#### 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 Cell containing value	• Trees	Grid of positions	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



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 date relationship is most important in your story, then look at the different types of chart





#### Change over Time

M

Columns and and a set of the set

-

An original making danga mikeng danga mikeng na uniganami antar uniganami antar

~

terente di les con el fisiolos

5

#### Magnitude



The data of the da

As an observed interest of others for interest of other interest interest of others for int

II.

density the color and supporter of class of the case lines - on long. at the cities of









山靈

Casel for develop a result of the develop a

inspect of the Ar after dies in Ar after dies in Anter die diese die Anter die die

#### Part-to-whole











Upe for transitional sam includiose missionetifipe, contine citize di normati scher-france ano manufaciali





















Spatial

eate for boro rapi en und de pete buforco georali enerte e deservore eporen











A ......

Flaw



P Pages C dar h

























# Magnitude





#### Bar Chart Variants



Vertical Bar Chart / Column Chart





#### Grouped Bar Chart







200k



# Comparison of bar chart types







#### Stacked bar chart

Streit & Gehlenborg, PoV, Nature Methods, 2014

Rank	op Player	Current League	T Current Club	Position		# Age	#
	↓ <sup>A</sup> <sub>Z</sub> <b>Q T …</b>	$\downarrow_{z}^{A} \square \mathbf{T} \cdots$	↓ <sup>A</sup> <sub>Z</sub> Q T …	$\downarrow_{\mathrm{Z}}^{\mathrm{A}}$ $\square$ $\intercal$	$\downarrow^{\mathrm{A}}_{\mathrm{Z}}$ $\square$ $\intercal$	$\downarrow_{9}^{1}$ T	
							-
				CE W	both left right	0.00 38.00	160.00
10	Dmitri Poloz	PLIS - Premier Liga	Zenit St. Detersburg	w w	right	0.00 38.00	100.00
2 0	Emiliano Pigoni	RUS - Premier Liga	Zenit St. Petersburg	W	both		
3 0	Sebastián Driussi	RUS - Premier Liga	Zenit St. Petersburg	CE	both		
4 0	Aleksandr Kokorin	RUS - Premier Liga	Zenit St. Petersburg	CF	right		
5 0	Anton Zabolotnyi	RUS - Premier Liga	Zenit St. Petersburg	CF	right		
6 🗆	Quincy Promes	RUS - Premier Liga	Spartak Moscow	W	both		
7 0	Pedro Rocha	RUS - Premier Liga	Spartak Moscow	W	right		
8 0	Lorenzo Melgareio	RUS - Premier Liga	Spartak Moscow	W	left		
9 0	Zelimkhan Bakaev	RUS - Premier Liga	Spartak 2 Moscow	W	left		
10 🗆	Luiz Adriano	RUS - Premier Liga	Spartak Moscow	CE	right		
11 0	Zá Luíc	PLIS - Promior Liga	Spartak Massow		loft		
12 0	Abmod Musa	RUS - Premier Liga		CF	both		
12 0	Ender Chalov	RUS - Premier Liga	CSKA Moscow	CF	right		
14 0	Timur Zhamalatdinov	RUS - Premier Liga	CSKA Moscow	CF	right		
15 0	Wandorson	RUS - Premier Liga	EK Kraspadar		right		
16 0	loãozinho	RUS - Premier Liga	EK Krasnodar	W	loft		
17 0	Androi Ivon	RUS - Premier Liga	EK Krasnodar	W	right		
18 0	Ricardo Labordo	RUS - Premier Liga	EK Krasnodar	W	right		
19 0	Magamad-Shapi Sulovn	RUS - Premier Liga	EK Krasnodar	W	loft		
20	Endor Smolov	RUS - Premier Liga	EK Krasnodar	CE	right		
20 0	Ivan Ignatiev	RUS - Premier Liga	FK Krasnodar	CF	right		
22 □	Alan Kasaev	PLIS - Premier Liga	Lokomotiv Moscow	W	right		
23 □	lefferson Farfán	PLIS - Premier Liga	Lokomotiv Moscow	W	right		
24 🗆	Arshak Korvan	RUS - Premier Liga	Lokomotiv Moscow	W	right		
25 □	Éder	RUS - Premier Liga	Lokomotiv Moscow	CE	both		
26 □	Δri	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	CE	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		
0							







#### 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







[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	

0.2

0.7

0.5

0.7

0.5

#### heatmap

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



#### 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







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





### 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	<b></b>				
				Binders				
	Copiers	Accessories						
				Art	Appliances		Furnishings	Bookcases





## 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.





# Histogram









80

#### Box Plots

- aka Box-and-Whisker Plot
- 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





#### 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



[Corell et al, InfoVis 2019]

(b

### 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



#### Surplus/deficit filled area

Same as before.





