8	26 3/8	+ 1/2			504,600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

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

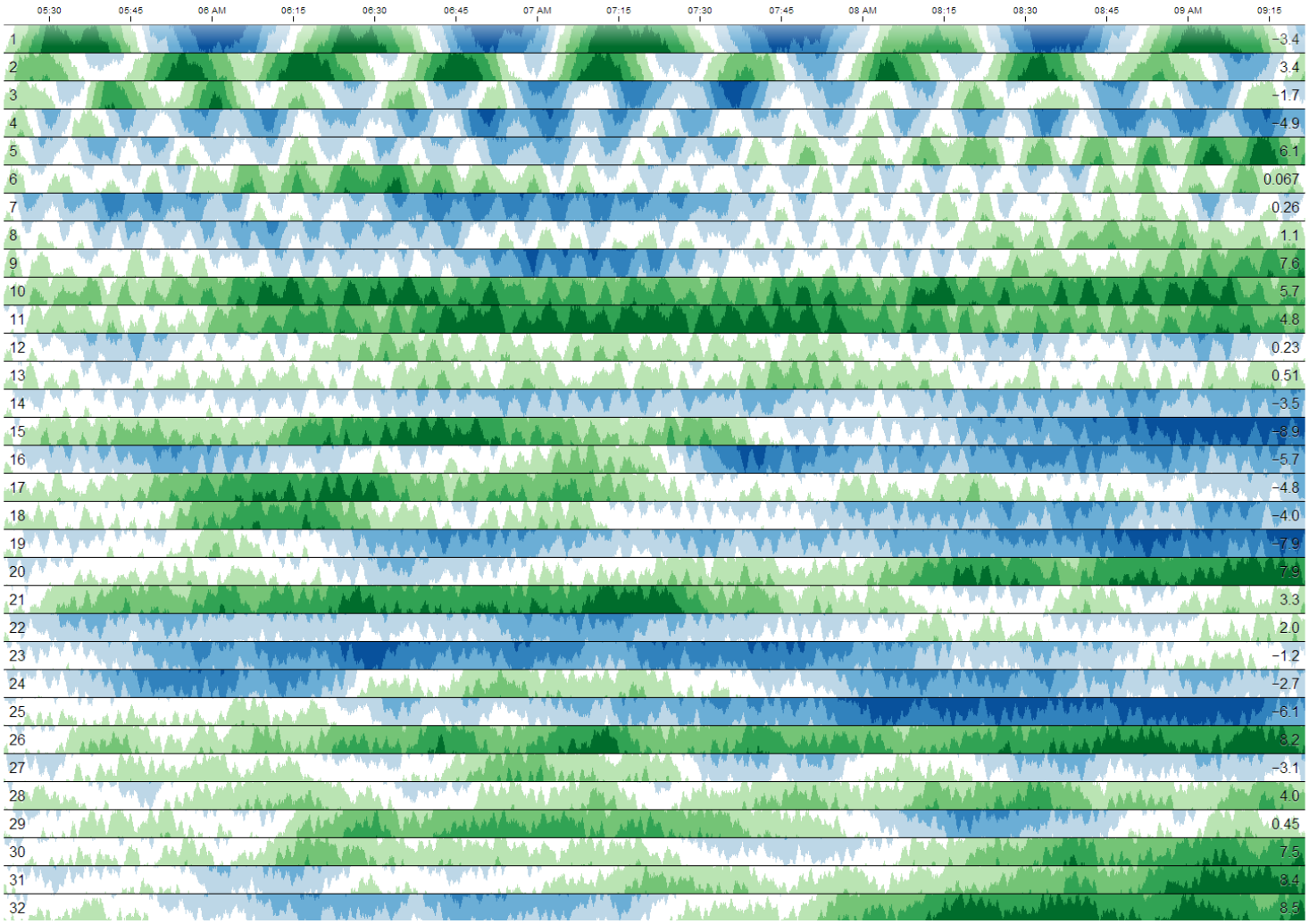
Alibaba stock is at 5 yr high 93.89  152.11 as of July 2017

