Alexander Lex
@alexander_lex
http://alexander-lex.net

THE UNIVERSITY OF UTAH

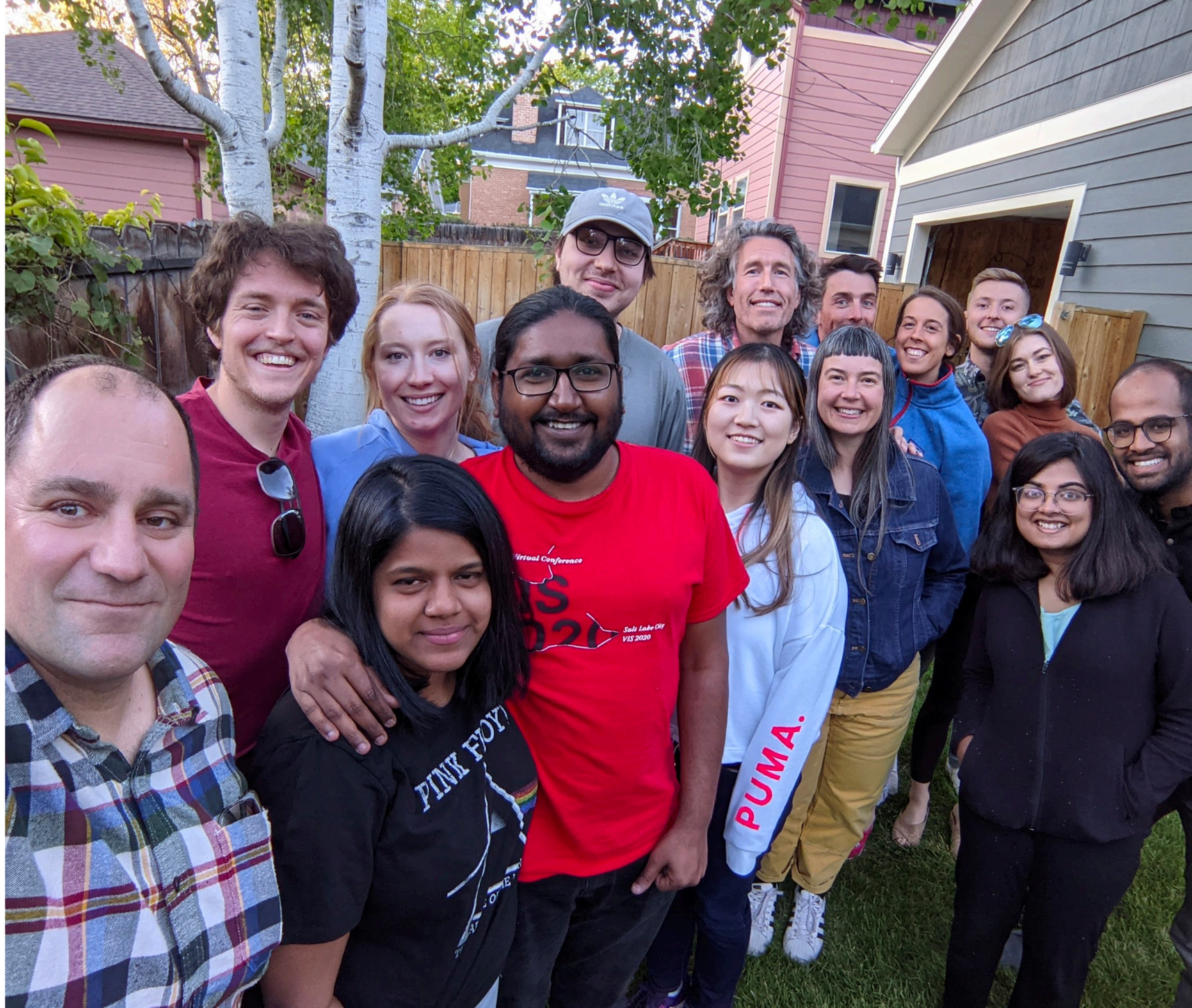
Opportunities for Understanding Semantics of User Interactions

visualization design lab

SCI
www.sci.utah.edu

visualization
design lab

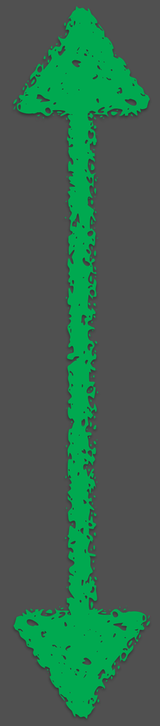
http://vdl.sci.utah.edu/



PREMISE: UNDERSTAND WHAT ANALYST IS TRYING TO DO / SEE / UNDERSTAND IN A VA SYSTEM

CAPTURING INFORMATION

Semantically Poor



Semantically Rich

Simple Logs
(Clicks, Keystrokes, Buttons)

Functional Logs & Provenance
(Functions, Operations)

Pattern Based Intents
(Types of Patterns)

Higher Level Intents
(Context, Thought Process)

Rational and Explanations
(Explanatory Text, Notebooks, Methods Sections)

Human Context (Gaze, Biomedical signals, Attention)

CAPTURING INFORMATION

Machine Readable

Difficult to Extract

Rich Semantics != Useful for ML

Pattern Based Intents

Functional Logs & Provenance

Simple Logs

Higher Level Intents

Rational and Explanations

WAYS OF CAPTURING INFORMATION

SIMPLE LOGS

SIMPLE LOG APPROACHES

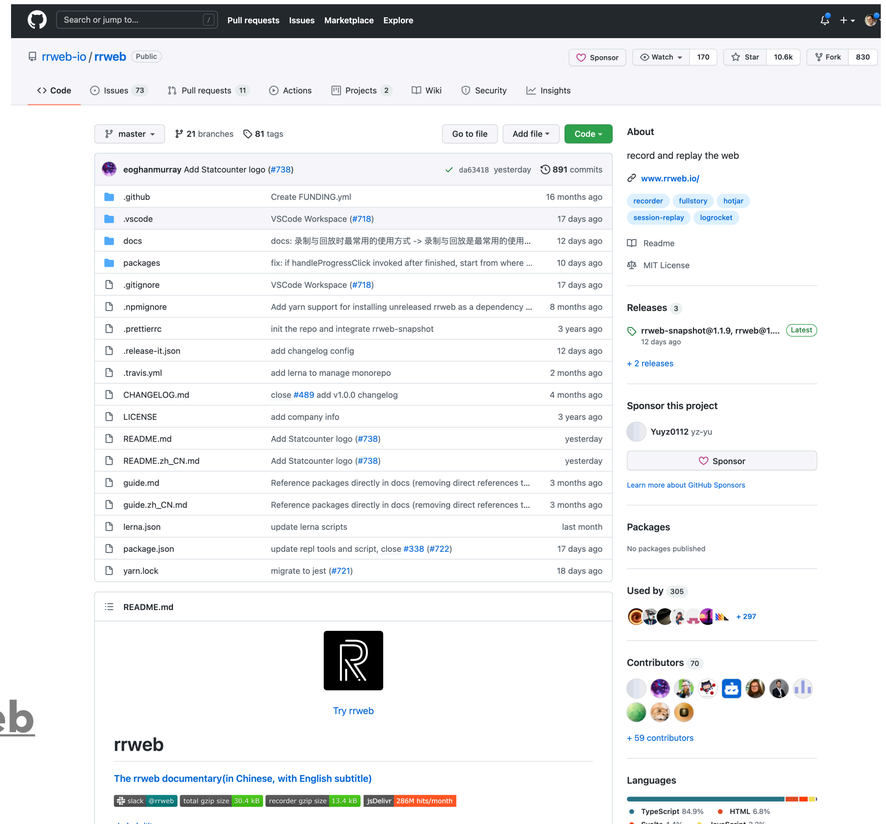
Input (clicks, buttons, etc.)

Events (website loaded)

Log messages

Capturing the state (DOM)

https://github.com/rrweb-io/rrweb



SIMPLE LOGS

Pro:

- Easy to do
- Generic solutions feasible

Con:

- Analysis is difficult
- Not much about the why

WAYS OF CAPTURING INFORMATION

FUNCTIONAL LOGS & PROVENANCE

FUNCTIONAL LOGS & PROVENANCE TRACKING

Not just a record of selected things, but full provenance

- **Can be used for undo/redo, reproducibility, storytelling, collaboration, and post-hoc analysis**

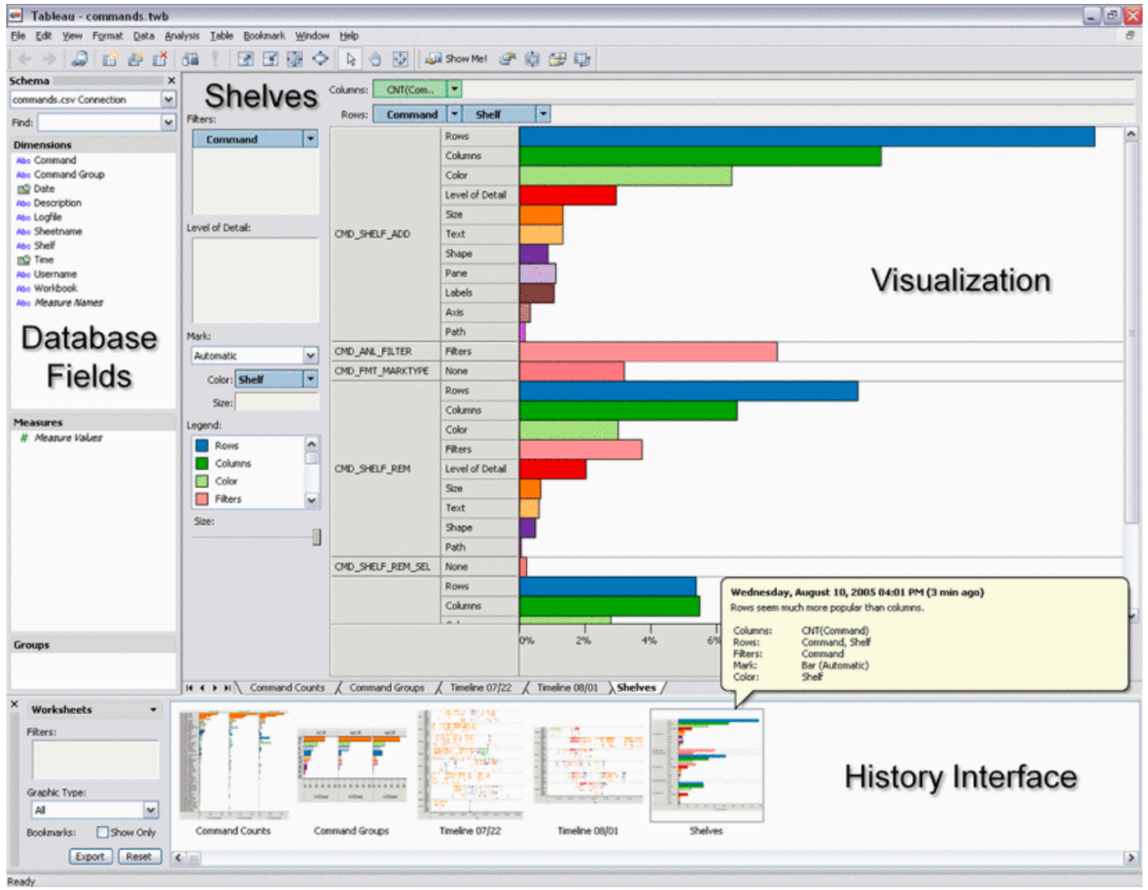
Opportunity to capture semantics:

- **Not “button X” was clicked, but “Filter applied to dataset Y”**

K. Xu, A. Ottley, C. Walchshofer, M. Streit, R. Chang, and J. Wenskovitch. **Survey on the Analysis of User Interactions and Visualization Provenance**. Computer Graphics Forum, 2020.

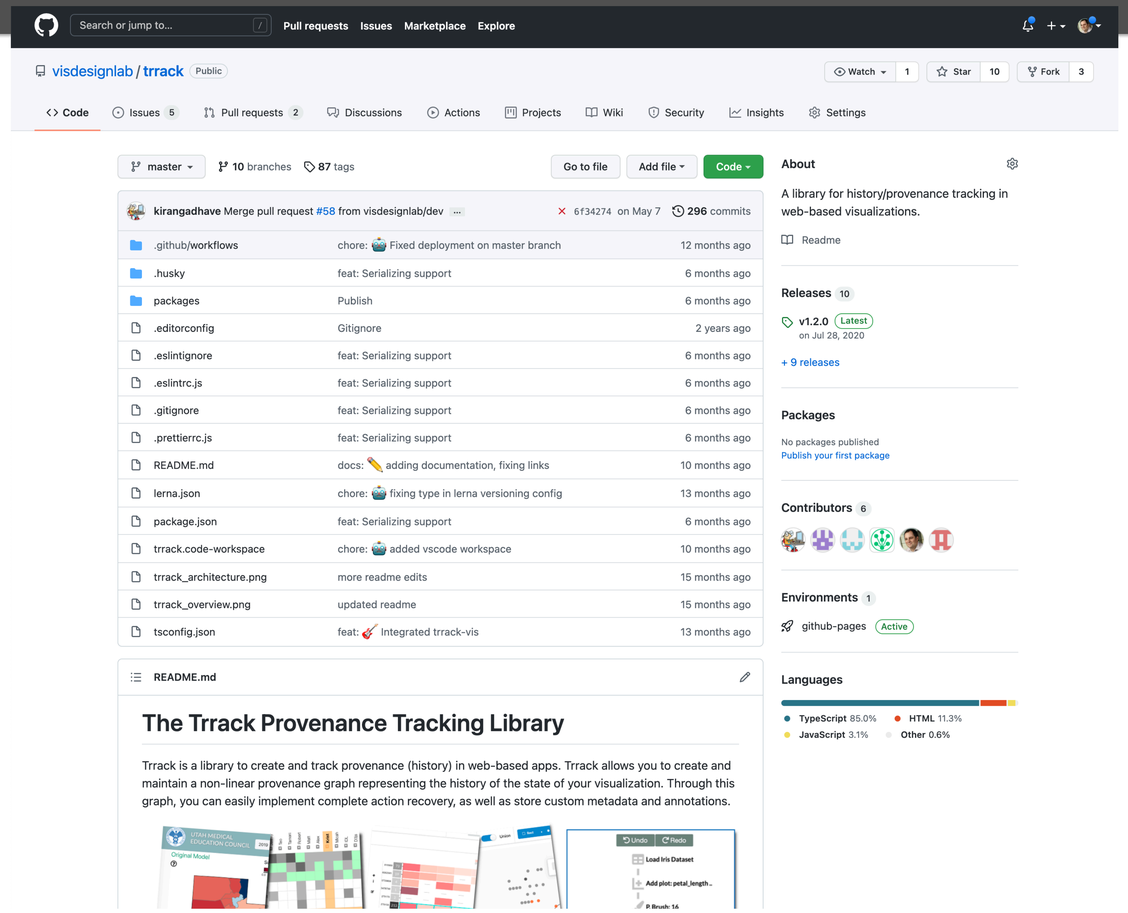
E.Ragan, A.Endert, J.Sanyal, andJ.Chen. **Characterizing Provenance in Visualization and Data Analysis: An Organizational Framework of Provenance Types and Purposes**. IEEE Transactions on Visualization and Computer Graphics (VAST ’15), 22(1):31–40, 2016.

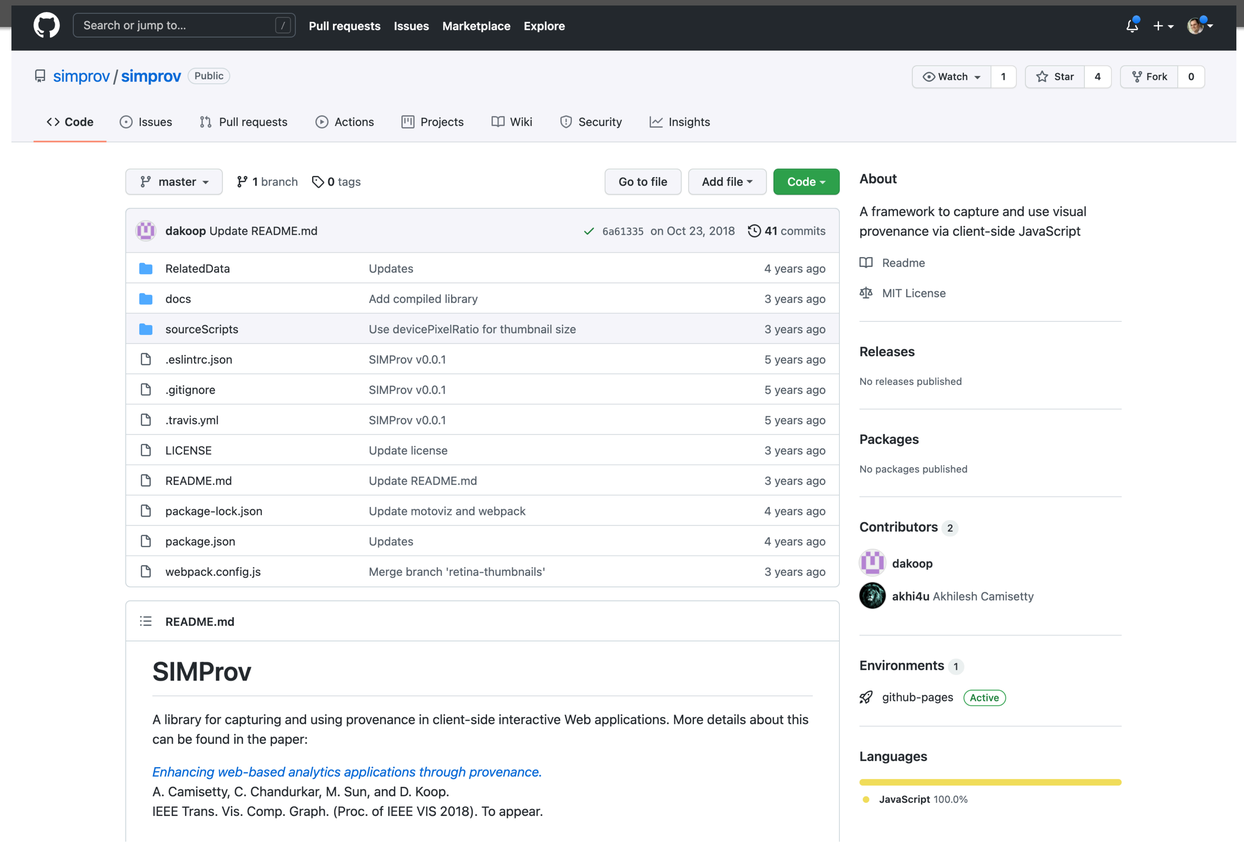
S.Gratzl, A.Lex, N.Gehlenborg, N.Cosgrove, and M.Streit. **From Visual Exploration to Storytelling and Back Again.** Computer Graphics Forum, 35(3):491–500, 2016.



J.Heer, J.Mackinlay, C.Stolte, and M.Agrawala. **GraphicalHistories for Visualization: Supporting Analysis, Communication, and Evaluation.** IEEE Transactions on Visualization and Computer Graphics (InfoVis '08), 14(6):1189–1196, 2008.

HOW TO CAPTURE PROVENANCE?





https://github.com/visdesignlab/trrack

Zach Cutler, Kiran Gadhave, Alexander Lex. Trrack: **A Library for Provenance-Tracking in Web-Based Visualizations.** IEEE Visualization Conference (VIS), 116-120, doi:10.1109/VIS47514.2020.00030, 2020.

https://github.com/simprov/simprov

A.Camisetty, C.Chandurkar, M.Sun, and D.Koop. **Enhancing Web- based Analytics Applications through Provenance.** IEEE Transactions on Visualization and Computer Graphics, 25(1):131–141, 2019. doi: 10.1109/TVCG.2018.2865039

FUNCTIONAL LOGS

Pro:

Lots of good things happen if you do this.

Undo/Redo, Reproducibility, Sharing/Collaboration, Storytelling, Meaningful Labels for Actions, etc.

