Designing, Developing, & Democratizing **Guidance for Visual Analytics**

Dr. Arpit Narechania

Assistant Professor

arpit@ust.hk

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Designing, Developing, & Democratizing
Guidance for Visual Analytics

Georgia Tech

Computer Science Ph.D.

Dr. Arpit Narechania

Assistant Professor

arpit@ust.hk

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Designing, Developing, & Democratizing **Guidance for Visual Analytics**

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

Worked with the **Government of India**

>

Georgia Tech

Computer Science Ph.D.

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Designing, Developing, & Democratizing **Guidance for Visual Analytics**

BS (B.Tech.)
Mechanical Engineering

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

Worked with the **Government of India**

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY | DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Designing, Developing, & Democratizing **Guidance for Visual Analytics**

Dr. Arpit Narechania

Assistant Professor

arpit@ust.hk

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DataVisards Lab

Data Visualization & Visual Analytics

Human-Computer Interaction (HCI)

Artificial Intelligence (AI)

Decision making is inevitable today!

Data is ubiquitous

(to aid decision making)

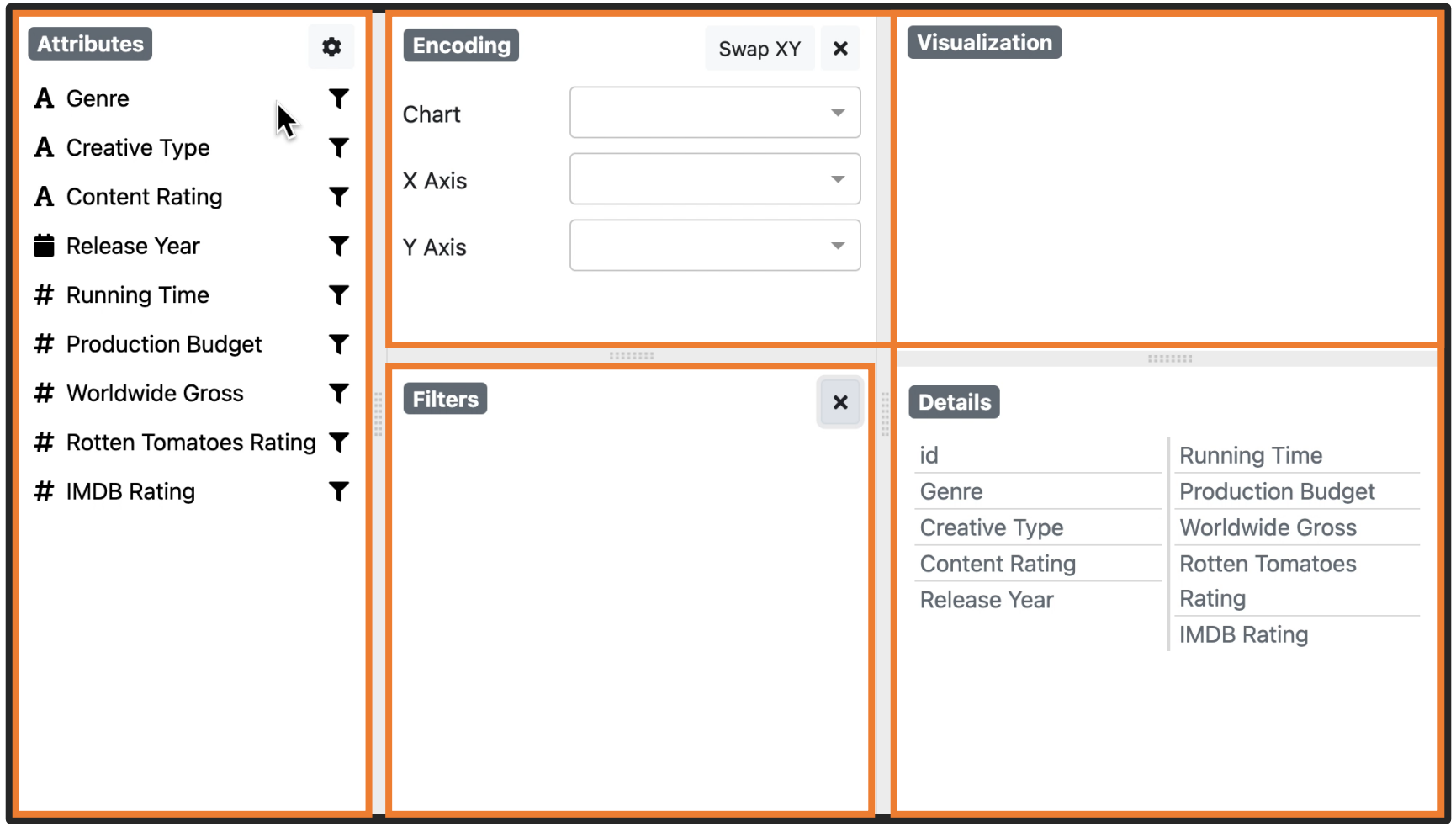
Data can be hard to analyze

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

Data Visualizations can help

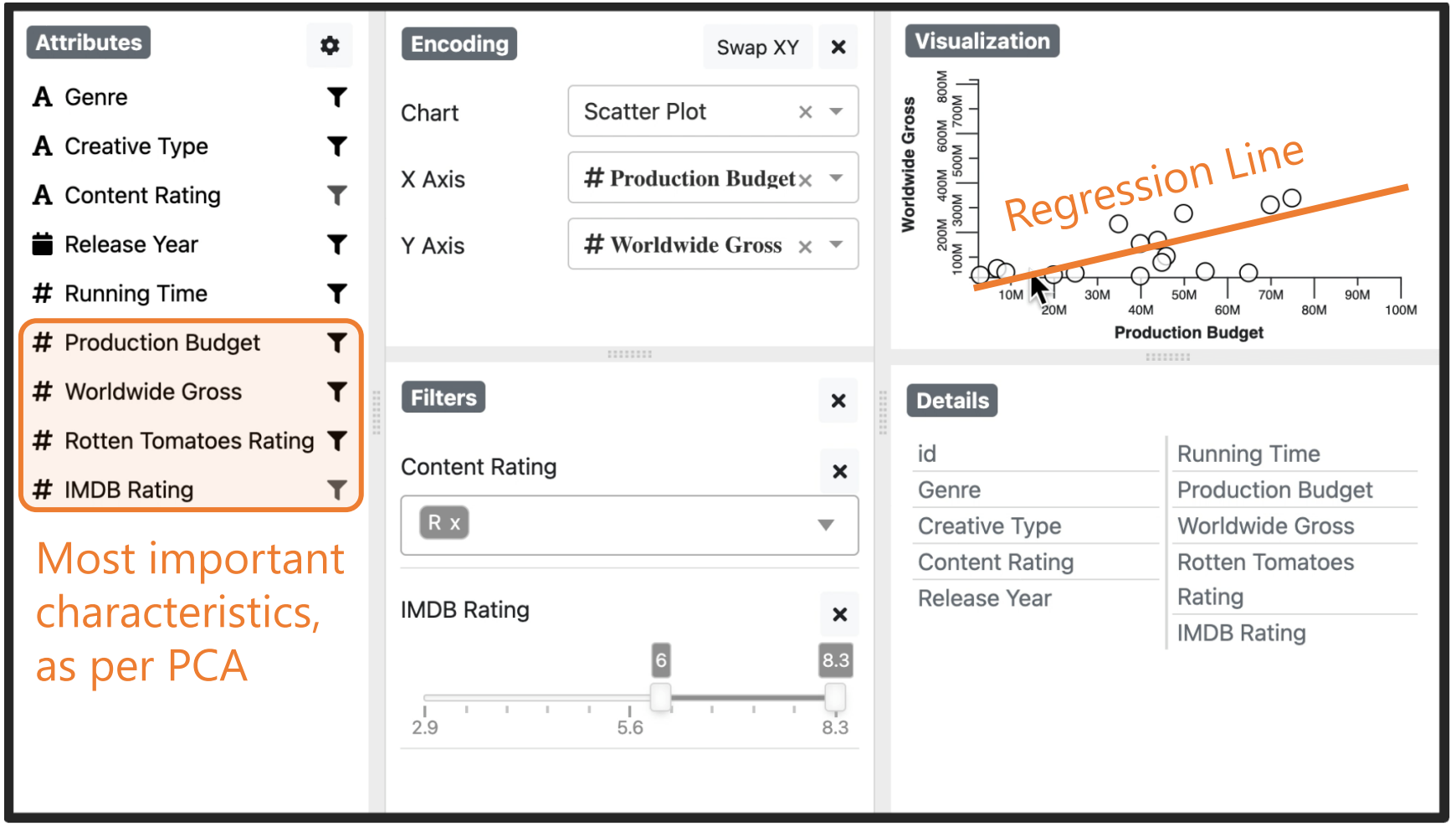
to amplify cognition.

Visual Data Analysis



Visual Analytics can help scale analysis

"by combining automated analysis techniques with interactive visualization …

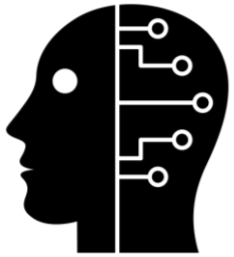


We need both humans and AI

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

Analytic Skills

Perceptual Capabilities

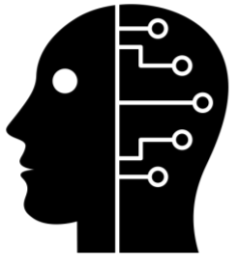


Computational Powers

Working Memory

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)

Knowledge gap

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

Achieving Desirable Human AI Collaboration!

DataVisards Lab

Guidance for Visual Analytics (and Human AI Collaboration)

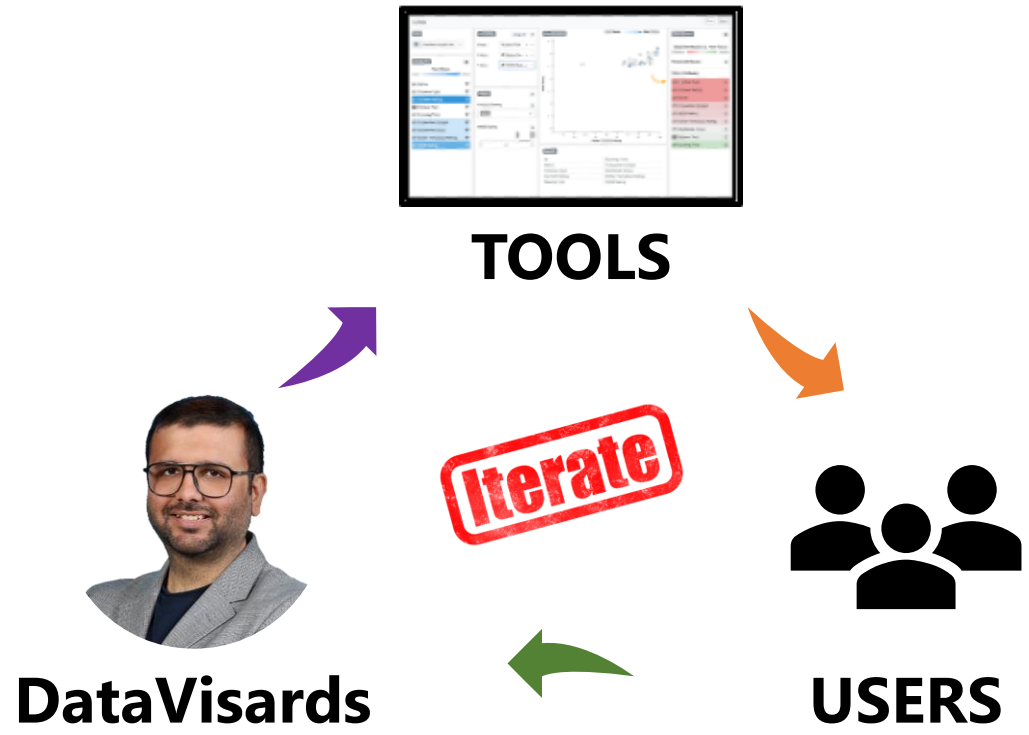
Design

Develop

Democratize

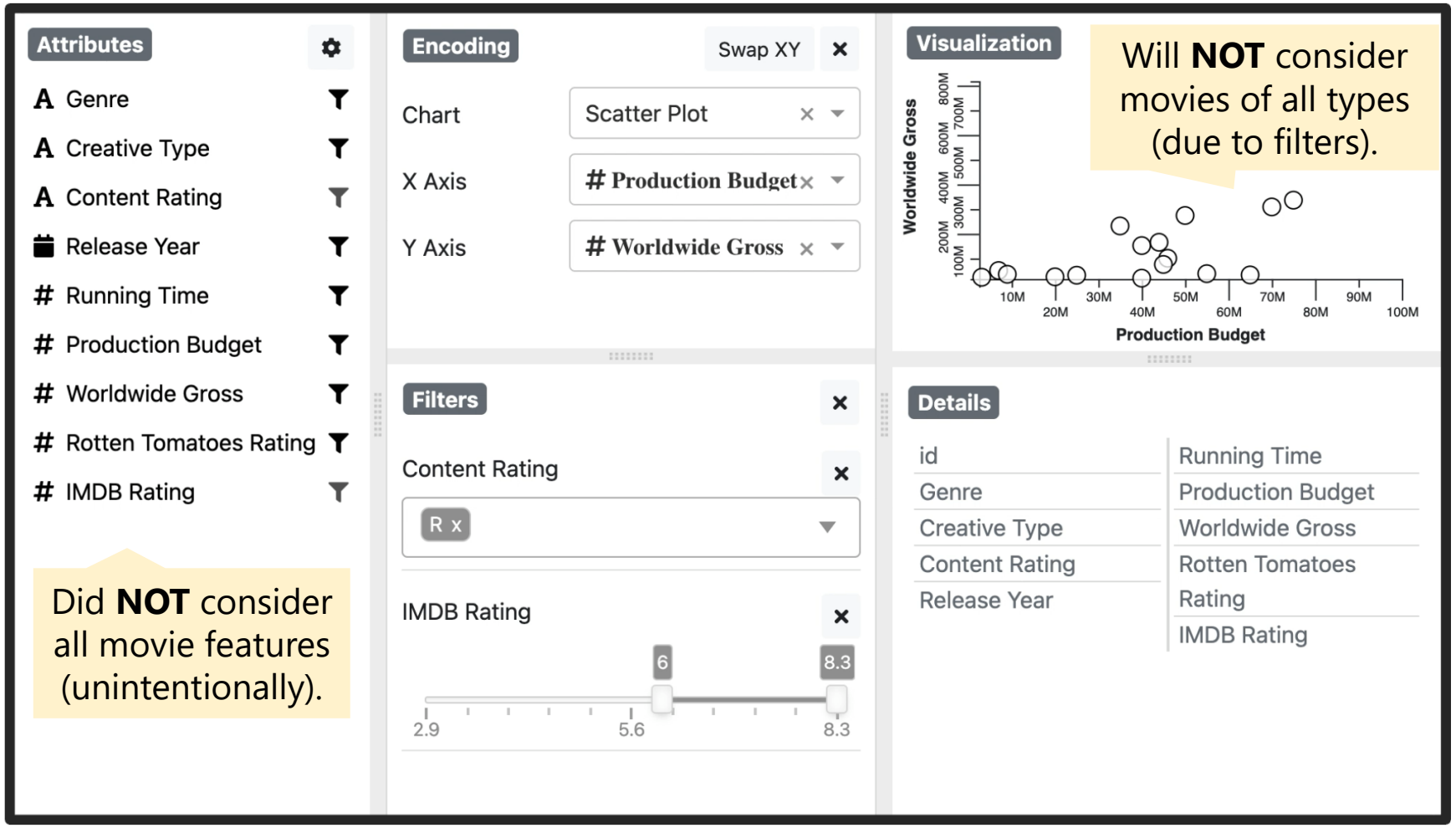
Research Methodology

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



Recall Visual Data Analysis. What could go wrong?

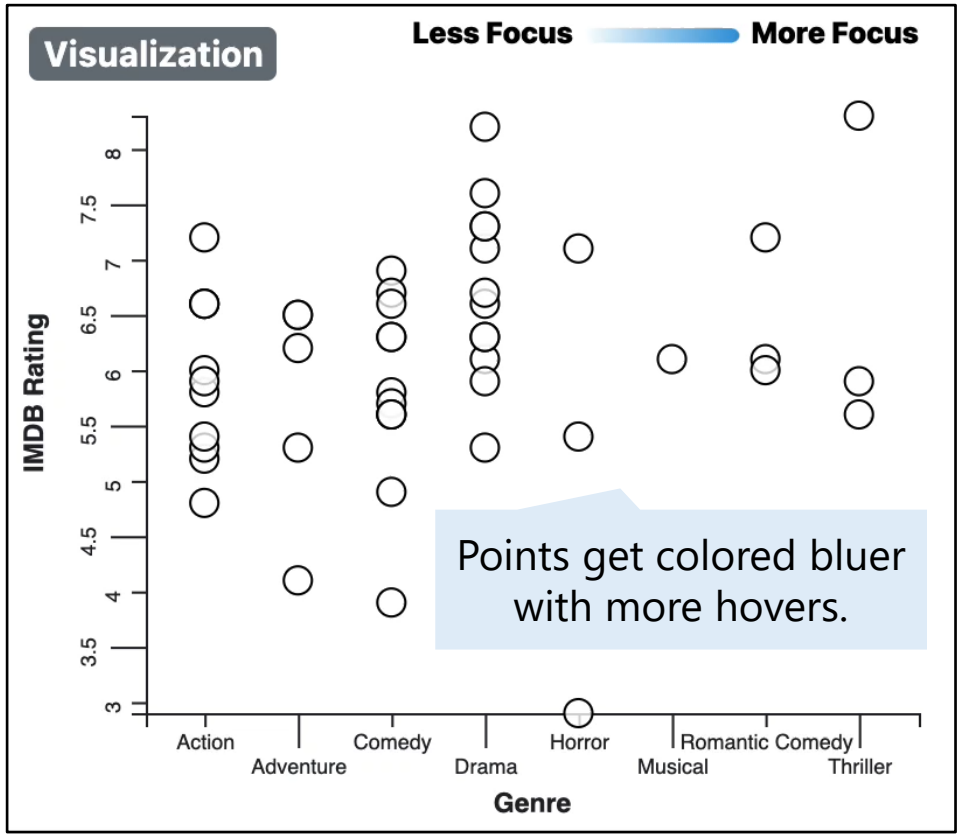
Did **NOT** consider all movie features (unintentionally).

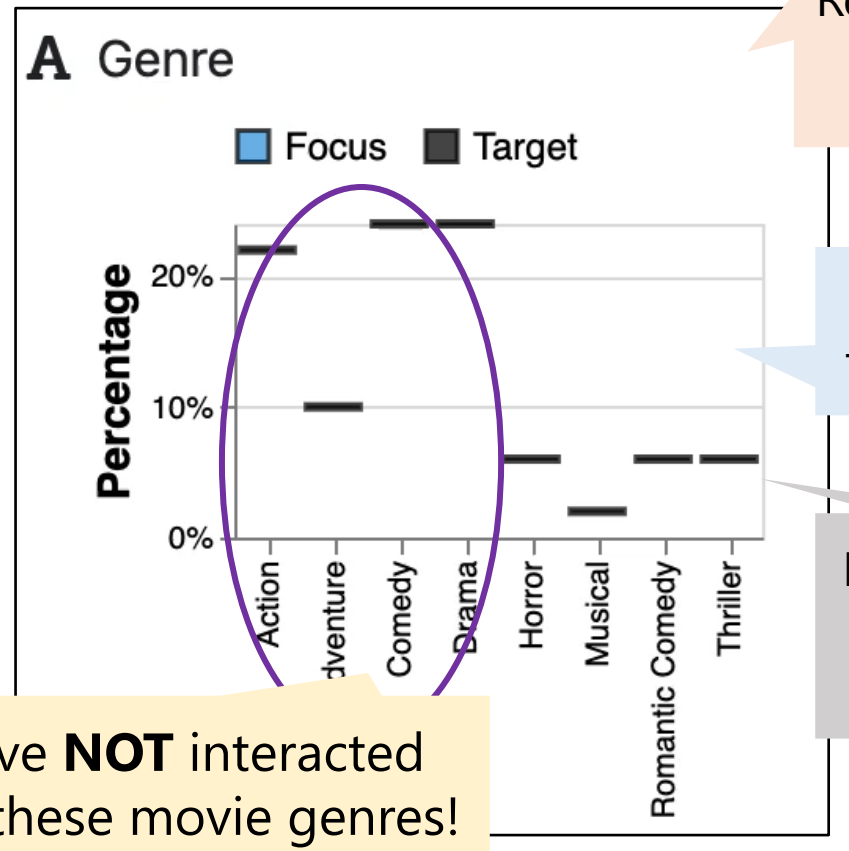


How do we increase awareness of users' analytic behavior during visual data analysis?

Interaction Traces

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





I have **NOT** interacted with these movie genres!

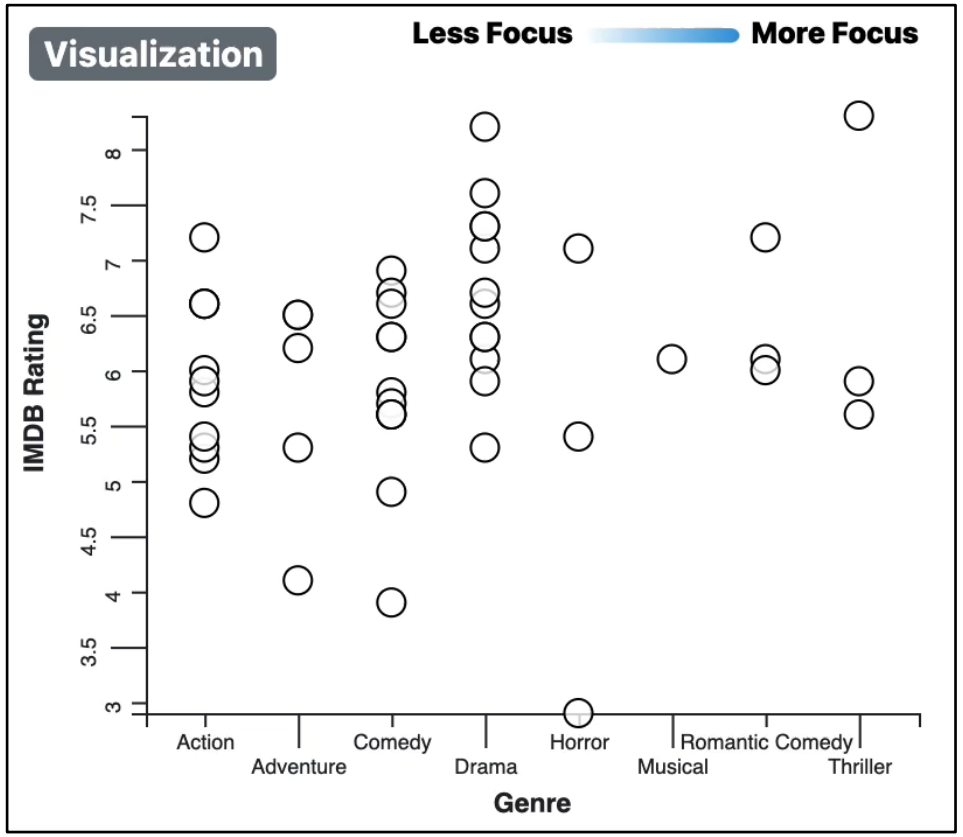
Red banner means high bias (black minus blue)

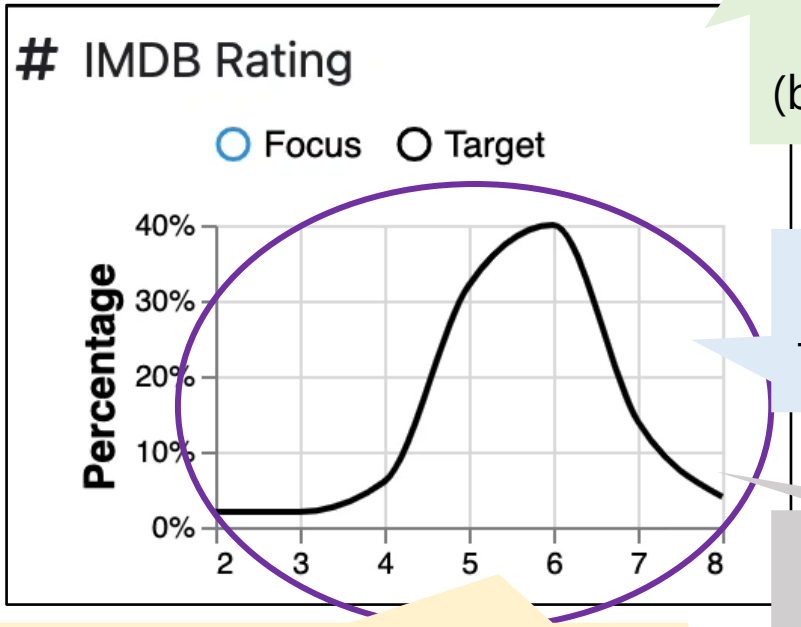
Blue bars show the **user's focus**.

Black strips show the **underlying data** as target.

Interaction Traces

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





Green banner means low bias (black minus blue)

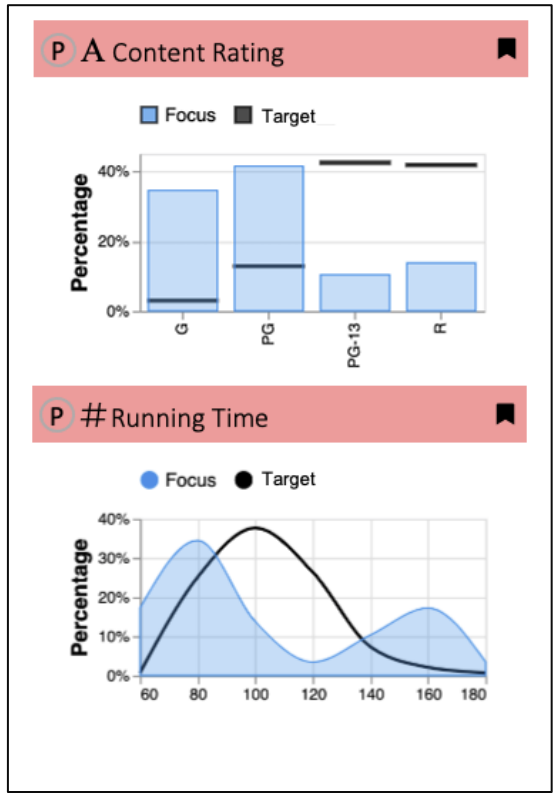
Blue area shows the **user's focus**.

Black line show the **underlying data** as target.

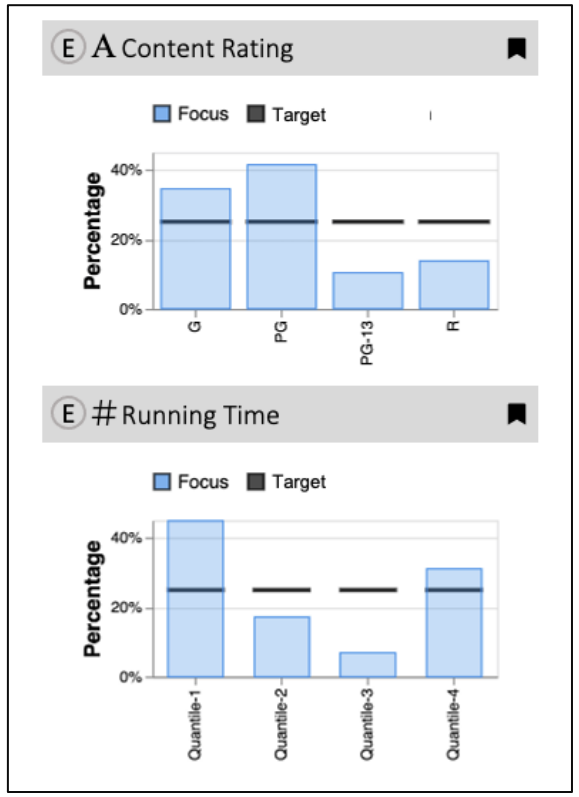
My interactions are proportional to the underlying data!

Configuring Custom Baselines / Targets

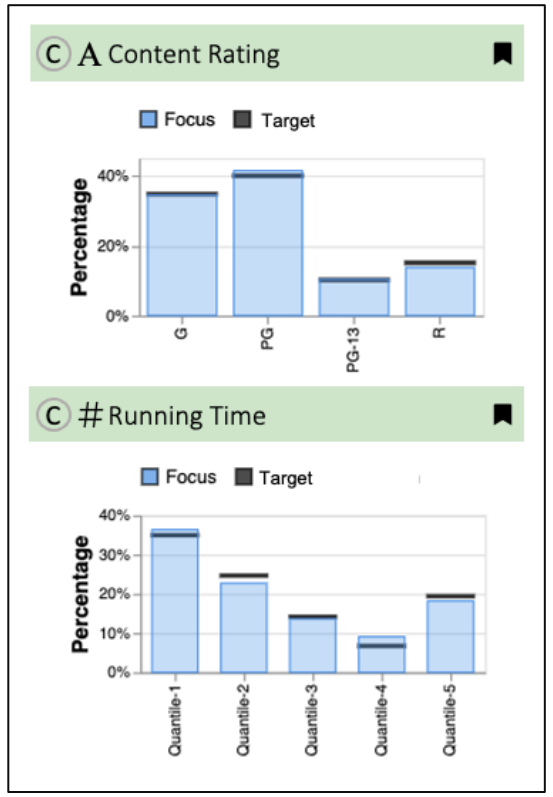
Proportional
(per data distribution)



Equal
(equal focus on all)



Custom
(per work requirement)

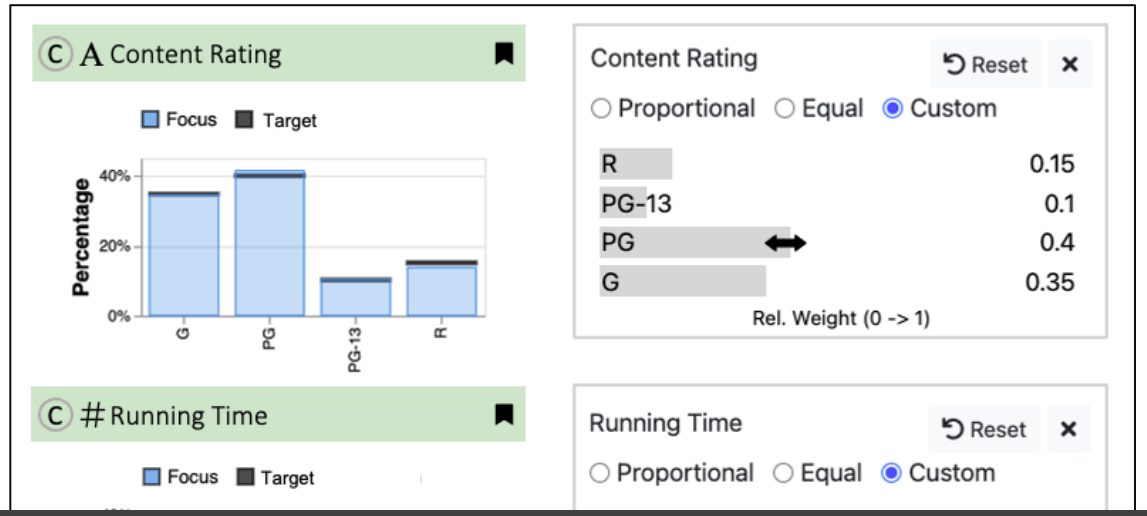


Configuring Custom Baselines / Targets

Proportional
(per data distribution)

Equal
(equal focus on all)

Custom
(per work requirement)



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

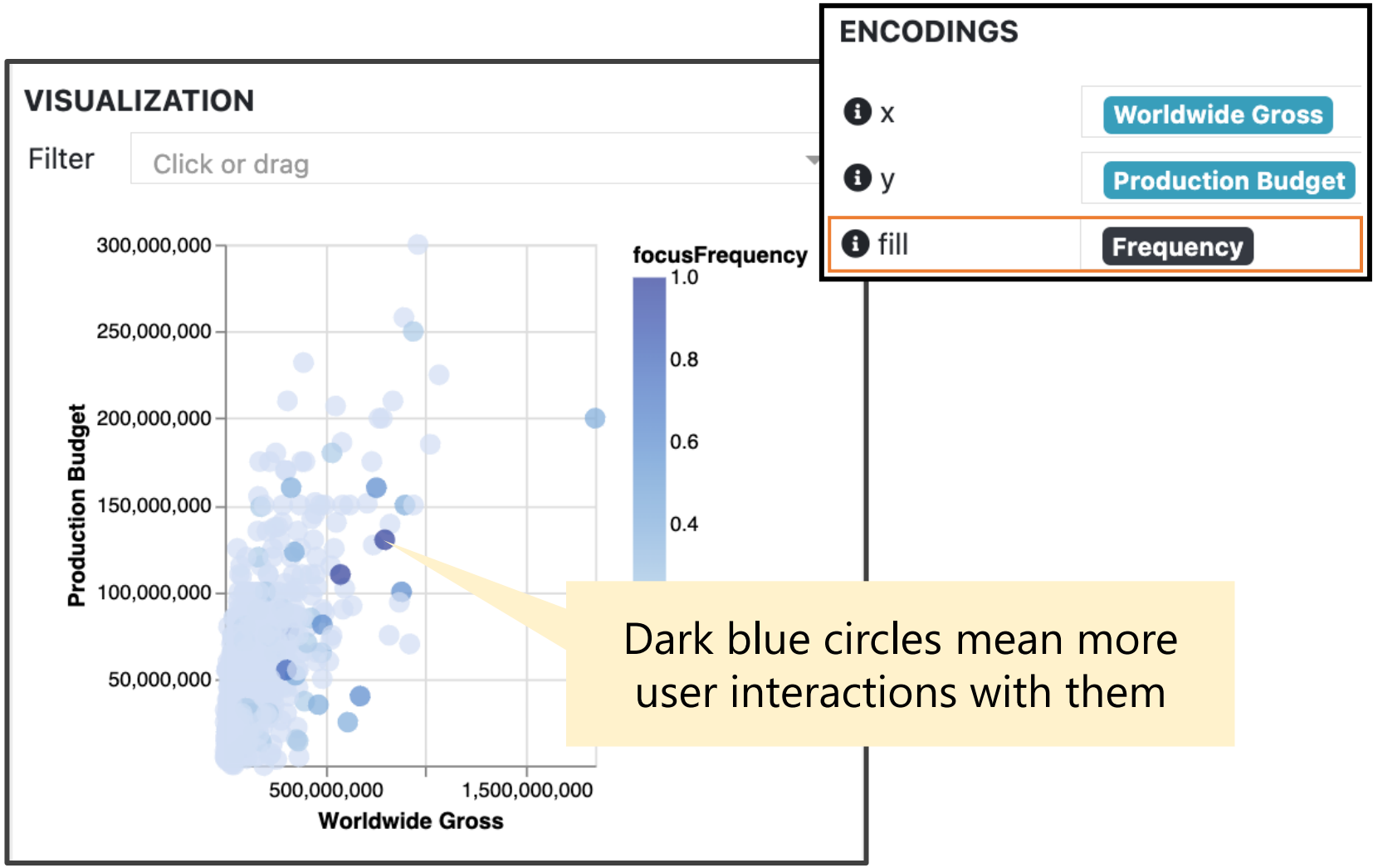
Alex Endert

lumos-vis.github.io

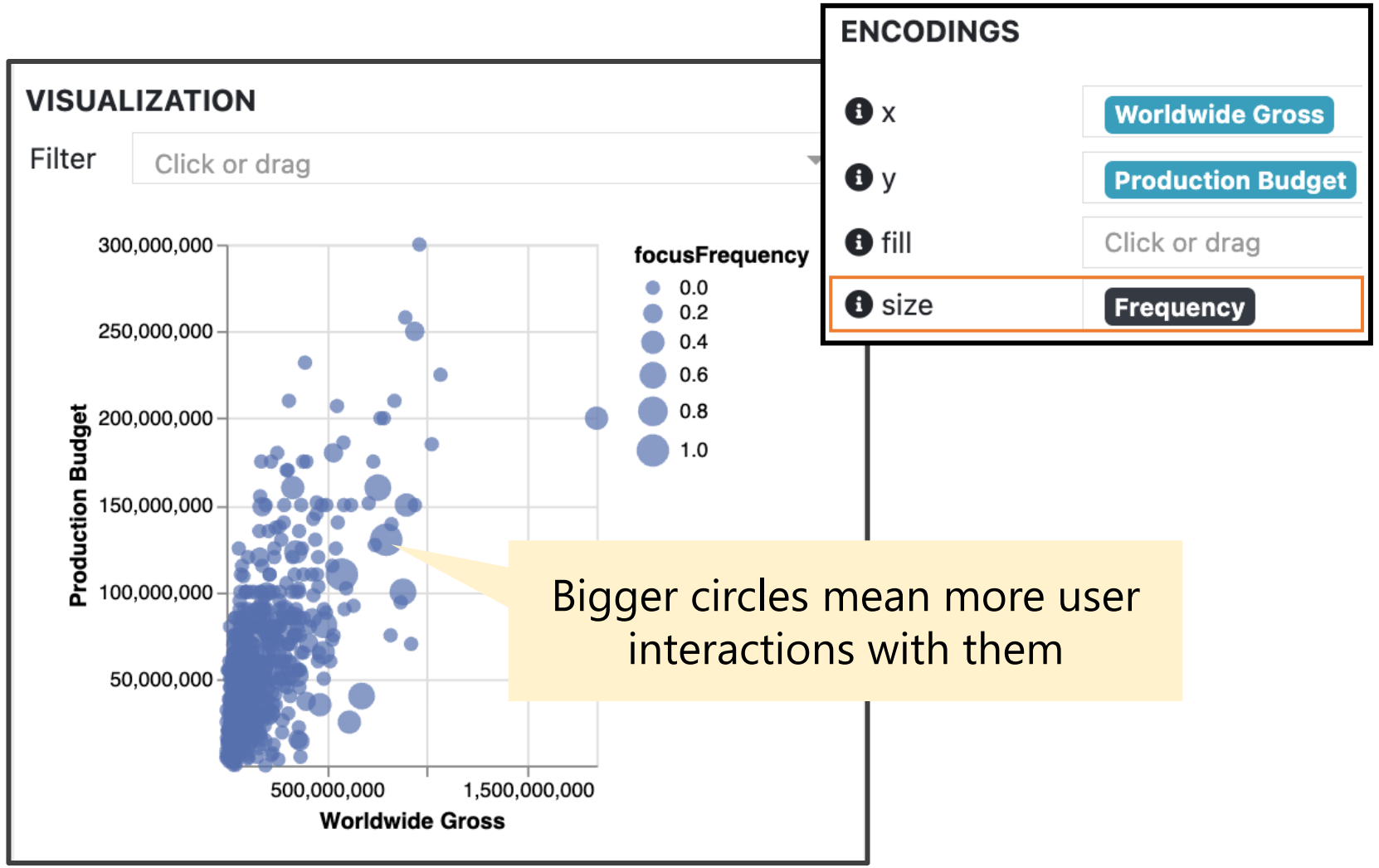
VIS 2021

Why only Color?

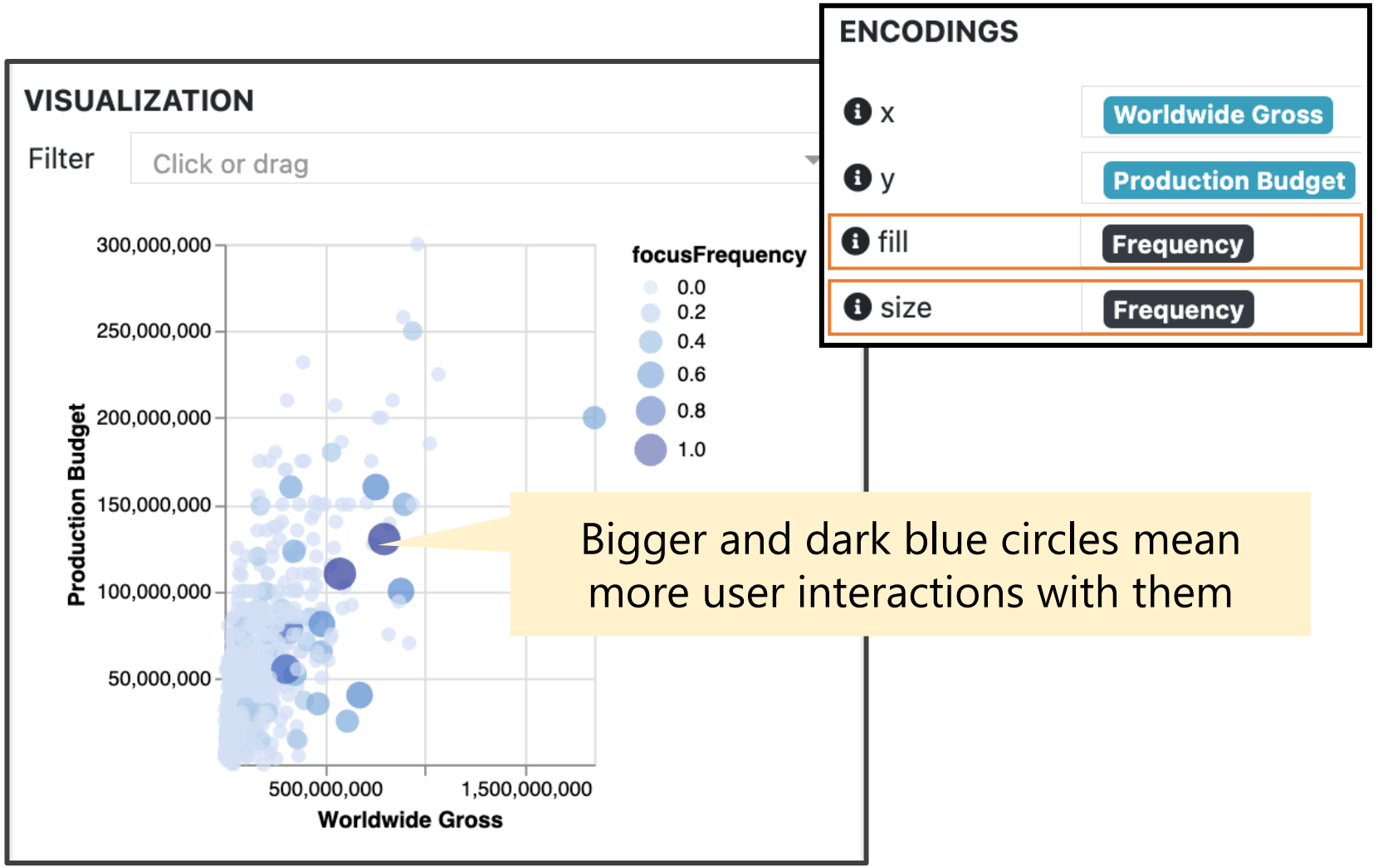
Mapping my focus to “fill” (color)



If not “fill” (color), how about “Size”?



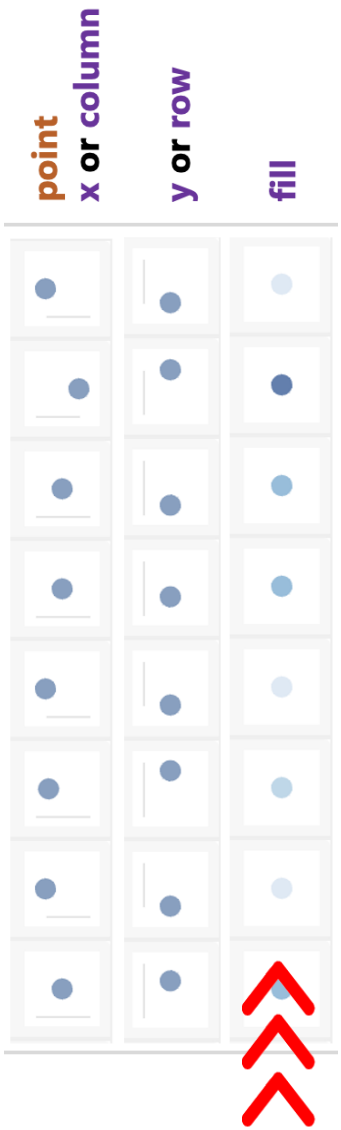
How about both “fill” (color) and “Size”?



How about ANY encoding?

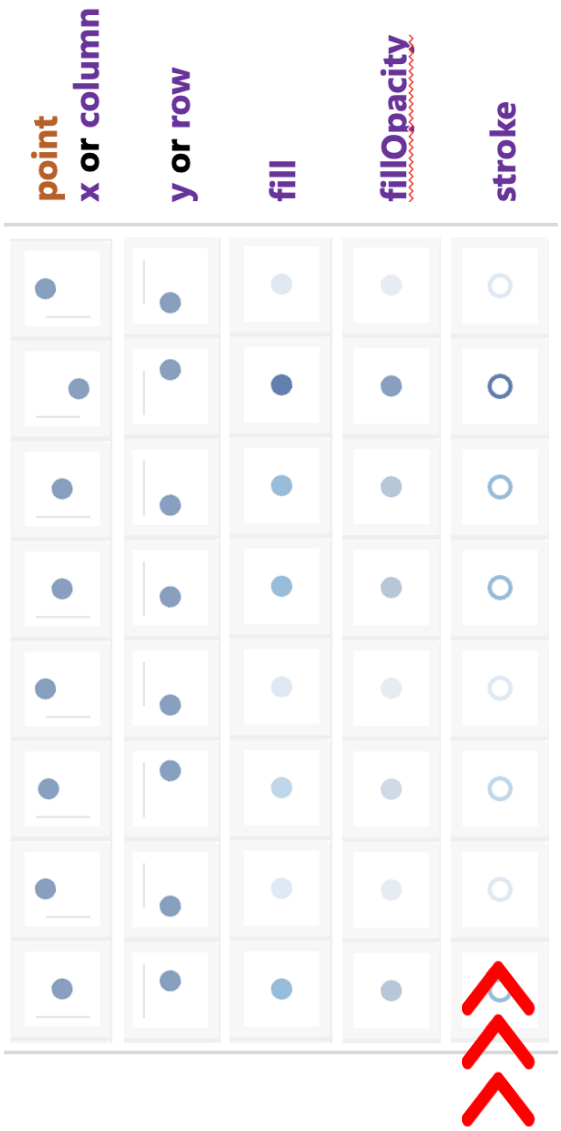
How about ANY encoding?

How about ANY encoding?

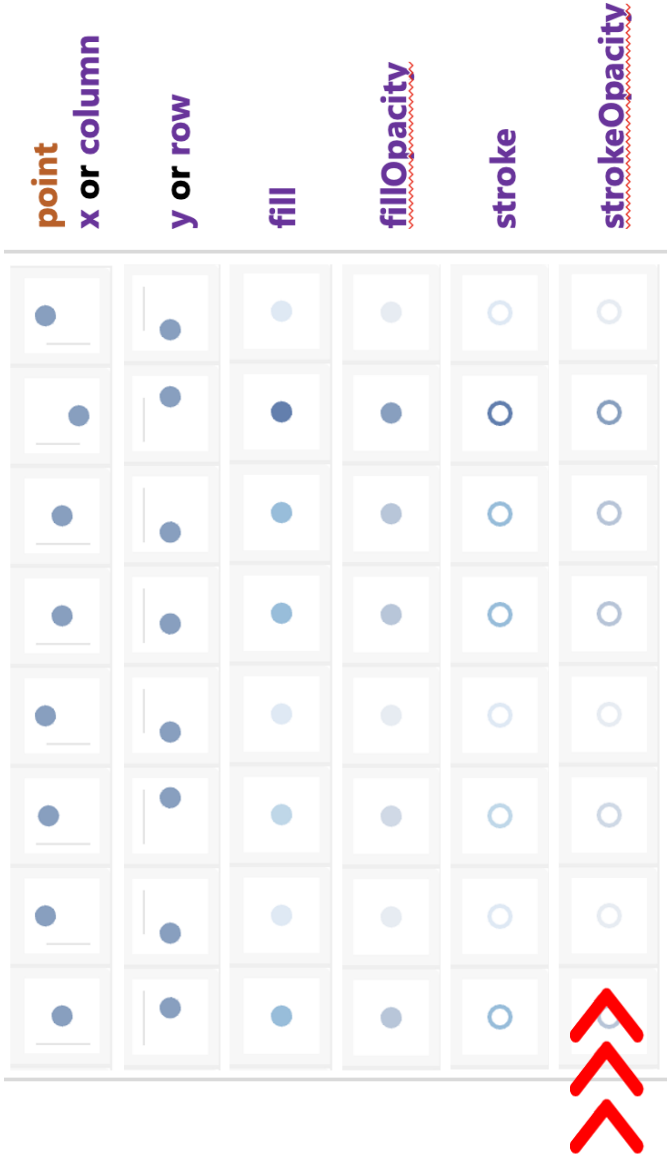


How about ANY encoding?

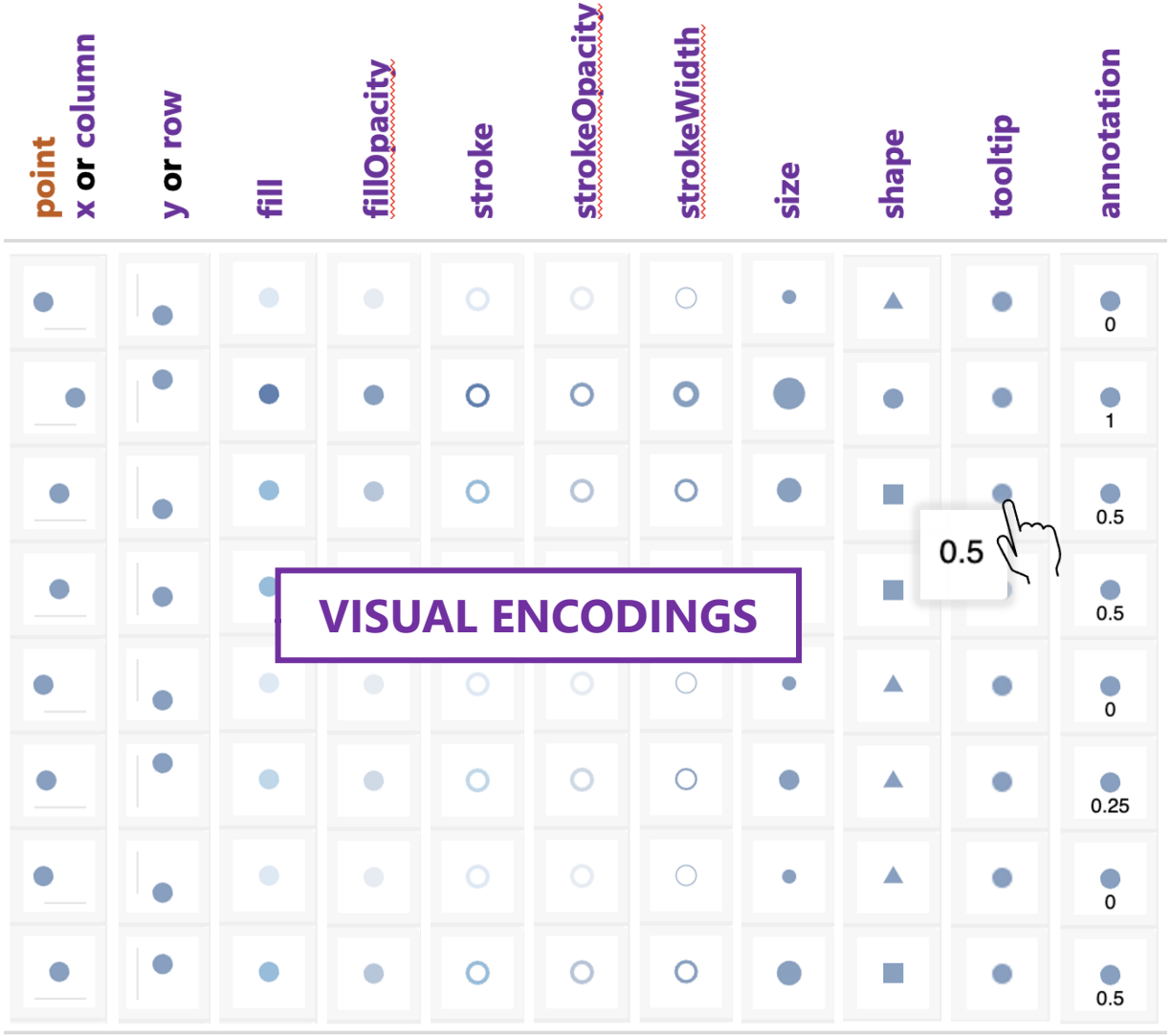
How about ANY encoding?



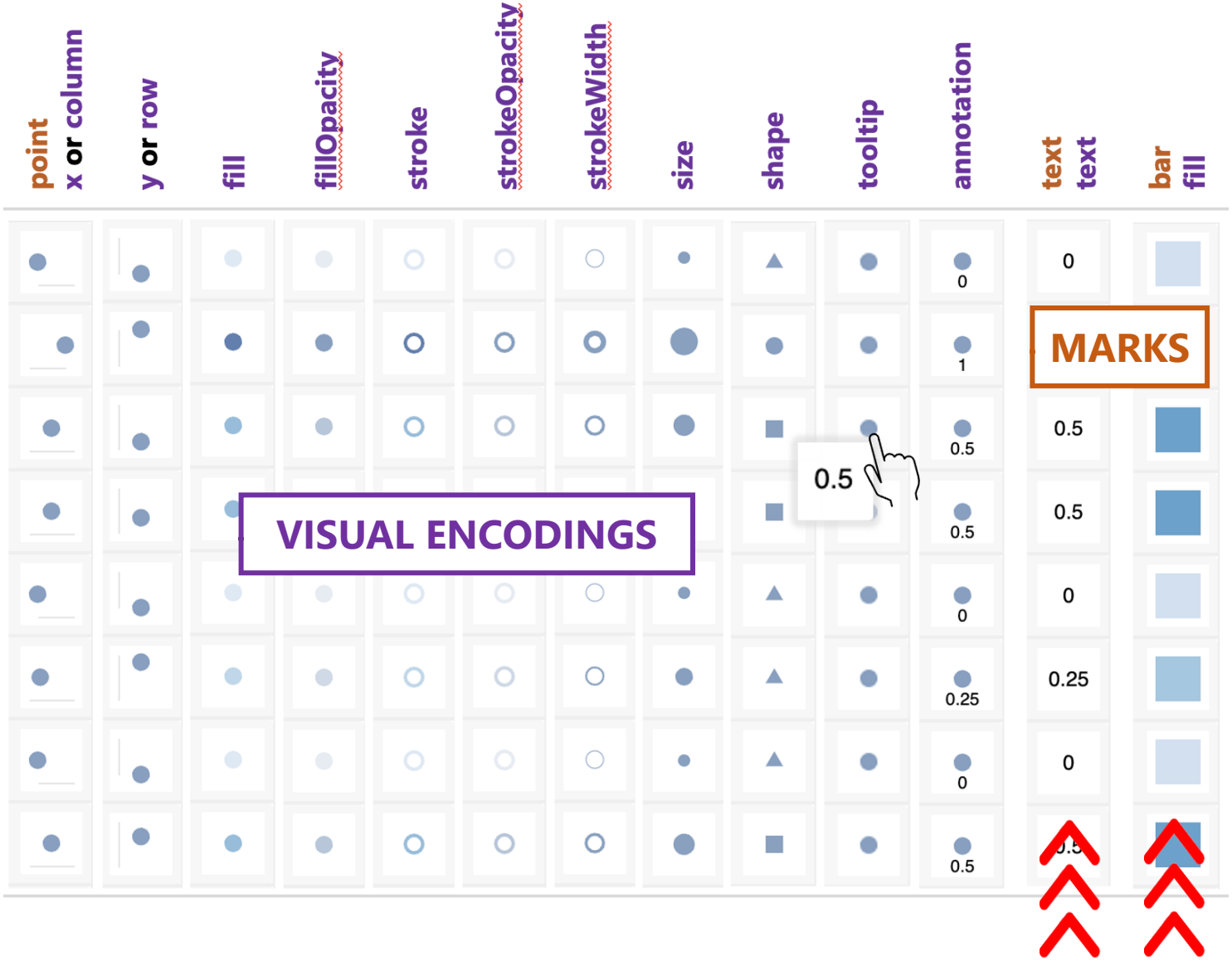
How about ANY encoding?



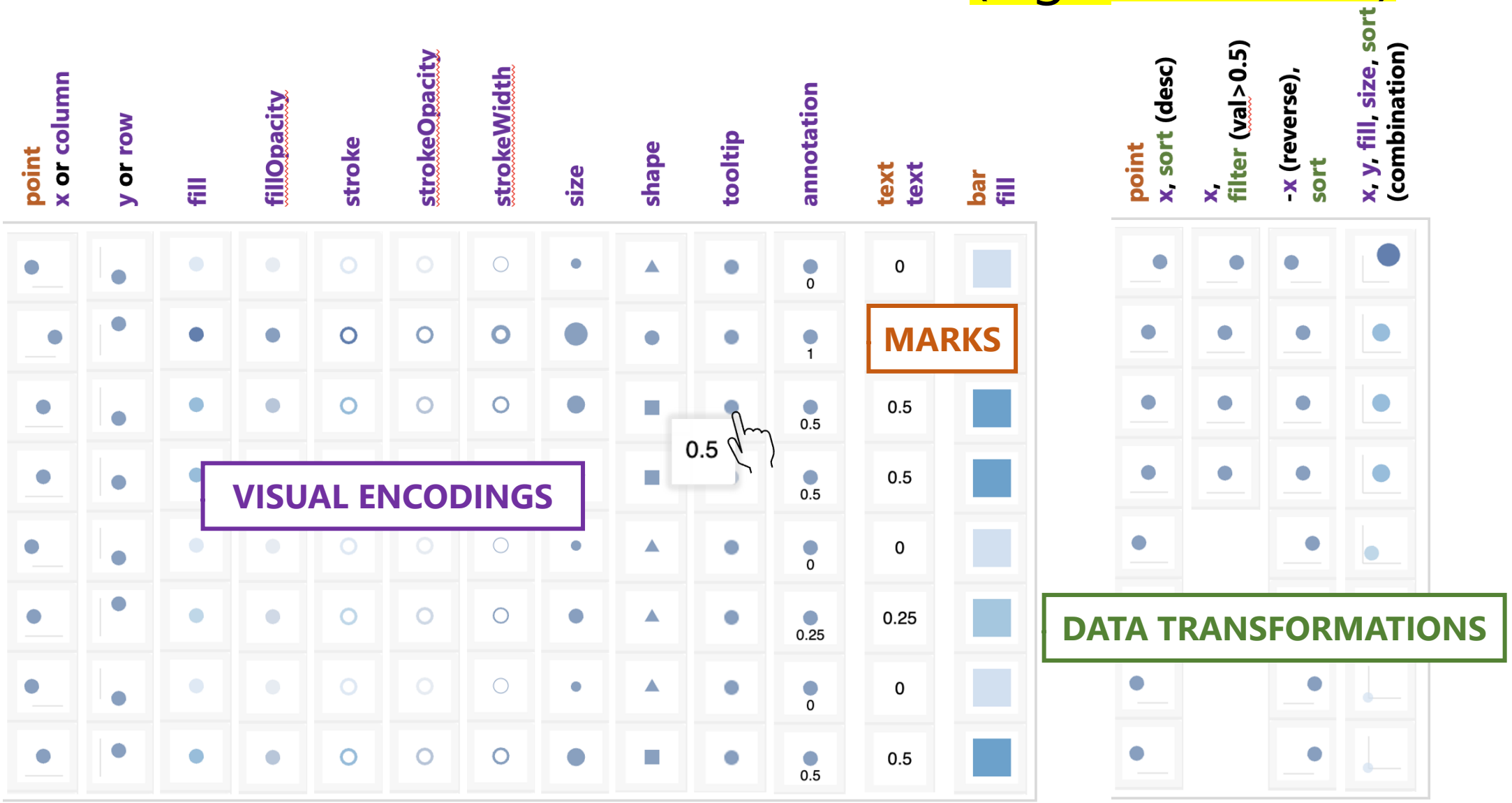
How about ANY encoding?



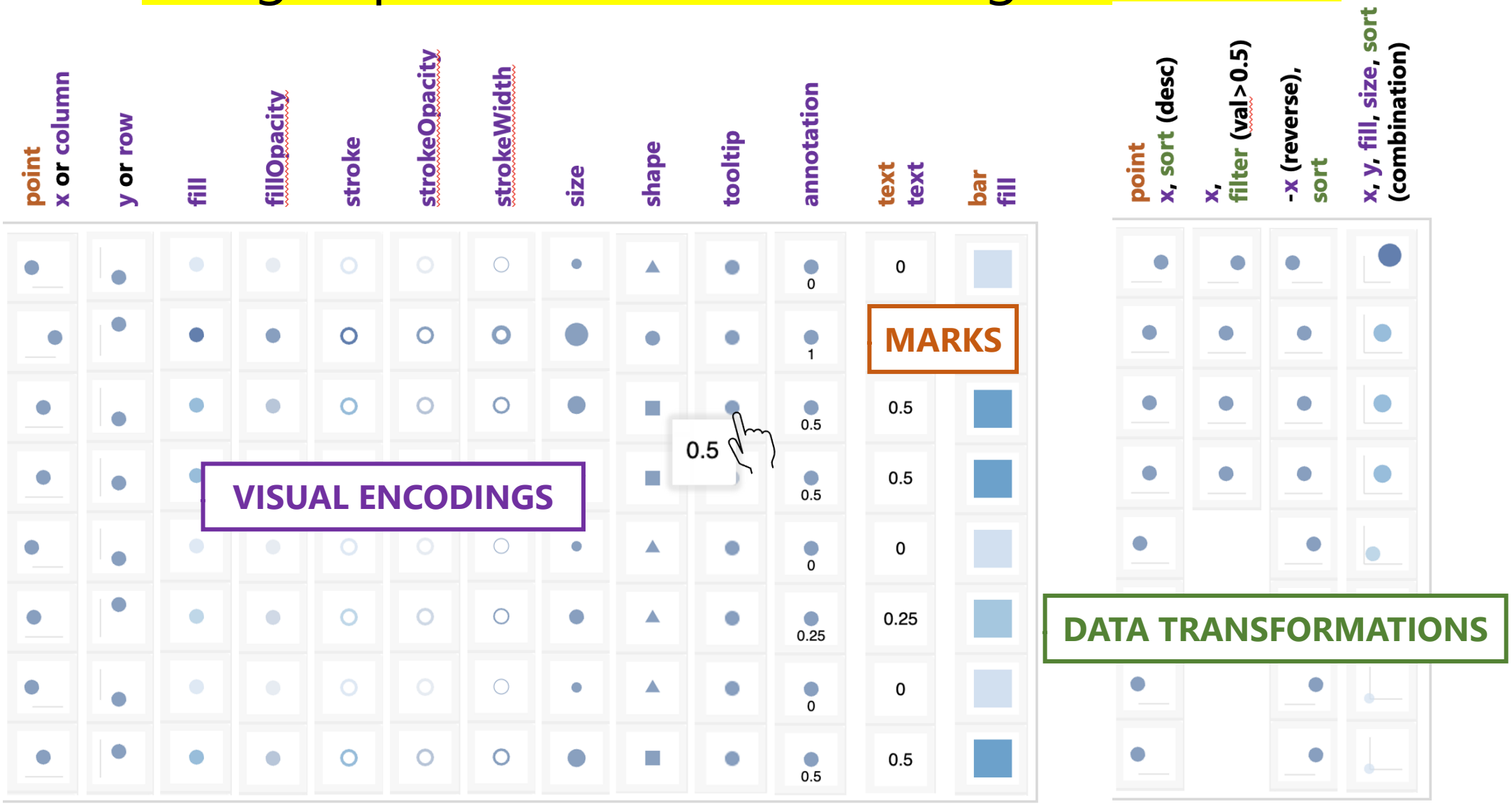
How about ANY mark (e.g., text, bar)?



How about Data Transformations (e.g., Sort, Filter)



Design Space for Communicating Provenance



Adobe

PATENTED

ProvenanceLens

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

Arpit Narechania

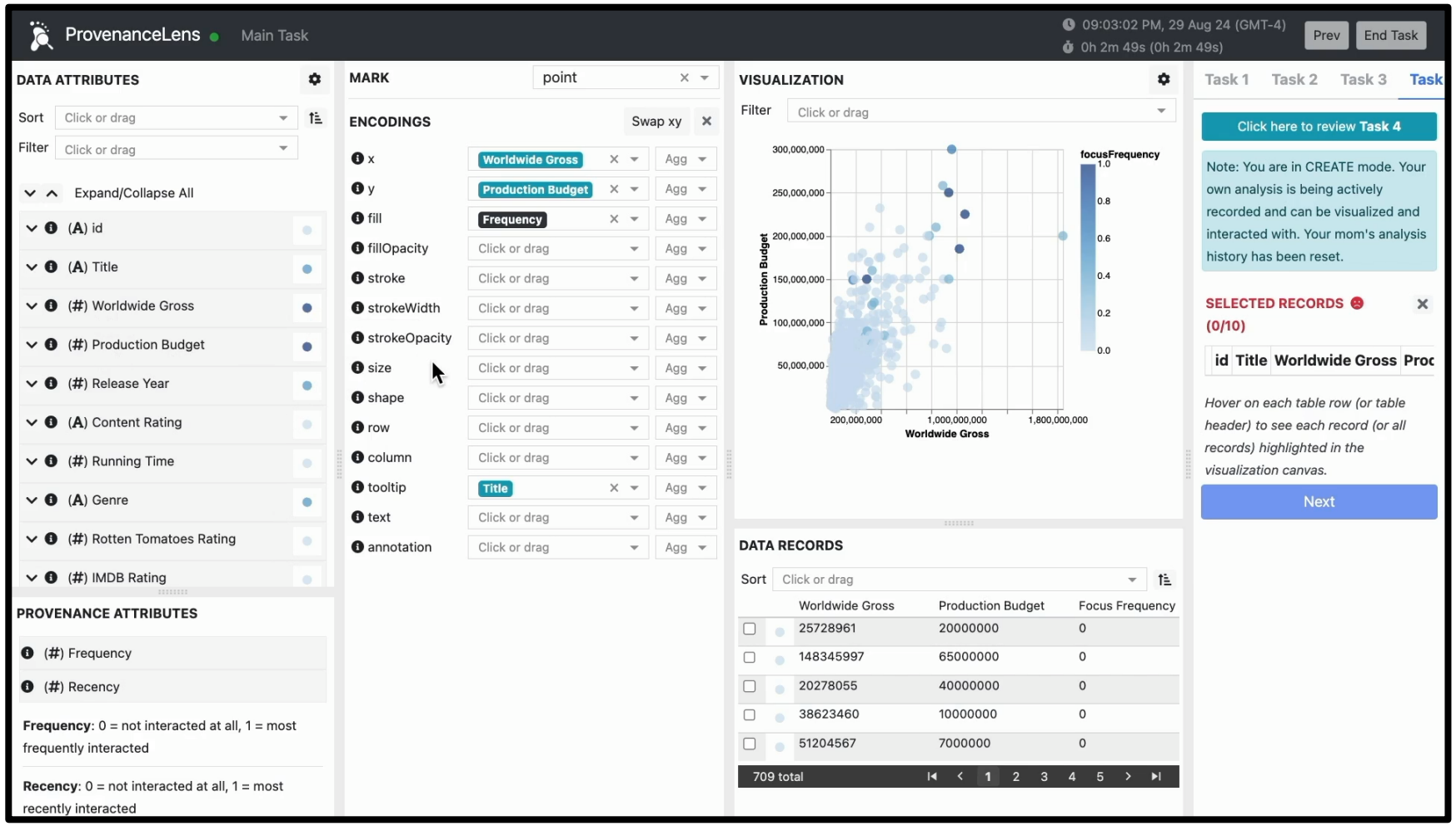
Shunan Guo

Eunyee Koh

Alex Endert

Jane Hoffswell

IEEE TVCG 2025



How about Multimodal Guidance (why only Visual)?

BiasBuzz

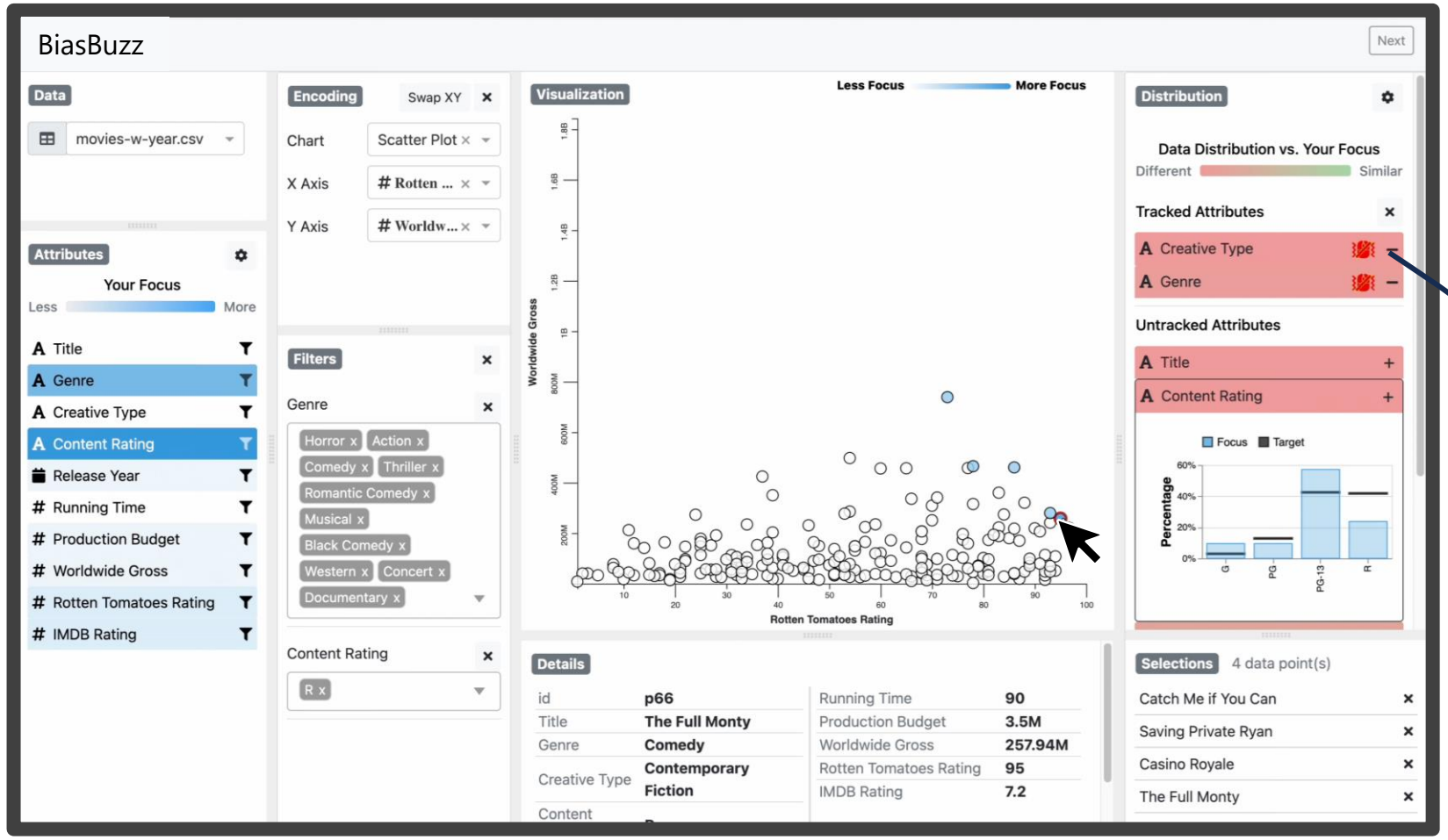
Combining Visual Guidance and Haptic Feedback to Increase Awareness of Analytic Behavior during Visual Data Analysis

Jamal Paden*

Arpit Narechania*

Alex Endert

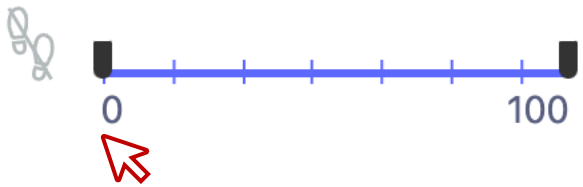
CHI 2024



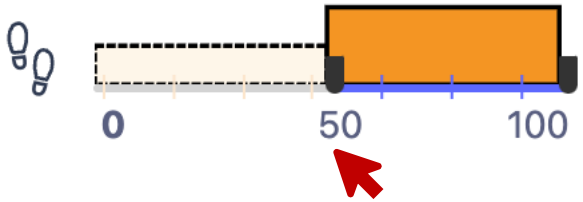
*equal contribution

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

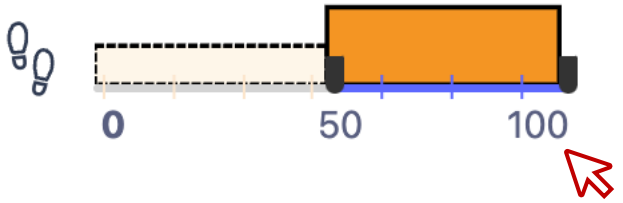
Range Slider



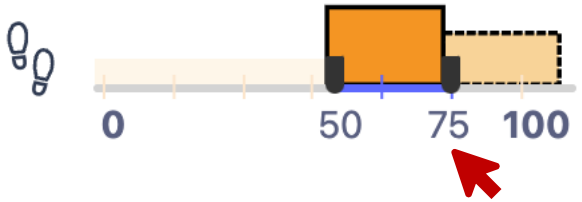
Range Slider



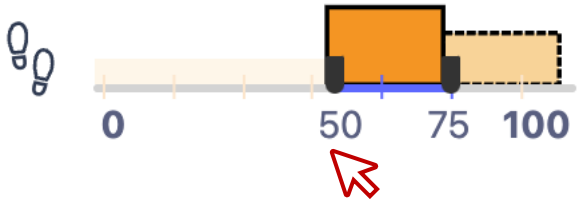
Range Slider



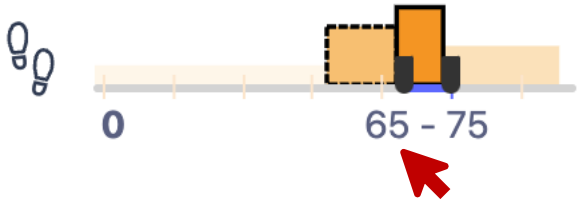
Range Slider



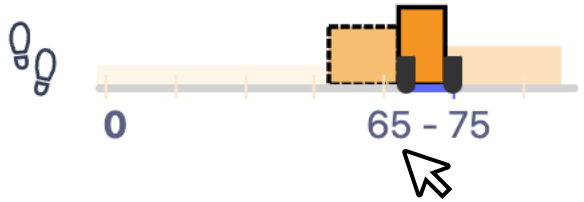
Range Slider

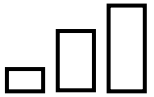


Range Slider



Range Slider

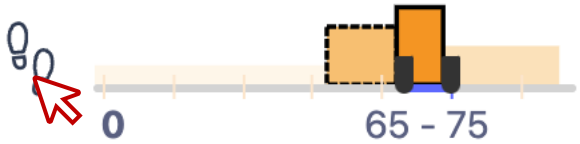




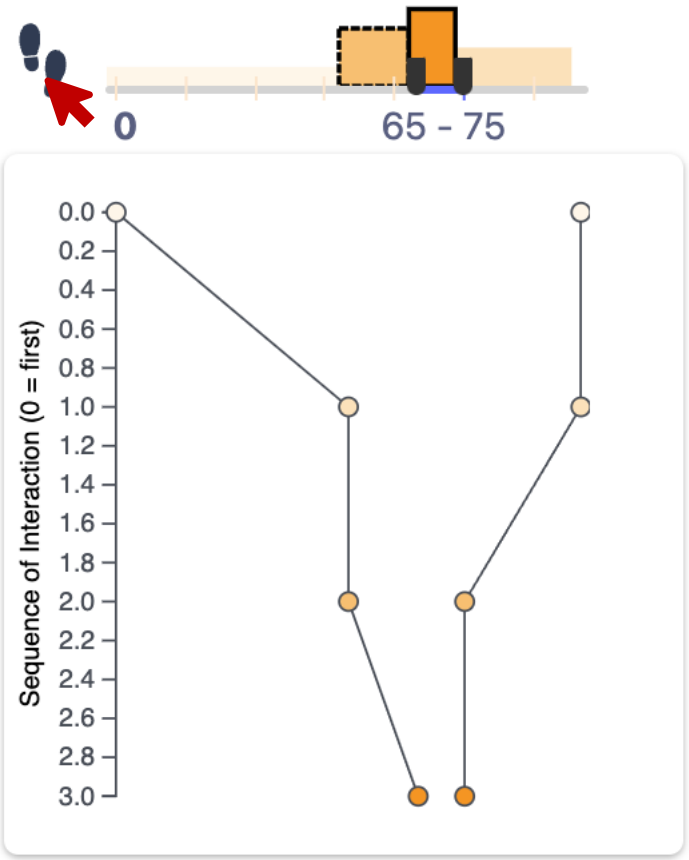
height = frequency

color = recency

Range Slider



Range Slider



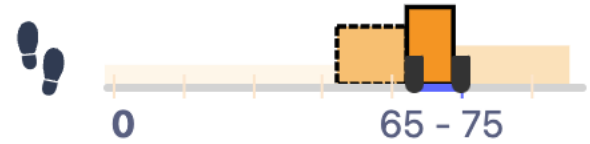
Range Slider

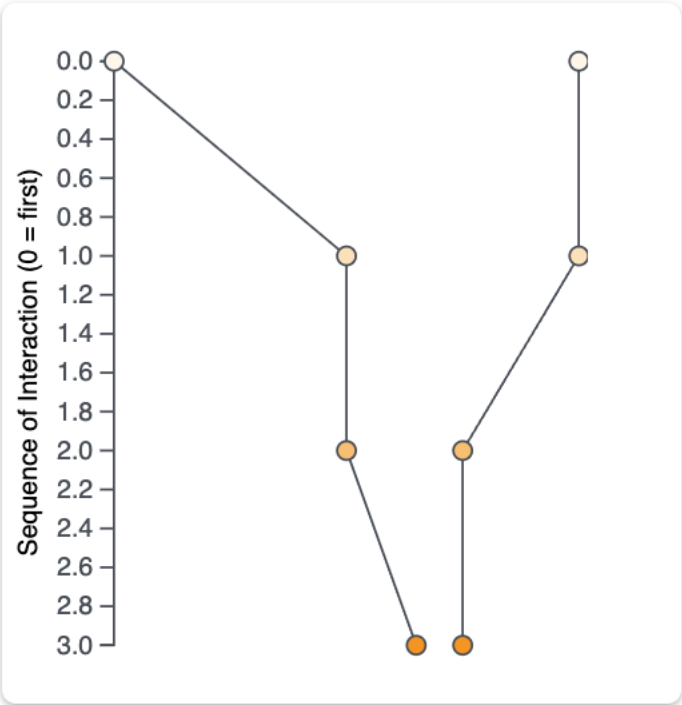
initial state
[0, 100]

slider handle state = ◯

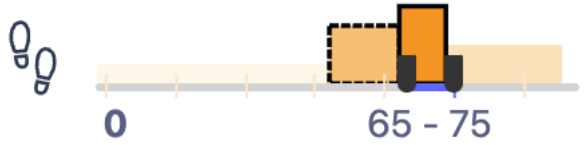
recency = ● ● ●

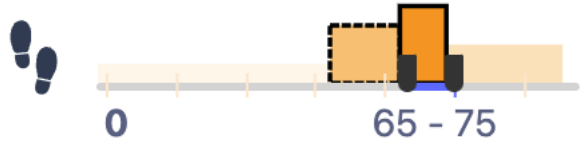
current state
[65, 75]

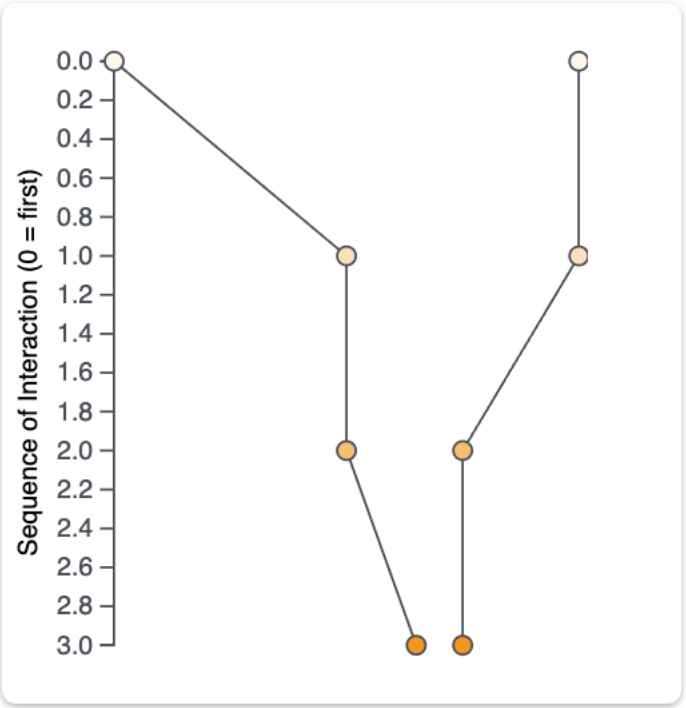




Such capabilities can also help increase user awareness during analysis.







ProvenanceWidgets

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

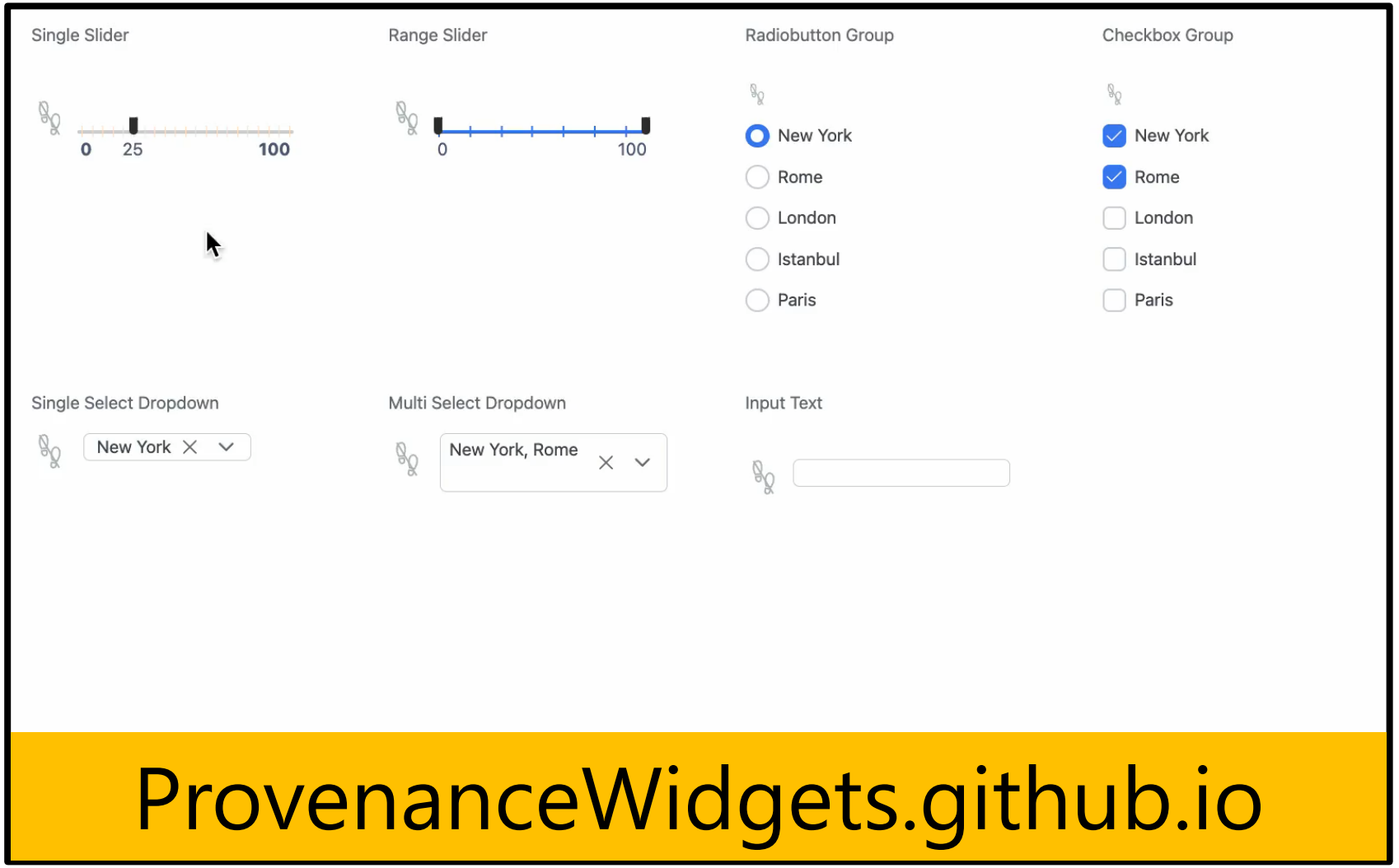
Arpit Narechania

Kaustubh Odak

Mennatallah El-Assady

Alex Endert

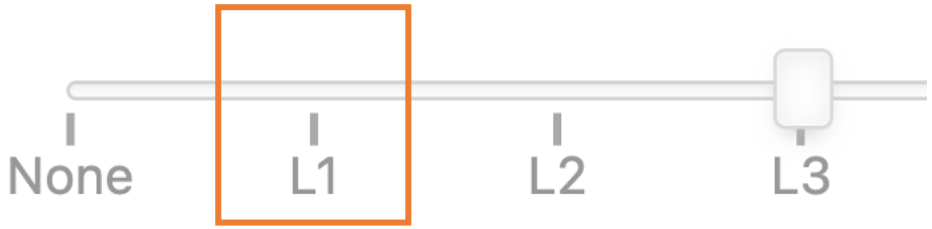
VIS2024



How about different amounts/levels of Guidance?

No Guidance

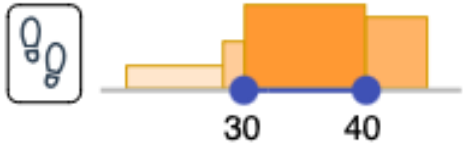
Range Slider



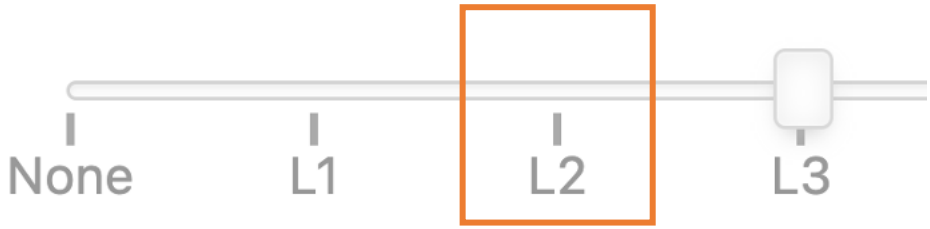
No Guidance

Level 1 Guidance

Range Slider



ProvenanceWidgets

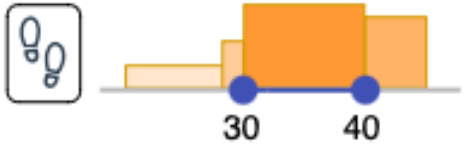


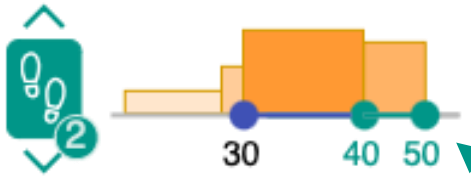
No Guidance

Level 1 Guidance

Level 2 Guidance

Range Slider



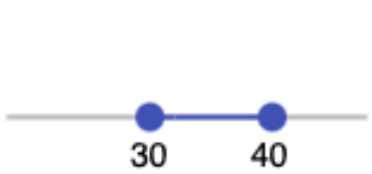


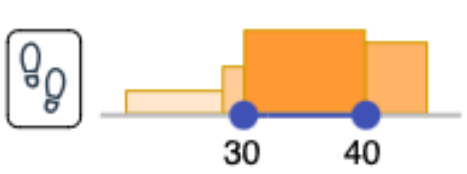
Range to Select

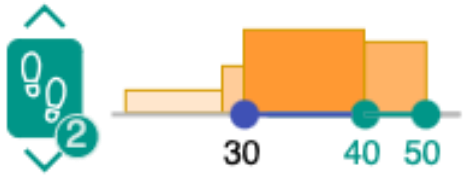
Multiple options available as guidance

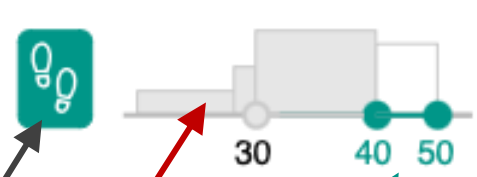
No Guidance | Level 1 Guidance | Level 2 Guidance | Level 3 Guidance

Range Slider







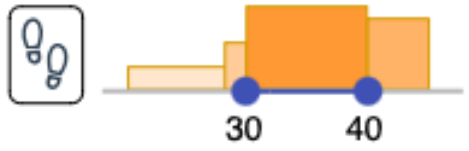
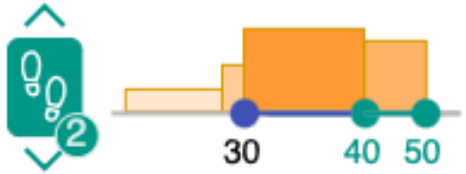
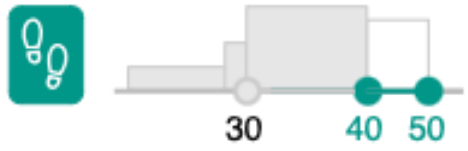


Only one option available as guidance

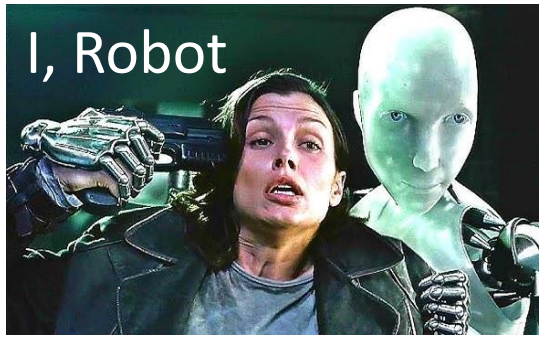
Range to Select

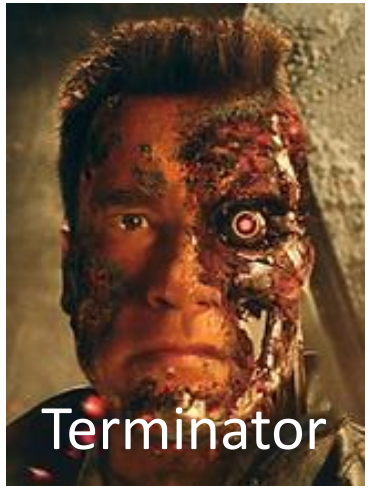
Disabled range

GuidanceWidgets [coming soon]

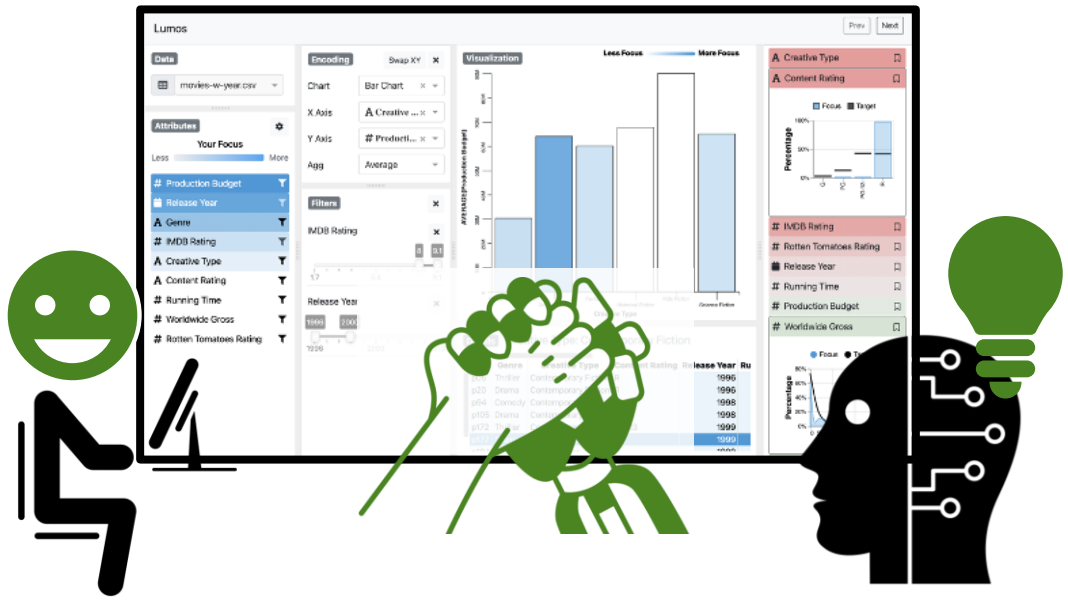
	No Guidance	Level 1 Guidance	Level 2 Guidance	Level 3 Guidance
Range Slider				

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)

Design

- Interaction Traces
- Provenance Design Space

Develop

- Lumos
- BiasBuzz
- ProvenanceLens

Democratize

- ProvenanceWidgets
- GuidanceWidgets

Past and Ongoing **Industry** Collaborations

Ford	Minimize Risks to Make Cars Safer 
Adobe	Improve Data Lake Navigation 
	Provide Adaptive Guidance 
	Analyze PDF Documents 
Microsoft Research	Debug Natural Language to SQL Tasks 
Alibaba.com	Storytelling of eCommerce Data
Non-Profit Mapping Orgs	Educate How Maps can Lie 

Past and Ongoing **Industry** Collaborations

Ford — Minimize Risks to Make Cars Safer

Adobe — Improve Data Lake Navigation

Provide Adaptive Guidance

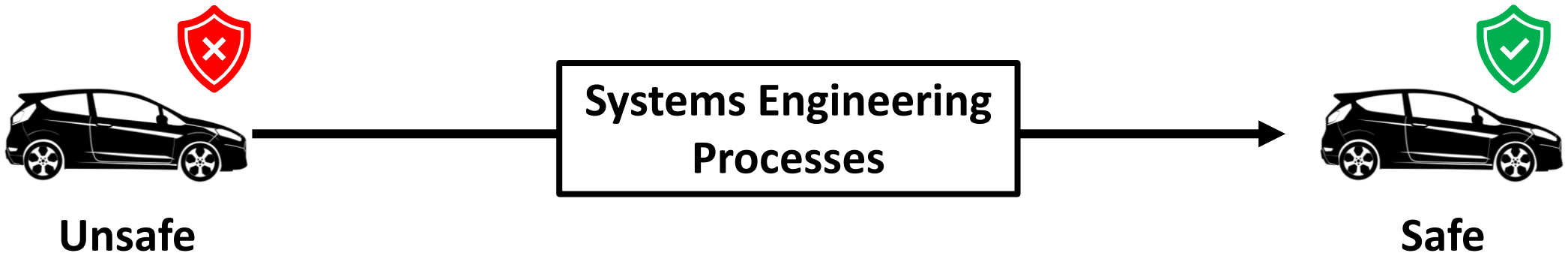
Analyze PDF Documents

Microsoft Research — Debug Natural Language to SQL Tasks

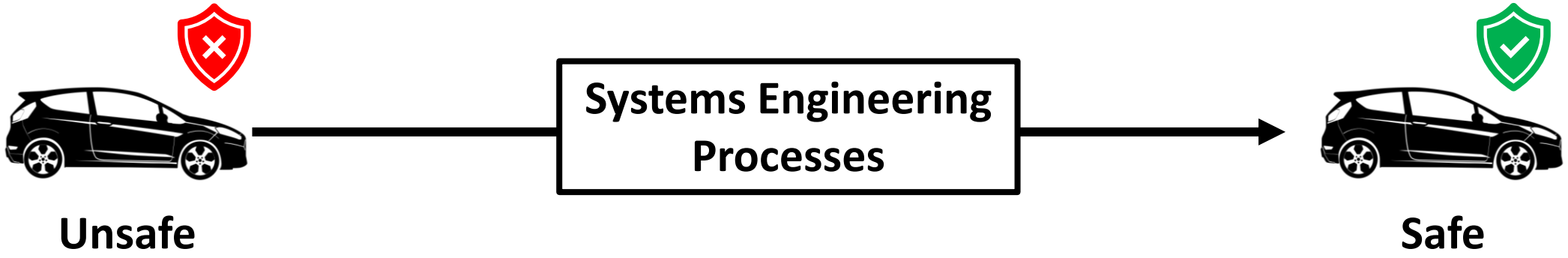
Alibaba.com — Storytelling of eCommerce Data

Non-Profit Mapping Orgs — Educate How Maps can Lie

Minimize Risks to Make Cars Safer

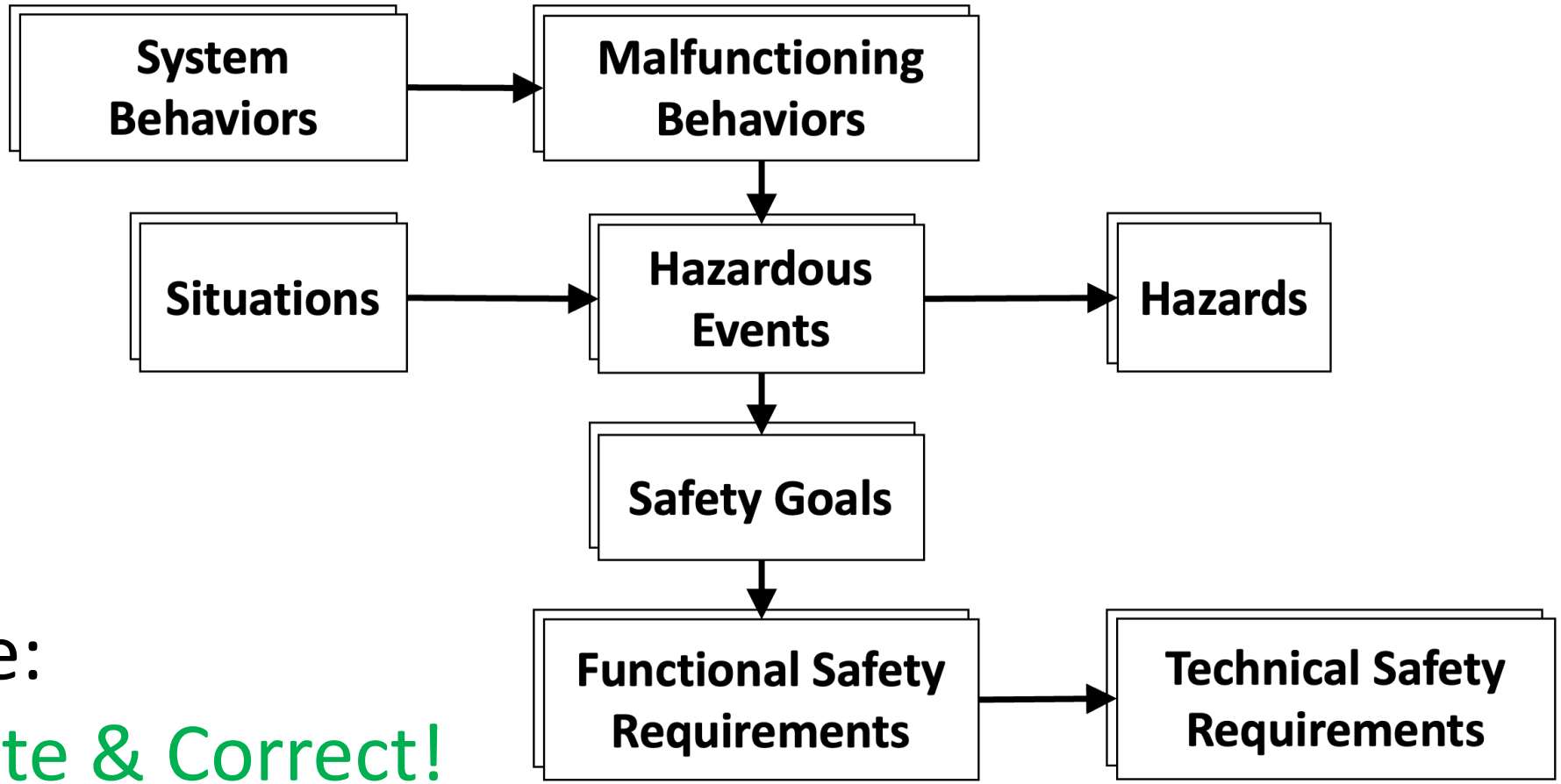


Minimize Risks to Make Cars Safer



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

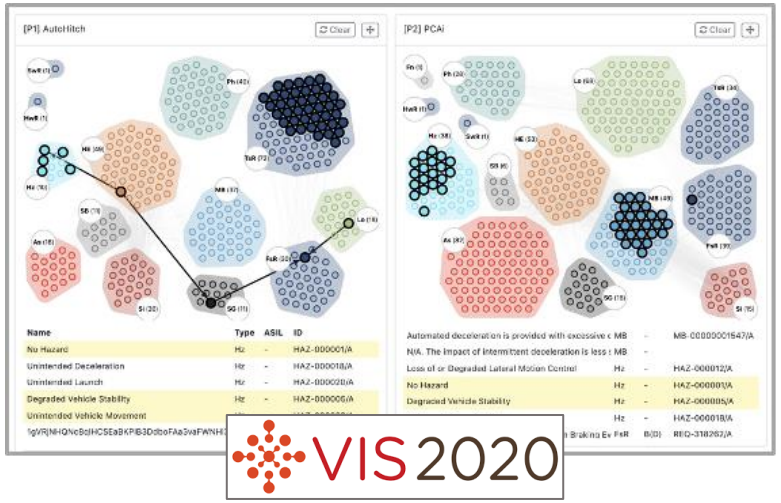
Functional Safety Analysis (e.g., of Adaptive Cruise Control)



Must be:

Complete & Correct!

Design Study w/ **10+ Safety Engineers, Analysts**

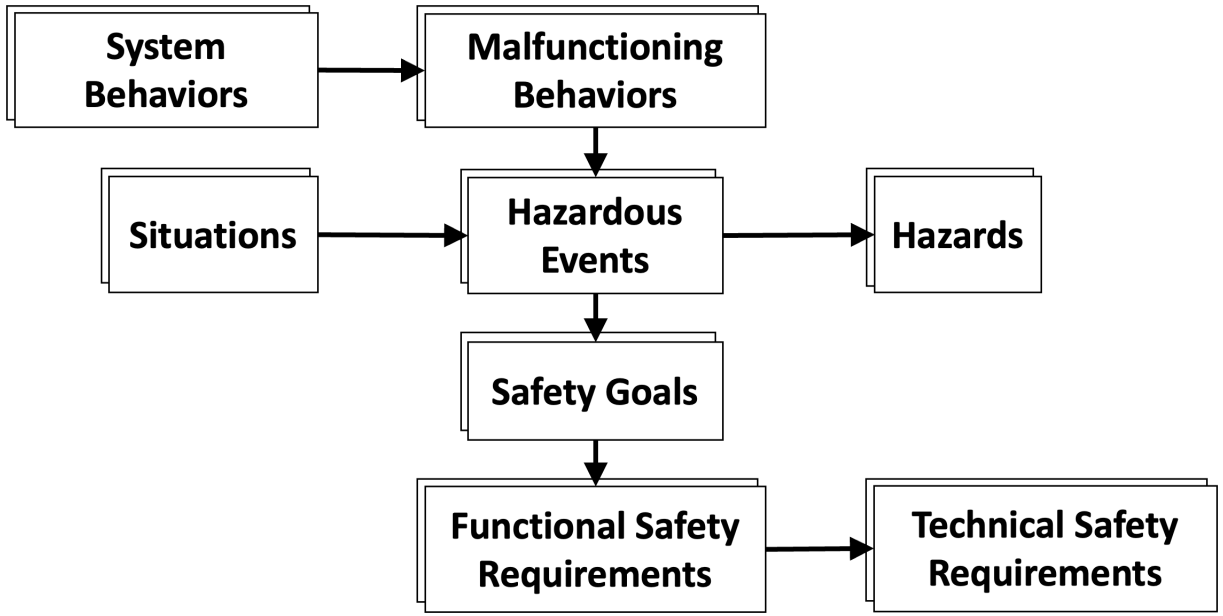


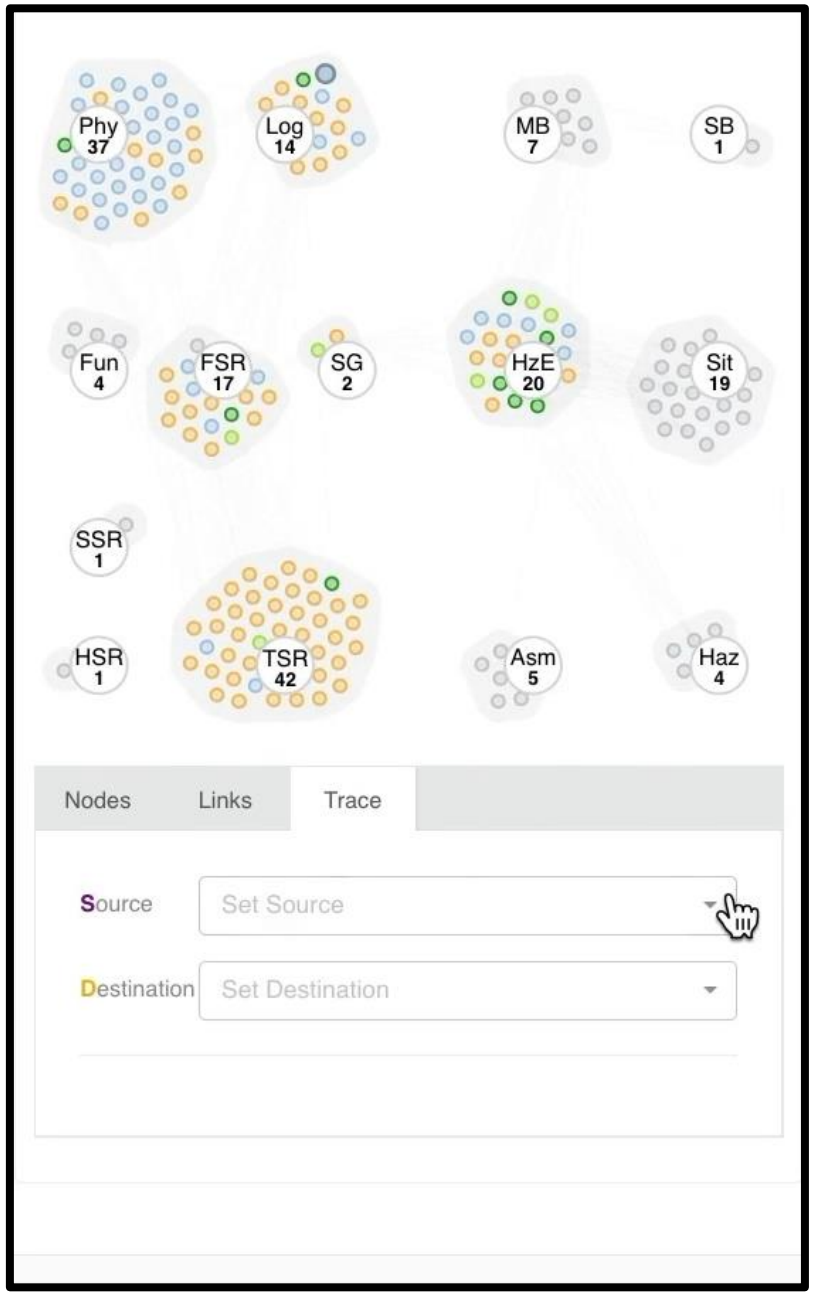
SafetyLens 🚀

Visual Data Analysis of **Functional Safety** of Vehicles

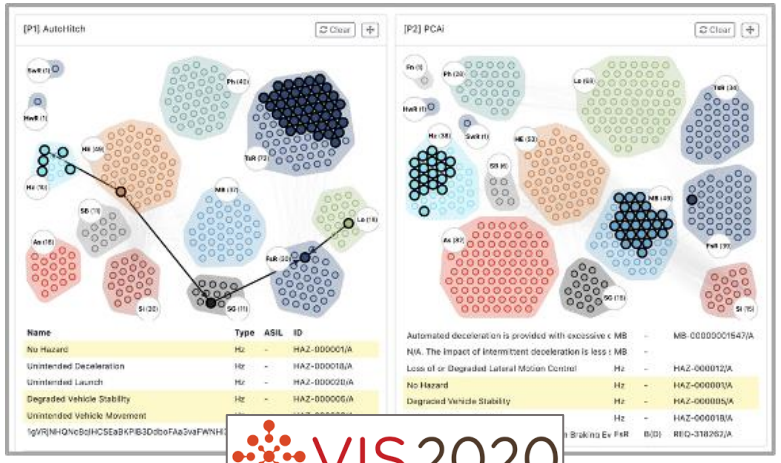
A. Narechania, A. Qamar, A. Endert

Functional Safety





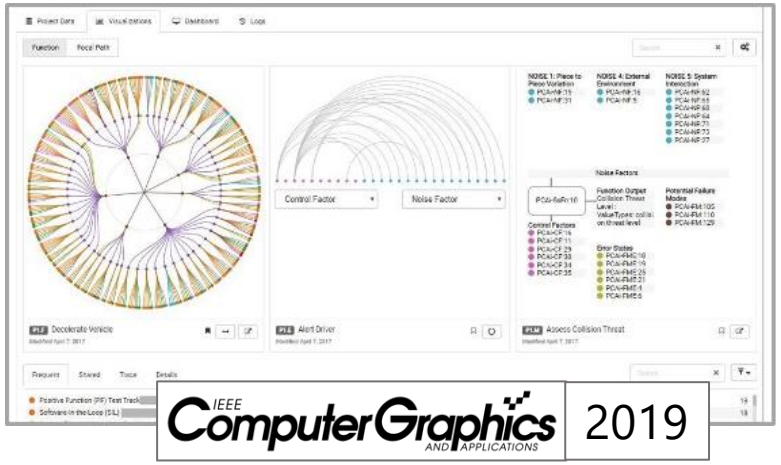
Design Study w/ **10+ Safety Engineers, Analysts**



SafetyLens 🚀

Visual Data Analysis of **Functional Safety** of Vehicles

A. Narechania, A. Qamar, A. Endert



DataHawk 🚀

Understanding **FMEA Data** Using Interactive Visual Analytics

R. Basole, A. Qamar, B. Pal, M. Corral, M. Meinhart, **A. Narechania**

Past and Ongoing **Industry** Collaborations

Ford — Minimize Risks to Make Cars Safer

Adobe — Improve Data Lake Navigation

Provide Adaptive Guidance

Analyze PDF Documents

Microsoft Research — Debug Natural Language to SQL Tasks

Alibaba.com — Storytelling of eCommerce Data

Non-Profit Mapping Orgs — Educate How Maps can Lie

Adobe Design Study w/ 14 data engineers & scientists

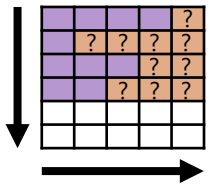
*"Is the data complete? correct? unbiased? ... showing **data quality insights** may instil feelings of curiosity and care in the user."*

*"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

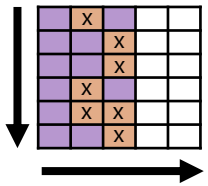
missing (?)



completeness

% frequency of non-missing values in the data

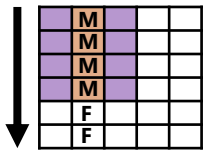
incorrect (x)



correctness

% frequency of correct values in the data

skewed (Male, Female)



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

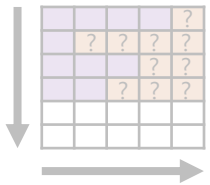
in-visualizations

% of users that visualized the attribute

We modeled data quality and usage insights

1. quality

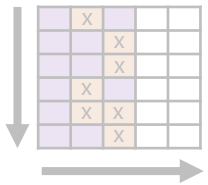
missing (?)



completeness

% frequency of non-missing values in the data

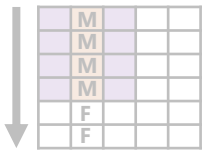
incorrect (x)



correctness

% frequency of correct values in the data

skewed (Male, Female)



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

DataPilot

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

Arpit Narechania

Fan Du

Atanu Sinha

Ryan Rossi

Jane Hoffswell

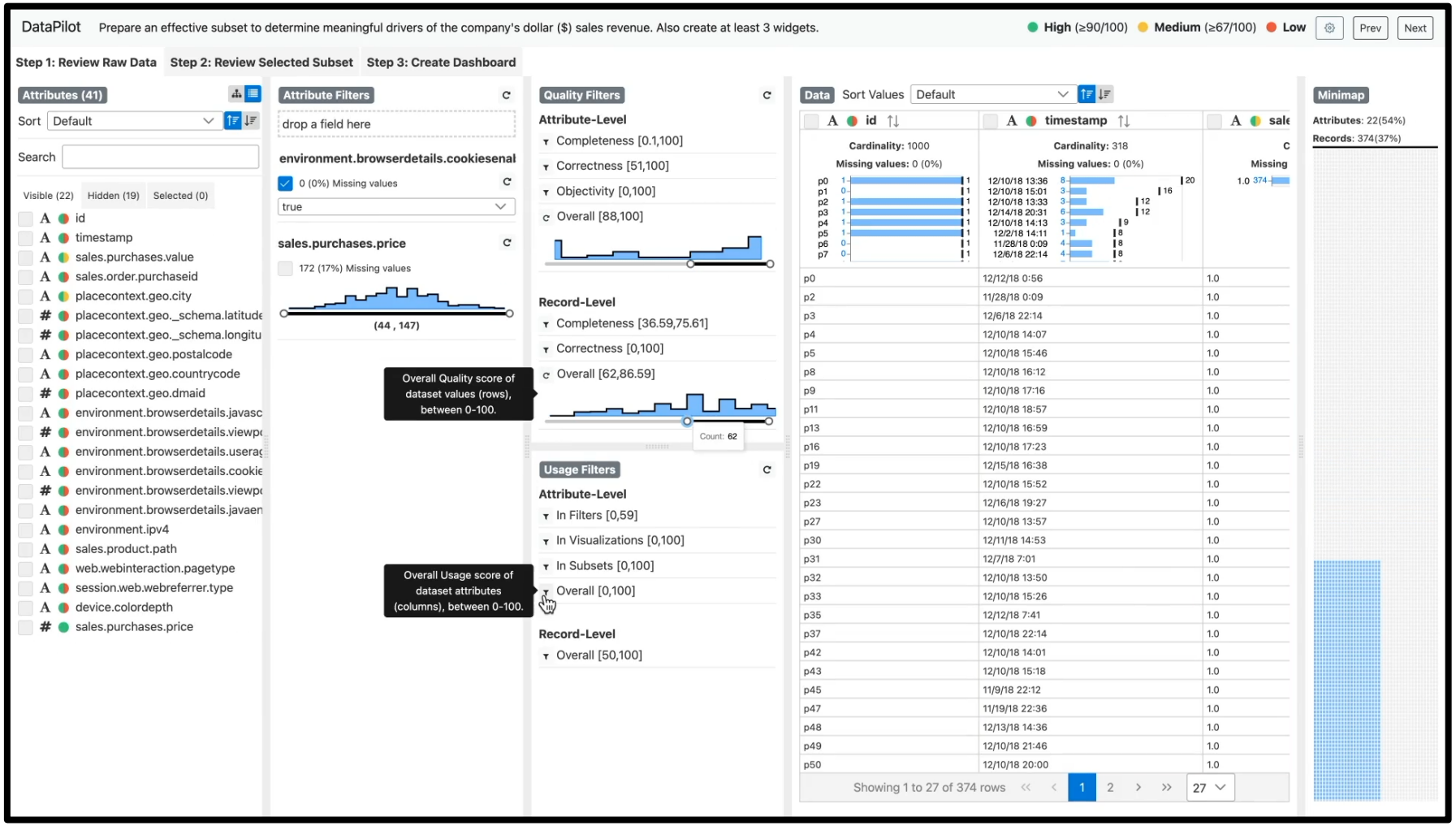
Shunan Guo

Eunyee Koh

Shamkant Navathe

Alex Endert

CHI23
Hamburg, Germany | Hybrid
April 23-28, 2023
reCHInnecting



PATENTED

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

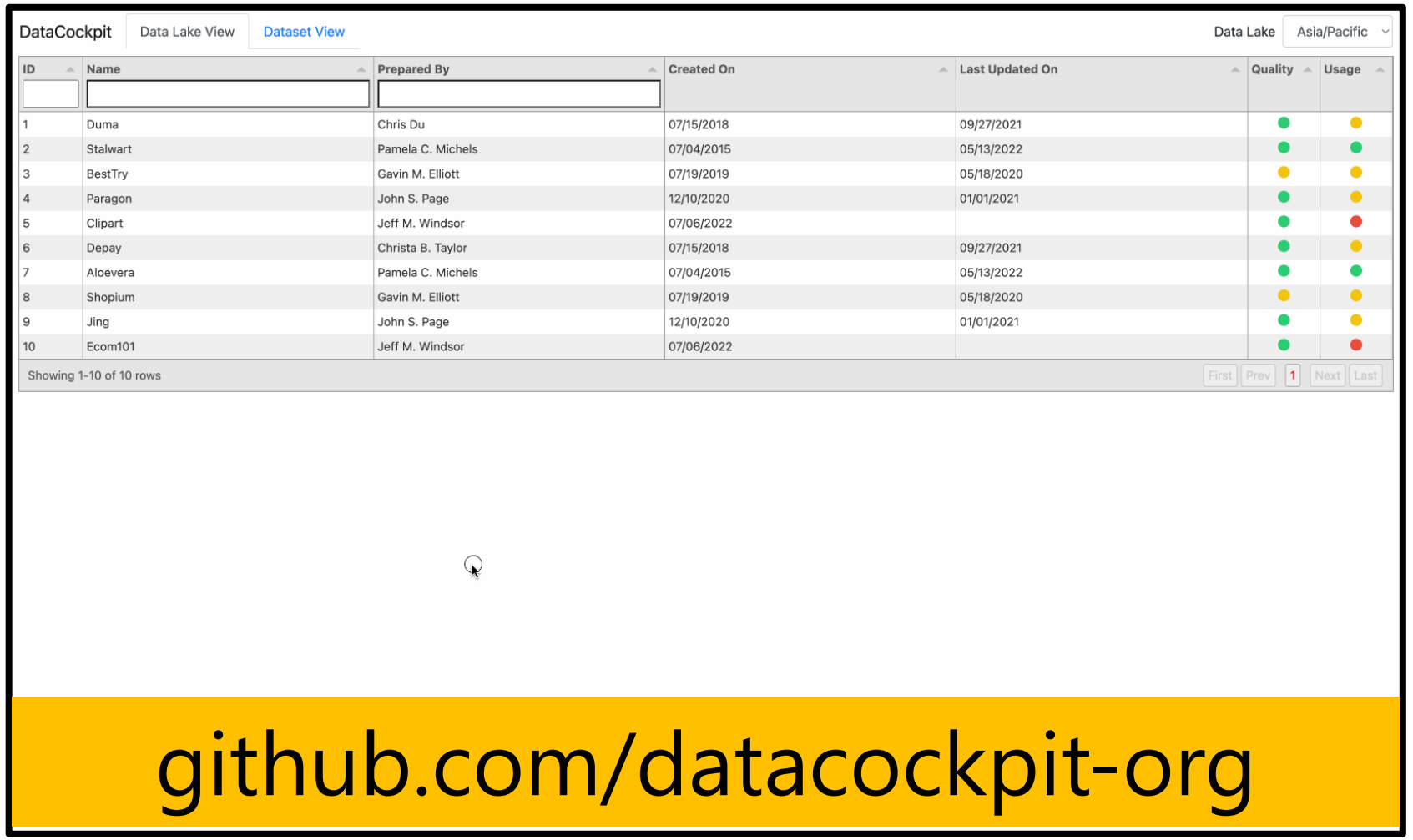
Shunan Guo

Eunyee Koh

Alex Endert

Shamkant Navathe

IEEE BigData



PATENTED

PATENTED

Past and Ongoing **Industry** Collaborations

Ford	Minimize Risks to Make Cars Safer
Adobe	Improve Data Lake Navigation
	Provide Adaptive Guidance
	Analyze PDF Documents
Microsoft Research	Debug Natural Language to SQL Tasks
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

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING