Visualizing Multi-Dimensional Data

Chen He

Data abstraction: Multi-dimensional data



1-dimensional data -- Word clouds

Show words used in text data

- Pros quickly **perceive prominent** words; draw attention;
- Cons contradiction between size and length; lose context;



1-dimensional data -- Box plot

Use quartiles to graphically depict a group of numerical data.

Interquartile Range

- Pros Summarization;
 - Easy comparison between datasets.
- Cons Details are invisible;





1-dimensional data -- Box plot



IQR: interquartile range

Suspected outliers: > 1.5 IQR + 3rd Quartile; < 1st Quartile - 1.5 IQR.

Outliers:

- > 3 IQR + 3rd Quartile;
- < 1st Quartile 3 IQR.

"I've stopped using box plots."



"I've stopped using box plots."









1-dimensional data -- Histogram

Depict the frequency distribution of a univariate data set.

By convention, an interval = [lower bound, upper bound)



1-dimensional data -- Histogram

Unequal bin sizes (not recommended)

Frequency (the area) = Class Width x Frequency Density.

Density is not easy to understand.



2-dimensional data -- Bar chart

Quantitative Benefits from both dependent position (top of bar) variable and length (size of bar) Quantitative • Position Position Discrete/nominal Length Density Angle independent variable Slope / Hue Area • • Length Density • • • Angle Saturation ••• Slope

Nominal Ordinal • Position •.• ... Hue ∠ Saturation ••• Density • • Saturation Shape . . . _ ∠ Length _ 2 1/ Angle 1-Hue ••• Area Slope .. Shape • • • Shape ● ▲ ■ Area ..

2-dimensional data -- Line chart



Benefits from position but not length

2-dimensional data -- Scatter plot



2-dimensional data -- Gantt chart



2-dimensional data -- Gantt chart

		Q1 2009			Q2 2009			Q3 2009	
lask Name	Dec '08	Jan '09	Feb '09	Mar '09	Apr '09	May '09	Jun '09	Jul '09	Aug
Planning		<i></i>							
Research			S						
Design				<i></i>					
Implementation									
Follow up							Ø		

2-dimensional data -- Table

Discrete/nominal Benefits from independent position only variable (notice the lateral inhibition flashing?) Discrete/nominal independent variable

2-dimensional data -- What chart to use?

Don	Quantitative Continuous	Bar	Line
Dep.	Quantitative Discrete	Bar	Bar
Ind	Quantitative Continuous	Gantt	Scatter
ina.	Nominal or Q. Discrete	Table	Gantt
		Nominal or Q. Discrete	Quantitative Continuous
		Indepe	endent

Chart source: Sala-Burbaré, A., Peltonen, J., Pyhältö, K., & Castelló, M. (2018). Doctoral candidates' research writing perceptions: A cross-national study. *International Journal of Doctoral Studies*.

How about this chart?



Figure 2. Doctoral candidates' research writing profiles

So far - 1D data

word clouds



box plot



histogram



So far - 2D data

Bar chart, scatter plot, line chart, gantt chart, and table.

Don	Quantitative Continuous	Bar	Line
Dep.	Quantitative Discrete	Bar	Bar
Ind	Quantitative Continuous	Gantt	Scatter
ina.	Nominal or Q. Discrete	Table	Gantt
		Nominal or Q. Discrete	Quantitative Continuous
		Indepe	endent

https://www.gapminder.org/tools/#\$chart-type=bubbles

Multi-dimensional data -- Add additional dimensions on top of 2D charts



Quantita	tive	Ordina	al	Nomin	al
Position	•••	Position	•••	Position	•••
Length	=	Density	•••	Hue	•••
Angle	2	Saturation		Density	•••
Slope	1-	Hue		Saturation	
Area	• •	Length	—	Shape	• • •
Density	• • •	Angle	2	Length	—
Saturation		Slope	1-	Angle	2
Hue		Area	••	Slope	11
Shape	• • =	Shape	• • •	Area	• •

Multi-dimensional data -- Add additional dimensions on top of 2D charts



https://medium.com/@jshlbrd/hunting-for-powershell-using-heatmaps-69b70151fa5d

Multi-dimensional data -- Scatter plot matrix

N variables mean n*n plots.

Diagonal maps the same variable twice.

Each pair is plotted twice, once on each side of the diagonal.

Allows convenient sequential browsing of one variable **compared** to all other variables.



Brushing is the process of interactively selecting a subset of data items from a visual representation.

Brushing & linking cause the brush effect (highlighting, etc.) to be applied on those points in the other plots that represent the same data items.

Useful for exploring relationships in multi-dimensional data.

Brushing & linking scatter plots



Multi-dimensional data -- Parallel coordinates

https://www.coursera.org/learn/datavisualization/lecture/2v40S/2-2-2-parallel-coordinates





Interacting with parallel coordinates

Sort an axis; Reorder axes; Single axis brushing; Angular brushing.