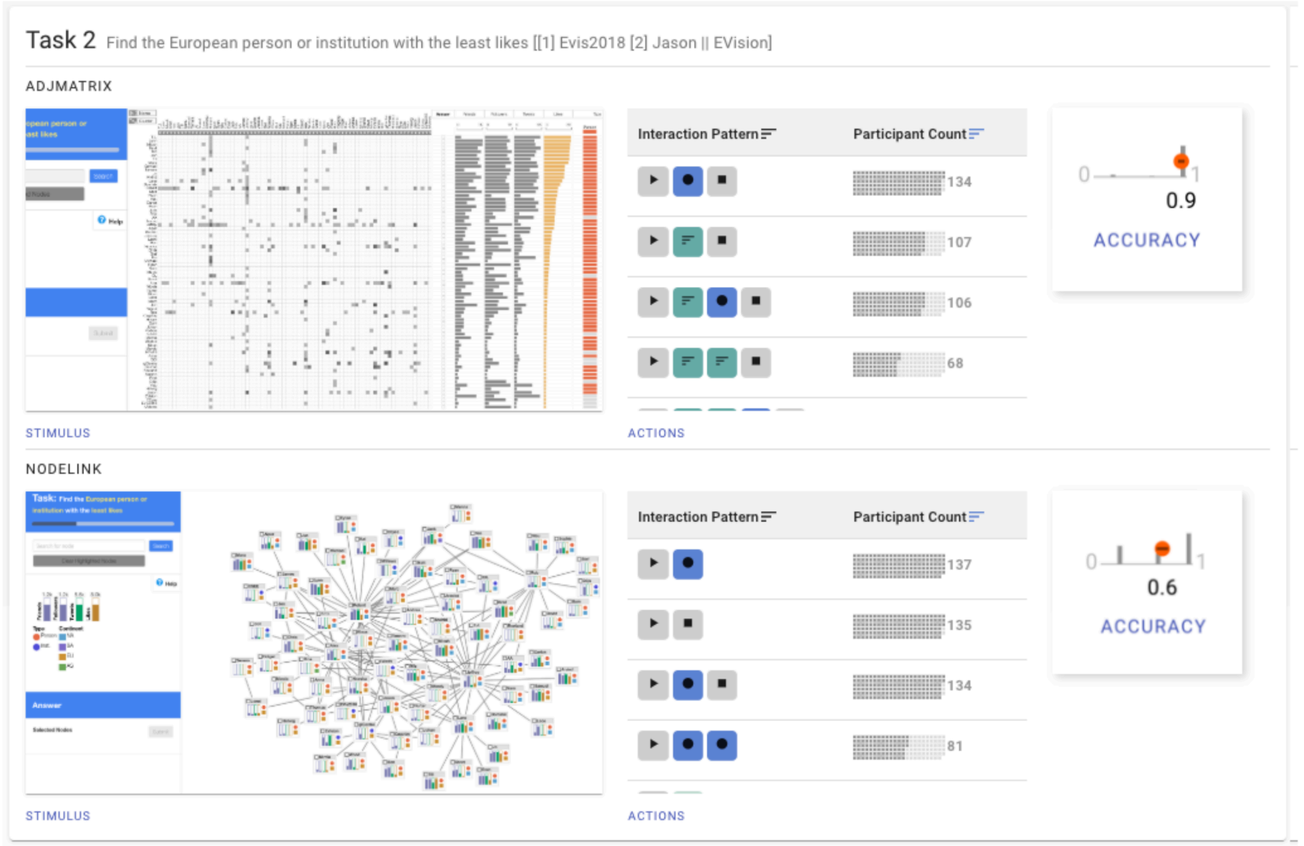
Con:

Increased development effort

But mitigated by libraries

Still some UI challenges

OPPORTUNITY: DISCOVERING USAGE PATTERNS



Carolina Nobre, Dylan Wootton, Zach Cutler, Lane Harrison, Hanspeter Pfister, Alexander Lex. **reVISit: Looking Under the Hood of Interactive Visualization Studies.** SIGCHI Conference on Human Factors in Computing Systems (CHI), 1-13, doi:10.1145/3411764.3445382, 2021.

OPPORTUNITIES

What are key states?

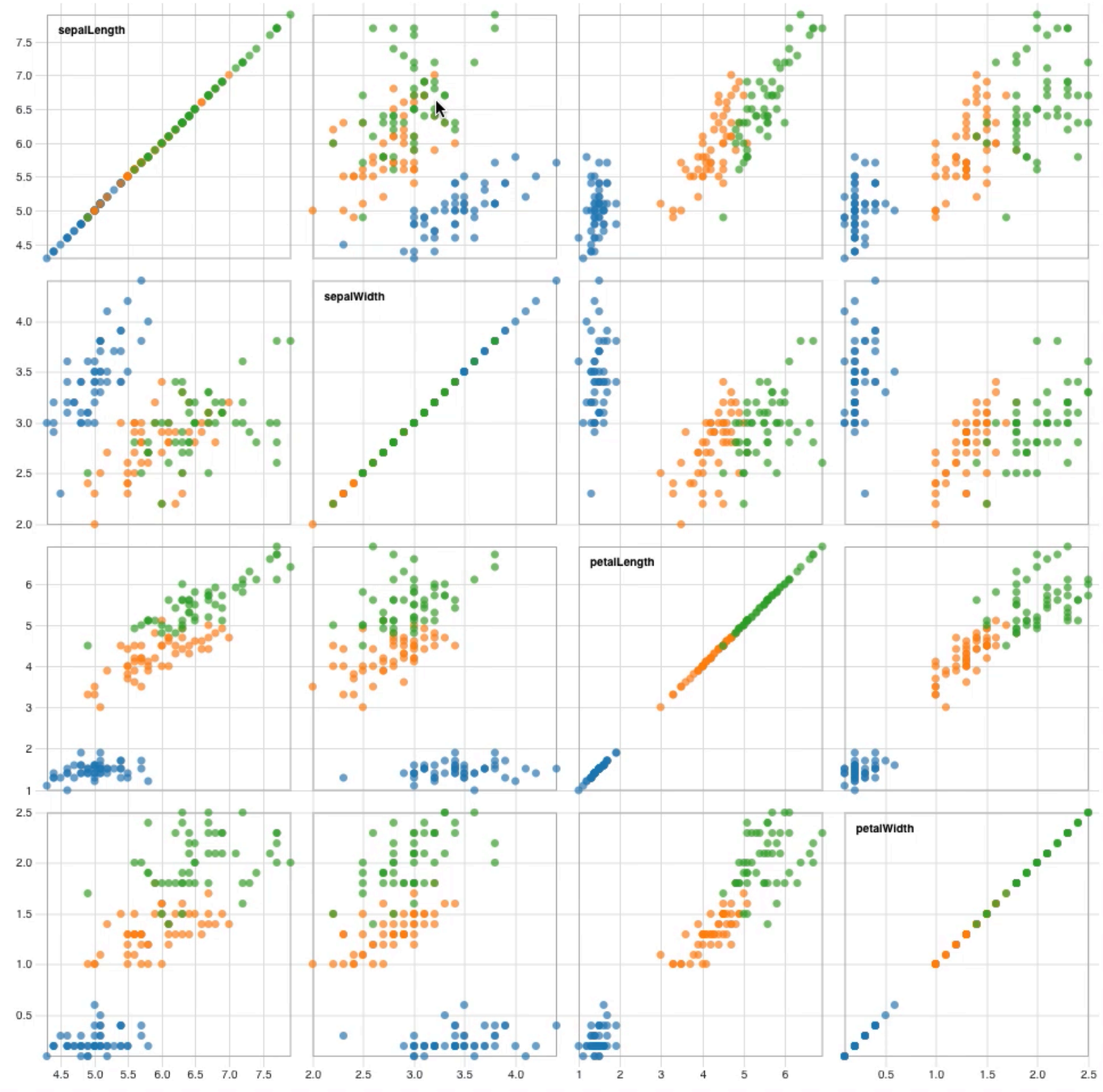
How can I find states?

Use provenance for guidance & tutorials

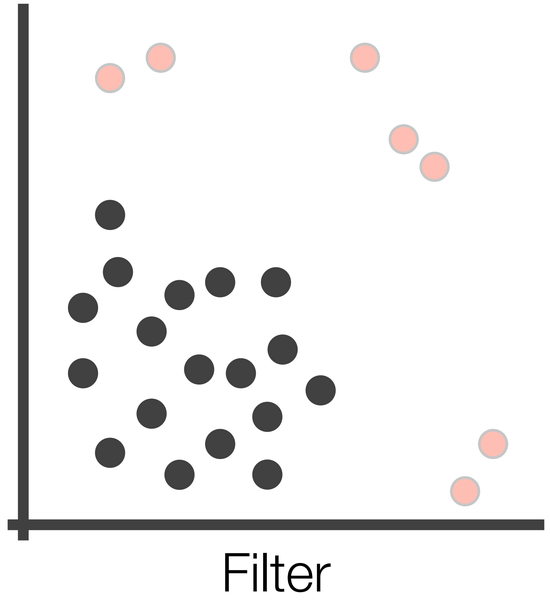
CAPTURING PATTERN BASED INTENTS

Kiran Gadhave, Jochen Görtler, Zach Cutler, Carolina Nobre, Oliver Deussen, Miriah Meyer, Jeff Phillips, Alexander Lex. **Predicting Intent Behind Selections in Scatterplot Visualizations.** Information Visualization, 2021

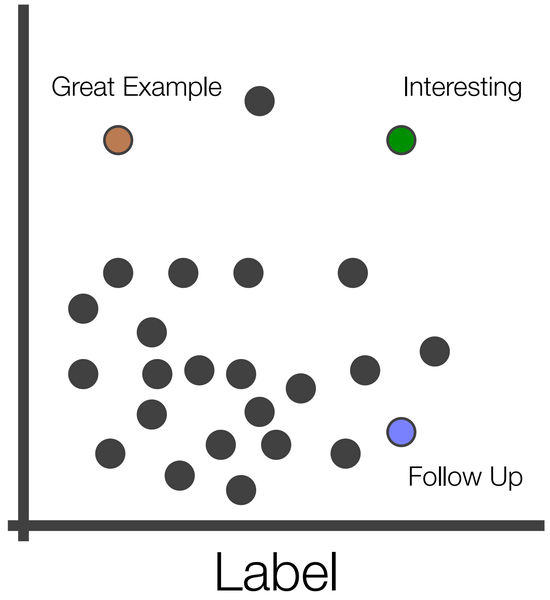
SELECTIONS?



