

Dr. Arpit Narechania

Assistant Professor

arpit@ust.hk





Computer Science Ph.D.



Dr. Arpit Narechania

Assistant Professor arpit@ust.hk



Co-founder, Founding Engineer of **4 FinTech, EdTech Startups**

Worked with the **Government of India**



Computer Science Ph.D.





BS (B.Tech.) Mechanical Engineering Co-founder, Founding Engineer of **4 FinTech, EdTech Startups**

Worked with the **Government of India**





Dr. Arpit Narechania

Assistant Professor

arpit@ust.hk







Data Visualization & Visual Analytics



Human-Computer Interaction (HCI)



Artificial Intelligence (AI)

Decision making is inevitable today!



Data is ubiquitous

(to aid decision making)



Data can be hard to analyze

due to its volume, velocity, variety, veracity, value, ...



Data Visualizations can help

to amplify cognition.

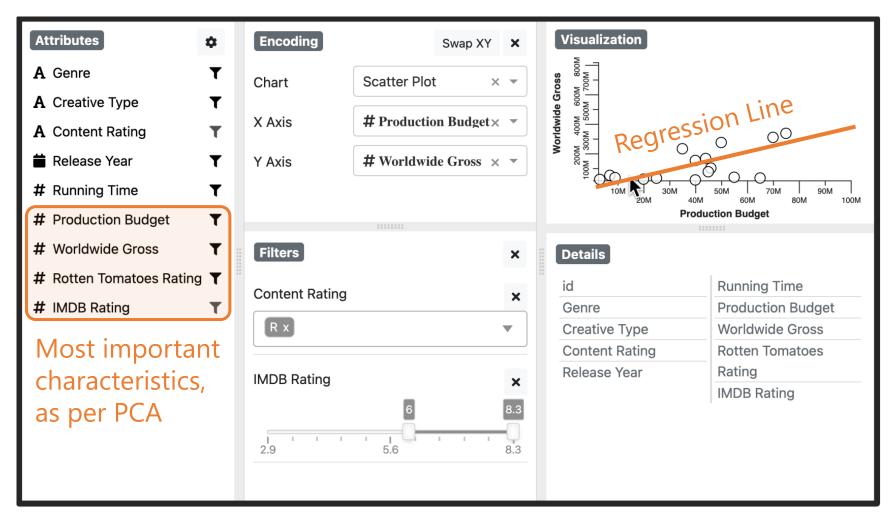


Visual Data Analysis

Attributes		Encoding	Swap XY 🗙		Visualization	
A Genre 💦 🕇	c	Chart	•			
A Creative Type						
A Content Rating	Х	X Axis	•			
🛱 Release Year 🛛 🕇	Y	Y Axis	▼]			
# Running Time 🛛 🕇						
# Production Budget 🛛 🕇						
# Worldwide Gross 🛛 🕇		Filters	×		Details	
# Rotten Tomatoes Rating $oldsymbol{ au}$				11	id	Running Time
# IMDB Rating T					Genre	Production Budget
					Creative Type	Worldwide Gross
					Content Rating	Rotten Tomatoes
					Release Year	Rating
						IMDB Rating

Visual Analytics can help scale analysis

"by combining automated analysis techniques with interactive visualization ...



We need both humans and AI

"A computer can never be held accountable. Therefore, a computer must never make a management decision" – an IBM Slide (1979)



HCI can facilitate this collaboration

"A field that studies how people interact with computers..." – Card, Moran, Newell (1983)



...but there's a problem – "Knowledge Gap"

Humans and AI systems do not know each other's capabilities and goals.



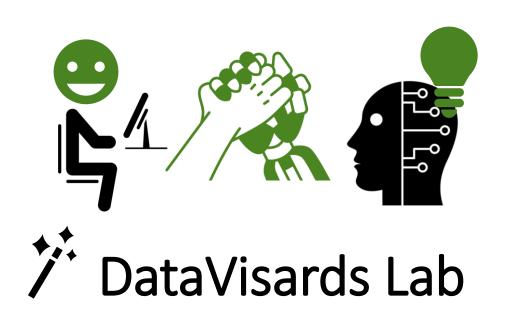
Guidance can help optimize this collaboration

"A computer-assisted process that aims to actively resolve a knowledge gap...." – Ceneda et al. (2016)



...by providing the right **kind** of assistance at the right **time** in the right **place**.

Achieving Desirable Human AI Collaboration!



Guidance for Visual Analytics (and Human AI Collaboration)





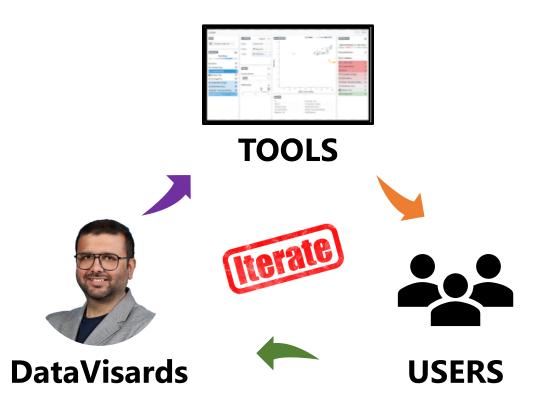
Develop



Democratize

Research Methodology

"We **build** tools, **evaluate** with users, and **refine** based on feedback."



Recall Visual Data Analysis. What could go wrong?

A Genre **A** Creative Type

Attributes

A Content Rating

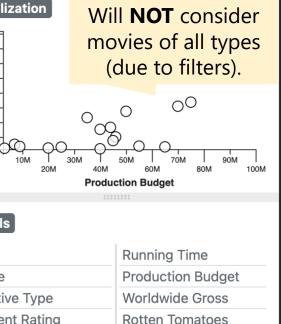
苗 Release Year **#** Running Time **#** Production Budget

Worldwide Gross

Rotten Tomatoes Rating **T** # IMDB Rating

Did **NOT** consider all movie features (unintentionally).

\$	Encoding	Swap X	YX	Visualization
T T	Chart	Scatter Plot	× •	e Gross
Т Т	X Axis	# Production Budge	t× •	400M
T	Y Axis	# Worldwide Gross	× •	
T T				10M T 30 20M
T	Filters		×	Details
T	Content Rating		×	id
T	R x		•	Genre Creative Type
			·	Content Rating
r	IMDB Rating		×	Release Year
	2.9	6 1 I 5.6	8.3 	



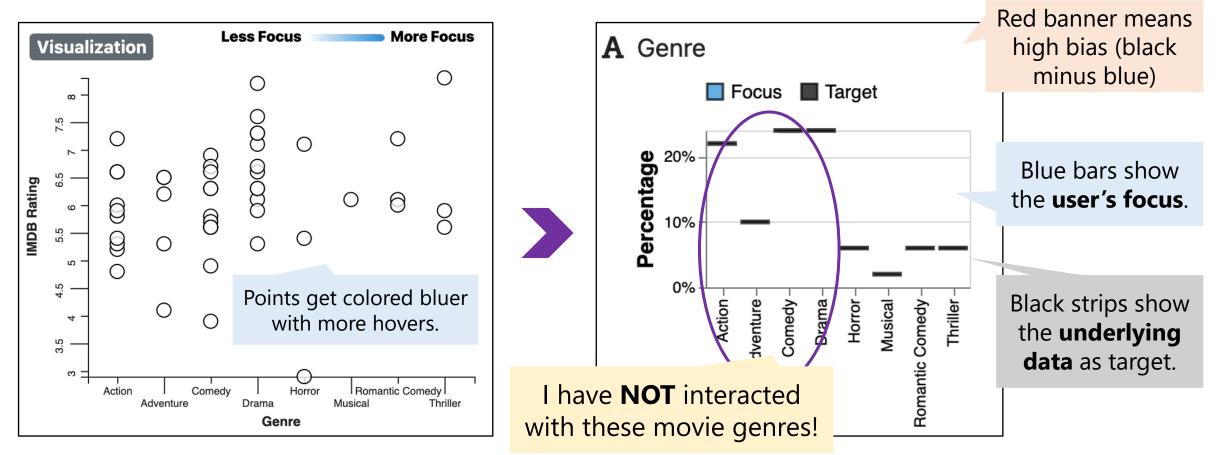
Rating

IMDB Rating

How do we increase awareness of users' analytic behavior during visual data analysis?

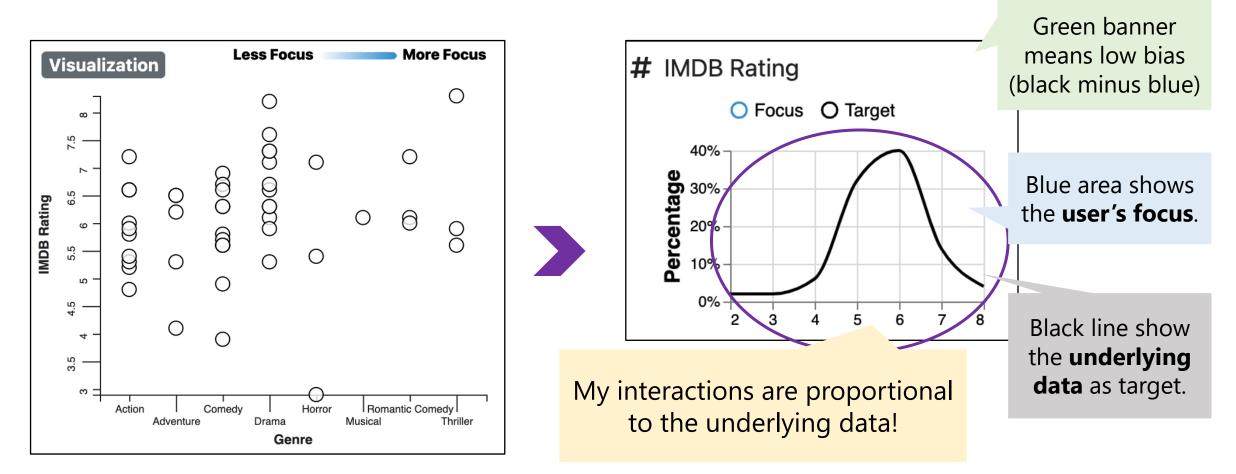


Visual feedback of the user's analytic behavior in the UI





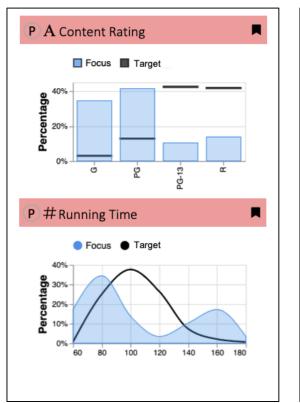
Visual feedback of the user's analytic behavior in the UI

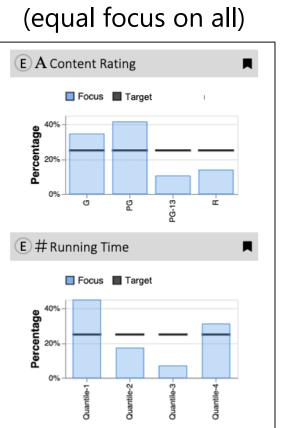


Configuring Custom Baselines / Targets (¢)

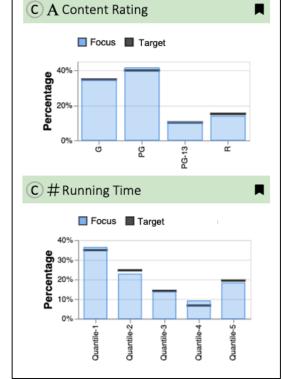
Equal

Proportional (per data distribution)

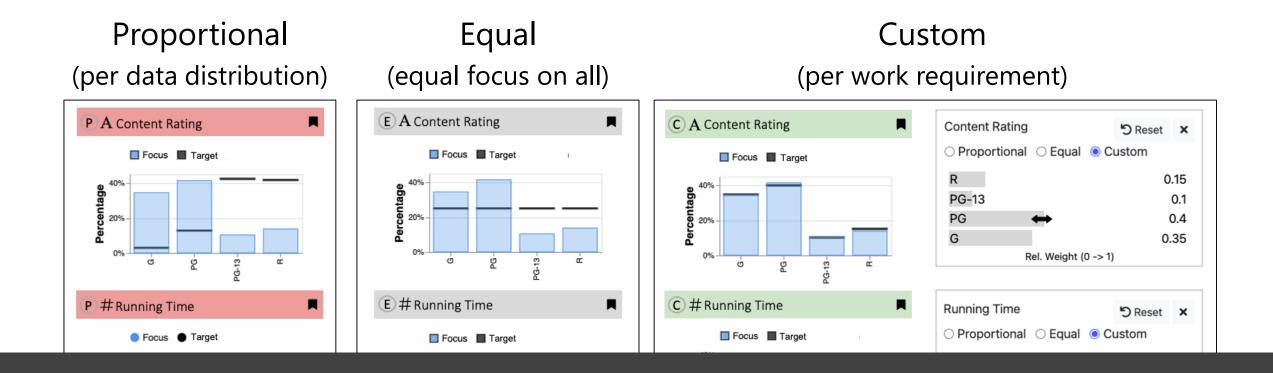




Custom (per work requirement)



Configuring Custom Baselines / Targets



In this way, user and system guide each other towards the goal. [Co-adaptive guidance process – Sperrle et al. (2021)]

- Lumos Increasing Awareness of Analytic Behavior during Visual Data Analysis



Arpit Narechania

Adam Coscia

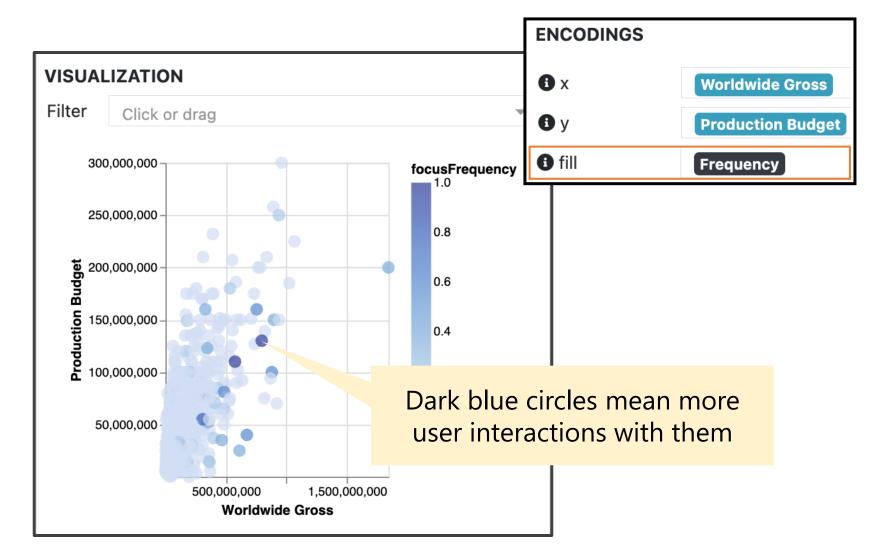
Emily Wall

Alex Endert

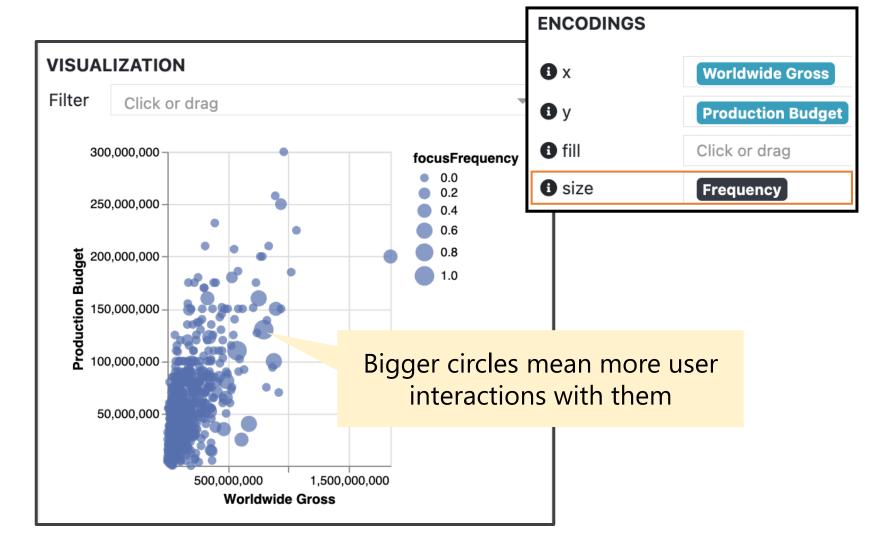


Why only Color?

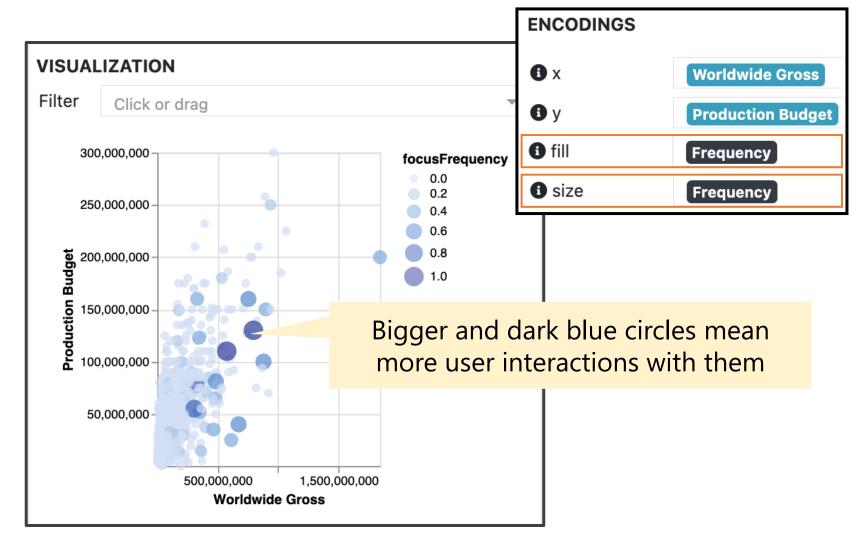
Mapping my focus to "fill" (color)



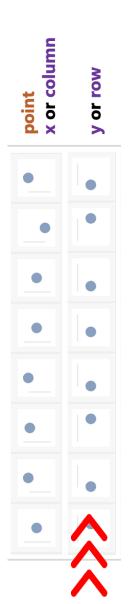
If not "fill" (color), how about "SiZe"?

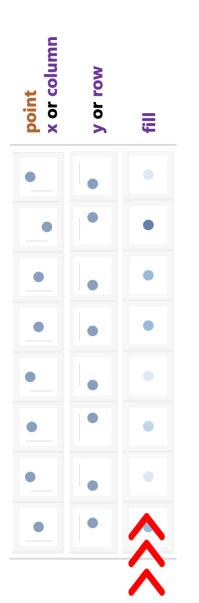


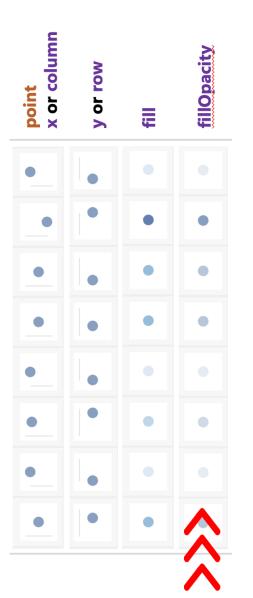
How about both "fill" (color) and "SiZe"?

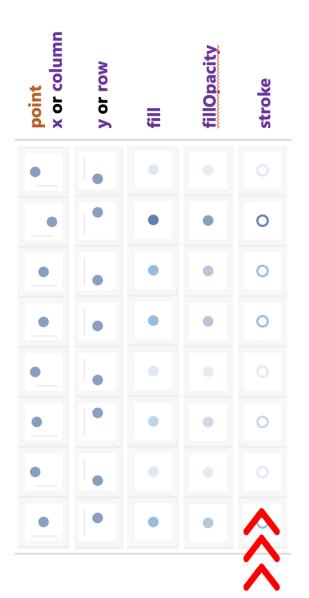


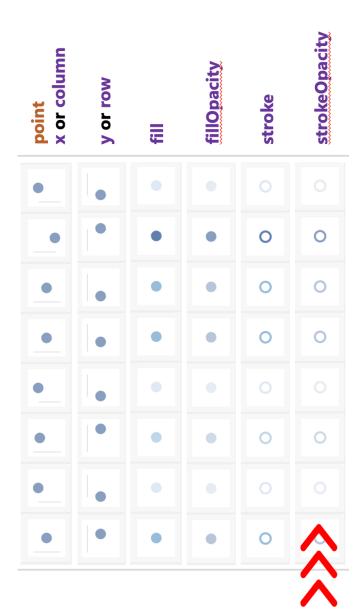




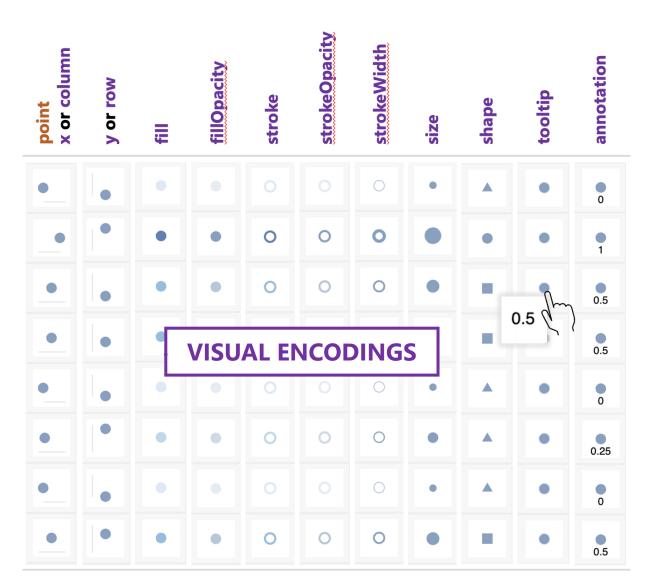




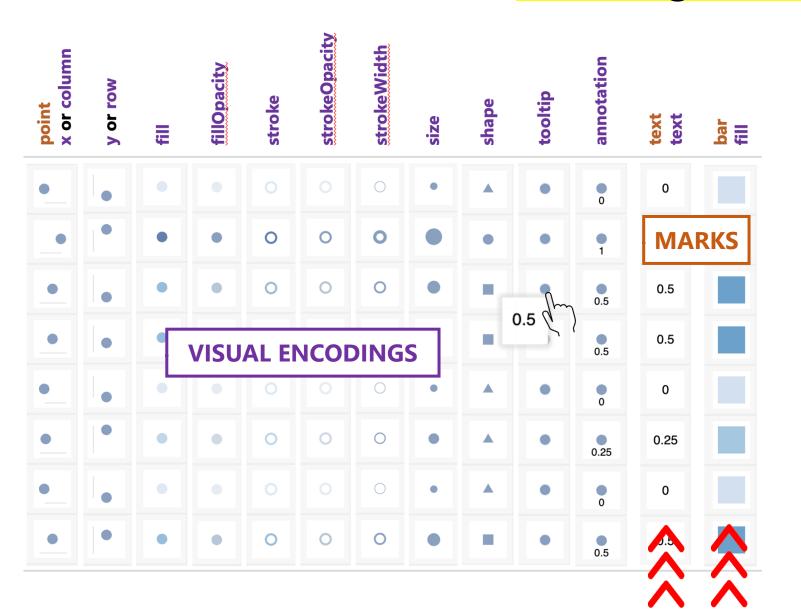




How about ANY encoding?



How about ANY mark (e.g., text, bar)?



	Hc	W	abc	but	Da	ita	Tra	nsf	orn	nat	ions	s <mark>(e.</mark>	g.,	Soi	rt, F	Filte	<mark>er)</mark>
point x or column	y or row	fill	fillOpacity	stroke	strokeOpacity	strokeWidth	size	shape	tooltip	annotation	text text	bar fill		point x, sort (desc)	x, filter (val>0.5)	-x (reverse), sort	x, y, fill, size, sor (combination)
•	•					0	•		•	0	0					•	
•		•	•	0	0	0		•	•	•	MA	RKS		•	•	•	
•	•	•	•	0	0	0	•		- Am	0.5	0.5			•	•	•	
•	•	-	VISU	AL EI	νςοι	DING	s		0.5 2	0.5	0.5			•	•	•	
•	•	•		0	0	0	•		•	0	0			•		•	
•		•	•	0	0	0	٠		٠	0.25	0.25		DA		RANS	FOR	MATIONS
•	•					0	•		•	0	0			•		_	
•		•	•	0	0	0	•	•	•	0.5	0.5			•		_	-

Design Space for Communicating Provenance

point x or column	y or row	fill	fillOpacity	stroke	strokeOpacity	strokeWidth	size	shape	tooltip	annotation	text text	bar fill		point x, sort (desc)	x, filter (val>0.5)	-x (reverse), sort	x, y, fill, size, soı (combination)	7
•	•					0	•		•	0	0				•	•		
•		•	•	0	0	0	•	•	•	•	MA	RKS		•	•	•		
•	•	•	•	0	0	0	٠		- Am	0.5	0.5			•	•	•		
•	•	-	VISU	AL EI		DING	s		0.5 C	0.5	0.5			•	•	•		
•	•	•		0	0	0	•		•	0	0			•		•		
•		•	•	0	0	0	•		•	0.25	0.25		DAT	A TF	RANS	FOR	ΜΑΤΙΟ	ONS
•	•					0	•		•	0	0			•		-	-	
•		•	•	0	0	0	•		•	0.5	0.5			•				

Adobe X ProvenanceLens

Utilizing Provenance as an Attribute during Visual Data Analysis Promotes Self-Reflection

• 09:03:02 PM, 29 Aug 24 (GMT-4)



濸 ProvenanceLens 💿 Main Task Prev End Task 🗴 0h 2m 49s (0h 2m 49s) DATA ATTRIBUTES MARK point ۰. X Ŧ VISUALIZATION ۵. Task 1 Task 2 Task 3 Task Filter Click or drag Sort Click or drag → 11 ENCODINGS Swap xy X Click here to review Task 4 Filter Click or drag Ŧ 300,000,000 focusFrequency **6** x Worldwide Gross X Ŧ Agg 🔻 Note: You are in CREATE mode. Your O v own analysis is being actively roduction Budget Agg 🔻 ✓ ∧ Expand/Collapse All 250,000,000 recorded and can be visualized and 6 fill Frequency X Ŧ Agg 🔻 ∨ 0 (A) id interacted with. Your mom's analysis a 200,000,000 0.6 fillOpacity Click or drag -Agg 🔻 history has been reset. V 🛈 (A) Title stroke Click or drag 0.4 -Agg 🔻 5 150,000,000 -SELECTED RECORDS × ✓ Ⅰ (#) Worldwide Gross strokeWidth Click or drag Agg 🔻 0.2 å 100,000,000-(0/10)strokeOpacity Click or drag -Agg 🔻 ✓ ● (#) Production Budget 0.0 id Title Worldwide Gross Proc 50,000,000 size Click or drag Agg 🔻 ✓ ● (#) Release Year Click or drag 6 shape -Agg 🔻 Hover on each table row (or table (A) Content Rating 200.000.000 1 000 000 000 1.800.000.000 header) to see each record (or all row Click or drag Agg 🔻 $\overline{\mathbf{w}}$ Worldwide Gross records) highlighted in the Column Click or drag Agg 🔻 ✓ ① (#) Running Time visualization canvas tooltip Title × × Agg 🔻 V 0 (A) Genre text Click or drag Agg 🔻 * ✓ ● (#) Rotten Tomatoes Rating DATA RECORDS annotation Click or drag 🔻 Agg 🔻 ✓ Ⅰ (#) IMDB Rating . Sort Click or drag Worldwide Gross Focus Frequency Production Budget PROVENANCE ATTRIBUTES 25728961 20000000 0 (#) Frequency 148345997 65000000 0 20278055 40000000 0 (#) Recency 38623460 10000000 0 Frequency: 0 = not interacted at all, 1 = most 51204567 7000000 0 frequently interacted I< < 1 2 3 4 5 > ►I 709 total Recency: 0 = not interacted at all, 1 = most recently interacted

Arpit Narechania

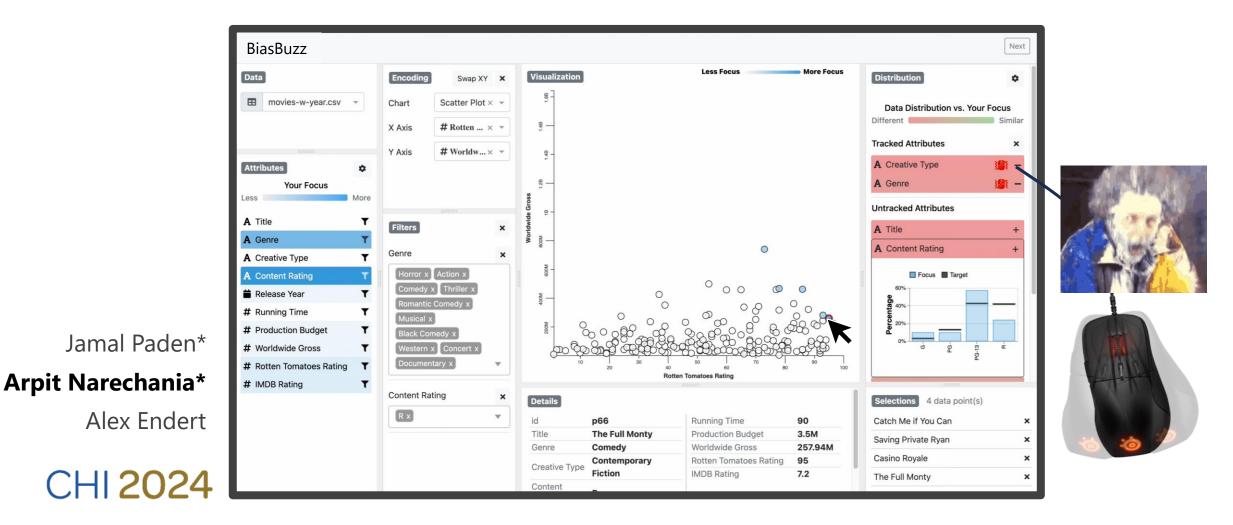
Shunan Guo Eunyee Koh Alex Endert Jane Hoffswell

IEEE TVCG 2025

How about Multimodal Guidance (why only Visual)?