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

HORIZONTAL GRAPH

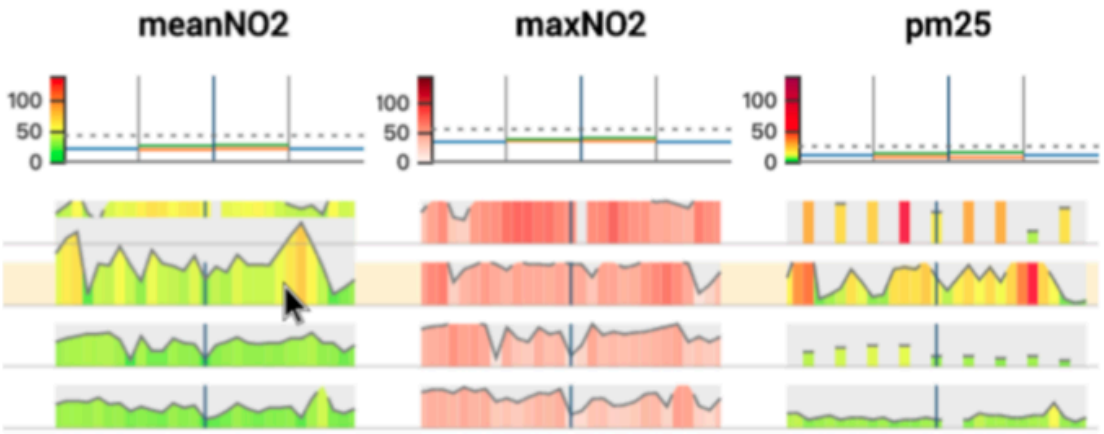


HORIZONTAL GRAPH

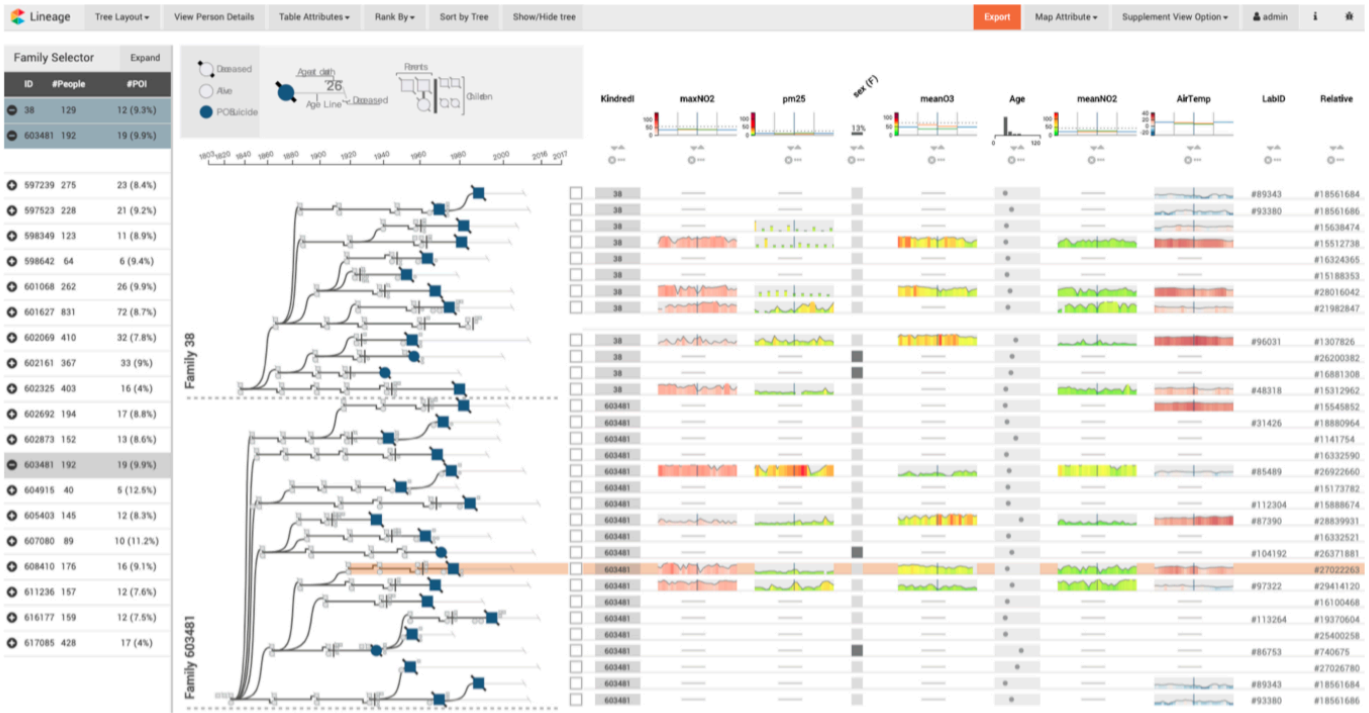


http://square.github.io/cubism/

CLIPPED GRAPHS



https://youtu.be/IxoaFe6yBMg

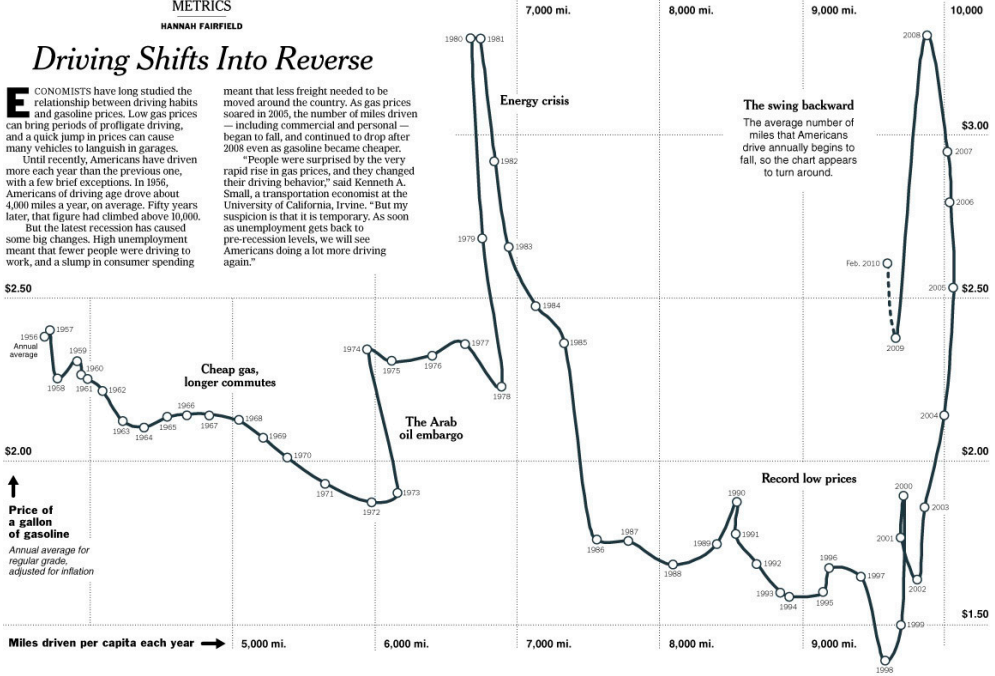