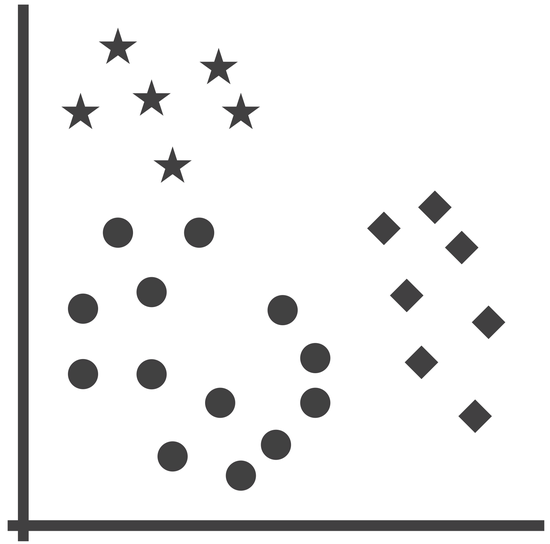
FROM SELECTIONS TO ADVANCED OPERATIONS

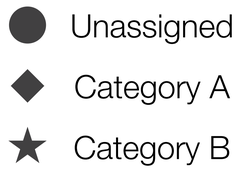


Filter

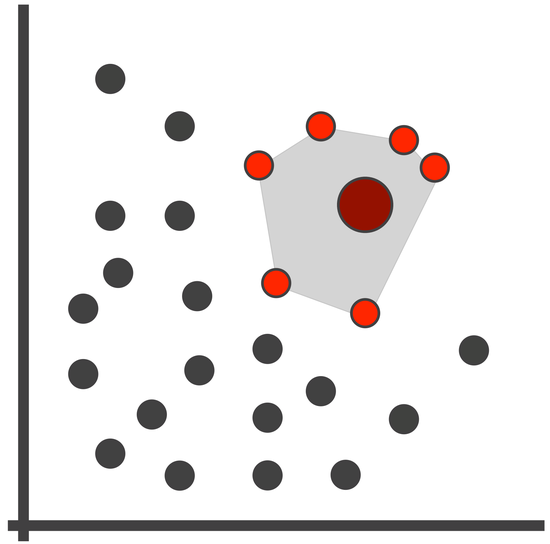


Label





Categorize



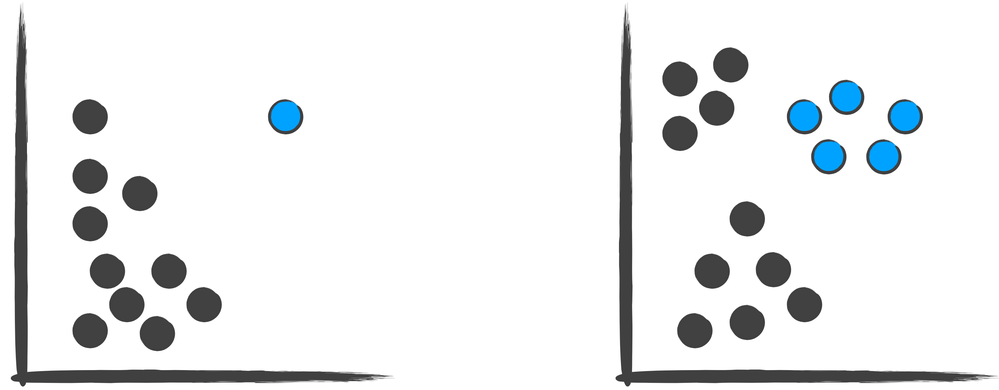
Aggregate

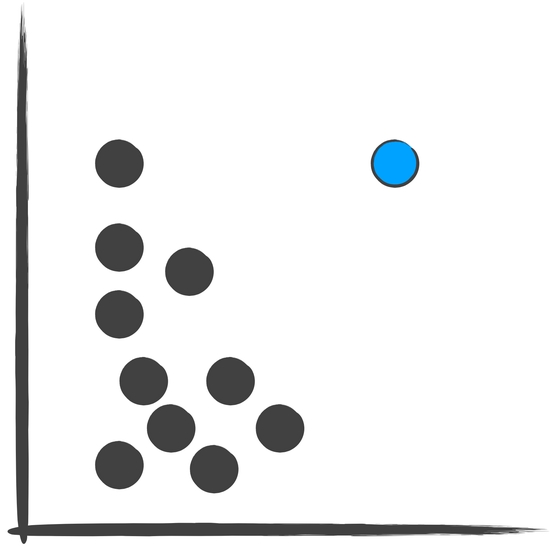
WHAT IS INTENT WHEN SELECTING?

Intent is the user's reason for performing a brush with a visualization.

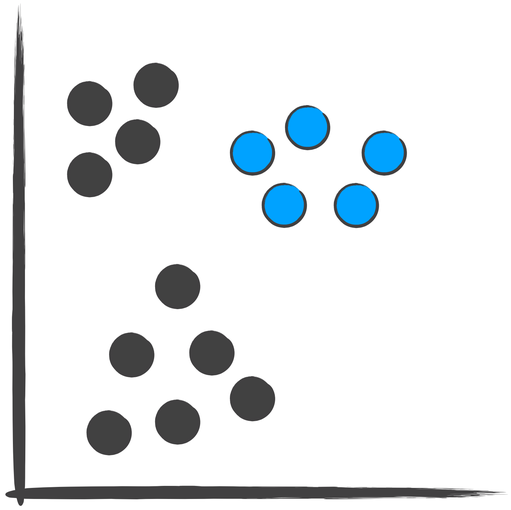
Domain Specific Intent: Capture through Annotation

Pattern-Based Intent: Capture Automatically

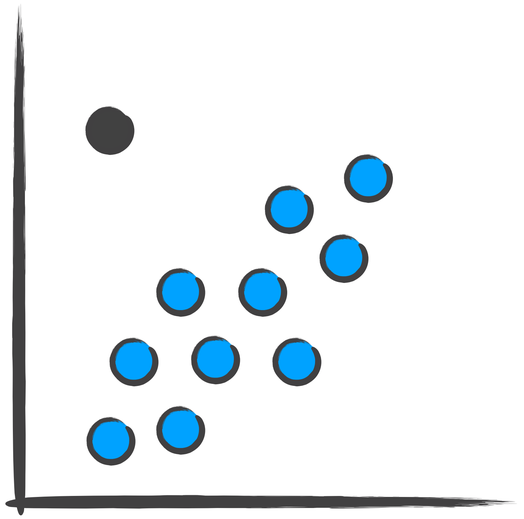




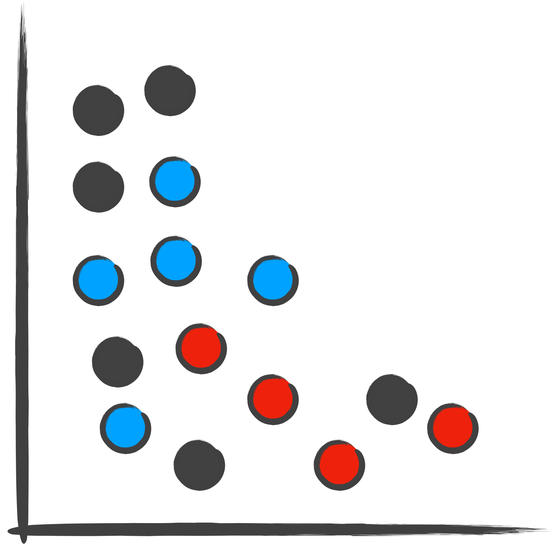
Outlier



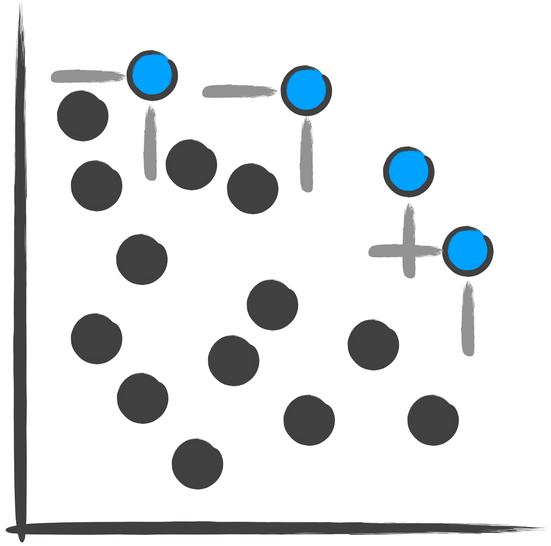
Clusters



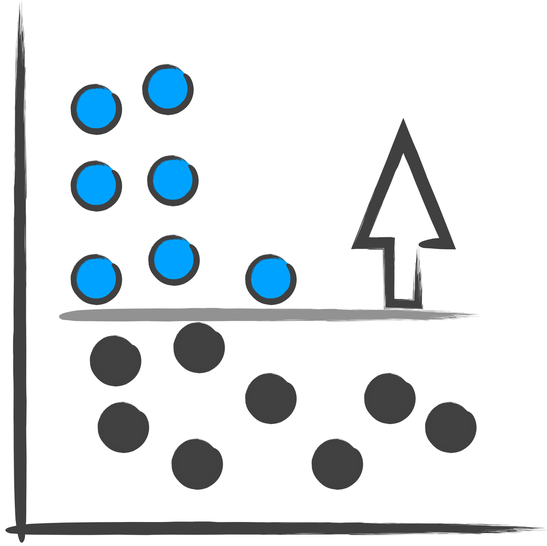
Correlation



Categories



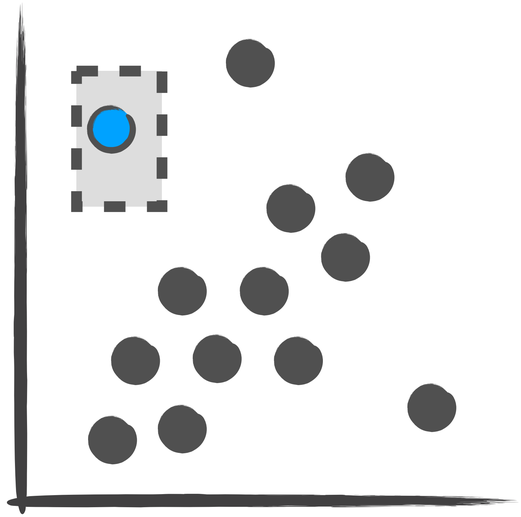
Multivariate Optimization



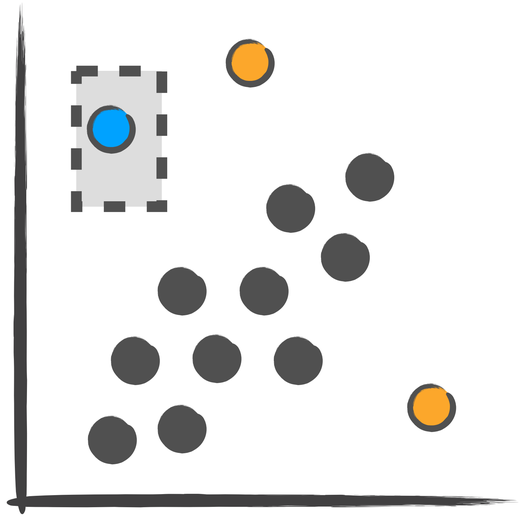
Ranges

WHY DO WE CARE?

Simplify complex selections



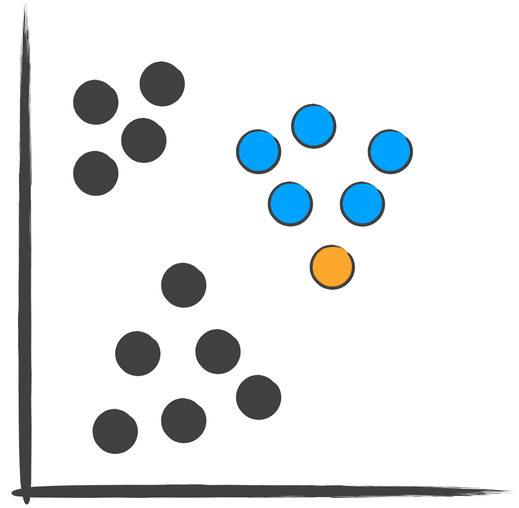
Selection



Outliers?

WHY DO WE CARE?
R & R

ID Based Selection:
Selected Elements: 7, 9, 13, 18, 22

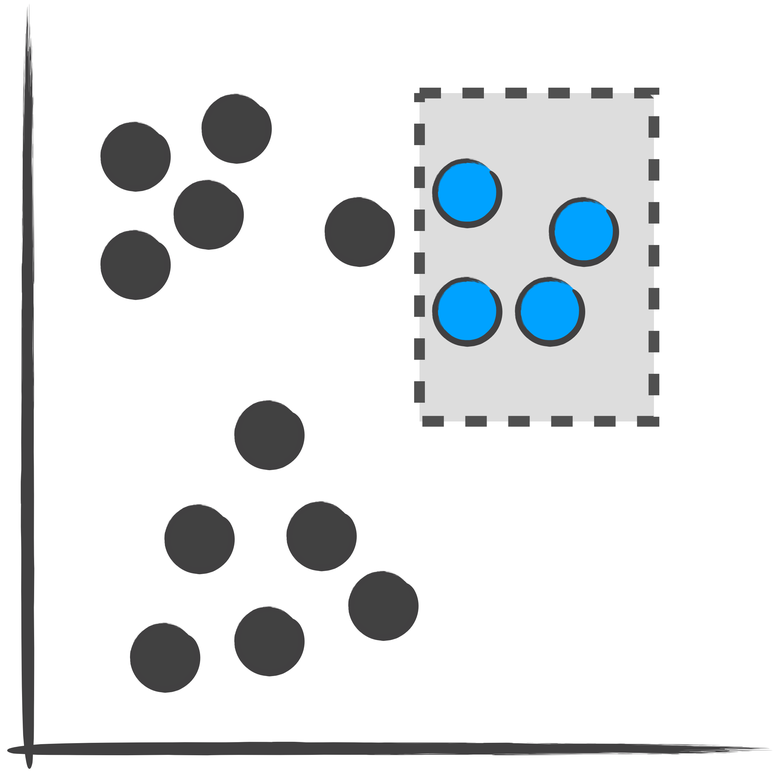


Semantic Selection:
Elements in K-Means cluster centered at [2, 3]

Meaningful, higher level concept:
improves reproducibility

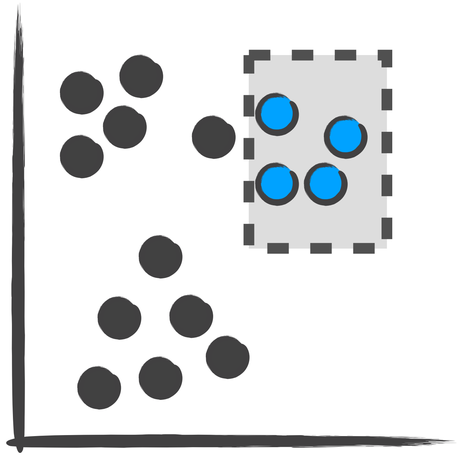
Robust to changes and updates in dataset:
enables re-usability

HOW DO WE INFER INTENT?

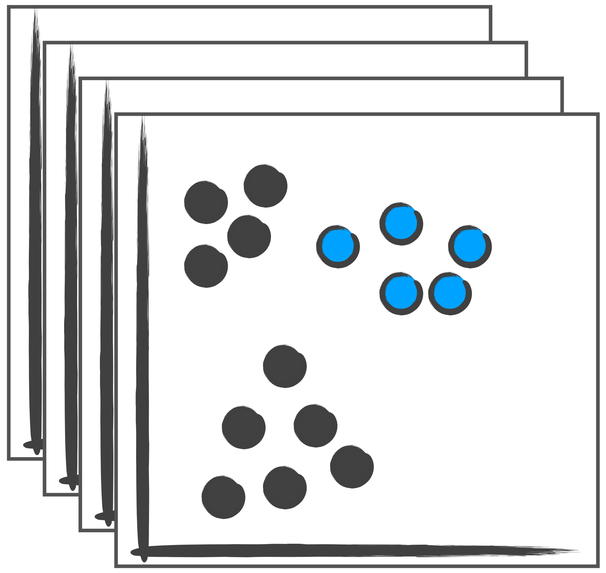


Selection

HOW DO WE INFER INTENT?



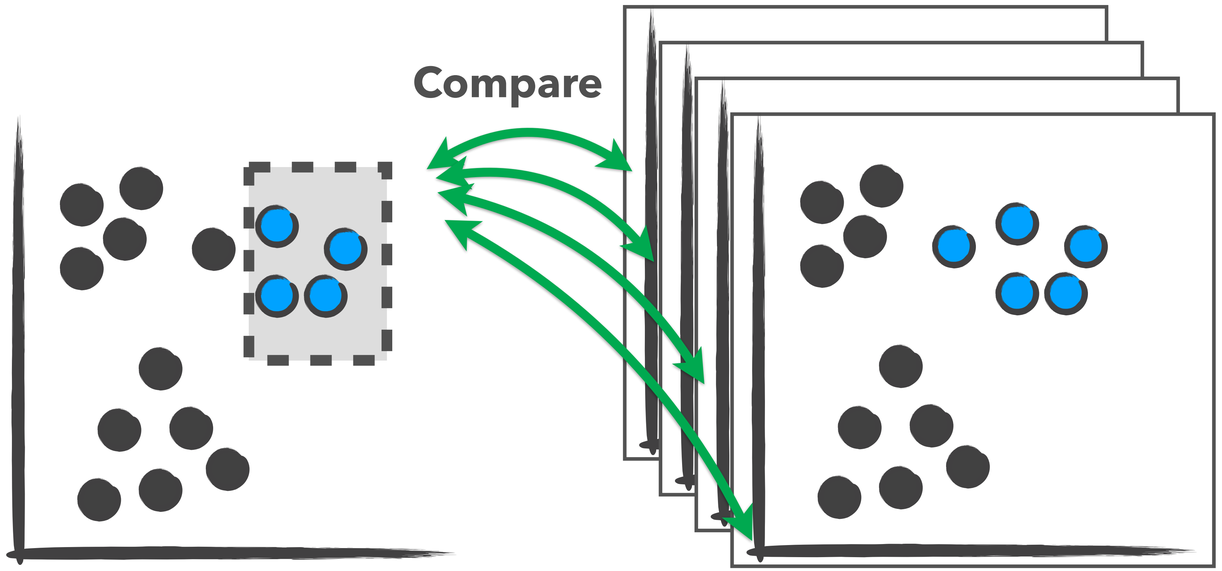
Selection



Predictions

K-Means
DBScan
Regression
Outlier Detection
Skyline
Decision Trees / Ranges
Categories

HOW DO WE INFER INTENT?

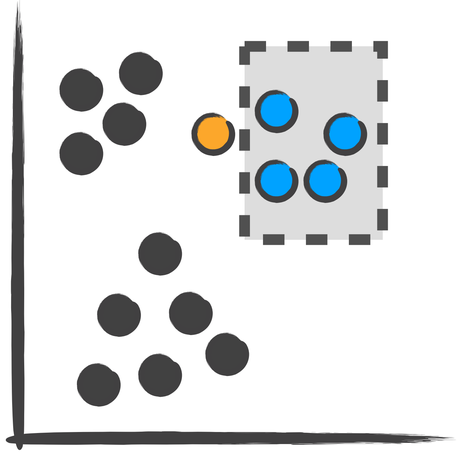


Selection

Predictions

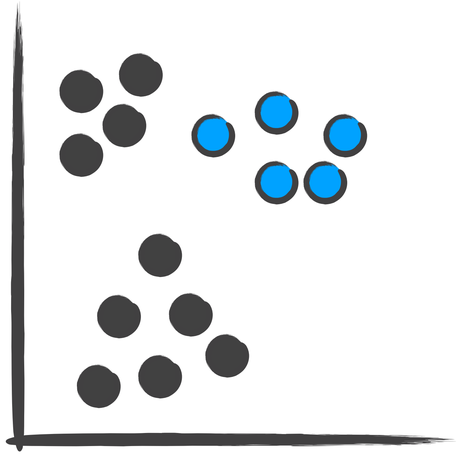
K-Means
DBScan
Regression
Outlier Detection
Skyline
Decision Trees / Ranges
Categories

1. Range
2. Cluster
3. Outlier



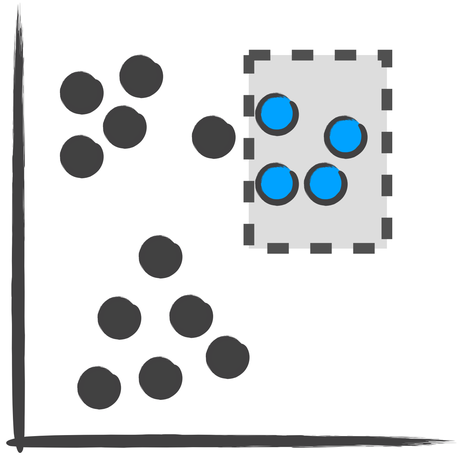
Ranking

Jaccard Distance
Naive Bayes
Classifier
Heuristic
Measures

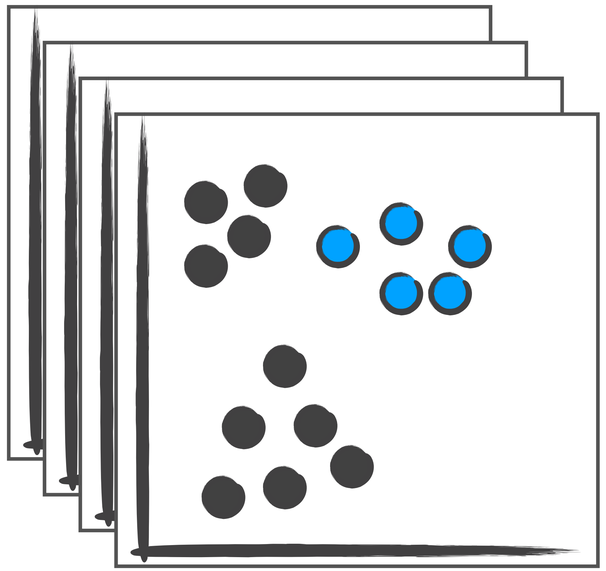


$$J(S,C) = \frac{|S \cap C|}{|S \cup C|}$$

HOW DO WE INFER INTENT?