Juxtaposing Two Variables (male/female)


### Change over Time



### Line Chart

- Simple
- Familiar
- Accurate
- Fairly Scalable







### 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





#### much may hinder size, and orientation , 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

2M -

1M







#### 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 hlub mandull 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 13/16	6 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	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	Jan Marine and Street	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 a second	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	Ť	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 15/16	33 13/16	+ 7/16	Ť	, and the hard of the second	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	Margare Provide Street	389,200	14 1/4	131/8	<mark>-3.11%</mark>	-44	1,363	-9.17%	-138
GNET	102	103 3/16	101 5/16	+ 6 1/8	î	MINING WITH AND	307,600	108	97	+6.43%	613	10,131	+130.269	5,731
RNWK	67	67 1/4	67	+ 2 3/4	î	Non and a state of the state	1,233,900	69	64 15/16	+4.28%	275	6,700	+79.87%	2,975
MSFT	1731/8	1731/4	173 5/16	+1.3/4	Ŷ	and a second second	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	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 sub-		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/

By Peter Zelchenko









30

32

## 0.23 -1.20.45

#### 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 -
Baseline 0	0.4 -
Container Width 400	Press and hold to stack!
Pow Height 40	-0.4 -
	-0.6 -
Match Row Height	-0.8 -



#### **Explanation Chart**



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



### Clipped Graphs





## 10 R 5 0

**Clipped Graphs** 

#### [Lin 2019]



### Clipped Graphs







rt by Tree	Show/Hide tr	ee						Export	Map Attribute - S	Supplement View Option +	🌢 admin	i #
51												
Childen			Kindredl	maxN02	pm25	set	mean03	Age	meanNO2	AirTemp	LabID	Relative
aj				100 50		13%	100 50 6		100 60	40 0 0		
	00 and an	1	w.ek	44	44		ψA	0	120 vgrulie			
1980 20	2010 2010		0	0	0	0	0	0	0	0	0	0
-			-									
5		H-	38								#89343	#18561684
		H	38		la di si di						#93380	#16538474
- C		H	38	1 miles			The second		Company of the			#15512738
-		n	38						_			#16324365
			38									#15188353
			38	wy			mar		m			#28016042
			38		····							#21982847
			38	man	mon				- A Anon		#96031	#1307826
			38									#26200382
-			38									#16881308
	· · · · · · · · · · · · · · · · · · ·		38	mon						-	#48318	#15312962
			603481							ALC: NOT THE R.		#15545852
			603481								#31426	#18880964
			603481									#1141754
-			603481									#16332590
			603481								#85489	#26922660
-			603481									#15173782
0			603481								#112304	#15888674
		<u>H</u>	603481	m			V1 1 1		Chan the second	THE R. LEWIS CO., NAME	#87390	#28839931
		H-	603481			-						#16332521
_		<u>H</u>	603481	COULT and a					and the second second		#104192	#26371881
		H	603481									#27022263
		8	603481	11 11 11			and annual		E Y M		097322	#29414120
	-	8	603481									#16100468
		H	603481								#113264	#19370604
		H	603461			100					496753	#25400258
		H	603461			12					#80/53	#740070
		H	603481								#80343	#18561684
		H	603481							a second a second	#03380	#18561696
		hand .	000401							and the second	#93380	#10301000



#### 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





ericans spent more time in their as highways networks anded and more workers muted from new, far-flung urbs. The number of commuters as more women joined the k 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 age concumption 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.



asoline remained cheap for more nan a decade, and the average umber of miles Americans drove nnually jumped by more than 000. Economists observed that onsumers became less sensitive o small gas-price changes as ousehold incomes rose.



and growth in uriving faithered as as prices started to climb. But much of the sharp reduction in friving was caused by the long ecession and its high unemployment rate. A small but growing number of thrifty and carbonrouscious commuters switched to icycles and public transportation.

THE NEW YORK TIMES

es: Energy Information Administration; Federal Highway Administration; Brookings Institution

### 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)



1965

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

### Waterfall Chart

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

#### A history of the European Union

How the European Economic Comunity's territory grew and shrank throughout the years, up until Brexit





### Ranking





U

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

### Temporal Rankings

Paris SG       S3       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         5       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       18       11       24       18       6         9       Nice       29       7       8		······		Pts ▽	W	D	L	Но	Aw	GF	GA	GD
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         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       ■	1 -	Paris SG	\$	38	11	5	3	20	18	36	12	24
3       Marseille       38       12       2       5       17       21       24       20       4         4       Rennes       32       10       2       7       16       16       29       24       5         5       Lorient       32       31       8       7       4       19       12       32       29       3         6       Valenciennes       31       8       7       4       19       12       32       29       3         6       Valenciennes       29       8       5       6       21       8       31       24       7         7       Bordeaux       4       29       6       11       2       15       14       21       14       7         8       11       29       7       8       4       18       11       24       18       6         9       Nice       4       29       7       8       4       21       8       26       26       0	2 🗆	Lyon	\$	13	11	5	3	25	13	33	17	16
4       ■       Rennes       32       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	3 -	Marseille		38	12	2	5	17	21	24	20	4
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	4 🗆	Rennes	*	32 18	10	2	7	16	16	29	24	5
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	5 =	Lorient	4	31	8	7	4	19	12	32	29	3
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	6 .	Valenciennes	8	29	8	5	6	21	8	31	24	7
8  Lille  29  7  8  4  18  11  24  18  6	7 🗖	Bordeaux	4	29	6	11	2	15	14	21	14	7
9 🛡 Nice 🆈 29 7 8 4 21 8 26 26 0	8 🗖	Lille	4	29	7	8	4	18	11	24	18	6
	9 🗖	Nice	\$	29	7	8	4	21	8	26	26	0





### Table Lens

 Interactive tablebased representation







				Business Objects
Displacement	Horsepower	Weight	Acceleration	Model_Year

#### Rao & Card 1994

### LineUp

				_	Customized	d Combina	ition				
De	Ochool Marrie	Question	23.30%	9.85%	16.00%	10.35%	10.94%	29.55%		Deels	
ка	Filter: <none></none>	Filter: 1 out of 71	Academic reputati	Facult	Arts & Hum	Internat	Internat	World University Ranki	Separato	капк	AC
1.	Harvard University	United States								1.	
2.	Yale University	United States	Sec. 1							2.	
3.	Massachusetts Institute of Technology (MIT)	United States								3.	
4.	Princeton University	United States	100 (1)		90.7 (0.91) 73.08	(0.73) 85.5 (0	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								7.	
8.	Stanford University	United States								8.	
9.	Columbia University	United States								9.	
10.	University of Pennsylvania	United States							$ X^{}$	10.	
11.	Cornell University	United States								11.	
12.	University of Michigan	United States							A	12.	
13.	Johns Hopkins University	United States							$\langle \rangle$	13.	
14.	New York University (NYU)	United States								14.	
15.	Duke University	United States							X /	15.	
16.	University of Wisconsin-Madison	United States		0						16.	
17.	University of California, Los Angeles (UCLA)	United States							$\wedge$	17.	
18.	Northwestern University	United States								18.	
19.	University of Illinois at Urbana-Champaign	United States								19.	
20.	Brown University	United States								20.	
21.	Purdue University	United States							4	21.	
22.	University of Texas at Austin	United States		-						22.	
23.	Boston University	United States							$\rightarrow$	23.	
24.	Georgia Institute of Technology	United States								24.	
25.	University of North Carolina, Chapel Hill	United States								25.	
26.	Ohio State University	United States							0 10	26.	
27.	University of Pittsburgh	United States								27.	

als see \_\_\_\_

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





#### Video at http://lineup.caleydo.org



### Correlation





U

#### What is Correlation

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











-1









### Axis-Based Techniques

Drama

65





#### Physics Dance 95 70

Table

90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90



Math

85

#### Parallel Coordinates



### Scatterplots





U

### 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





Ľ

### 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







Ш

	V1	V2	<b>V</b> 3	V4	V5
D1	7	3	4	8	1
D2	2	7	6	3	4
D3	9	8	1	4	2





U





#### 4 8 1









#### 4 2

## show correlation

- positive correlation: 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?





1homicides 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







Ш

## 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





ANC CONCORD.