Each data value placed somewhere along the line, scaled between the minimum at the bottom and the maximum at the top.

Ordering of the axis is fundamental to draw insights.

Multi-dimensional data -- Edge bundling for parallel coordinates





Radial coordinates / Radar chart

- + Useful for seeing which variables have similar values or if there are any outliers amongst each variable.
- + Visually interesting.



Comparing Cars

Radial coordinates / Radar chart

- Visually cluttered with too many polygons;
- Not easy to compare values along circular grids;



Radial coordinates / Radar chart

- When axes have different scales, values are not comparable between axes.



Parallel sets for categorical dimensions

Show data frequencies rather than individual data points.





Titanic Survivors

Chernoff faces

Invented by Herman Chernoff in 1973.

Encode different variables to different facial features, like the shape of the head and the size of the eye.

Assumption: object-like appearance is useful for perception.



Unpredictable effects from emergent expressions.

Chernoff faces -- Not generally adopted in practice

Human differences in sensitivities to different features make the perceptual spaces likely to be non-linear.

E.g. eye size and eyebrow-slant were found to carry significant weight.



This example shows Chernoff faces for lawyers' ratings of twelve judges.

So far - 1D data

word clouds



box plot



histogram



So far - 2D data

Bar chart, scatter plot, line chart, gantt chart, and table.

Don	Quantitative Continuous	Bar	Line
Dep.	Quantitative Discrete	Bar	Bar
Ind	Quantitative Continuous	Gantt	Scatter
ina.	Nominal or Q. Discrete	Table	Gantt
		Nominal or Q. Discrete	Quantitative Continuous
		Indepe	endent

So far - Multi-dimensional data

Add additional dimensions on top of 2D charts



Scatter plot matrix

Parallel coordinates





So far - Multi-dimensional data

Radar chart



Parallel sets



Chernoff faces



Multi-dimensional data -- Composite visualization

Juxtaposed views

Superimposed views

Overloaded views

Nested views

$A \otimes_{_{jux}} B = A B$
$A \otimes_{sup} B = A B$
A ⊗ _{ovl} B = A B
$A \otimes_{\text{nst}} B = A^{\mathbb{B}}$

Juxtaposed views (coordinated views) -- ComVis

Place visualizations side-by-side in one view.



Juxtaposed views (coordinated views) -- Bohemian bookshelf Coordinate surprising aspects, like cover color and page numbers, to support serendipitous discovery.



The Bohemian bookshelf, Thudt et al., 2012



Juxtaposed views

Implicit visual linking

Benefits

Little clutter to the resulting display. Scalable.

Drawback

Not easy to see implicit visual links.

Apply for

heterogeneous datasets;

where different independent visualizations need to be combined.



Superimposed views -- Overlay two visualizations in one view.



Superimposed views

Benefits

Direct comparison.

Drawbacks

Visual occlusion;

High visual clutter;

Multiple views need to share the same spatial mapping.

Apply for where

comparison is common,

the component views need to be as large as possible (potentially the entire available space).



Overloaded views -- Utilize the space of one visualization for another. Scatting points in parallel coordinates

Parallel coordinates are difficult to comprehend.

Scatter plots are more intuitive to discover correlations.





Overloaded views

Benefits

Views do not have to share the same coordinate space.

More flexibility and control over visual clutter.

Drawbacks

Visual clutter is increased

The visual design dependencies between components are significant.

Apply for situations where one visualization can be folded into another to yield a compact (and complex) result.

Nested views -- Nest the content of one visualization inside another visualization.

2	Agrtu_1611948	Agrtu_1589125	Fratu_6670835	Parac_4644728	Acine_5008393	Pyrae 1831256	Bacha_150157	Buchn_156167	Nosto_1723256	Desvu_4056227	0eosu_399981	Riopr_1580457	Oceih_230995;		Theth_4010083	Agitu_1588791	Helpy_1584541	Brume_179876	Nosto_1723156	Sinme_159654		Bacsu_160794	Staau_1592543	reipe_10123/2	Pseae_1000010	Braja_2737602	Baoth_2934848			Glovi 3752113 Fusnu 1970358 Desps 5124521	Haein 1627364 Haein 1627230 Glovi 3752113	Leixy 5095488 Laola 1567230 Helpy 1564618 Helpy 1564555 HelpJ 1561194	Meslo 1347076 Lisin 1680150 Lisin 1680053	Meslo 1347409 Meslo 1347330 Meslo 1347328	Meslo 1347584 Meslo 1347584 Meslo 1347452	Metth 1567850 Mettac 2009010 Mesto 1347401	Pasmu 156019 Oniye 3993862 Oniye 3993897 Oceih 2309833	Agrtu 1588816 Pasmu 156035	Agrtu 1589041 Agrtu 1588872 Agrtu 1588795 Agrtu 1589143 Agrtu 1611964	Agrtu 1589174 Agrtu 1589073 Agrtu 1611968 Agrtu 1589176 Agrtu 1589073	Agrtu 1611928 Agrtu 1589182 Agrtu 1589177	Bifle 2346552 Baosu 1608043 Baosu 1608043 Arofu 1146919	Borbr 3360352 Borbr 3360228 Borbr 3360196 Borbr 3360162 Borbr 3360114	Braja 2738202 Braja 2738160 Braja 2737909 Braja 2737920	Camje 157922 Brume 179880 Braja 2737842	Corgl 1955186 Cloae 1589586 Cloae 1589427	Vesps D12441t Desps 5124541 Dehet 5723476
A-4-46440400	0	õi	•	2	8	бі -	: 8	8		ü	10	ø	6	8	8 3	3 10	ā	ස	Ñ	4	8	6	5 4	6 8	4	8	8	-	Meslo_13470759	10	2 9 2	29.29.8	844	8 জ ম	2 7 2	5684	8 6 8 8	94 12 8	00100	07670	0007	5 % 2 0	04284	0-000	5 8 8 8	52878	× 77 76 16
Agitu_10119490		1		-	-	+	-	-	-			-			-			H	-		-	-	-		-	+			Mesio_13473259 Mesio_13473305 Mesio_13474007	-	Ľ,									-			1111				
Fratu 56708350				-	-	-	-	-	-					r	17				-	-	-	-	-	-	-	+			Meslo_13474522 Meslo_13475817		E									ш						111	
Parao 46447291					-	+	+	-	+									h		-	+	-	+	+	┢	+			Meslo_13475942 Meslo_13474014													1.1					
Acine_50083933						+		+						r		r	т	E		-	-	+	+	+	+	+	+		Metao_20090100 Metth_15678506		12				14			į.		**	**		1.		11 14	4.	
Pyrae_18312565	-	-	-	-				t	+	1								E	-	\vdash	-	+		-	-	-	+		Oceih_23098332 Oniye_39938976	Ť1	E										••					11	
Salen_16762471	-	-	-					T	+					т		г	т	E	\vdash		+	+	+	+	-	+	+		Oniye_39938620 Pasmu_15601990	2				1	- 11. - 11.		i a la										
Bacha_15615723		-																			+	+	+	+	+	+	\square		Pasmu_15803594 Agrtu_15888162 Acine_50085668														10.11				144
Buchn_15616768	-				Т	Т	1		Т					т		г	Т				+			+	╈				Agrtu_16119640 Agrtu_15891435																		
Nosto_17232565						+						\square			IC.									\top					Agrtu_15887957 Agrtu_15888729		Ц		н			-							**			-	
Desvu_46562273												П																	Agrtu_15890417 Agrtu_15890732 Agrtu_15891761		Ľ.		н	-		-	-						**		H	HH	
Geosu_39998119																													Agrtu_16119689 Agrtu_15890737				H		Ħ	11							Ħ				
Ricpr_15604579												-																	Agrtu_15891743 Agrtu_15891779				п		п	п							т				
Oceih_23099576												-	_	_															Agitu_16119287 Agitu_16119287 Arcfu_11499199	÷.													1.11		1 44		11
Symth_61892208													_	-															Bacsu_16080435 Bacsu_16080423																		
Theth_46199838																													Biflo_23465528 Borbr_33601144	-				4 -													
Theel_22299247															_										L				Borbr_33601020 Borbr_33601951 Borbr 33602284	- 11										**	••						#
Agrtu_15887919								e.							۰.							4			P				Borbr_33603622 Braja_27378209				1					ĩ	***	-							
Helpy_15645416					η.		٩.	P			_										۰.								Braja_27379091 Braja_27381509	-																	
Brume_17987643		-		\rightarrow	+	+		Ļ	+							P	ι.,	-			_			+	+	_			Braja_27382020 Braja_27382710 Braja_27378421																		
Nosto_17231552				-		I	1	I	I													I	I	I	I				Brume 17988945							1											
Sinme					1		Г							Г							Г																1	- Г						i i		4.4.1	
					L							L						-								-													~	-							
					L											-	/																			0	1		/								
																																	L														
																					1																										



Nested views

Benefits

Very compact representation; Easy correlation.

Drawbacks

Limited space for the client visualizations;

High visual clutter;

Visual design dependencies are high.

Applications

Call for augmenting a particular visual representation with additional mapping.



Recap - 1D data

word clouds



box plot



histogram



Recap - 2D data

Bar chart, scatter plot, line chart, gantt chart, and table.

Don	Quantitative Continuous	Bar	Line
Dep.	Quantitative Discrete	Bar	Bar
Ind	Quantitative Continuous	Gantt	Scatter
ina.	Nominal or Q. Discrete	Table	Gantt
		Nominal or Q. Discrete	Quantitative Continuous
		Indepe	endent

Recap - Multi-dimensional data

Add additional dimensions on top of 2D charts



Parallel coordinates



Radar chart



Scatter plot matrix



Parallel sets



Recap - Multi-dimensional data

Chernoff faces



Composite views

Juxtaposed views

Superimposed views

Overloaded views

Nested views





	В	=	A	B	
--	---	---	---	---	--

	B	
--	---	--

Applications of multi-dimensional data visualizations

Magnet metaphor









https://www.youtube.com/watch?v=bm59Y8QYbAQ



Domino: Relations across multiple tabular datasets.



Item Type 1 ____ Item Type 2





(c) Parallel Sets [18]











(f) Scatterplot Matrix (SPLOM) [4]



Domino: Relations across multiple tabular datasets



https://www.youtube.com/watch?v=iFgCBI4T8ks

		Addition of the local division of the local	a larger further al. have	and the same	a state have seen to be a set of	and the second second second second
			100	T AND THE	The Transfer Care 1	and in land on
transferrer 1	Sec. 1	- termine makers	trans furgesters	a tan one of the first the	and dealers frame fragments	C Descourse bank, and here.
	1000		-	and a parent of the second sec		and the second se
		the second se	and the second second	· · · · · · · · · · · · · · · · · · ·	Contraction of the local division of the loc	and the second s
and the second se	100300					
Rosent August	Contractions (
The Property of Control of Contro	1000					and the second se
Dimently Children	140.040					A Designation of the local division of the l
pressy changes a	140740			1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		
Country Country	1.00			1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		
Calleria and A. P. Samuelle, Calleria	Date: Test					
FRANKS STRENGT,				16.6		And Personal Property lies of the local Person
an and and all approximite	1.00			1 Concentration (1997)		and the second se
allow Property constraints	- Berlinsen			100		Contraction of the local division of the loc
	-		State Street			and the second se
And the second sec						
State of the local state of the						and the second se
and the second se	1000					
a constant of the second second						and the second se
and the second s						COLUMN TWO IS NOT
and the second sec				And a second sec	and the second	and the second s
And all the second second second second				And a second sec		and the second s
And and a Real Property lies and						Contract of Contra
and the second se	-					
Real Provide Contractory of the State						
and a second sec						
and the state of t	and the second second		and the second se			
A higher that has a second second			and the second se	1< 0.0	and in the second se	6
Participant and	Contraction of the			- 100 H		
francisco di Internetti	1000				and the second se	
NAMES OF TAXABLE PARTY AND ADDRESS OF TAXABLE	Contractor 1					
The Real Property.	- Internet				the second se	
Strength of Statement	and the second second					
Period and the second s	in the lines					
of the set of the sector in these sectors	in the second second		and the second se	Contraction of the second seco		
And and a first summary in	and the second			1 V/V 112	and the supervised of the local division of	
Concise on Annalysis	interesting 1			Loco and	the second se	
Contract Include	and the second			DOCO, Inter	and the second s	
manufactory in residual.	industrian 1		and the second se	1.00 C 144	and the second se	
manifold Bardinate of Chemistrate	in the literal			12 States	and the second s	
Dissection of Annalisation (Speed)	interactions.		and the second se	1.00044	Address of the owner of the owner of the owner of the owner owner owner owner owner owner owner owner owner own	
this streets.	in the lines			1000		
Adda and a standay	in the state of			122		

References

Kosara, R., 2011, January. Indirect multi-touch interaction for brushing in parallel coordinates. In Visualization and Data Analysis 2011 (Vol. 7868, p. 786809). International Society for Optics and Photonics.

Matkovic, K., Freiler, W., Gracanin, D. and Hauser, H., 2008, July. ComVis: A coordinated multiple views system for prototyping new visualization technology. In *Information Visualisation, 2008. IV'08. 12th International Conference* (pp. 215-220). IEEE.

Javed, W. and Elmqvist, N., 2012, February. Exploring the design space of composite visualization. In *Visualization Symposium (PacificVis), 2012 IEEE Pacific* (pp. 1-8). IEEE.

N. Elmqvist, T.-N. Do, H. Goodell, N. Henry, and J.-D. Fekete. ZAME: Interactive large-scale graph visualization. In Proceedings of the IEEE Pacific Symposium on Visualization, pages 215–222, 2008.

Kosara, R., Bendix, F. and Hauser, H., 2006. Parallel sets: Interactive exploration and visual analysis of categorical data. IEEE transactions on visualization and computer graphics, 12(4), pp.558-568.

Yuan, X., Guo, P., Xiao, H., Zhou, H. and Qu, H., 2009. Scattering points in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, *15*(6), pp.1001-1008.

http://caleydo.org/