Selection



Predictions

K-Means
DBScan
Regression
Outlier Detection
Skyline
Decision Trees / Ranges
Categories

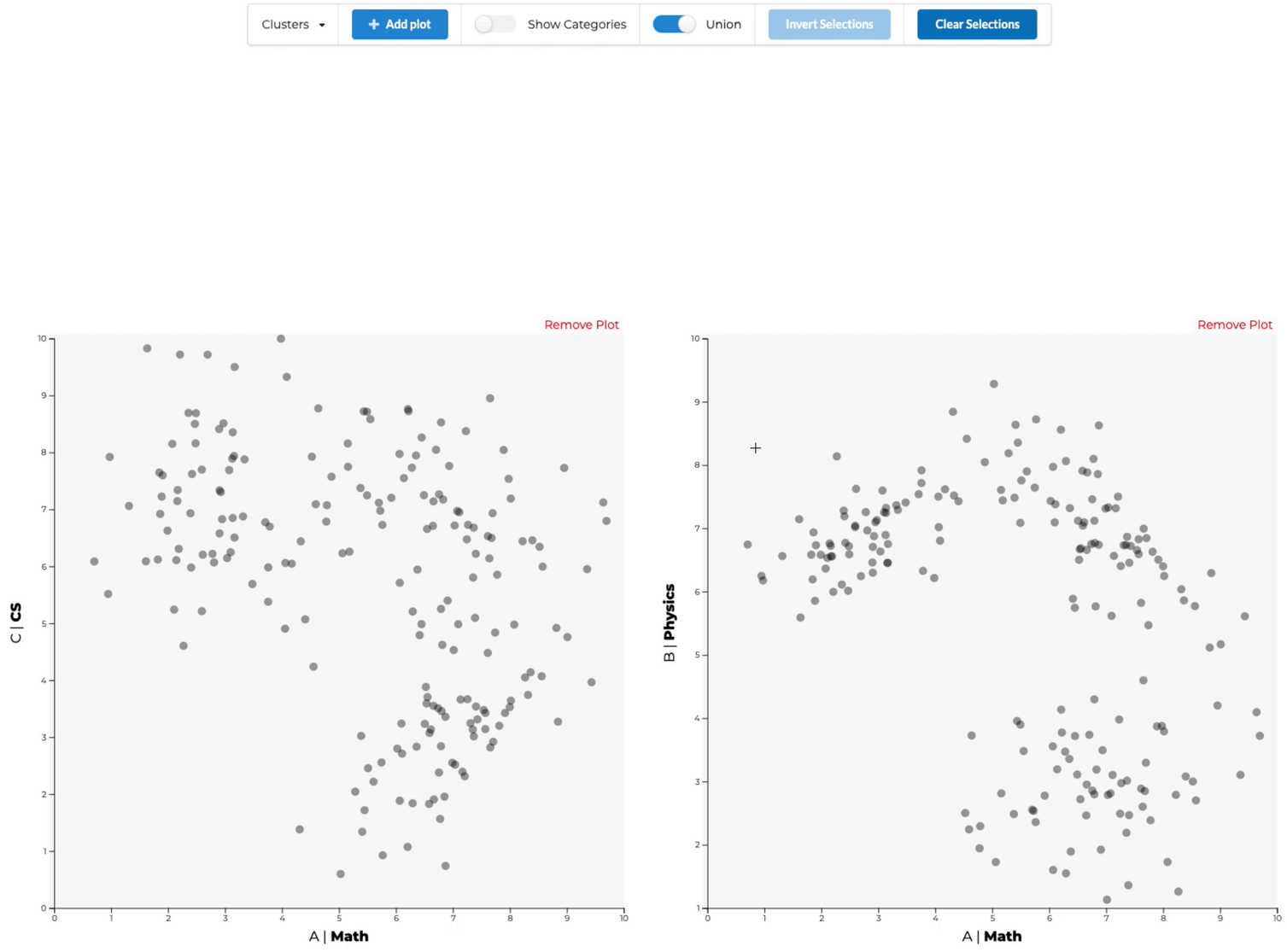
1. Range
2. Cluster
3. Outlier

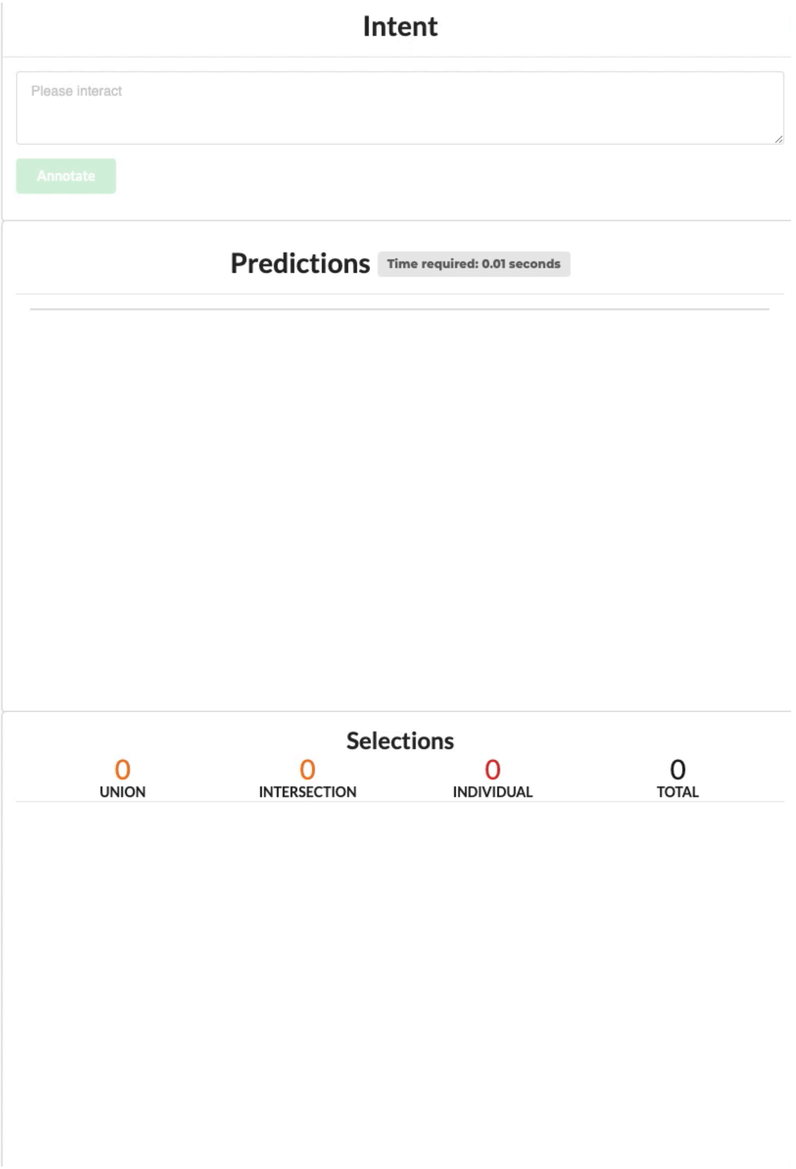
I think this cluster…

Ranking

Jaccard Distance
Naive Bayes
Classifier
Heuristic
Measures

Confirming Intent & Annotation





Visualization and Selection

Annotation of Intent and Predictions

http://vdl.sci.utah.edu/predicting-intent/

PATTERN BASED INTENTS

Pro:

- Capturing more semantics!
- Better understanding of intent.
- Helps manage deluge of provenance data.

Con:

- Technically challenging
- Needs to be adapted to different vis techniques and data types

OPPORTUNITIES

Understand what patterns users are interested in

Mixed initiative: auto complete and correction

Pattern based intents for other Vis techniques and data types

Combine with workflows (next)

WAYS OF CAPTURING INFORMATION

CAPTURING HIGHER LEVEL INTENTS

HIGHER LEVEL INTENTS

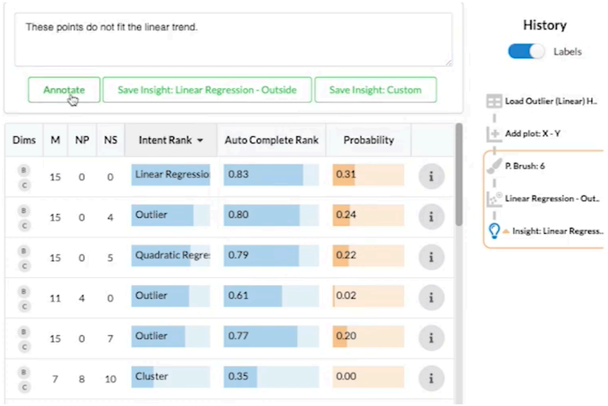
In a VA System

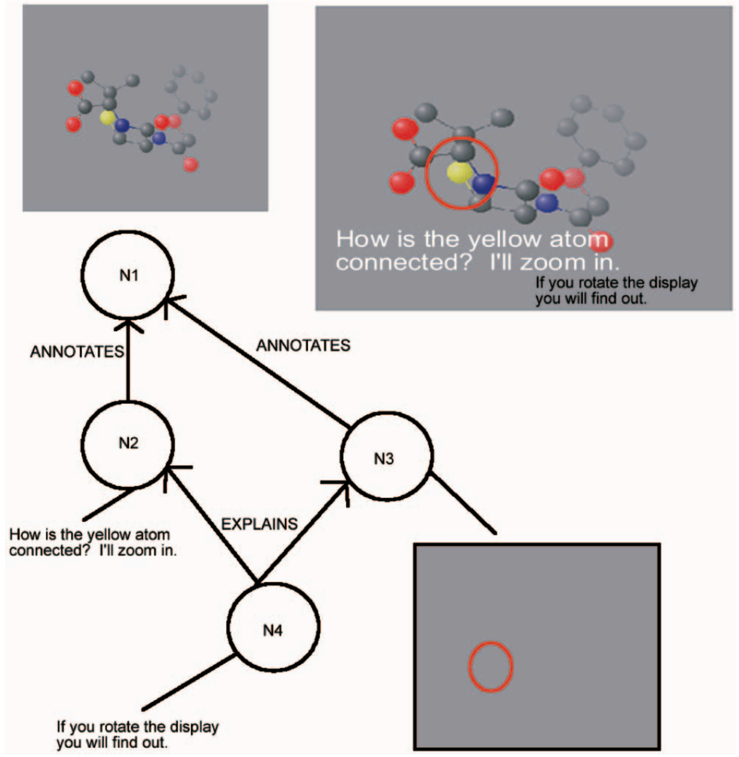
Annotations

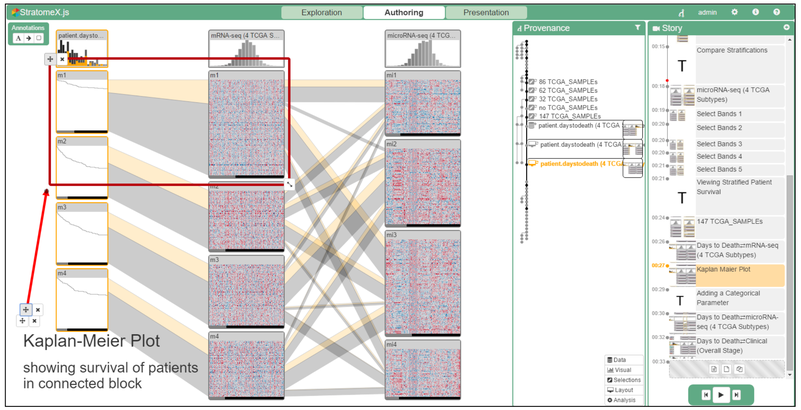
Workflows

All pose a burden on the user

ANNOTATIONS







Groth, D. P., & Streefkerk, K. (2006). Provenance and annotation for visual exploration systems. IEEE transactions on visualization and computer graphics, 12(6), 1500-1510.