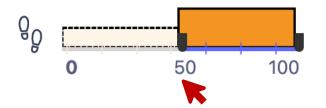
Combining Visual Guidance and Haptic Feedback to Increase Awareness of Analytic Behavior during Visual Data Analysis

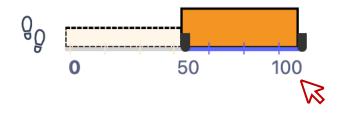


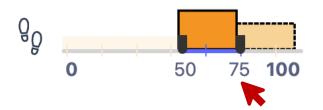
*equal contribution

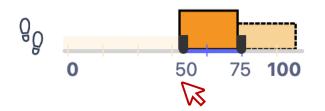
How about guidance through other UI elements (not visualizations)?

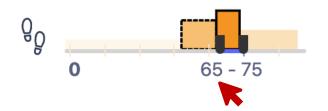


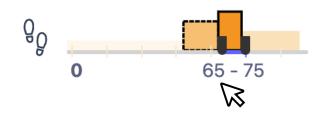


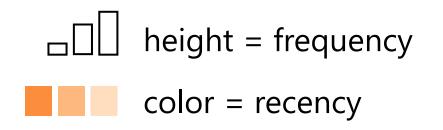


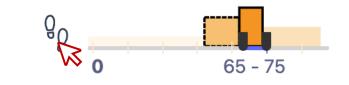


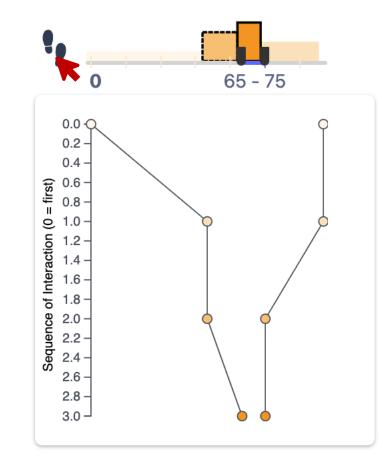


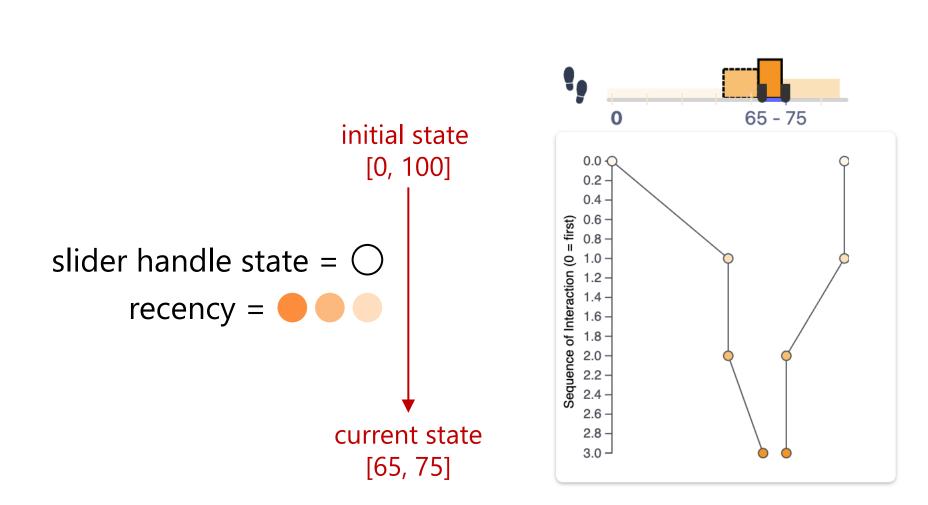




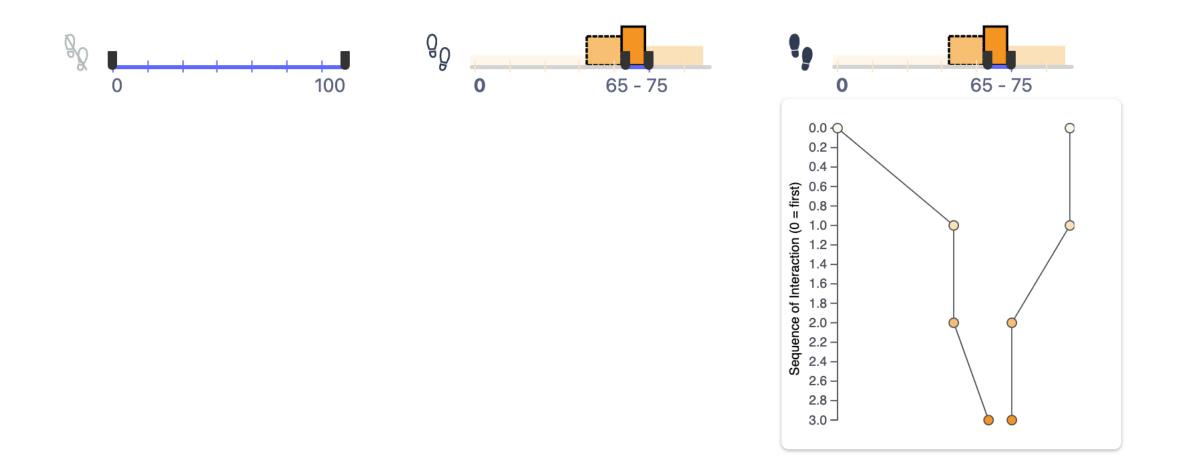




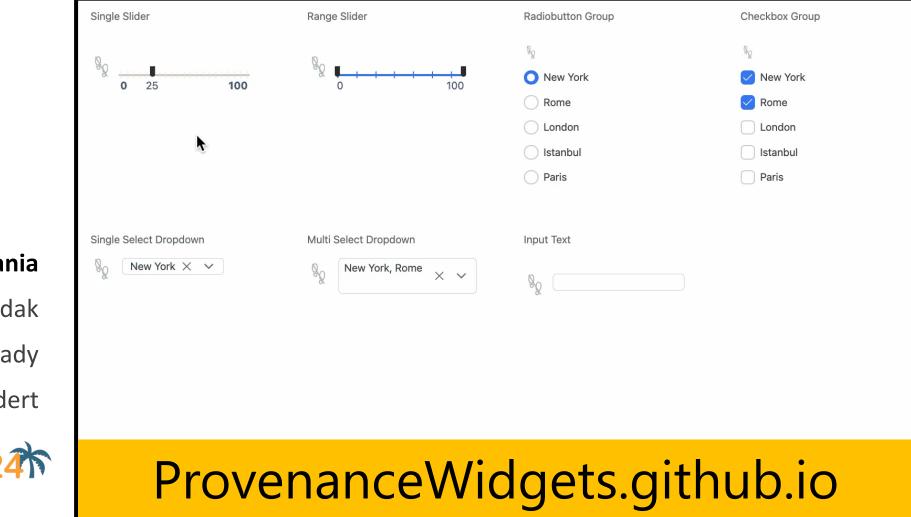




Such capabilities can also help increase user awareness during analysis.



ProvenanceWidgets A Library of UI Control Elements to Track and Dynamically Overlay Analytic Provenance