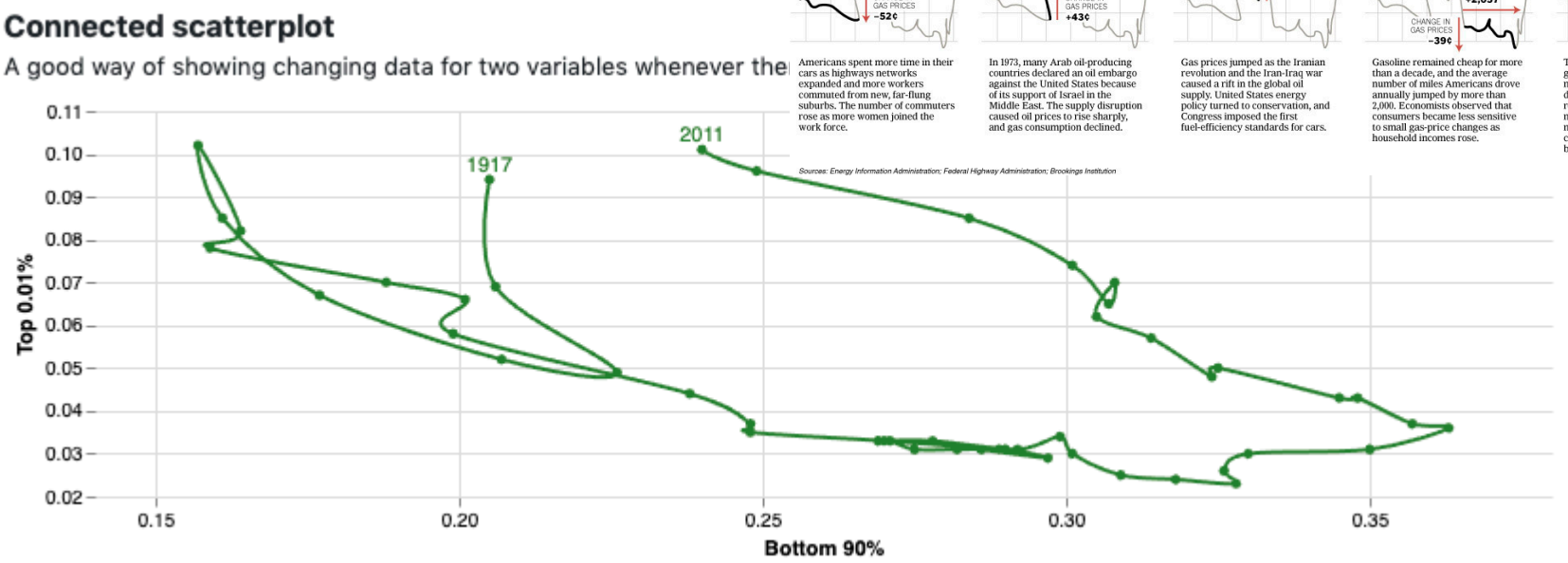
CONNECTED SCATTERPLOT

Two Variables + Time
Only one per Chart!
Labels important

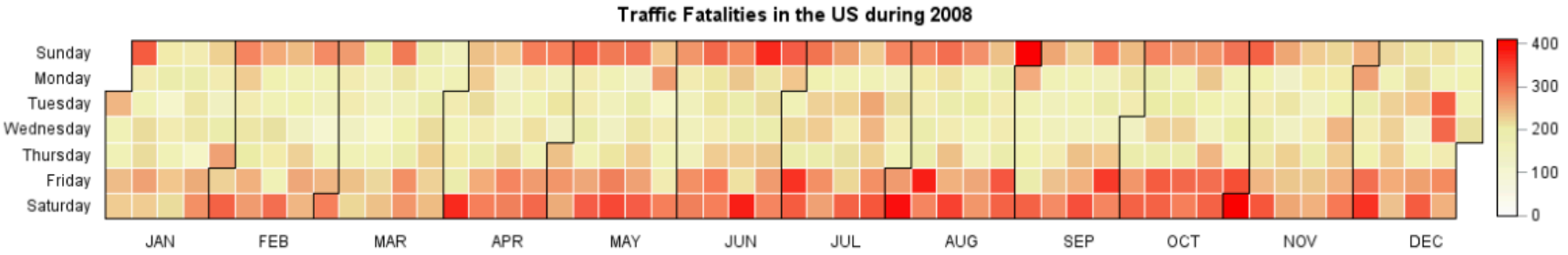
http://steveharoz.com/research/connected_scatterplot/

http://www.thefunctionalart.com/2012/09/in-praise-of-connected-scatter-plots.html