S.Gratzl, A.Lex, N.Gehlenborg, N.Cosgrove, and M.Streit. **From Visual Exploration to Storytelling and Back Again.** Computer Graphics Forum, 35(3):491–500, 2016.

WORKFLOWS

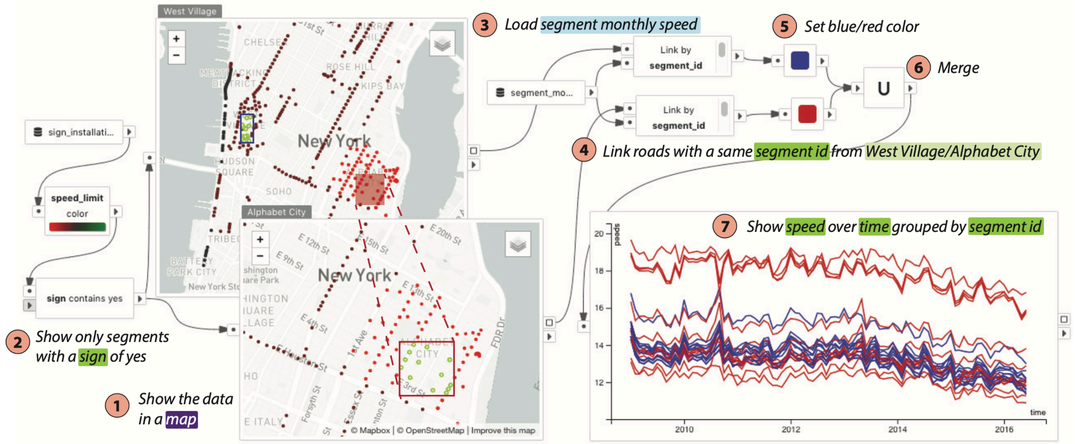
A sequence of actions with a purpose

"For two given dimensions, filter out the outliers"

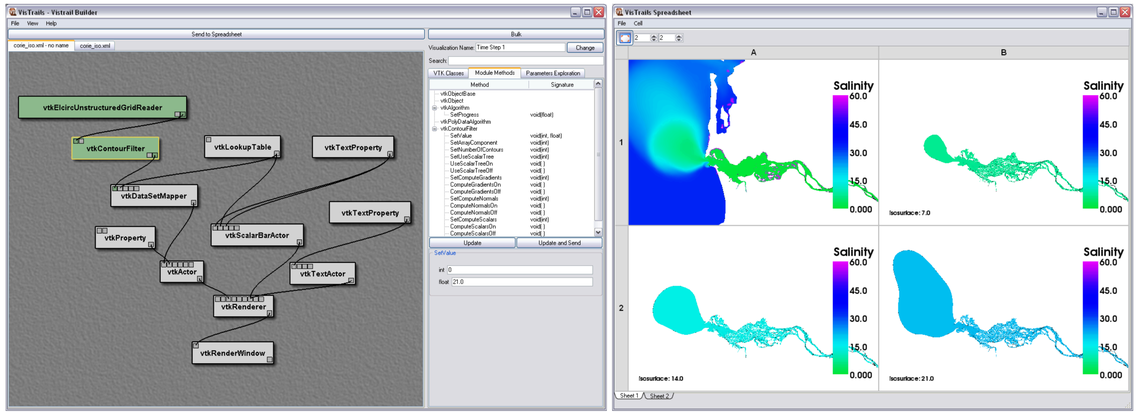
Explicit Workflow Modeling

Post-Hoc Workflow Extraction

EXPLICIT WORKFLOWS

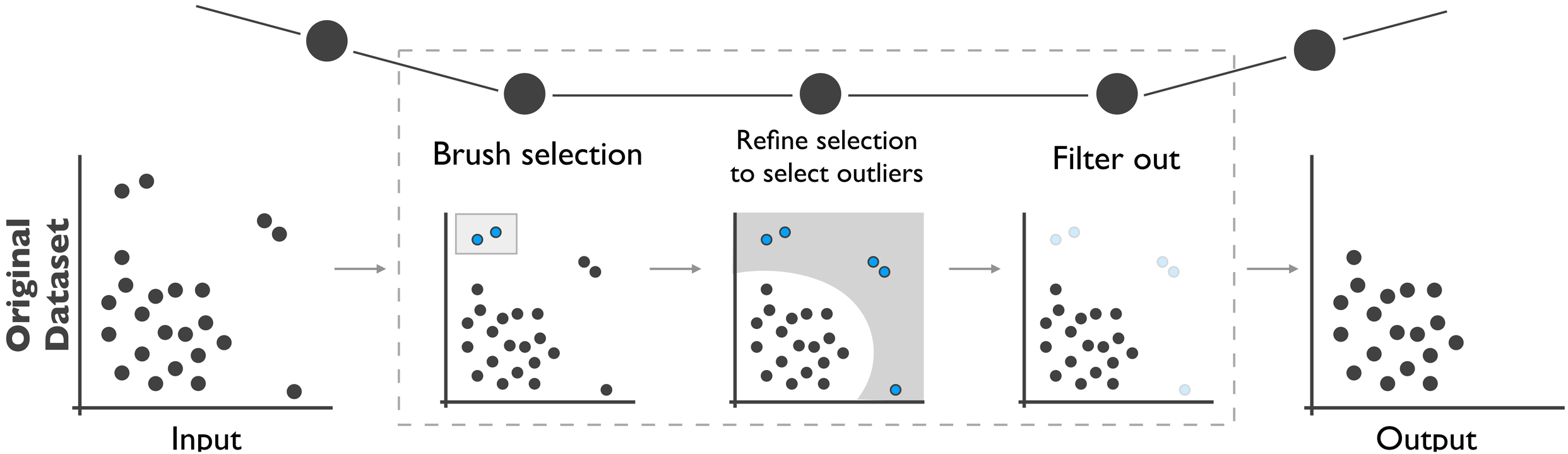


B. Yu and C. T. Silva, "**FlowSense: A Natural Language Interface for Visual Data Exploration within a Dataflow System**," in IEEE TVCG, vol. 26, no. 1, pp. 1-11, Jan. 2020, doi: 10.1109/TVCG.2019.2934668.



VisTrails: visualization meets data management
SP Callahan, J Freire, E Santos, CE Scheidegger, CT Silva, HT Vo
Proceedings of the 2006 ACM SIGMOD

POST-HOC SEMANTIC WORKFLOWS



Robust “Filter Outliers” Workflow

USING WORKFLOW IN A COMPUTATIONAL NOTEBOOK

```
# Installing the reapply-workflows adds a module called backend
# This module exposes the Reapply class which initializes the library

from backend import Reapply
```

```
# Here we load the reapply_workflows library.
r = Reapply()

# We add a workflow from our workflow database.
workflow = r.load_workflow("workflow16178086816 20")
```

```
# Print the workflow name
print("Workflow: ", workflow.name, "\n")

# Description of the workflow and the operations in it
workflow.describe
```

```
Workflow:  Deleting Cluster

| Root
 +--| Add Plot
     +--| Added brush to: X-Y
         +--| Cluster Selection
             +--| Filter: Out
```

```
# Prints the reapply results for all interactions, along with review status.
```

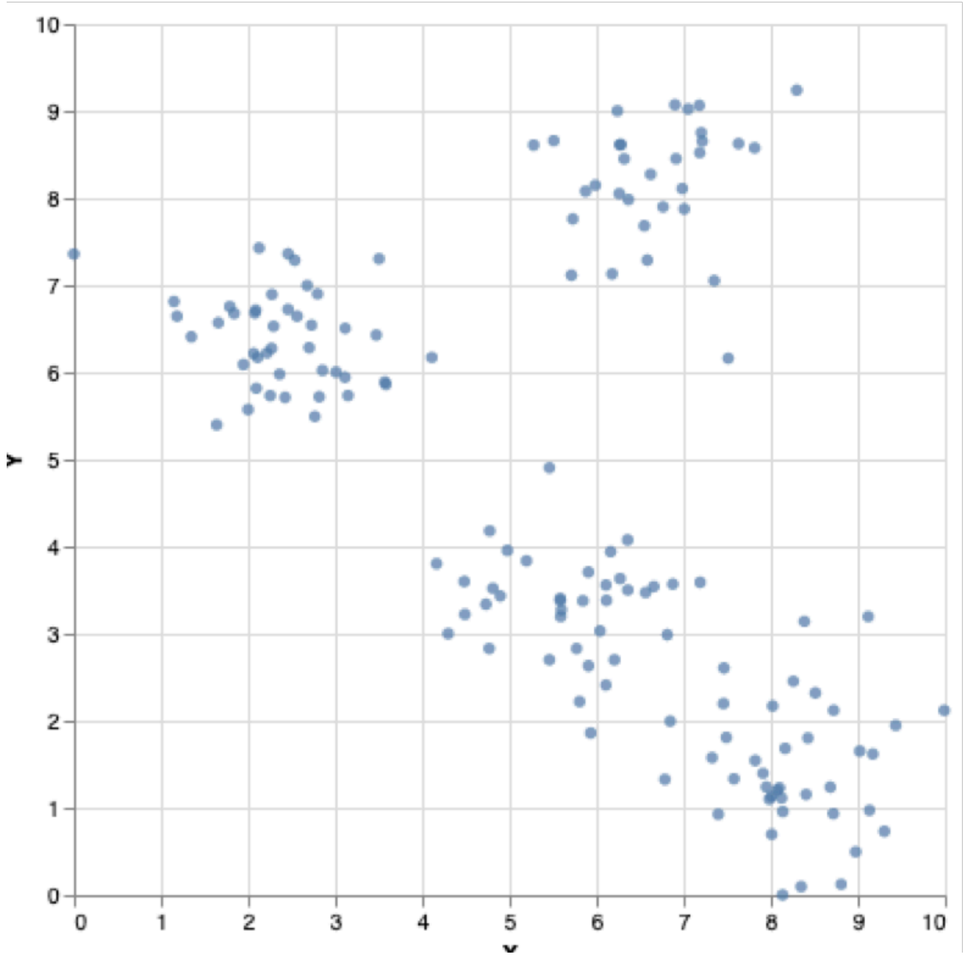
```
# Apply the workflow to target dataset.
# apply function requires the target dataset
# and the label column as arguments.
res = workflow.apply(target, "Label")

# Results is an array of datasets for each interaction
# we grab the final one.
result_dataset = res.results[-1]['data']
result_dataset
```

```
This workflow has not been reviewed for all interactions.
Please go to following url: https://reapply-workflows.gith
```

