Paul Rosen

paul.rosen@utah.edu @paulrosenphd https://cspaul.com

Visualization for Data Science DS-4630 / CS-5630 / CS-5630 / CS-6630

Tabular Data Visualization



THE UNIVERSITY OF UTAH

dataset types

Tables	Networks & Trees	Fields	Geometr
Items	Items (nodes)	Grids	Items
Attributes	Links	Positions	Positions
	Attributes	Attributes	
Attributes (columns) Items (rows) Cell containing value Cell containing value Attributes (rows) Cell containing value	• Trees	Cell Attributes (columns) Value in cell	PC





osition



Arrange Tables







Many Keys Recursive Subdivision





arrange is the focus of all four design choices for tabular data







spatial channels are the most effective for all attribute types

Techniques and Tasks

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







:	Charles of a read of
	Dath Ares
	Carrier alertics

States of Long

design of the second se











開 1

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



Change over Time

M

Antige States are in the audit of the international of the international of the international of the

demping administration in the delta control of additional tests of a code to additional tests of

Aproficial Real Incident Strategy Real Realized Strategy of Strategy of Realized Strategy of Strategy of Realized Strategy of Strategy of

N Arrastor

interior anima de anima di Anarana 1 Redak

3

Magnitude



F include are sufficient print, and additional total affects ratios.

In the second

F

nerg He contant active of security provides that a sec-





LN E

Ξ. remove a cross



Part-to-whole























tale peta logio nel alti pe dei peta logio co giore eferio i grane entro dori















Spatial



And the second second

Flaw



-II MAR



































Magnitude







Bar Chart Variants



Vertical Bar Chart / Column Chart







Grouped Bar Chart



Horizontal Bar Chart



Comparison of bar chart types







Streit & Gehlenborg, PoV, Nature Methods, 2014

Rank	op Player	Current League	T Current Club	Position	- Foot	# Age	#
	↓ ^A _Z Q T ···	$\downarrow_z^A \boxminus \mathbf{T} \cdots$	↓ ^A z Q T …	$\downarrow_z^{\scriptscriptstyle A} \boxminus {\color{black}{\intercal}} {\color{black}{\dashv}} {\color$	$\downarrow_z^{\scriptscriptstyle A}$ \boxminus \checkmark	↓ ¹ ₉ ▼ …	↓ ¹ ₉
							_
10	Duvitui Dalan	DUC Dramier Line	Zanit Ot Datarahuma		both left right	0.00 38.00	160.00
20	Dmitri Poloz	RUS - Premier Liga	Zenit St. Petersburg	VV	ngnt		
20	Emiliano Rigoni	RUS - Premier Liga	Zenit St. Petersburg	VV OF	both		
30	Sebastian Driussi	RUS - Premier Liga	Zenit St. Petersburg	CF	both		
4 U 5 O	Aleksandr Kokorin	RUS - Premier Liga	Zenit St. Petersburg	CF	right		
50	Anton Zabolotnyi	RUS - Premier Liga	Zenit St. Petersburg		right		
70	Quincy Promes	RUS - Premier Liga	Spartak Moscow	VV	both		
/U	Pedro Rocha	RUS - Premier Liga	Spartak Moscow	VV	right		
0 0	Lorenzo Melgarejo	RUS - Premier Liga	Spartak Moscow	VV	left		
90	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		







3D Pitfall: Occlusion & Perspective

3 2 0 Which one is the tallest bar? What is the pattern in the data?





[Gehlenborg and Wong, Nature Methods, 2012]



3D Pitfall: Occlusion & Perspective

3 2 0 Which one is the tallest bar? What is the pattern in the data?





[Gehlenborg and Wong, Nature Methods, 2012]



IsoType Visualization







http://steveharoz.com/research/isotype

heatmap

- uses heatmap representation
 - matrix layout using keys
 - encode values with color
- often augmented with clustering



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

heatmap

- uses heatmap representation
 - matrix layout using keys
 - encode values with color
- often augmented with clustering





U

heatmap

- uses heatmap representation
 - matrix layout using keys
 - encode values with color
- often augmented with clustering
- here, used on genomic data





Eisen 1998



Bad Color Mapping



Normal Vision



Deuteranope Vision



("Red-Green Blindness")



Good Color Mapping



Normal Vision



Deuteranope Vision ("Red-Green Blindness")







Color is relative!







Part of Whole







Stacked Bar Chart

- Keys: Class, Survival Class is spatial Survival is color
- Left: absolute values
- Right: proportional values





U

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 about the data (eg. total)



Similar to a pie chart - but the centre can be a good way of making space to include more information





pie charts: take care with accuracy











TreeMap

Treemap

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

Phones	Machines		Supplies	Paper		Tables	Chairs
			Storage				
				Binde	rs		
	Copiers	Accessories					
				Art	Appliances	Furnishings	Pookosoo
							DUURLASES





Part of Whole for Time Series







U

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.





Histogram







Box Plots

- aka Box-and-Whisker Plot
- Bad for non-normal distributed data
- Especially bad for bi- or multimodal distributions





Ш

One Boxplot, Four Distributions



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





Violin Plot

• = Box Plot + Probability Density Function





http://web.stanford.edu/~mwaskom/software/seaborn/tutorial/plotting_distributions.html



A Collection of Univariate Plots





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

Surplus/deficit filled line

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









Response Ratio


Change over Time



Line Chart

- Simple
- Familiar
- Accurate
- Fairly Scalable





2001

2003

2002

2004

2005

2006



	1				8	1 1
007 2	2008	2009	2010	2011	2012	2013

don't use line charts for categorical attributes!







don't use line charts for categorical attributes!











Talbot 2011

Aspect Ratio [Cleveland 1994]

- Bank to 45°
 - The aspect ratio of a graph is an important factor for judging rate of change.
 - perceptual principle: most accurate angle judgment is at 45°





U

Counter-Point: Talbot 2012

- people use two different strategies to estimate slope—angle and height
- slope angle accuracy NOT minimized at 45° (closer to 60°)

Residuals (percentage points)

50

10°







Tick Placement

- Ticks help in user interpretation of data, but too much may hinder
- Automatic optimization of label formatting, font size, and orientation
 - placement based on simplicity, coverage, granularity, and legibility





Talbot 2010

Stacked Area Chart







100% Stacked Area Chart







U

Stacked Area vs. Line Graphs





leancrew.com & Practically Efficient



Can you spot the trends? Overall vs Individual Components





86 87 88 89 90 91 92 93 94 95 96 97 98 99 00 01 02 03 04 05 06 07 08 09 1







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 hluth manufull 152.11 as of July 2017

Symbol	Bid	Ask	Last	Change	Т	Chart	Volume	High	Low	∣Value C	hange	Value	Ga	in
DELL	89 3/4	89 13/10	89 3/4	+11/4	î	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	10,310,100	90 1/8	88 1/2	+1.41%	250	17,950	+273.729	13,147
CPQ	48 7/16	48 9/16	48 7/16	- 13/16		han and the second	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	Ŷ	Jun Margan	504,600	27 3/8	25 5/8	+1.93%	250	13,188	+133.159	7,531
COMS	46 1/2	46 9/16	46 9/16	- 25/32	ŧ	The second secon	3,191,100	47 15/10	45 3/4	-1.65%	-102	6,053	+29.79%	1,389
LU	111 5/8	111 11/10	111 9/16	+19/16		and the second second	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	ŧ	and the second s	3,787,800	381 3/16	280	+4.91%	431	9,213	-0.41%	-38
AOL	162 13/16	163	163	+ 8		and the second s	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	Ŷ	and a second and a s	1,323,800	98 1/2	93	+6.41%	705	11,700	+186.769	7,620
SPLN	33 13/16	33 15/10	33 13/16	+ 7/16	î		300,200	34 3/4	33 5/8	+1.31%	88	6,763	+94.60%	3,288
BEAS	131/2	13 5/8	13 5/8	- 7/16	î	Water and the second	389,200	14 1/4	131/8	-3.11%	-44	1,363	-9.17%	-138
GNET	102	103 3/16	101 5/16	+61/8	î	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	307,600	108	97	+6.43%	613	10,131	+130.269	5,731
RNWK	67	67 1/4	67	+ 2 3/4	î	Man and a start when the	1,233,900	69	64 15/16	+4.28%	275	6,700	+79.87%	2,975
MSFT	1731/8	173 1/4	173 5/16	+13/4	Ŷ	and the second second	13,284,500	174 7/16	170	+1.02%	175	17,331	+54.74%	6,131
INTC	133 3/4	133 13/10	133 13/16	- 3 1/8	ŧ	Marine marine	8,094,300	137 1/2	133 3/8	-2.28%	-625	26,763	+65.20%	10,563
TOTAL					Û	and a supervised of the superv		205,302	80,993	+1.63%	2,293	143,280	+79.41%	63,377



https://www.bram.us/2017/09/12/spark-a-typeface-for-creating-sparklines-in-text-without-code/









30

32

0.23 0.45

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



Horizon Chart Explanation

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.

Select Function $y = sin(x)$ ~	Horizon Chart
Mirror Negative Values	_
Include Bin Lines	0.8 -
Mod Height 0.25	0.6 -
Deceline 0	0.4 -
Daseillie	0.2 -
Container Width 400	Press and hold to stack!
Davullainht (0	-0.4 -
Row Height 40	-0.6 -
Match Row Height	-0.8 -



0

Explanation Chart



http://www.horizon-chart-explanation.devinlange.com/



Clipped Graphs





[Lin 2019]



Clipped Graphs







rt by Tree	Show/Hide I	tree						Export	Map Attribute - S	upplement View Option -	🛔 admin	i n
						.6	0					
Oliden			Kindredl	maxNO2	pm25	set	mean03	Age	meanNO2	AirTemp	LabID	Relative
-1				50 T		<u>139</u> .	100	_h.,	80			
	c0	\$7			w.e.		4.4	0 yet.	120	4.4	w.A.	-
1 20	1 10 10		0	0-	0	0	0…	0	0-	0	0	0
	~		38								#89343	#18561684
_			38								#93380	#18561686
			38									#15638474
	-5		38	with			1 THE STATE		Company and			#15512738
1			38									#16324365
			38									#15188353
			38	water			the work		- And			#28016042
2P			38		·····				, nerring	a		#21982847
~			38	man	man		THE R.	+	- nh-	and the second sec	#96031	#1307826
			38									#26200382
- 14			38		-							#16881308
	÷		38	man	man and					1 1 1 1 1 1 1 1	#48318	#15312962
			603461							ALC: NOT THE OWNER OF		#15545852
			603481	· · · · · · · · · · · · · · · · · · ·							#31426	#18880964
			603481									#1141754
			603481									#16332590
U.			603481				- market				#85489	#26922660
			603481									#15173782
-			603481								#112304	#15888674
			603481	mon	march		And the second s		anno	110.0	#87390	#28839931
			603481									#16332521
			603481								#104192	#26371881
			603481	when					- avair			#27022263
			603481		-mp				- Aller		#97322	#29414120
			603481									#16100468
	· · · ·		603481								#113264	#19370604
			603481									#25400258
			603481	Committee of the second							#86753	#740675
			603481									#27026780
			603481								#89343	#18561684
			603481								#93380	#18561686

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.





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





Americans spent more time in their cars as highways networks expanded and more workers commuted from new, far-flung suburbs. The number of commuters rose as more women joined the work force.



In 1973, many Arab oil-producing countries declared an oil embargo against the United States because of its support of Israel in the Middle East. The supply disruption caused oil prices to rise sharply, and are consumption declined.



Gas prices jumped as the Iranian revolution and the Iran-Iraq war caused a rift in the global oil supply. United States energy policy turned to conservation, and Congress imposed the first fuel-efficiency standards for cars.



Gasoline remained cheap for more han a decade, and the average number of miles Americans drove nnnually jumped by more than .000. Economists observed that consumers became less sensitive o small gas-price changes as household incomes rose.



the growth in driving faltered as as prices started to climb. But much of the sharp reduction in fiving was caused by the long eccession and its high unemploynent rate. A small but growing number of thrify and carbononscious commuters switched to icycles and public transportation.

THE NEW YORK TIMES

U

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.



Monitoring the U.S. Outbreak

Confirmed cases by state, ranked by latest full-day count

Daily confirmed cases per 100,000 residents





Note: Trend indicates whether a state had an increase or decrease in total number of cases in the past seven days compared with previous seven days. Last updated March 3, at 1:56 p.m. Sources: Johns Hopkins Center for Systems Science and Engineering; the Lancet; Associated Press; U.S.



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/

1965

Waterfall Chart

 Great way to show evolution of part of whole over time / events (non-linear time)





A history of the European Union

Ranking







Rankings are Popular









U

Ranking



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

<u>nener.github.io/visual-vocabulary-vega/#/Ranking/</u>

Temporal Rankings

1 2 3 4 0	Paris SG Lyon Marseille	9 8 4	38	11	5		1000	the second s		and the second state of th	
2 🗆 3 📕	Lyon Marseille	8	12			3	20	18	36	12	24
3	Marseille	-h	10.0	11	5	3	25	13	33	17	16
4 0		4	38	12	2	5	17	21	24	20	4
	Rennes	*	32 he	10	2	7	16	16	29	24	5
5 =	Lorient	4	31	8	7	4	19	12	32	29	3
6 .	Valenciennes	0	29	8	5	б	21	8	31	24	7
7 🗖	Bordeaux	4	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
41	/ Yer	5	ত্ব	75	10	V		10	20	64	







Table Lens

 Interactive tablebased representation







				Eusiness Objects
Displacement	Horsepower	Weight	Acceleration	Model_Year

Rao & Card 1994

LineUp

					Customized	I Combina	ation	_				
Ra	School Name Filter: <none></none>	Country Filter: 1 out of 71	23.30% Academic reputati	9.85% Facult	16.00% Arts & Hum	10.35% Internat	10.94% Internat	World	29.55% University Ranki	Separato	Rank	Aca
1.	Harvard University	United States								\mathbb{N}	1.	
2.	Yale University	United States								\sim	2.	
3.	Massachusetts Institute of Technology (MIT)	United States								\sim	3.	
4.	Princeton University	United States	100 (1)		90.7 (0.91) 73.08 (0.73) 85.5 ((0.86)	0.94			4.	
5.	University of Chicago	United States									5.	100
6.	University of California, Berkeley (UCB)	United States									6.	
7.	California Institute of Technology (Caltech)	United States								\mathcal{K}	7.	
8.	Stanford University	United States								\mathbb{N}	8.	
9.	Columbia University	United States									9.	
10.	University of Pennsylvania	United States								KX	10.	
11.	Cornell University	United States								ΓV	11.	
12.	University of Michigan	United States								A	12.	
13.	Johns Hopkins University	United States								X X	13.	
14.	New York University (NYU)	United States								\sim	14.	
15.	Duke University	United States								K /	15.	
16.	University of Wisconsin-Madison	United States								\mathbb{N}	16.	
17.	University of California, Los Angeles (UCLA)	United States									17.	
18.	Northwestern University	United States								X	18.	
19.	University of Illinois at Urbana-Champaign	United States								$\left \times\right $	19.	
20.	Brown University	United States									20.	
21.	Purdue University	United States								• /	21.	
22.	University of Texas at Austin	United States								$ \land \land$	22.	
23.	Boston University	United States								$\vdash \times$	23.	
24.	Georgia Institute of Technology	United States								\mathbb{N}	24.	
25.	University of North Carolina, Chapel Hill	United States								(\land)	25.	
26.	Ohio State University	United States								1 1	26.	
27.	University of Pittsburgh	United States									27.	

als and another

Engineering & T Life Sciences & Natural Science Employer reputa Citations per fac





Video at http://lineup.caleydo.org



Correlation







What is Correlation

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













Axis-Based Techniques

Drama







Physics

Table

Dance



Math

Parallel Coordinates



Scatterplots







Scatterplots

- Two orthogonal axis visualizing one dimension each.
- (see prior lecture)





U

Scatterplot Matrices (SPLOM)

- Matrix of size d*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:

 - Choosing dimensions
 - Choosing order



• Clustering & aggregating records
SPLOM Aggregation - Heat Map

Interactive Binned Scatterplot Matrix Dimensions: 5 V Bins: 20 V Data Points: 100k V



Powered by <u>Datavore</u> and <u>D3</u>.



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



SPLOM F+C, Navigation





[Elmqvist]



Parallel Coordinates



Parallel Coordinates (PC)

- Axes represent attributes
- Lines connecting axes represent items





Inselberg 1985



parallel coordinates







Protovis



Parallel Coordinates

- Each axis represents dimension
- Lines connecting axis represent records
- Suitable for
 - all tabular data types
 - heterogeneous data







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











4 8 1









1 4 2

show correlation

- <u>positive correlation</u>: straight lines
- <u>negative correlation</u>: lines cross at a single point



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



do you see any correlations?





homicides 63.00

Fua 1999



PC Limitation: Scalability to Many Dimensions





What is this?

A multidimensional explorer of nutrient data from the USDA. The parallel coordinates displays the nutrient content of foods in

Food Groups

327 Baby Foods 0-Baked-Products

Sample of 25 entries Search Foods.

Alcoholic Beverage, wine, table, red, Gamay Alcoholic beverage, distilled, whiskey, 86 proc



PC Limitations: Correlations only between adjacent axes

- Solution: Interaction
 - Brushing
 - Let user change order





U

PC Limitation: Ambiguity

- Solutions:
 - Brushing
 - Curves









Graham and Kennedy 2003

PC Limitation: Scalability to Many Items

- Solutions:
 - Transparency
 - Bundling
 - Clustering
 - Sampling









HIERARCHICAL PARALLEL COORDINATES

- goal: scale up parallel coordinates to large datasets
 - challenge: overplotting/occlusion











HPC: ENCODING DERIVED DATA

- visual representation: variablewidth 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







Fua 1999



Data-Scalable Parallel Coordinates

 Cluster into groups of homogeneous behavior and represent positive and negative correlations directly







(a) Conventional PCPs

(b) DSPCP using K-means clustering





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





http://bl.ocks.org/jasondavies/1341281

Star Plot

- Similar to parallel coordinates
- Radiate from a common origin







BUICK SKYLARK

1979 AUTOMOBILE ANALYSIS AMC PACER AMC SPIRIT AMC SPIRIT AUDI 2000 BMW 3201 BUICK CENTURY BUICK OPEL BUICK REGAL BUICK OPEL BUICK REGAL BUICK OPEL BUICK REGAL AUG <td

tl.nist.gov/div898/handbook/eda/section3/starplot.h

[Coekin1969]

What do you think?



http://www.visualcinnamon.com/2013/09/making-d3-radar-chart-look-bit-better.html

Zahlenergebnissen proportional ist. Auch können Verlänger= ungen der Nadien über die Peripherie hinaus hiezu benützt werden. Zweckmäßig wird auch hier die lineare Verbindung der Endpunkte der betreffenden Geraden vorgenommen.

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



Figur 4.

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

78

Pixel-based Representations



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









[Gehlenborg & Wong 2012]

HiVE example: London property partitioning attributes



house type neighborhoo d

sale time

encoding attributes

average price (color) number of sales (size)

results between neighborhoods,

different housing



Slingsby 2009



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?



Ħ			Ħ							Ħ			Ħ	Ħ				Ħ
	Ħ									Ħ			Ŧ					
										曲				Ш				
										Ħ								
						Ħ				Ħ				Ħ				
			Ħ			Ħ	Ħ					Ħ						
							Ħ			Ħ								
							Ħ							Ħ				
							Ħ	₩		₩								
		₩		₩	Ш	₩	曲	₩	₩	₩		Ш						
	₩	₩				₩	Ħ	₩	₩	₩	₩		Ħ					
		₩		₩	田	₩	₩	₩	₩	₩	₩		Ħ					
						₩	Ħ			Ħ								
							Ħ											
		誰			Ħ	Ħ	曲	Ħ	Ħ	Ħ	Ħ	Ħ	Ħ					
		Ħ								Ħ								
	Ħ	Ħ																
							Ħ			Ħ		Ħ	Ħ	Ħ				
				Ħ														

10

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











Keim, Kreigel, 1994









Keim, Kreigel, 1994