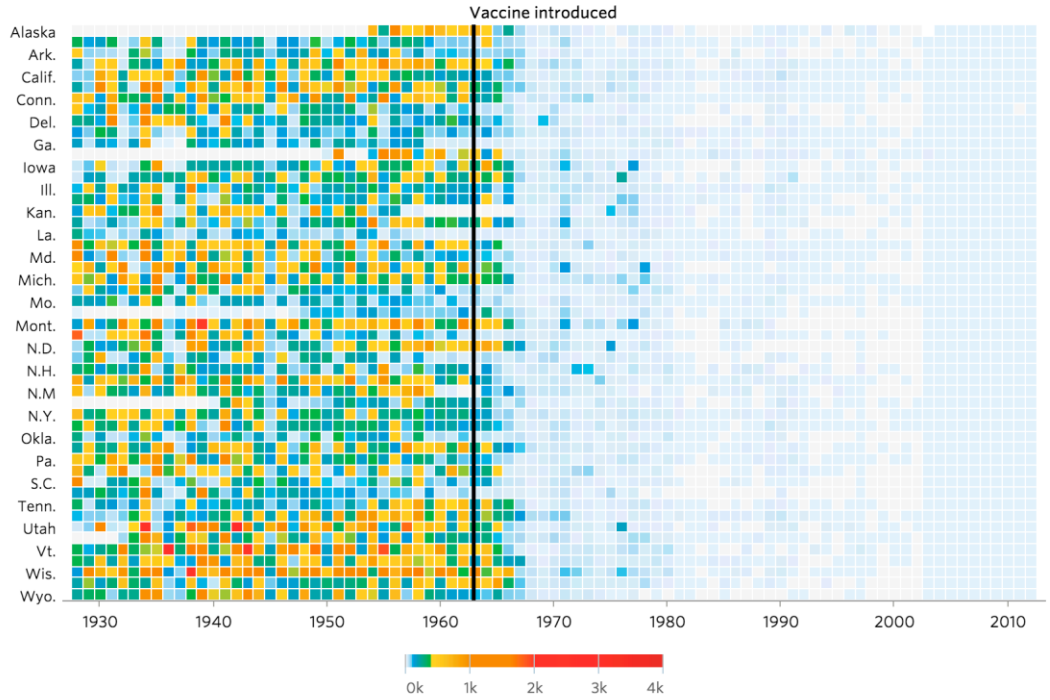


HEATMAP



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

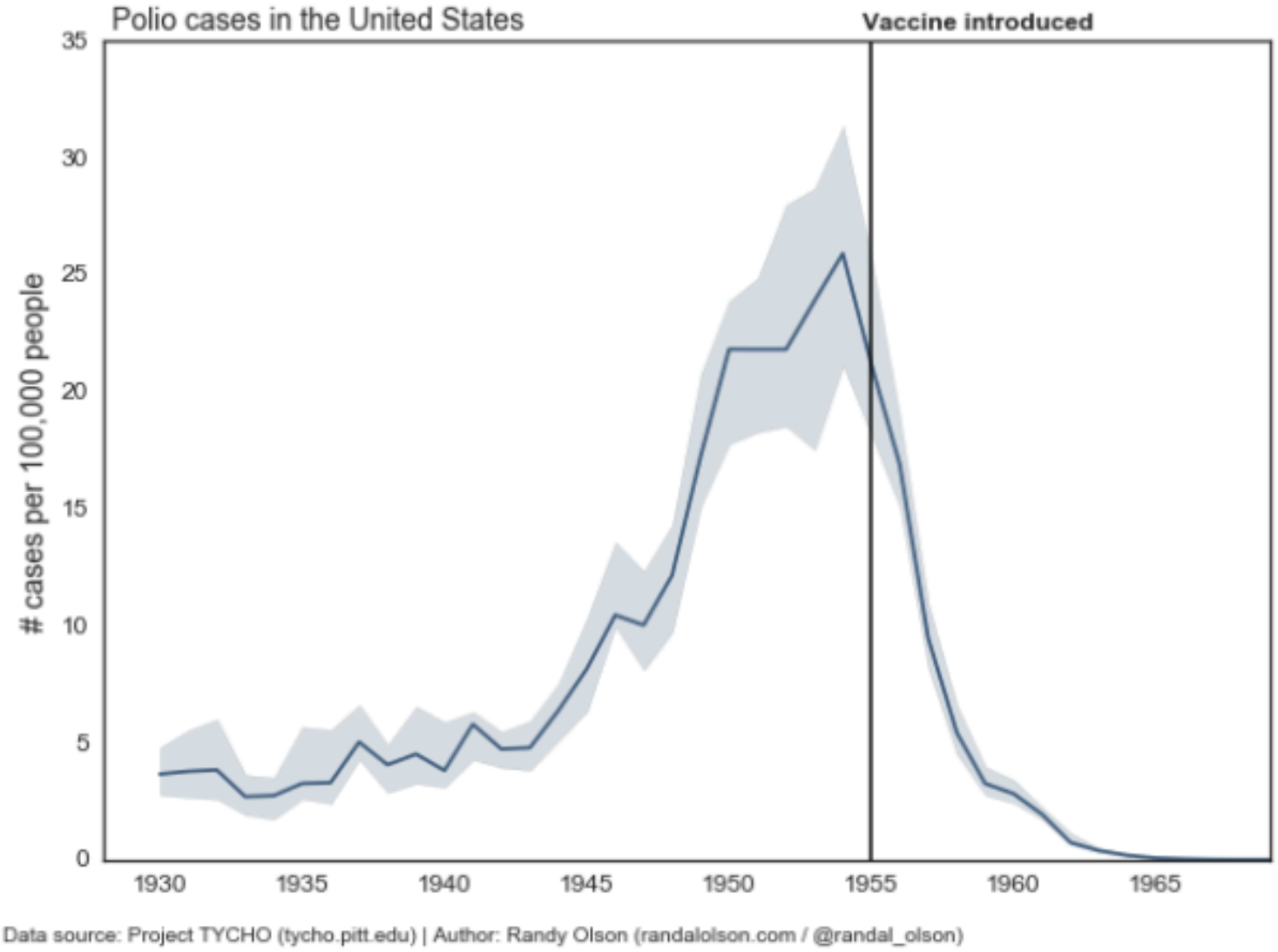
Measles



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.

https://www.informationisbeautifulawards.com/showcase/660-vaccines-and-infectious-diseases

SOMETIMES YOU CAN SHOW TOO MUCH DATA



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

RANKING

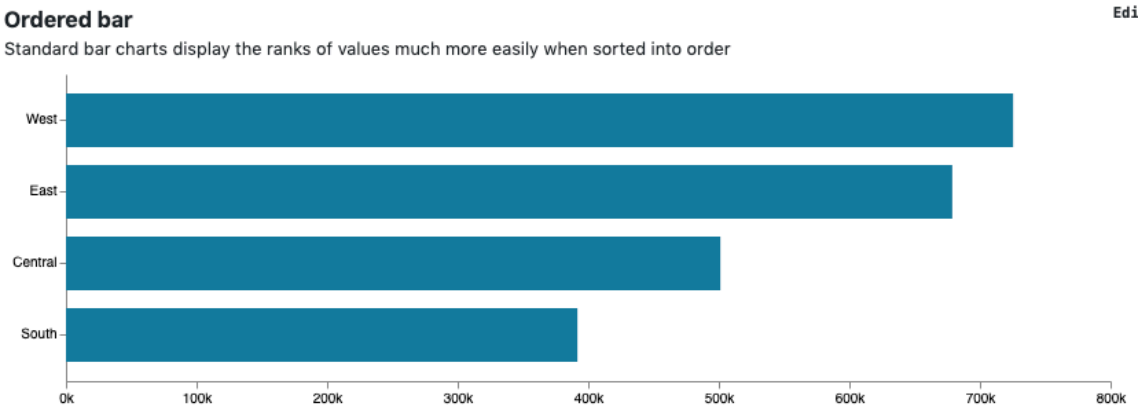
RANKING EXERCISE

	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

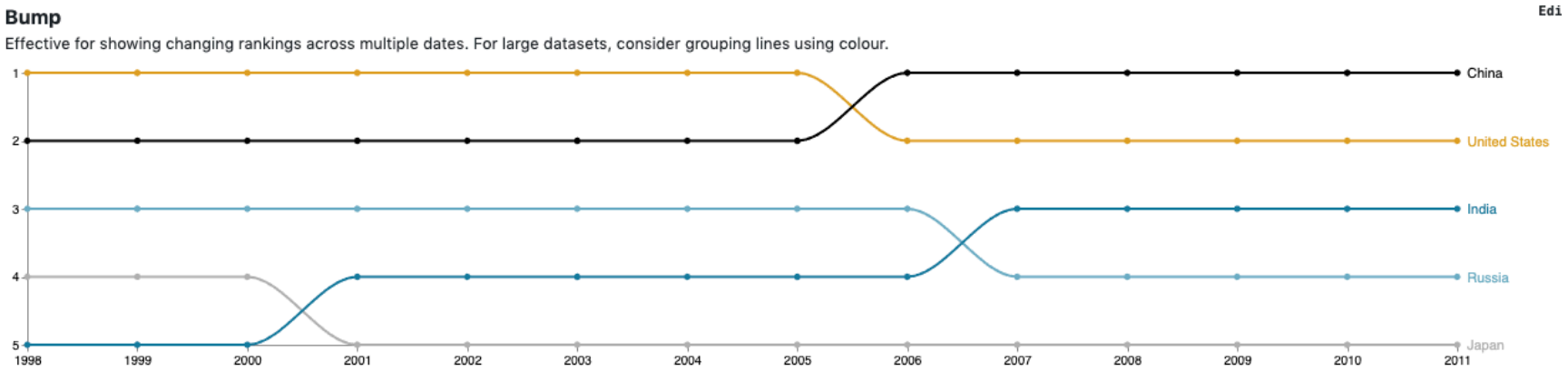
Design a visualization showing the ranking of these football clubs over time.

RANKING

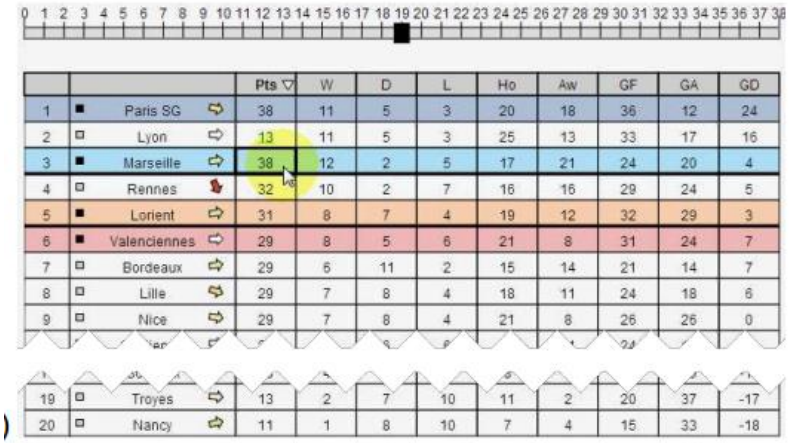
Magnitude Visualization + Sorting



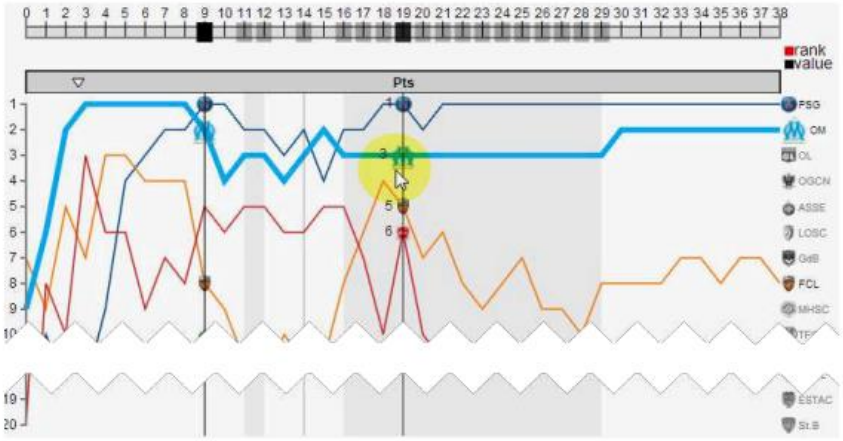
Bump Charts for Rankings over Time



TEMPORAL RANKINGS



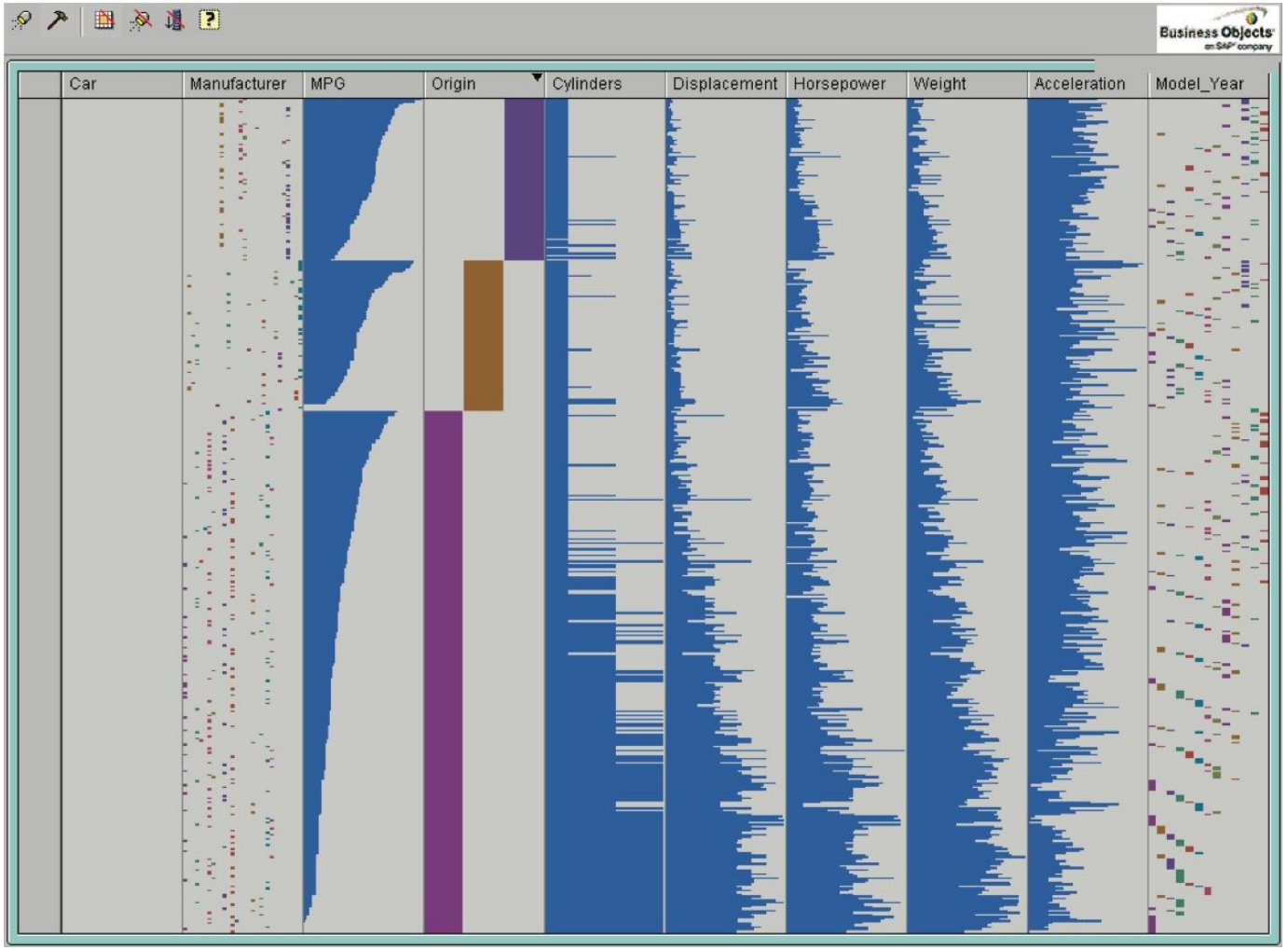
(b)



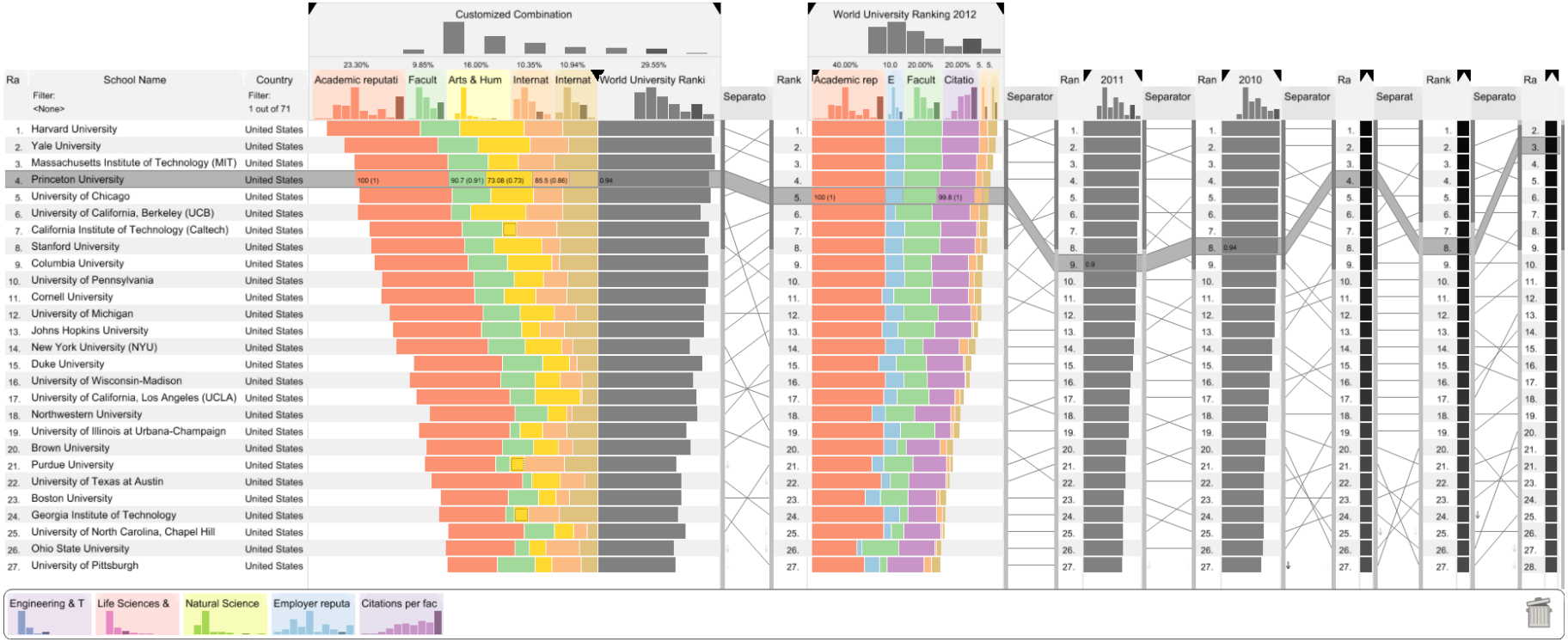
https://youtu.be/a0duFWu3Zp0

TABLE LENS

Interactive table based representation



LINEUP



http://lineup.caleydo.org

RANKING

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

MULTIPLE ATTRIBUTES

COMBINER FUNCTION

(Weighted) sum

Score = $w_a A + w_b B + w_c C$ → **Serial**

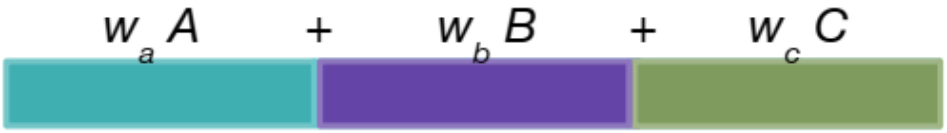
Maximum

Score = max (A, B, C) → **Parallel**

Product → **Complex**

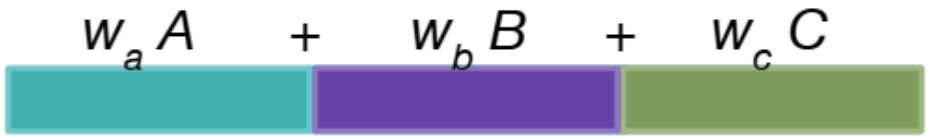
Nesting

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)

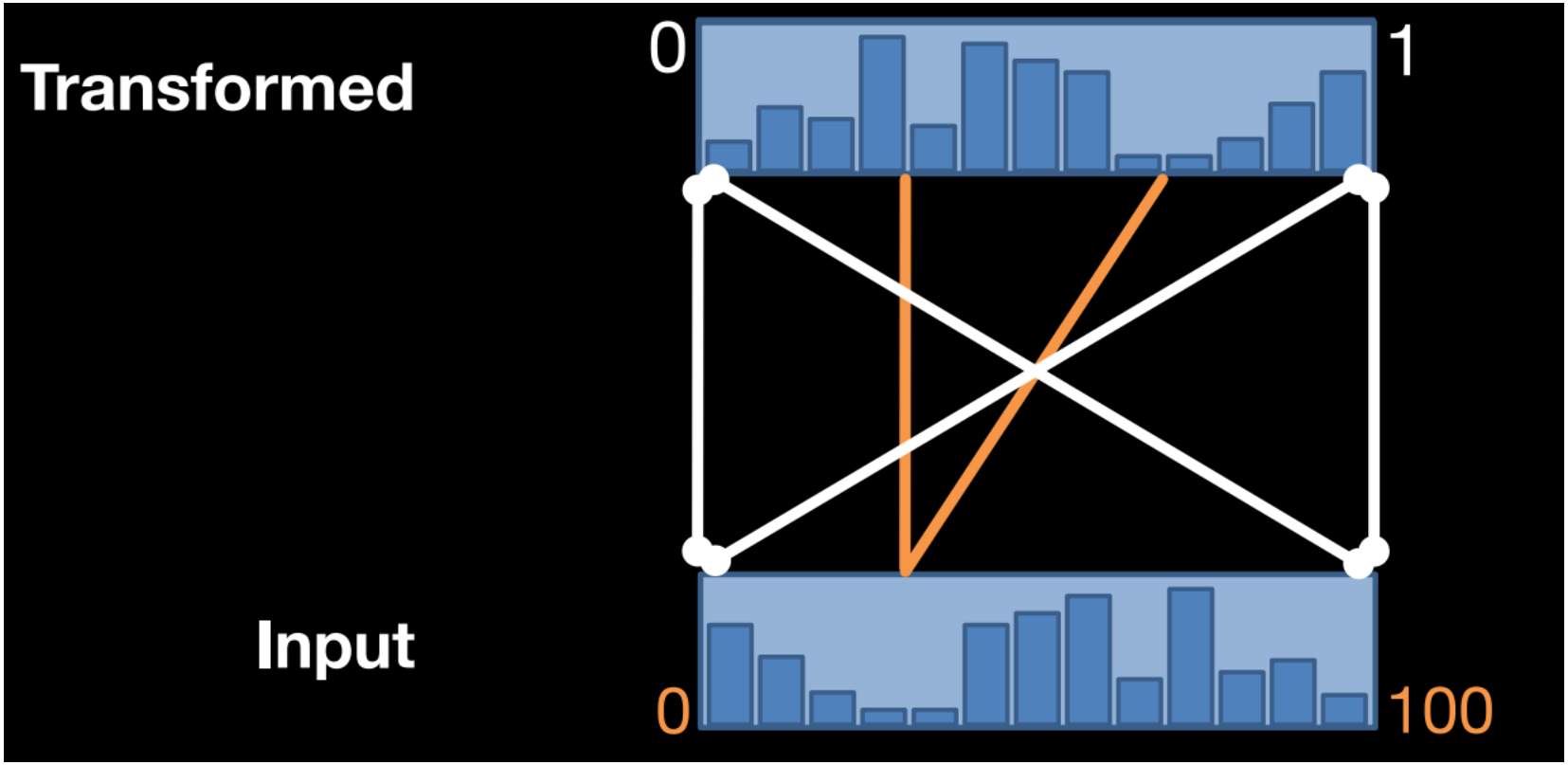


Rank	University	A	B	C
1.	MIT			
2.	Harvard			
3.	Princeton			
4.	Cambridge			
5.	Oxford			

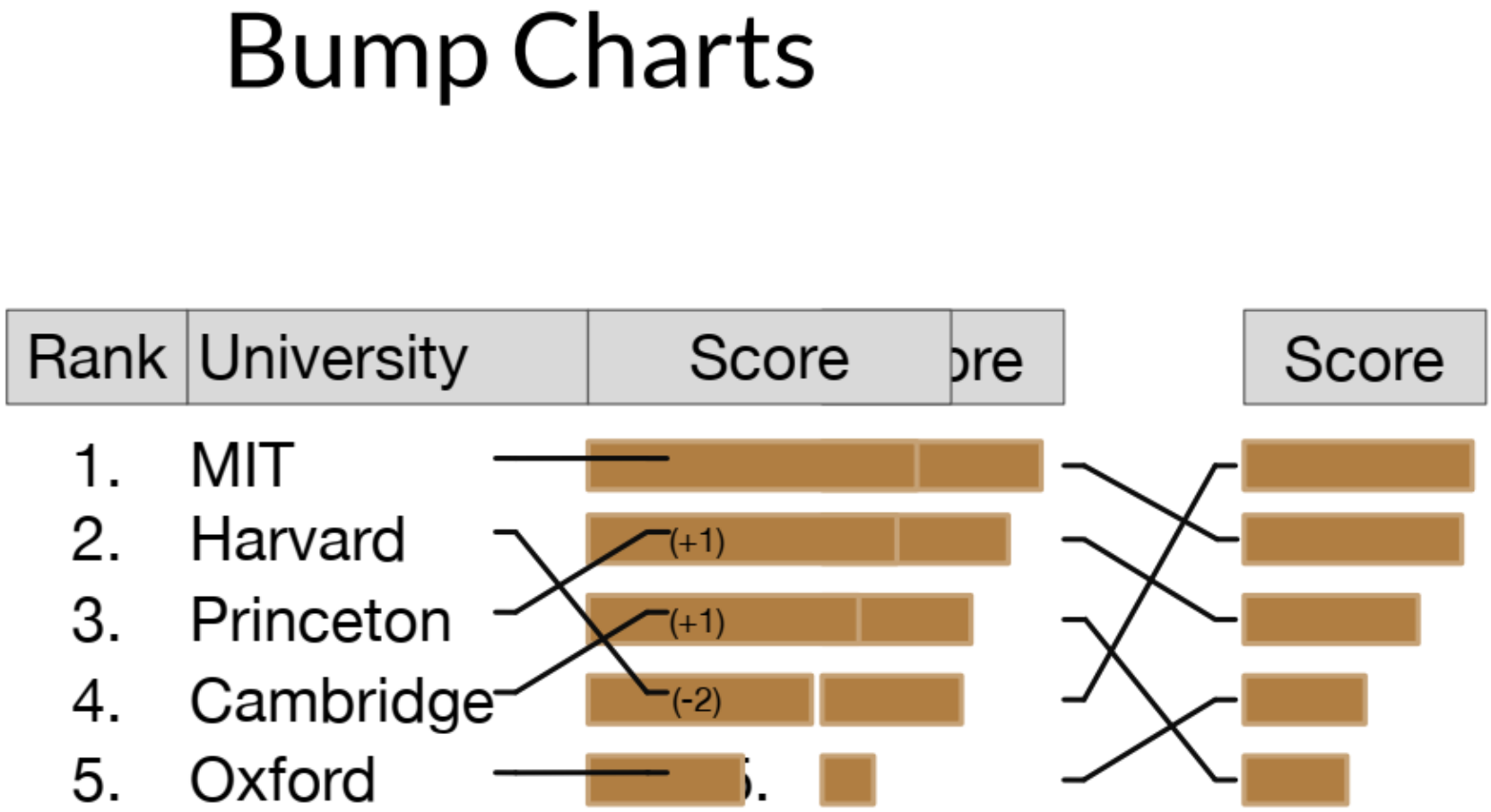
Serial Combiner (as Stacked Bar)

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

MAPPING ATTRIBUTES TO SCORES



BUMP CHARTS


Bump Charts
Rank
University
Score
ore
Score
1. MIT
2. Harvard
3. Princeton
4. Cambridge
5. Oxford
(+1)
(+1)
(-2)