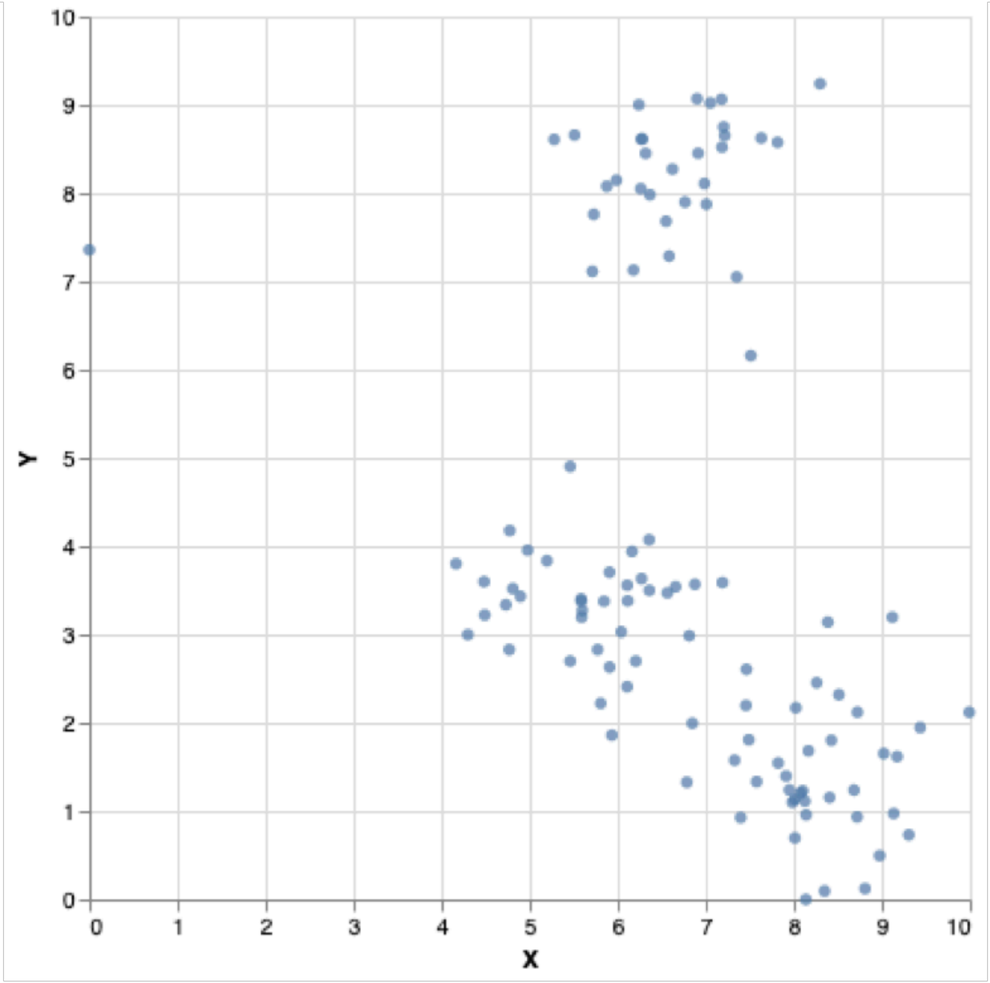
	Label	X	Y
3	P52	6.58351	7.28796
5	P171	4.77421	4.17980
8	P199	8.34966	0.09550
9	P183	8.42670	1.80299
10	P61	4.29760	2.99981
...	...	...	...
141	P138	7.35179	7.05215
142	P46	6.62171	8.27311

BEFORE AND AFTER

150 rows × 3 columns



108 rows × 3 columns



OPPORTUNITIES

Improving Annotations:

- **Think aloud protocol**
- **Sketching on screens**

Improving Workflows:

- **Automatically extracting workflows**
- **Learning workflows from many users**
- **Understanding relationships between semantics (annotation) and sequence of actions**
- **Adapting workflows to new data**

….

WAYS OF CAPTURING
INFORMATION

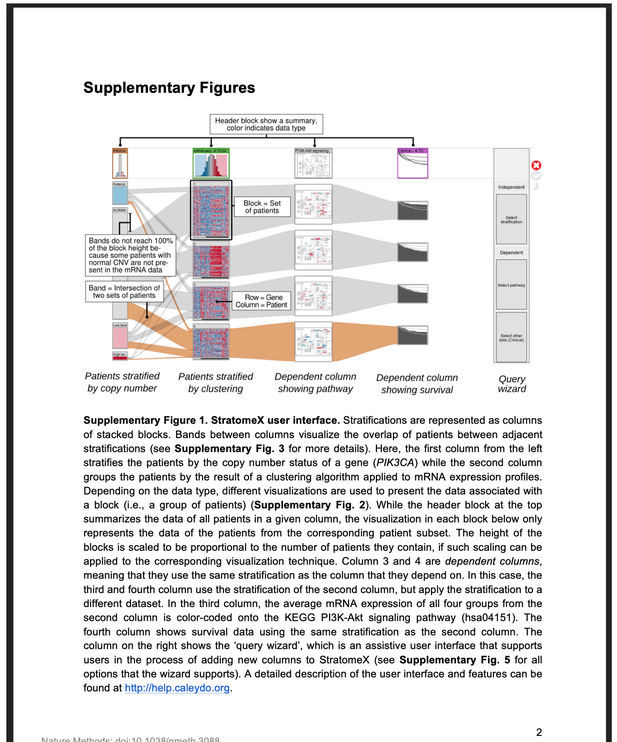
ANALYZING
RATIONAL AND EXPLANATIONS

RATIONAL AND EXPLANATIONS

Insights from VA are reported in papers/presentations/notebooks/ blog posts

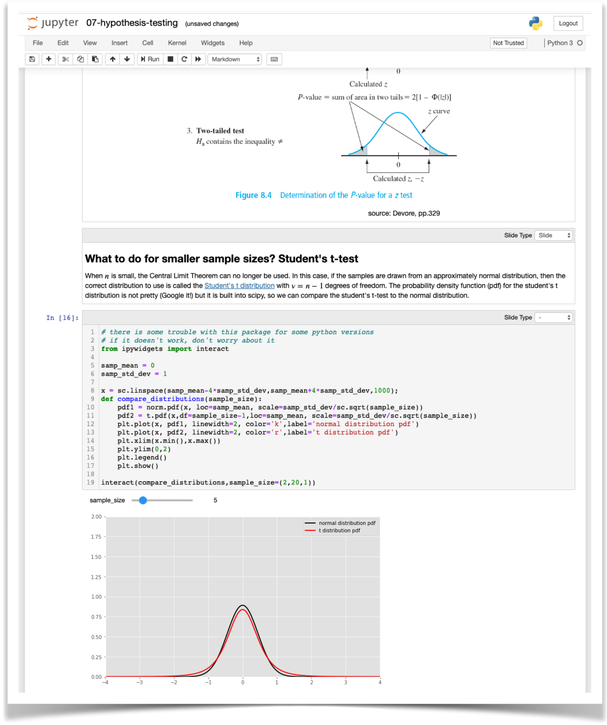
Contain high level information

Why? How? What?



Supplementary Figures

Supplementary Figure 1. StratomeX user interface. Stratifications are represented as columns of stacked blocks. Bands between columns visualize the overlap of patients between adjacent stratifications (see **Supplementary Fig. 3** for more details). Here, the first column from the left stratifies the patients by the copy number status of a gene (*PIK3CA*) while the second column groups the patients by the result of a clustering algorithm applied to mRNA expression profiles. Depending on the data type, different visualizations are used to present the data associated with a block (i.e., a group of patients) (**Supplementary Fig. 2**). While the header block at the top summarizes the data of all patients in a given column, the visualization in each block below only represents the data of the patients from the corresponding patient subset. The height of the blocks is scaled to be proportional to the number of patients they contain, if such scaling can be applied to the corresponding visualization technique. Column 3 and 4 are *dependent columns*, meaning that they use the same stratification as the column that they depend on. In this case, the third and fourth column use the stratification of the second column, but apply the stratification to a different dataset. In the third column, the average mRNA expression of all four groups from the second column is color-coded onto the KEGG PI3K-Akt signaling pathway (hsa04151). The fourth column shows survival data using the same stratification as the second column. The column on the right shows the 'query wizard', which is an assistive user interface that supports users in the process of adding new columns to StratomeX (see **Supplementary Fig. 5** for all options that the wizard supports). A detailed description of the user interface and features can be found at http://help.caleydo.org.



OPPORTUNITIES

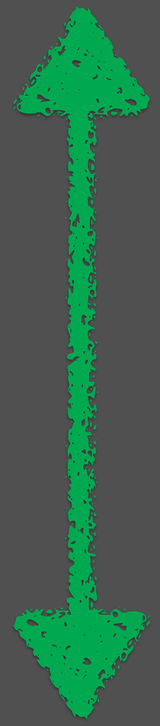
Connect text with visualizations (interactive figures)

Generate explanatory text from interactive visualizations

Understand relationships between text and visualizations

CAPTURING INFORMATION

Semantically Poor



Semantically Rich

Simple Logs

(Clicks, Keystrokes, Buttons)

Functional Logs & Provenance

(Functions, Operations)

Pattern Based Intents

(Types of Patterns)

Higher Level Intents

(Context, Thought Process)

Rational and Explanations

(Explanatory Text, Notebooks, Methods Sections)

Alexander Lex
@alexander_lex
http://alexander-lex.net

Thanks to: Kiran Gadhave, Zach Cutler, Carolina Nobre, Jen Rogers, Haihan Lin, Dylan Wootton, Jochen Görtler, Oliver Deussen, Miriah Meyer, Jeff Phillips, Samuel Gratzl, Holger Stitz, Marc Streit, Nils Gehlenborg, Lane Harrison, and many Others!

visualization
design lab