BUICK LE SABRE



BUICKSKYLARK

#### 1979 AUTOMOBILE ANALYSIS AMC PACER AMC SPIRIT AUDI 5000 AMC PACER AMC SPIRIT AUDI 5000 AUDI 5000

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

#### [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 neighborhood sale time

#### encoding attributes

average price (color) number of sales (size)

#### results

between neighborhoods, different housing distributions within neighborhoods, similar prices





#### 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?



	HH				H			HH	HH		H	H			H	HH	H	HH		H		HH		Ŧ
																						Ħ		
																								Η
																								Ħ
																								井
																								Ŧ
																								Η
											Ħ													
	<u>H</u>									ЩЩ								Щ						벽
																								Ŧ
	H									Ħ														
						Ħ			H						₩		H	H				##		
					Ш	Щ		Ш	Щ		Щ			<u>H</u>		ЩЩ	Щ	Щ	#					H
																								Ŧ
	H								Ħ	H	H	H			H		H					H		Ŧ
									₩		₩				₩	₩	₩							Н
		<u>H</u>		<u>H</u>		ш	Ш	Ш	Щ	Ш	Ш	LH -			Ш		Ш	Ш						
			▋キキ								Ħ			<b>H</b>		ĦĦ								
ŦŦ	ΗŦ	ĦŦ	<b>III</b>	ĦŦ		Ħ	Ħ	Ħ	Ħ	Ħ	Ħ	Î	I	ĦŦ	Ħ	ΗĦ	Ħ		H	ŦŦ	H	Ħ		Ē
										₩	⊞													Η
					Ш	Ш		Ш	Ш		Ш	H		▦	Ш	Ш	Ш							
			╏┼┼								Ħ				Ħ	Ħ								Ħ
<b>H</b>	ΗŦ	Ħ	ÎĦŦ	ĦŦ	İΠ	Ħ	Ħ	Ħ	Ħ	Ħ	Ħ	ITT	I	Ħ	Ħ	ΗĦ	Ħ	H	ŦĦ	ŦŦ	H	Ħ	<b>H</b> H	Ħ
									⊞	₽	₽													믭
															Ш									H
											Ħ			<b>H</b>		HH								
	<b>FT</b>	ĦŦ		ĦŦ	İΠ	Ħ	Ħ	Ħ	ΠĦ	Ħ	Ħ	itt	i i i i i i i i i i i i i i i i i i i	Ħ	Ħ	İΠ	Ħ	Ħ	Ħ	Ħ		Ħ		Ħ
									⊞		₩				⊞									
																$\blacksquare$								Η
											Ħ				Ħ	ĦĦ								
	F	Ħ		Ħ	Ħ	Ħ	Ħ	Ħ	Ħ		Ħ		H	Ħ	Ħ	Ħ	Ħ		ŦŦ				H	Ħ
																								믭
											Ħ													
	T	Ħ	i III		<b>H</b>	Ħ		Ħ	Π	Ħ	Ħ							Ħ						T
																								H
											H											Ħ		Ħ
																								H
											H													Ħ

### 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









			Ħ		Ħ	Ħ							
					H		Ħ						
					⊞								
									Ħ				
Ħ	Ħ	Ħ		Ħ	Ħ	Ħ	Ħ		Ħ				Ŧ
Ħ	Ħ	Ħ	Ħ		⊞	Ħ	Ħ	Ħ	Ħ				
		Ħ	Ħ	₩	₩								
		₩		₩	₩								
#		井	曲	₩	붜				Ħ				
	H	#	▦		₽	Щ			Ħ				
Ħ		▦	Ħ		▦	Ш				Ш			
Ħ		▦			▦								
		₽	Ħ		₽	▦							
Ħ	Ħ	Ħ	Ħ	Ħ	Ħ	Ħ	Ħ		Ħ	Ħ			
Ħ	Ħ	詌	Ħ	Ħ	苗	Ħ	Ħ		Ħ	Ħ			
Ħ	Ħ	詌	Ħ	艹	Ħ	Ħ	Ħ		卌	Ħ			ŦŦ
₩		₩		₩	₩	₩			⊞				
Ħ		₩	曲	#	曲	Ħ			曲				
#	H	₩	Ħ	₩	붜	Ħ			Ħ				
		Ш			Ш								
		▦			▦				Ħ				
	Ħ	Ħ	Ħ	ŦŦĒ	Ħ	Ħ	Ħ	Ħ					
Ħ	Ħ	Ħ	Ħ					Ħ		Ħ			

#### Keim, Kreigel, 1994