Arpit Narechania

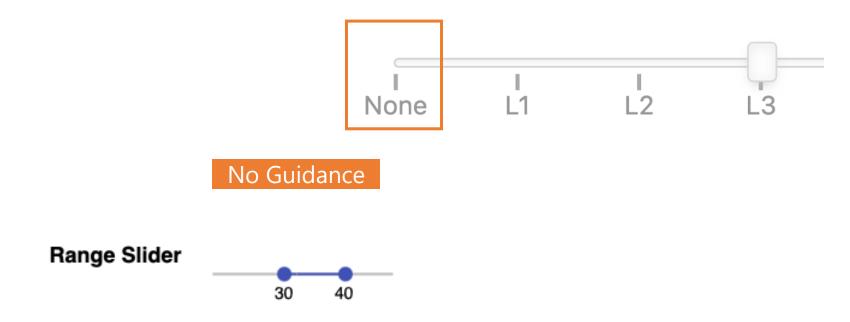
Kaustubh Odak

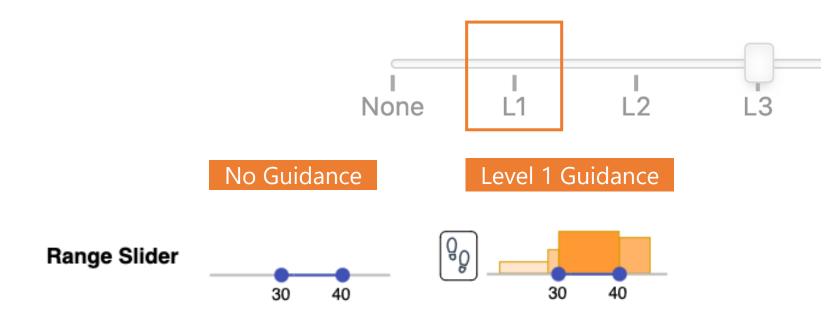
Mennatallah El-Assady

Alex Endert

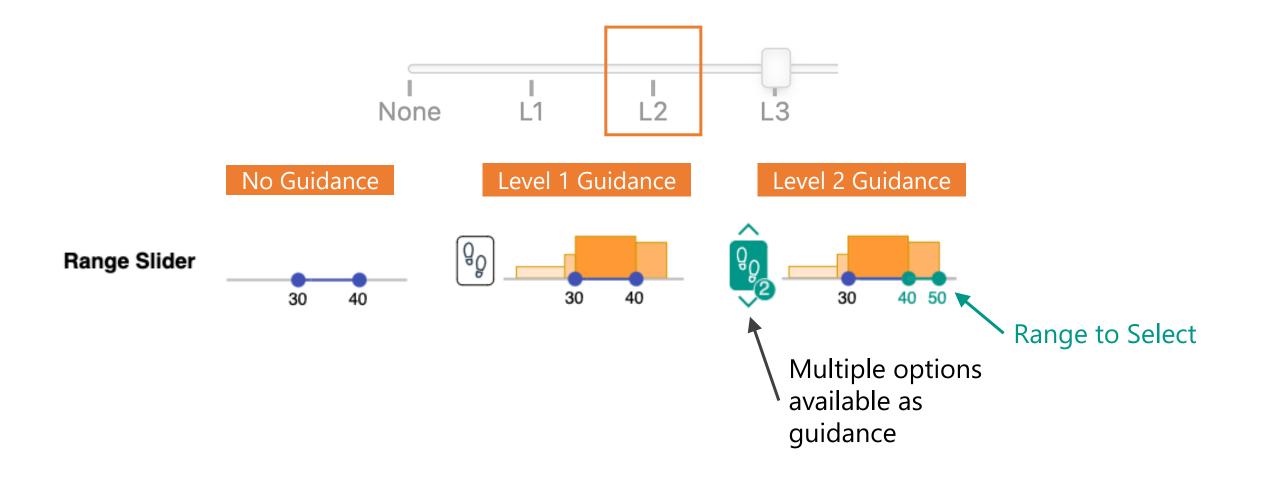


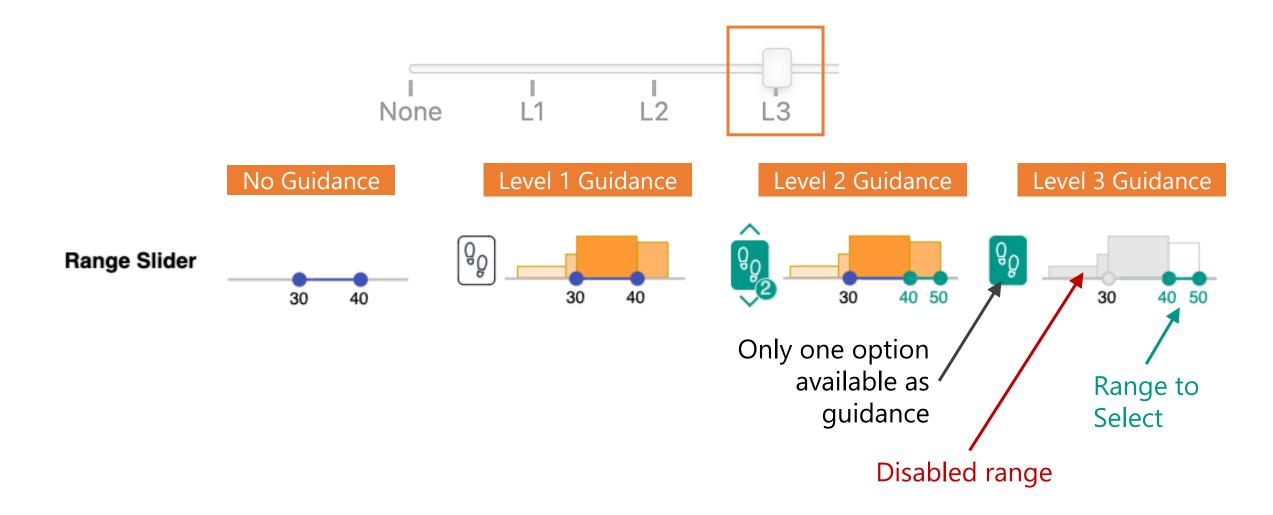
How about different amounts/levels of Guidance?



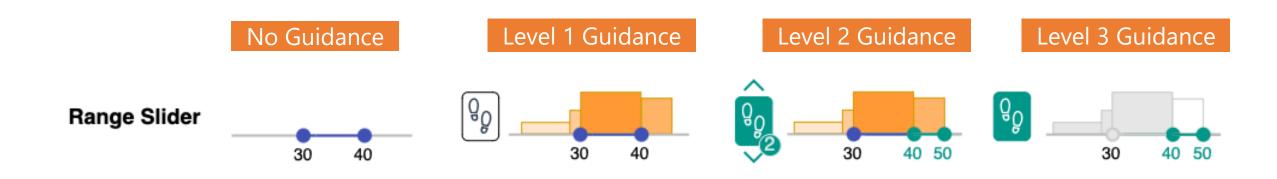


ProvenanceWidgets





GuidanceWidgets [coming soon]





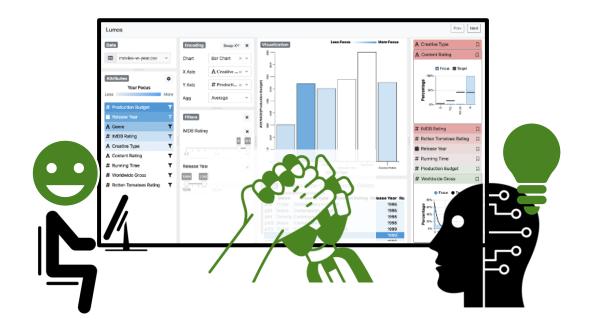
Should AI systems have "extreme" agency and control?





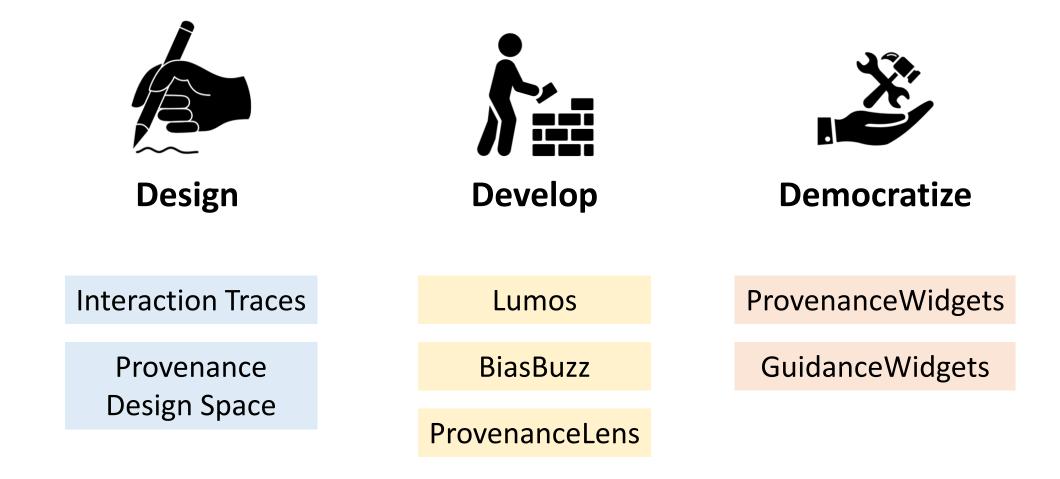


Hence, Catalyzing Human-AI Collaboration



"Building tools and fostering environments that *actively enhance* the way they work together."

Guidance for Visual Analytics (and Human AI Collaboration)



Past and Ongoing Industry Collaborations



Minimize Risks to Make Cars Safer



Improve Data Lake Navigation

Provide Adaptive Guidance

Analyze PDF Documents



Debug Natural Language to SQL Tasks



Storytelling of eCommerce Data

Non-Profit Mapping Orgs

Educate How Maps can Lie



Past and Ongoing Industry Collaborations



Minimize Risks to Make Cars Safer



Improve Data Lake Navigation

Provide Adaptive Guidance

Analyze PDF Documents



Debug Natural Language to SQL Tasks

EAlibaba.com

Storytelling of eCommerce Data

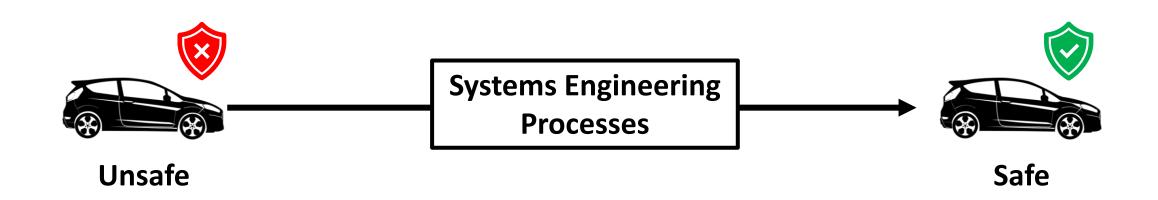
Non-Profit Mapping Orgs

Educate How Maps can Lie

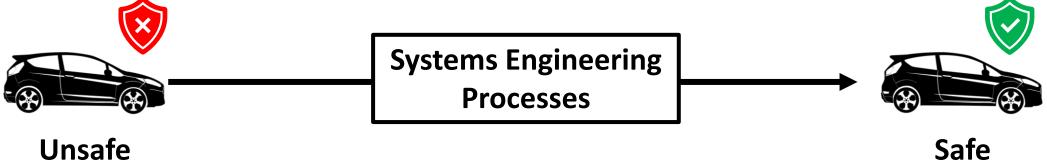




Minimize Risks to Make Cars Safer

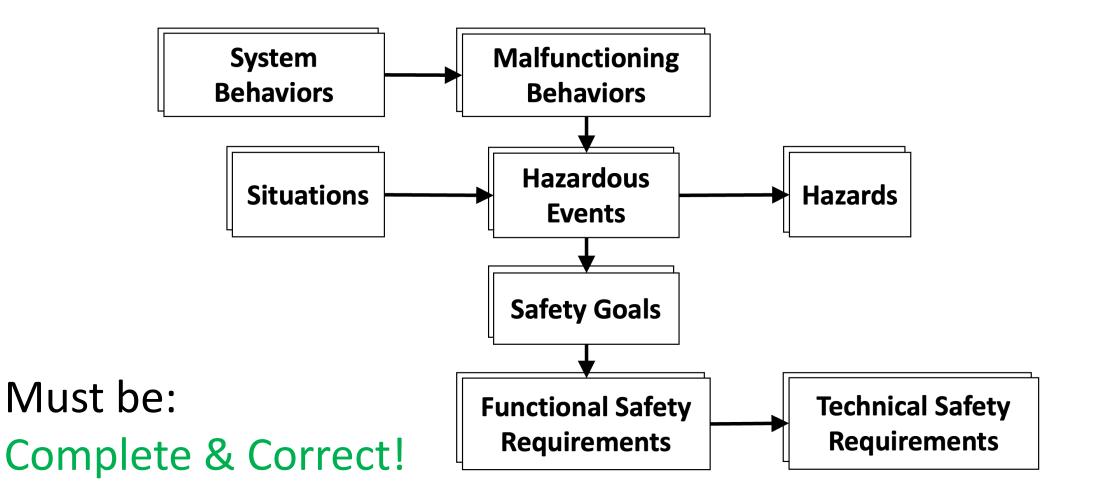






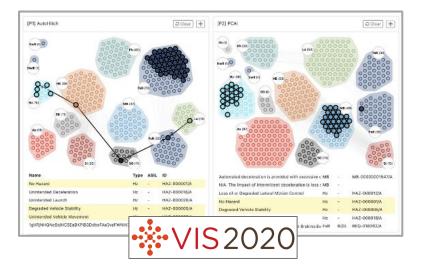
1. Functional Safety analysis helps minimize unreasonable risks associated with hazards.

Functional Safety Analysis (e.g., of Adaptive Cruise Control)





Design Study w/ 10+ Safety Engineers, Analysts

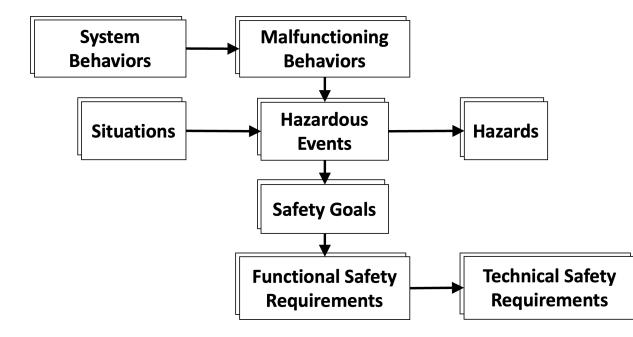


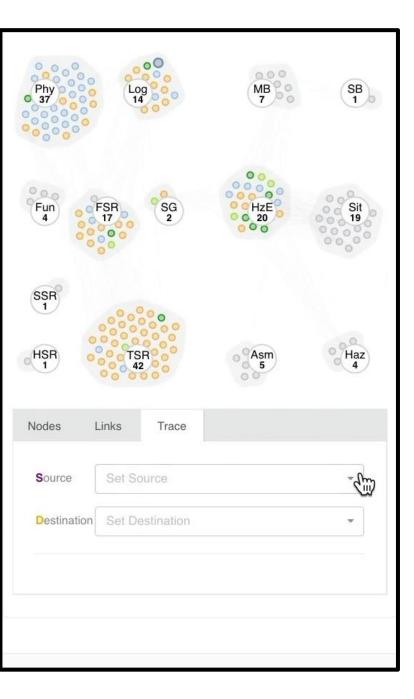


Visual Data Analysis of Functional Safety of Vehicles

A. Narechania, A. Qamar, A. Endert

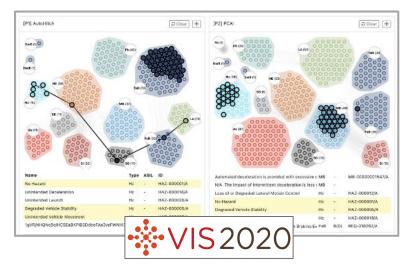
Functional Safety







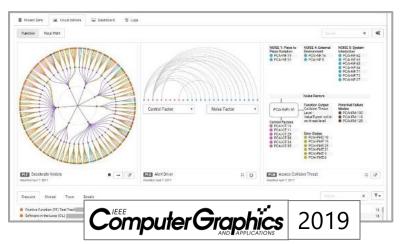
Design Study w/ 10+ Safety Engineers, Analysts





Visual Data Analysis of Functional Safety of Vehicles

A. Narechania, A. Qamar, A. Endert





Understanding **FMEA Data** Using Interactive Visual Analytics R. Basole, A. Qamar, B. Pal, M. Corral, M. Meinhart, **A. Narechania**

Past and Ongoing Industry Collaborations



Minimize Risks to Make Cars Safer



Improve Data Lake Navigation

Provide Adaptive Guidance

Analyze PDF Documents



Debug Natural Language to SQL Tasks

EAlibaba.com

Storytelling of eCommerce Data

Non-Profit Mapping Orgs

Educate How Maps can Lie



-

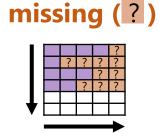
Adobe Design Study w/ 14 data engineers & scientists

"Is the data complete? correct? unbiased? ... showing data quality insights may instil feelings of curiosity and care in the user."

"There are several low-quality datasets just lying around. Knowing when, where, and by whom the datasets were last used, i.e., **data usage insights**, can help."

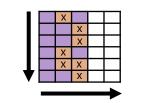
We modeled data quality and usage insights

1. quality



completeness

% frequency of nonmissing values in the data incorrect (x)



correctness

% frequency of correct values in the data



	M		
	Μ		
	Μ		
	Μ		
	F		
/	F		

objectivity

% extent that values conform to a target distribution



2. usage

in-subsets

% of users that select the attribute, record in their subset

in-filters

% of users that applied a filter on the attribute

in-visualizations

% of users that visualized the attribute

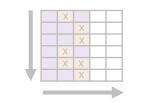
We modeled data quality and usage insights

1. quality



completeness

% frequency of nonmissing values in the data incorrect (X)



correctness

% frequency of correct values in the data

skewed (M ale, F emale)

M		
M		
M		
M		
F		
F		

objectivity

% extent that values conform to a target distribution



2. usage

in-subsets

% of users that select the attribute, record in their subset



% of users that applied a filter on the attribute



% of users that visualized the attribute

Adobe DataPilot Utilizing quality and usage information for subset selection during visual data preparation

● High (≥90/100) ● Medium (≥67/100) ● Low ③ Prev DataPilot Prepare an effective subset to determine meaningful drivers of the company's dollar (\$) sales revenue. Also create at least 3 widgets. Next Step 1: Review Raw Data Step 2: Review Selected Subset Step 3: Create Dashboard **Arpit Narechania** њ 🔳 Attribute Filters **Quality Filters** Data Sort Values Default ✓ 1 ↓ F Minimap Attributes (41) C C Attribute-Leve A 🌒 id 🏦 A 🌗 sale Sort Default ✓ 1 ↓ A 🌒 timestamp 🏦 Attributes: 22(54%) drop a field here Records: 374(37%) ▼ Completeness [0.1,100] Cardinality: 1000 Cardinality: 318 Fan Du Search environment.browserdetails.cookiesenal ▼ Correctness [51,100] Missing values: 0 (0%) Missing values: 0 (0%) Missing C 12/10/18 13:36 1.0 374 0 (0%) Missing values ▼ Objectivity [0,100] p1 p2 p3 12/10/18 15:01 Visible (22) Hidden (19) Selected (0) 12/10/18 13:33 \sim true 12/14/18 20:31 112 A 🌒 id c Overall [88,100] Atanu Sinha p4 p5 p6 p7 12/10/18 14:13 19 12/2/18 14:11 A 🔵 timestamp C sales.purchases.price 11/28/18 0:09 A 🌔 sales.purchases.value 12/6/18 22:14 172 (17%) Missing values A 🌒 sales.order.purchaseid p0 12/12/18 0:56 1.0 Ryan Rossi A 🌔 placecontext.geo.city p2 11/28/18 0:09 1.0 Record-Level # 🜒 placecontext.geo, schema.latitude p3 12/6/18 22:14 1.0 ▼ Completeness [36.59,75.61] (44, 147) # 🛑 placecontext.geo, schema.longitu 1.0 p4 12/10/18 14:07 Jane Hoffswell Correctness [0,100] A 🌒 placecontext.geo.postalcode p5 12/10/18 15:46 1.0 p8 1.0 A 🌒 placecontext.geo.countrycode 12/10/18 16:12 c Overall [62,86.59] Overall Quality score of 1.0 p9 12/10/18 17:16 dataset values (rows), # placecontext.geo.dmaid between 0-100. p11 12/10/18 18:57 1.0 A 🌒 environment.browserdetails.javasc Shunan Guo p13 12/10/18 16:59 1.0 # 🛑 er vironment.browserdetails.viewpo Count: 62 p16 12/10/18 17:23 1.0 Α 🌒 environment.browserdetails.userad p19 1.0 Usage Filters 12/15/18 16:38 A 🌒 environment.browserdetails.cookie C Eunyee Koh p22 12/10/18 15:52 1.0 # environment.browserdetails.viewpo Attribute-Leve p23 12/16/18 19:23 1.0 A 🌒 environment.browserdetails.javaen ▼ In Filters [0,59] 1.0 p27 12/10/18 13:57 A environment.ipv4 Shamkant Navathe In Visualizations [0,100] p30 12/11/18 14:53 1.0 A 🌒 sales.product.path p31 12/7/18 7:01 1.0 ▼ In Subsets [0,100] A web.webinteraction.pagetype Overall Usage score of p32 1.0 12/10/18 13:50 A session.web.webreferrer.type dataset attributes Overall [0,100] p33 12/10/18 15:26 1.0 Alex Endert A device.colordepth lumns), between 0-100. cm p35 12/12/18 7:41 1.0 # o sales.purchases.price Record-Level p37 12/10/18 22:14 1.0 p42 12/10/18 14:01 1.0 ▼ Overall [50,100] p43 12/10/18 15:18 1.0 p45 11/9/18 22:12 1.0 p47 1.0 11/19/18 22:36 p48 12/13/18 14:36 1.0 any I Hybrid 1.0 p49 12/10/18 21:46 April 23-28 2023 p50 12/10/18 20:00 1.0 reCHInnecting Showing 1 to 27 of 374 rows << > >> 27 ~

Adobe DataCockpit A Toolkit for Data Lake Navigation & Monitoring Utilizing Quality and Usage Information.

Arpit Narechania Surya Chakraborty Shivam Agarwal Atanu Sinha Ryan Rossi Fan Du Jane Hoffswell Shunan Guo Eunyee Koh Alex Endert Shamkant Navathe **IEEE BigData**

ID	A Name	Prepared By	 Created On 	A	Last Updated On	<u> </u>	Quality 🔺	Usage	
1	Duma	Chris Du	07/15/2018		09/27/2021		•	•)
2	Stalwart	Pamela C. Michels	07/04/2015		05/13/2022		•	•	
3	BestTry	Gavin M. Elliott	07/19/2019		05/18/2020		•	•	
4	Paragon	John S. Page	12/10/2020		01/01/2021		•	•	
5	Clipart	Jeff M. Windsor	07/06/2022				۲	•	
6	Depay	Christa B. Taylor	07/15/2018		09/27/2021		•	•	
7	Aloevera	Pamela C. Michels	07/04/2015		05/13/2022		٠		
8	Shopium	Gavin M. Elliott	07/19/2019		05/18/2020		•	•	
9	Jing	John S. Page	12/10/2020		01/01/2021		٠	•	
10	Ecom101	Jeff M. Windsor	07/06/2022				•	•	
		Q							

Past and Ongoing Industry Collaborations



Minimize Risks to Make Cars Safer



Improve Data Lake Navigation

Provide Adaptive Guidance

Analyze PDF Documents



Debug Natural Language to SQL Tasks

E Alibaba.com

Storytelling of eCommerce Data

Non-Profit Mapping Orgs

Educate How Maps can Lie



Designing, Developing, & Democratizing Guidance for Visual Analytics



Dr. Arpit Narechania arpit@ust.hk Rm 3552

narechania.com





DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING