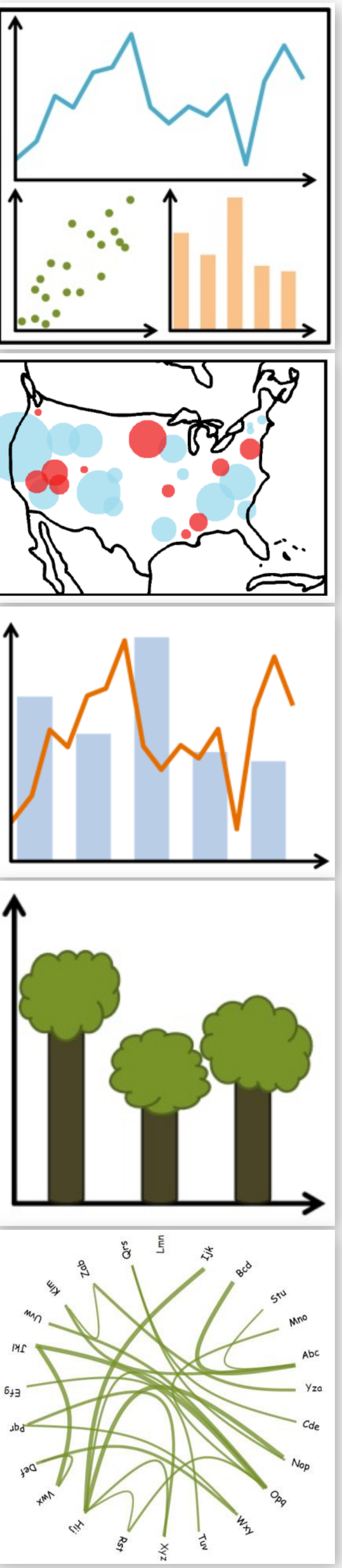


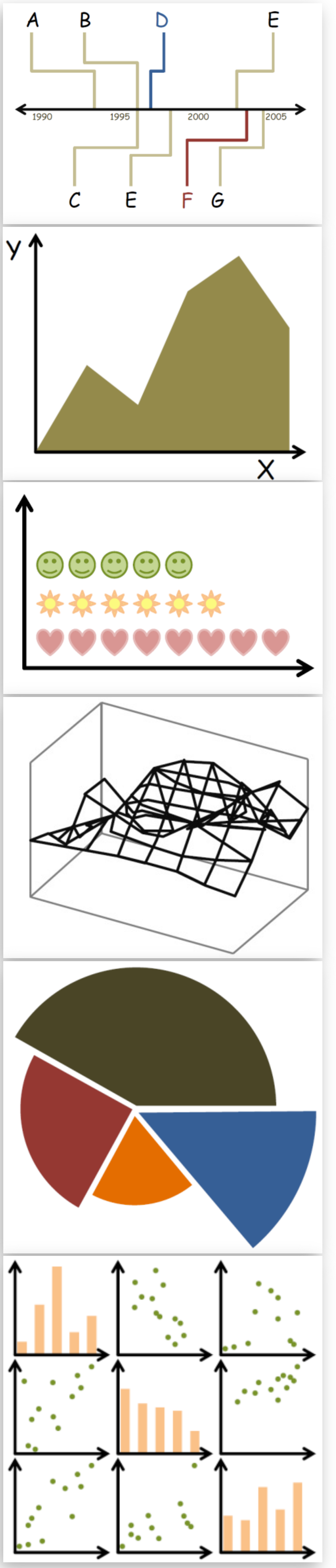
Lecture 02: Visualization Design Pipeline

DS 4200

SPRING 2023

Prof. Ab Mosca (they/them)

NORTHEASTERN UNIVERSITY



Slides and inspiration from Cody Dunne, Michelle Borkin, Remco Chang, Dylan Cashman, Krzysztof Gajos, Hanspeter Pfister, Miriah Meyer, Jonathan Schwabish, and David Sprague

Today

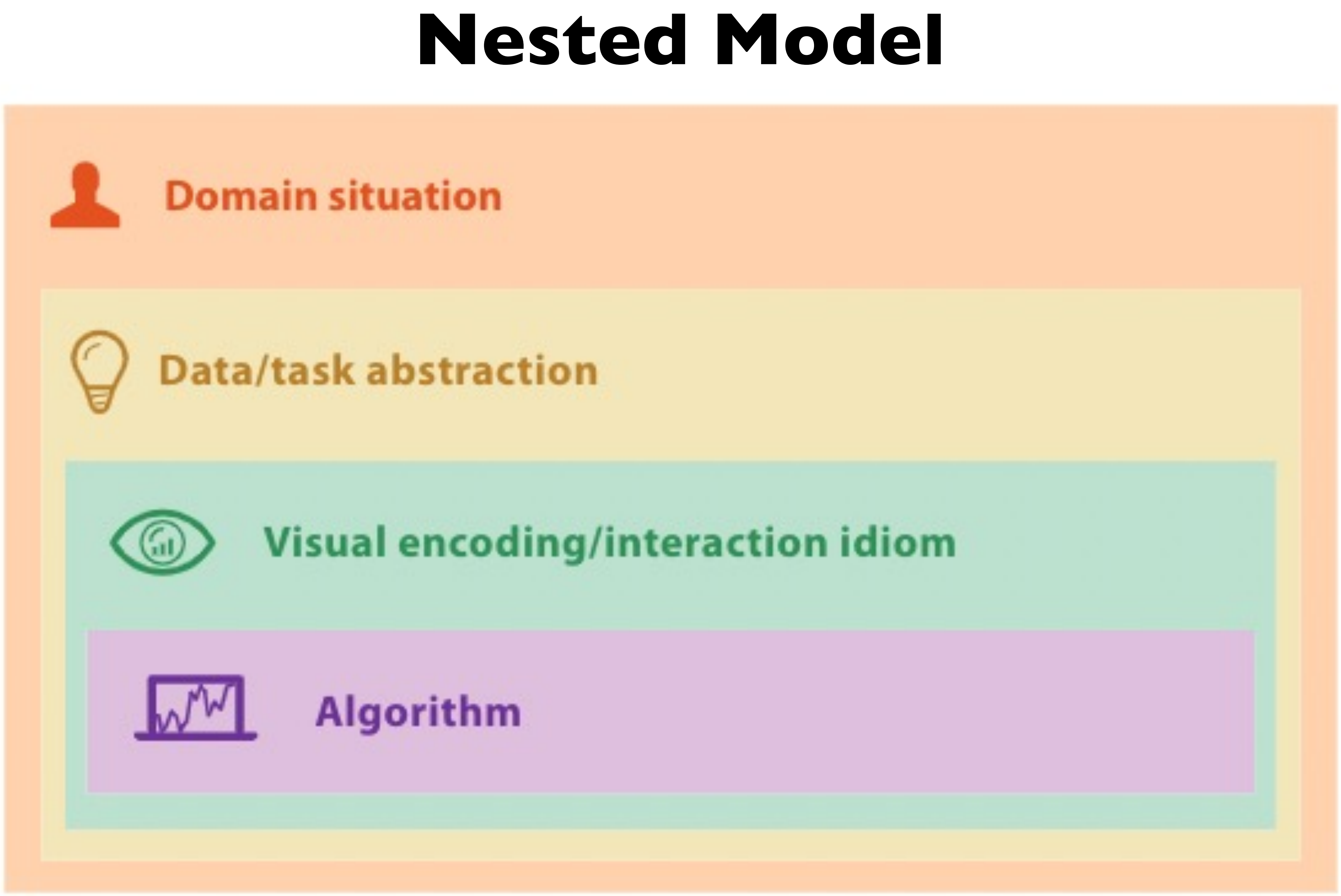
- Visualization Design Pipeline
- Finish ic-01

Note

- hw-01 is out today!
- It is a partner assignment – use the find a partner post on piazza, or the break in class to find a partner

Visualization Design Pipeline

Munzner's Nested Model



Understanding a Domain Situation

Domain Situation

Target Users

Purpose

Current Workflow

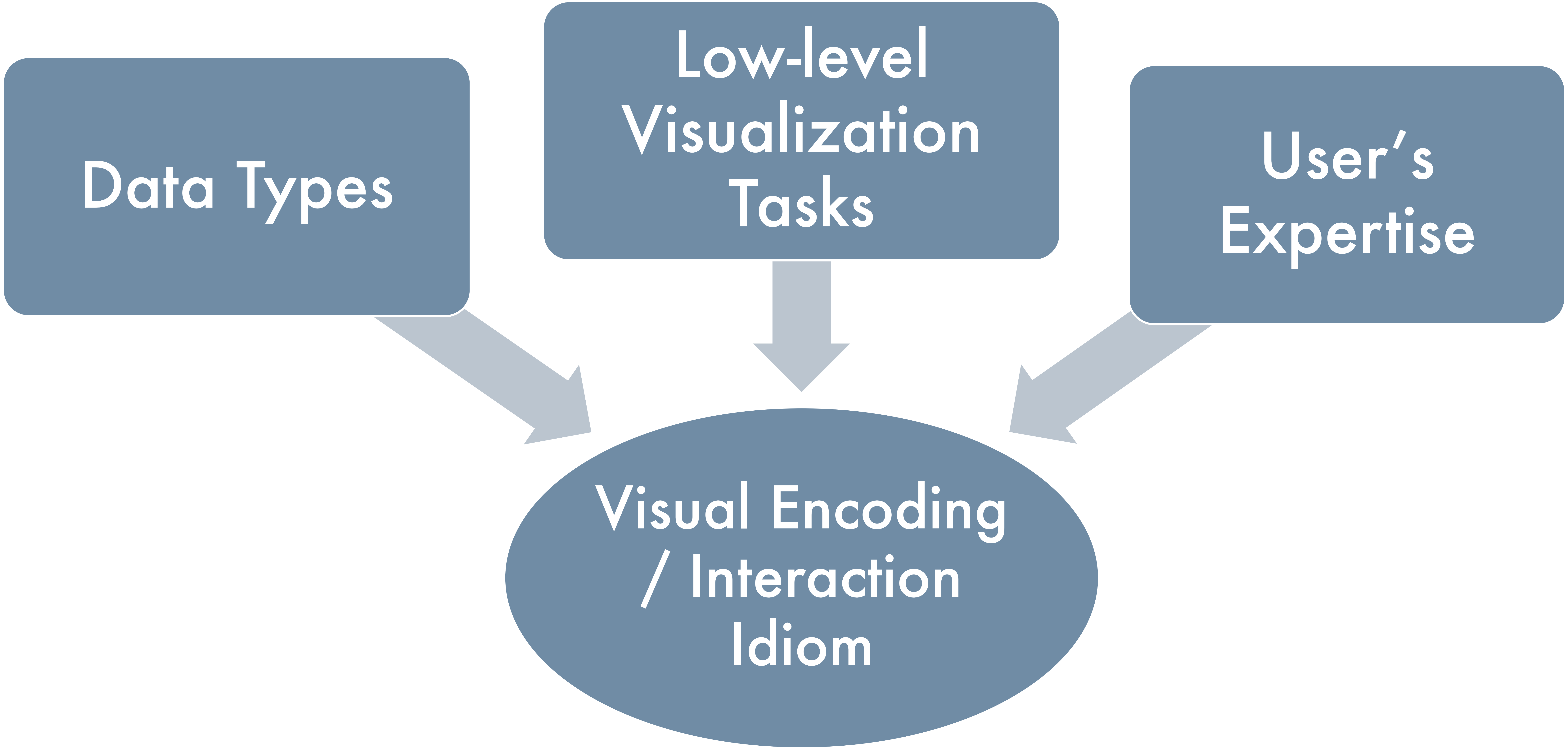
Expertise

Understanding Data / Task

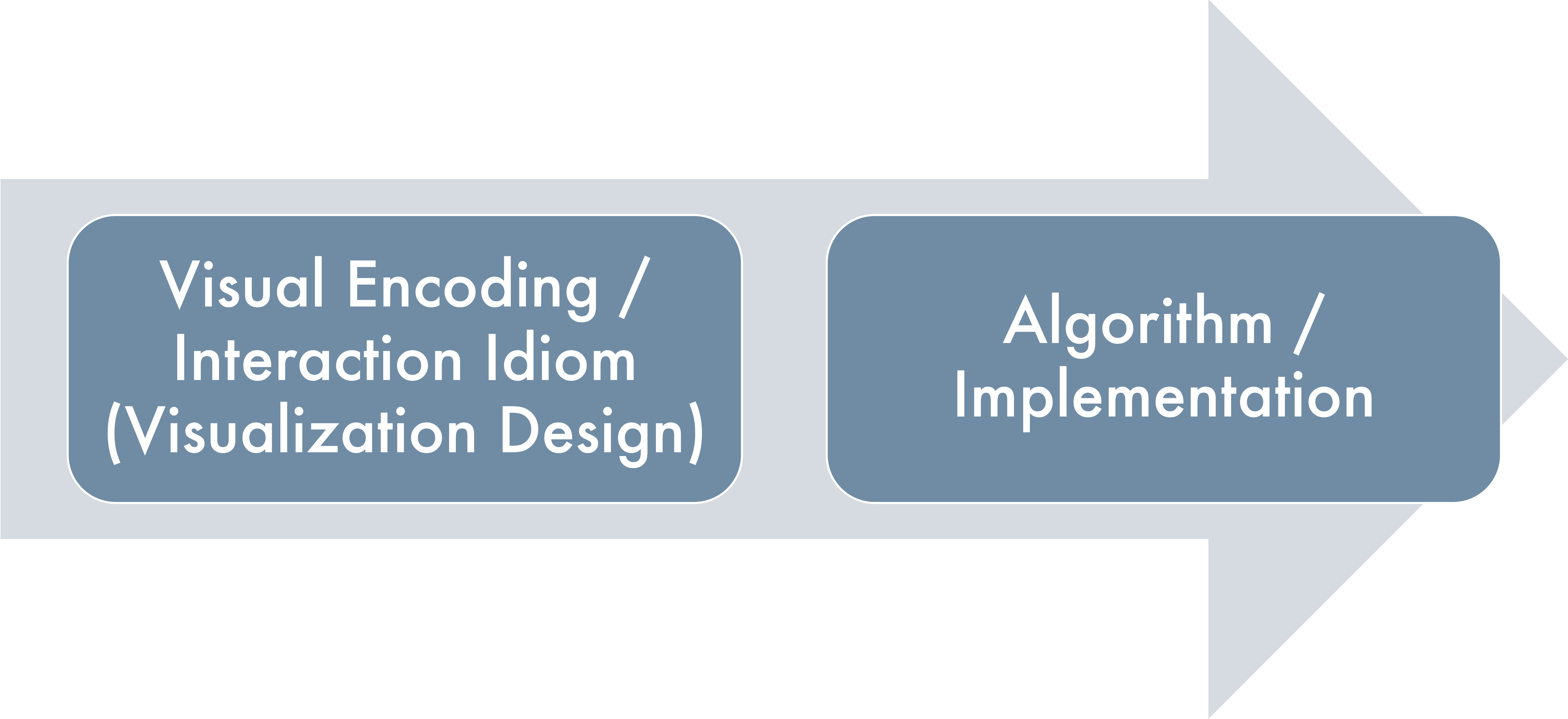
Data / Task Abstraction

Data				Task	
Source	Ethics / Biases	Types	Quality	Low-level workflow steps	Desired outcome

Visual Encoding / Interaction Idiom

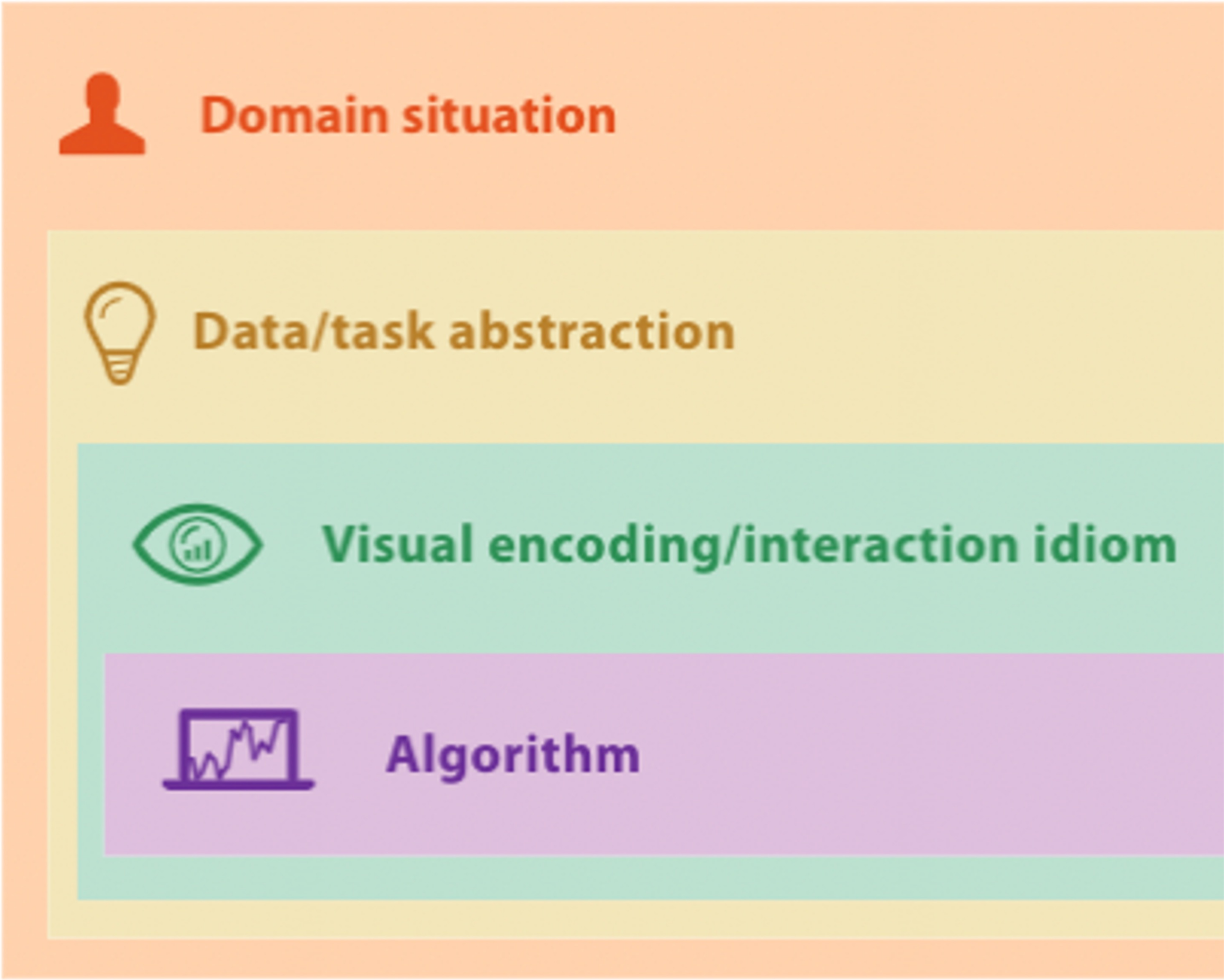


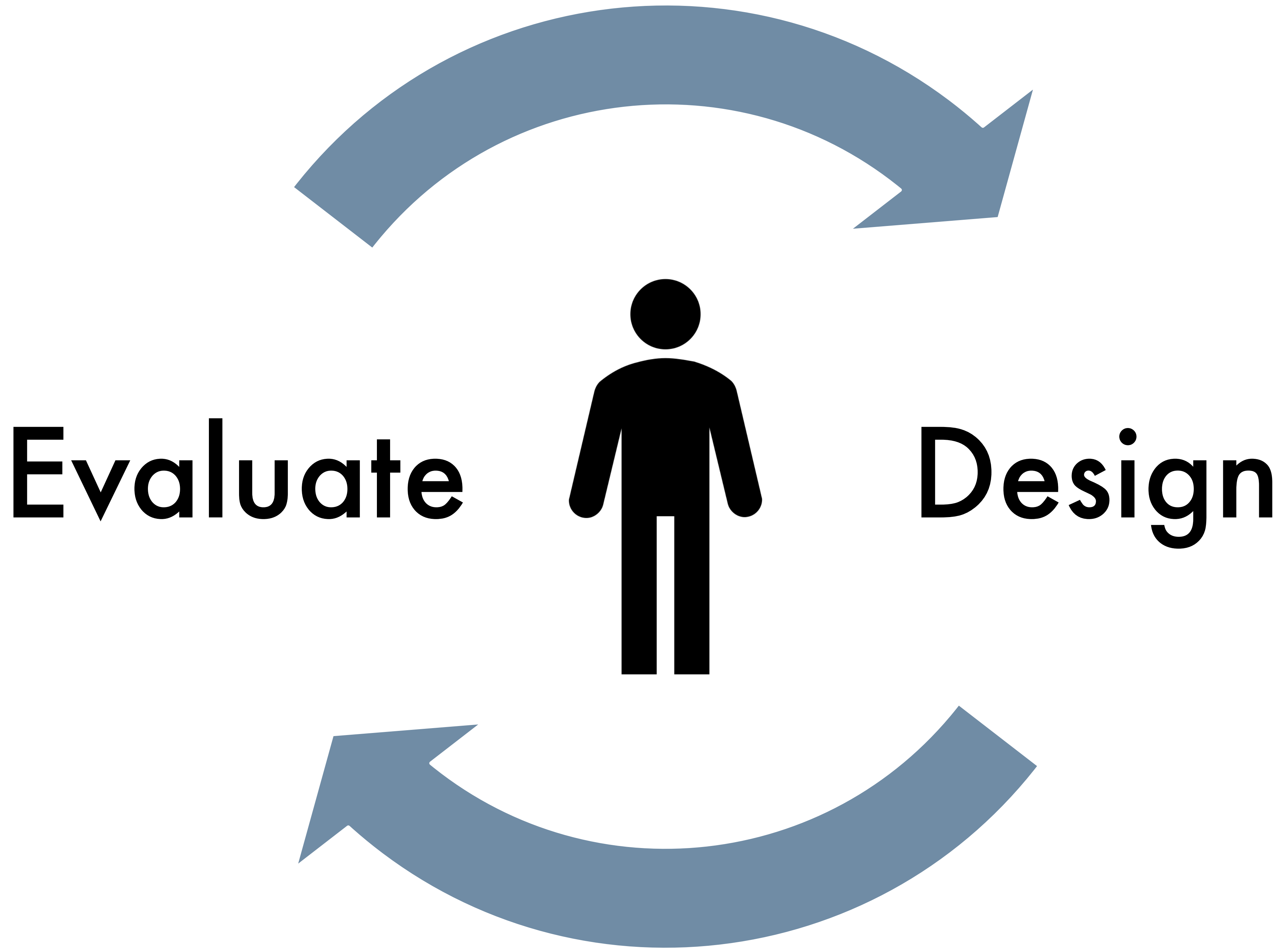
Algorithm


Visual Encoding / Interaction Idiom (Visualization Design)
Algorithm / Implementation

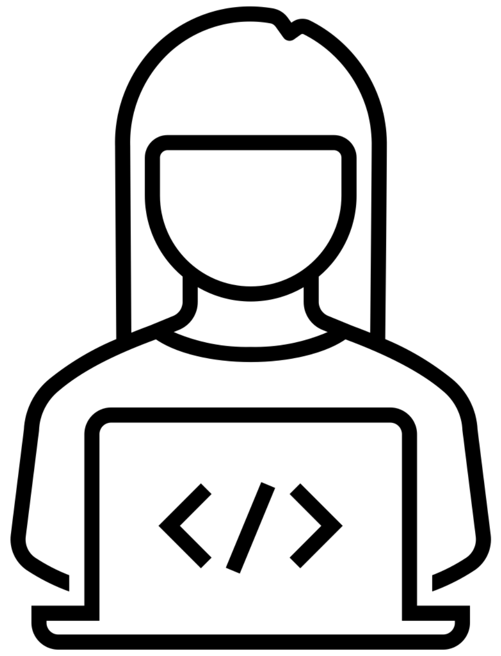
Munzner's Nested Model

Human-Centered, Iterative Design

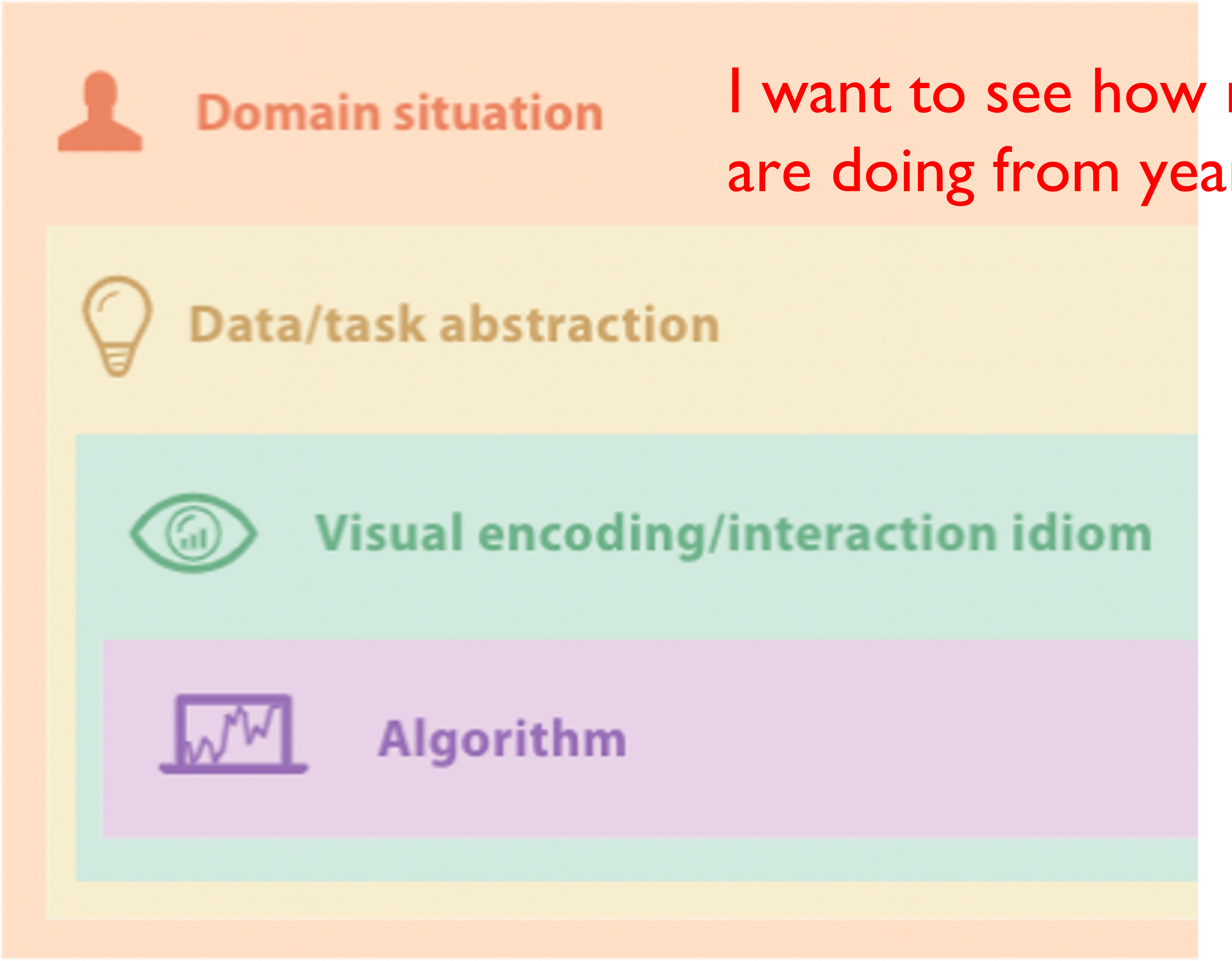




Human-Centered Design

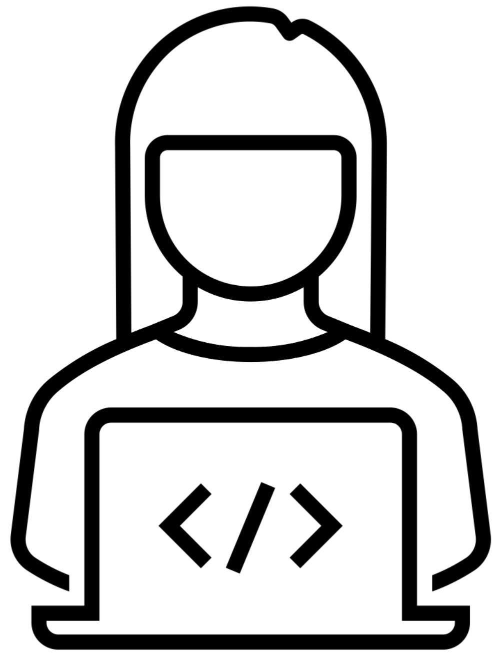


What do you need a visualization for?



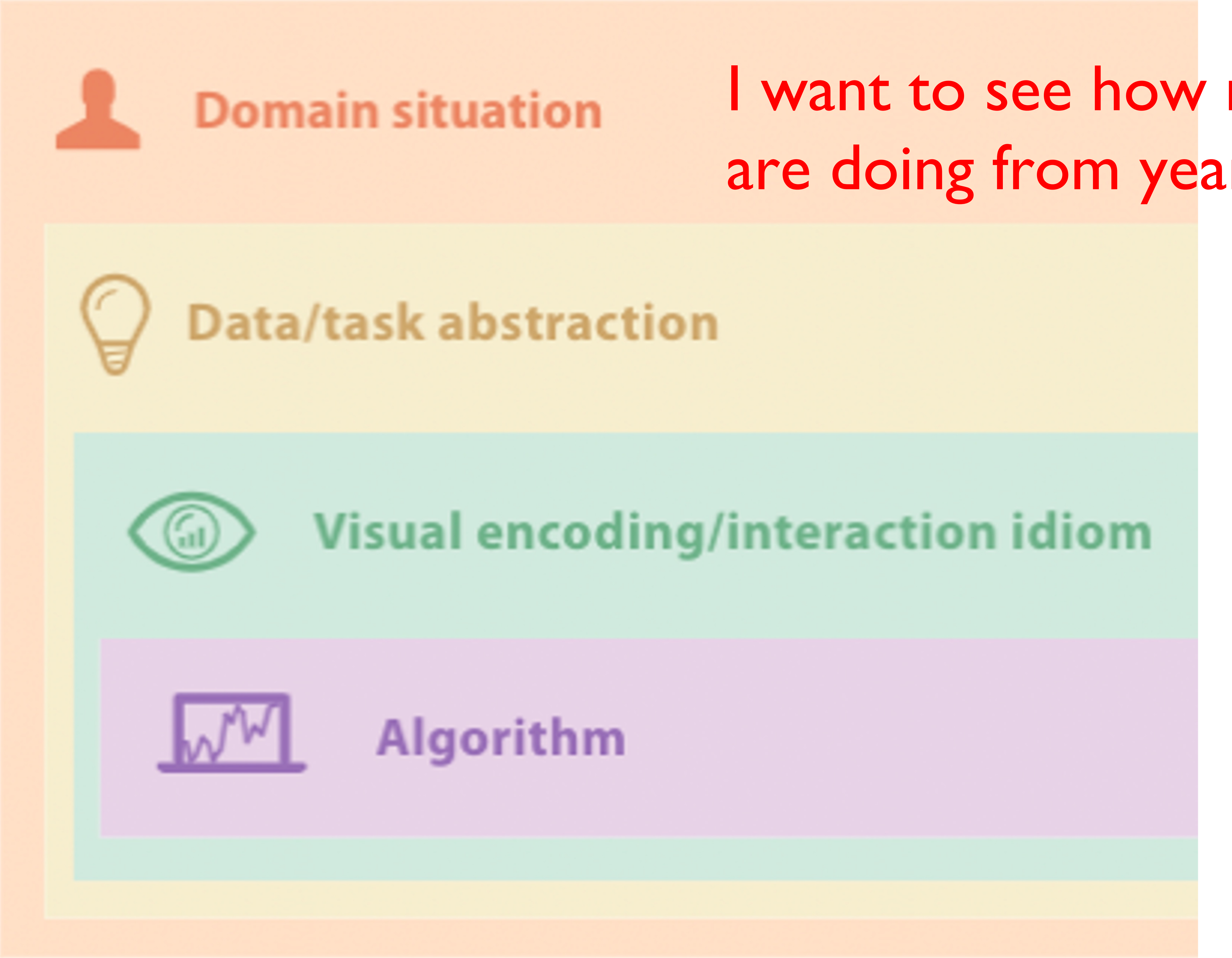
I want to see how my students are doing from year to year.

Human-Centered Design



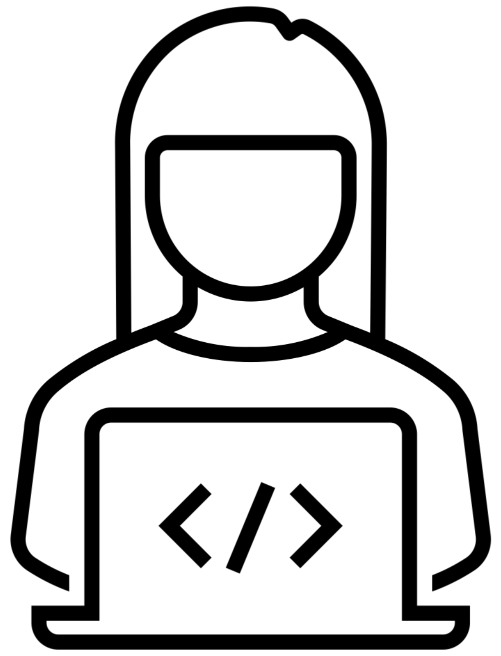
What do you need a visualization for?

How do you measure student success?



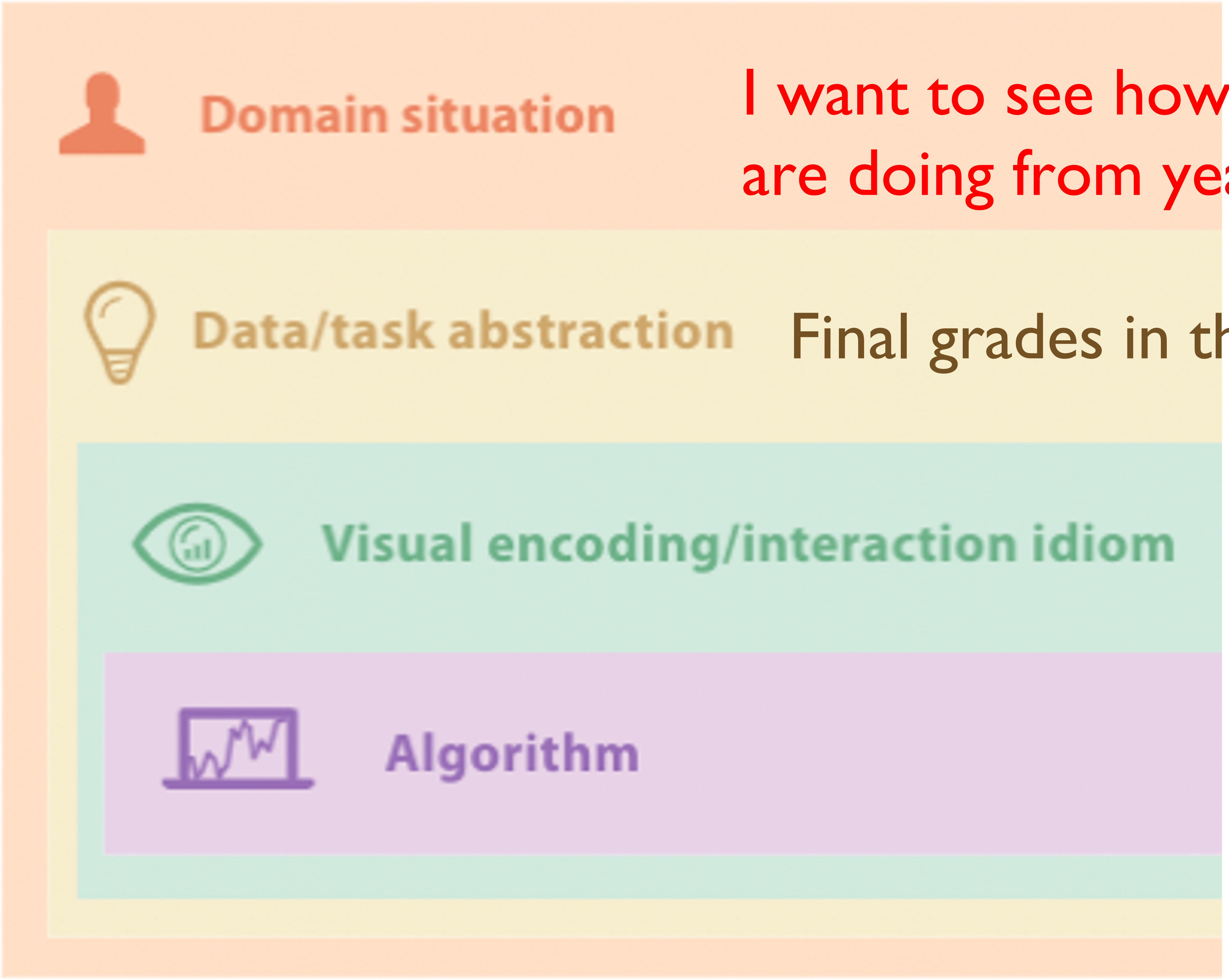
I want to see how my students are doing from year to year.

Human-Centered Design



What do you need a visualization for?

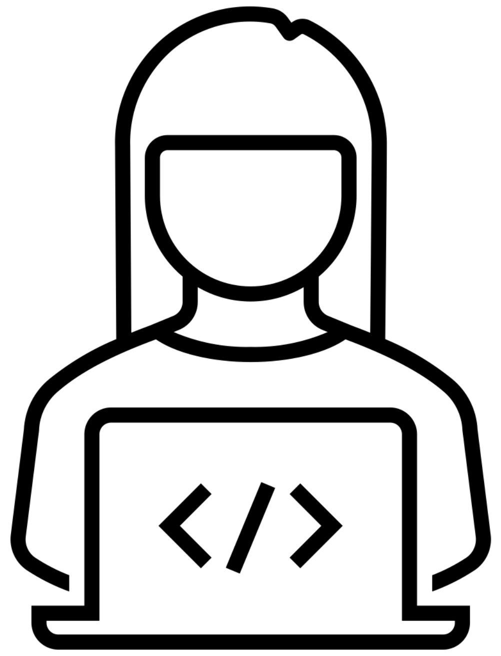
How do you measure student success?



I want to see how my students are doing from year to year.

Final grades in the class.

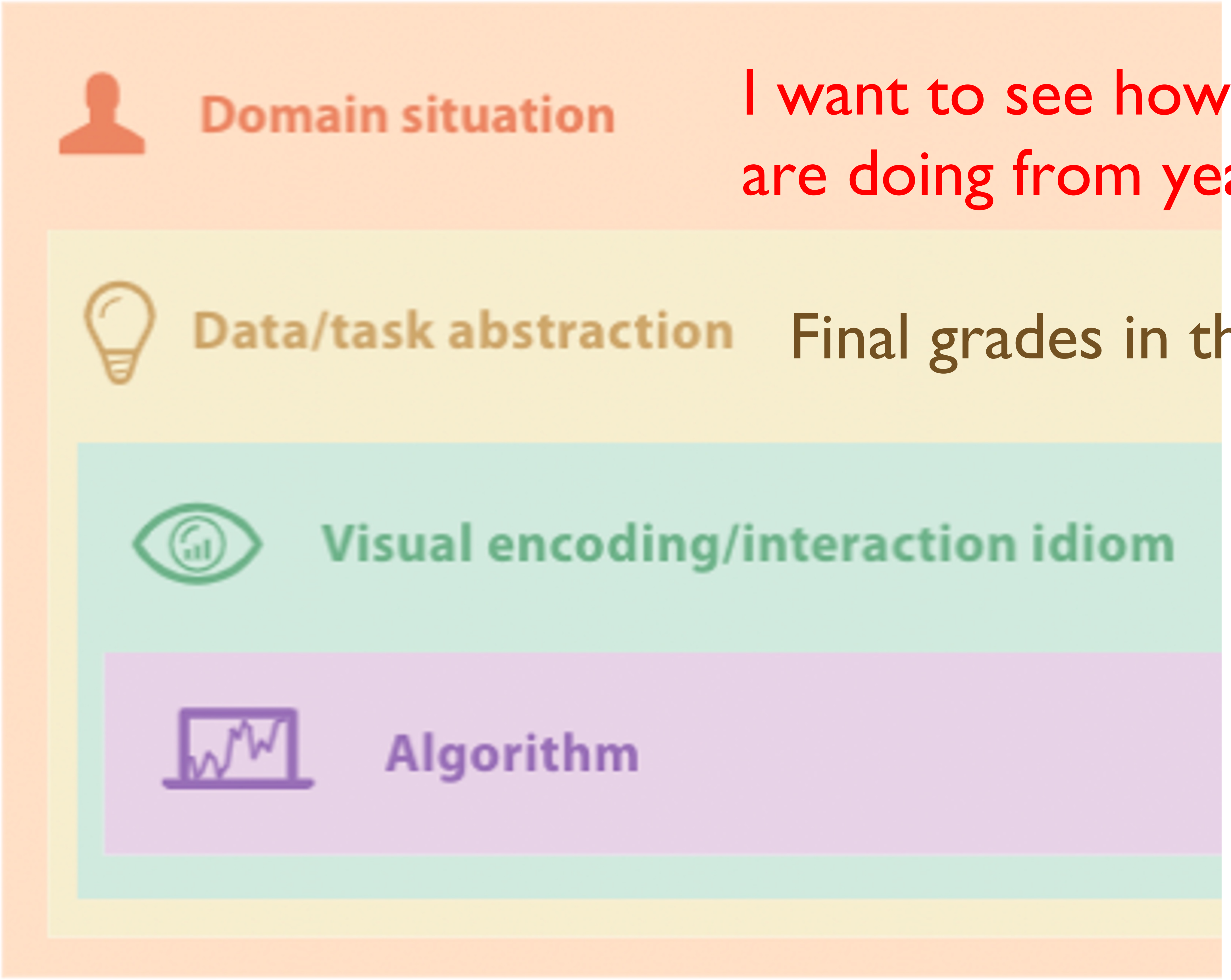
Human-Centered Design



What do you need a visualization for?

How do you measure student success?

What kinds of comparisons or statistics do you need?

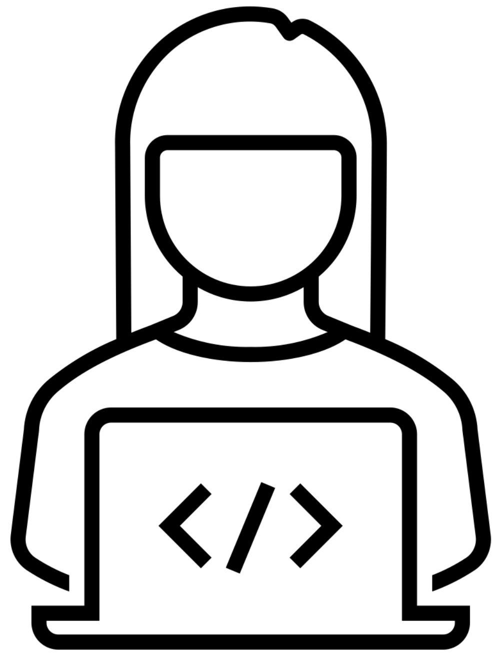


I want to see how my students are doing from year to year.

Final grades in the class.

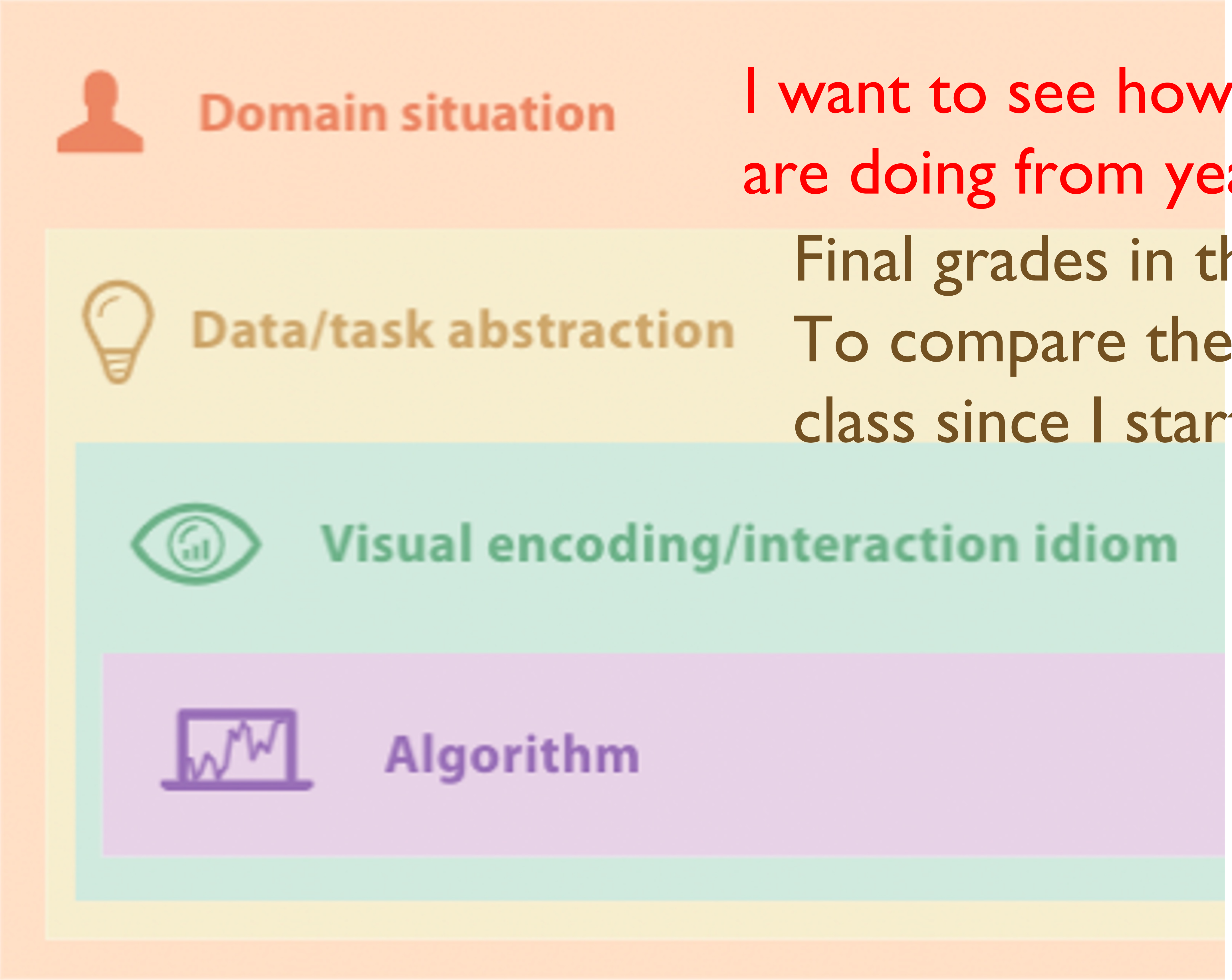
Human-Centered Design

What do you need a visualization for?



How do you measure student success?

What kinds of comparisons or statistics do you need?

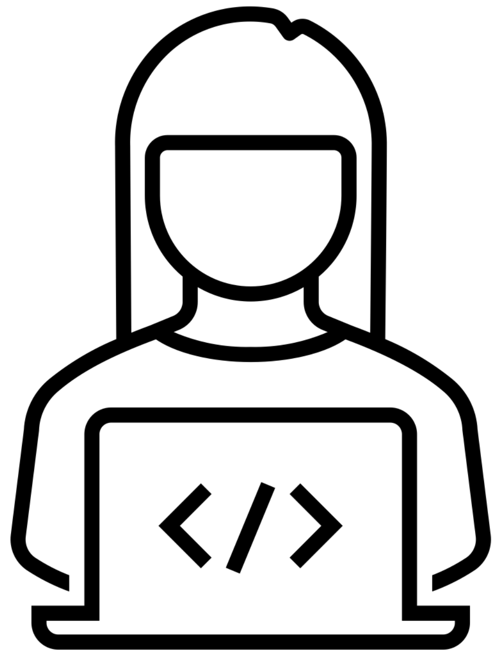


I want to see how my students are doing from year to year.

Final grades in the class.

To compare the average final grade of each class since I started teaching.

Human-Centered Design



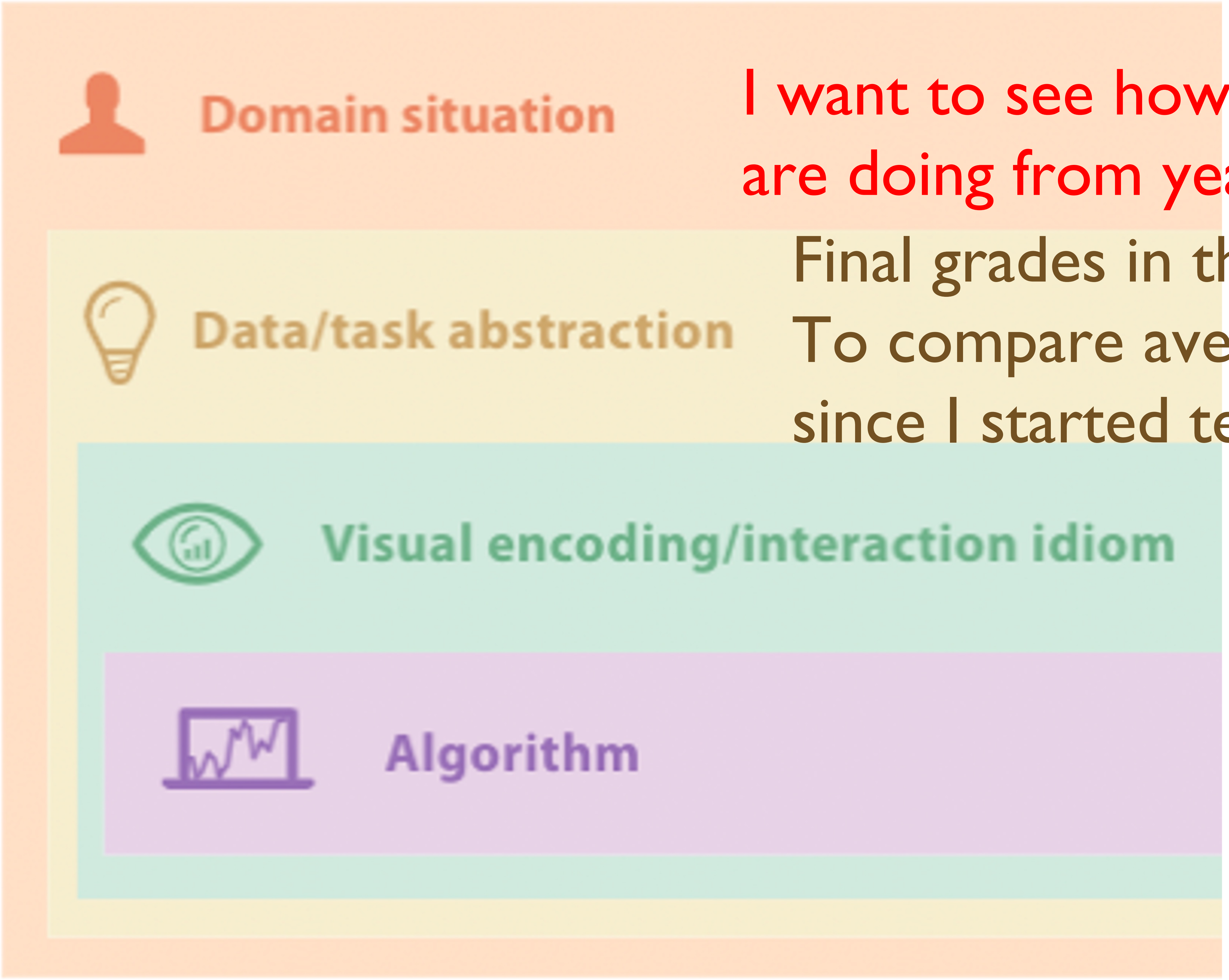
What do you need a visualization for?

How do you measure student success?

What kinds of comparisons or statistics do you need?

I'll start with a line chart showing average grade per year.

I'll code it in Tableau.



I want to see how my students are doing from year to year.

Final grades in the class.
To compare average grade of each class since I started teaching.

Human-Centered Design

- Ground our designs in human needs
- Gather an understanding of the domain situation from our end users
- Check our work by iterating, i.e. continually verifying and evaluating our proposed visualization tool

Example: PROACT

PROACT: Iterative Design of a Patient-Centered Visualization for Effective Prostate Cancer Health Risk Communication

Anzu Hakone, Lane Harrison, Alvitta Ottley, Nathan Winters, Caitlin Gutheil, Paul K. J. Han, Remco Chang

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS VOL. 23, NO. 1, JANUARY 2017

Example: PROACT

PROACT: Iterative Design of a Patient-Centered Visualization for Effective Prostate Cancer Health Risk Communication

Anzu Hakone, Lane Harrison, Alvitta Ottley, Nathan Winters, Caitlin Gutheil, Paul K. J. Han, Remco Chang

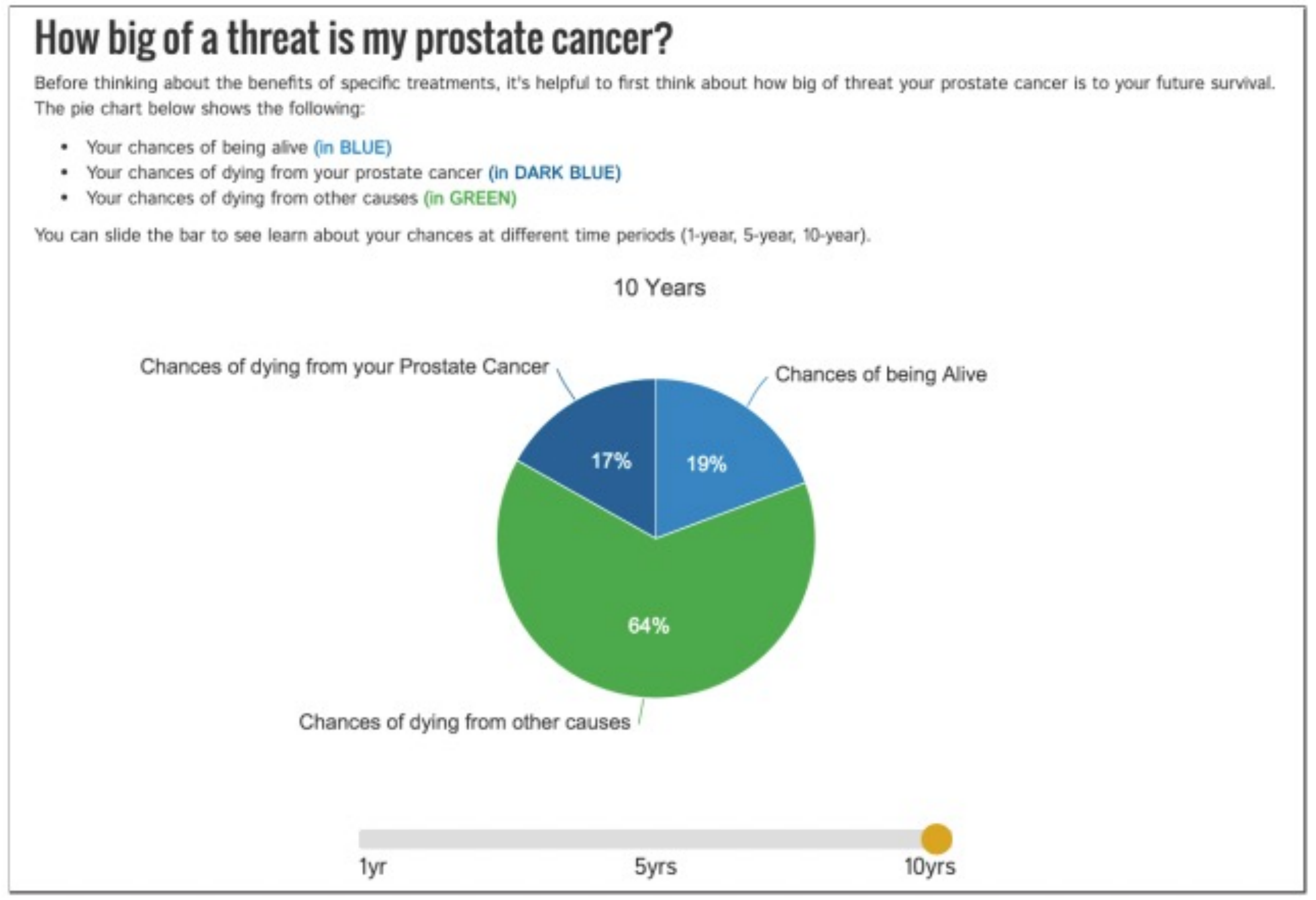
IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS VOL. 23, NO. 1, JANUARY 2017

Domain Situation: Doctors need to communicate expected prognoses following active vs. conservative treatment of localized prostate cancer.

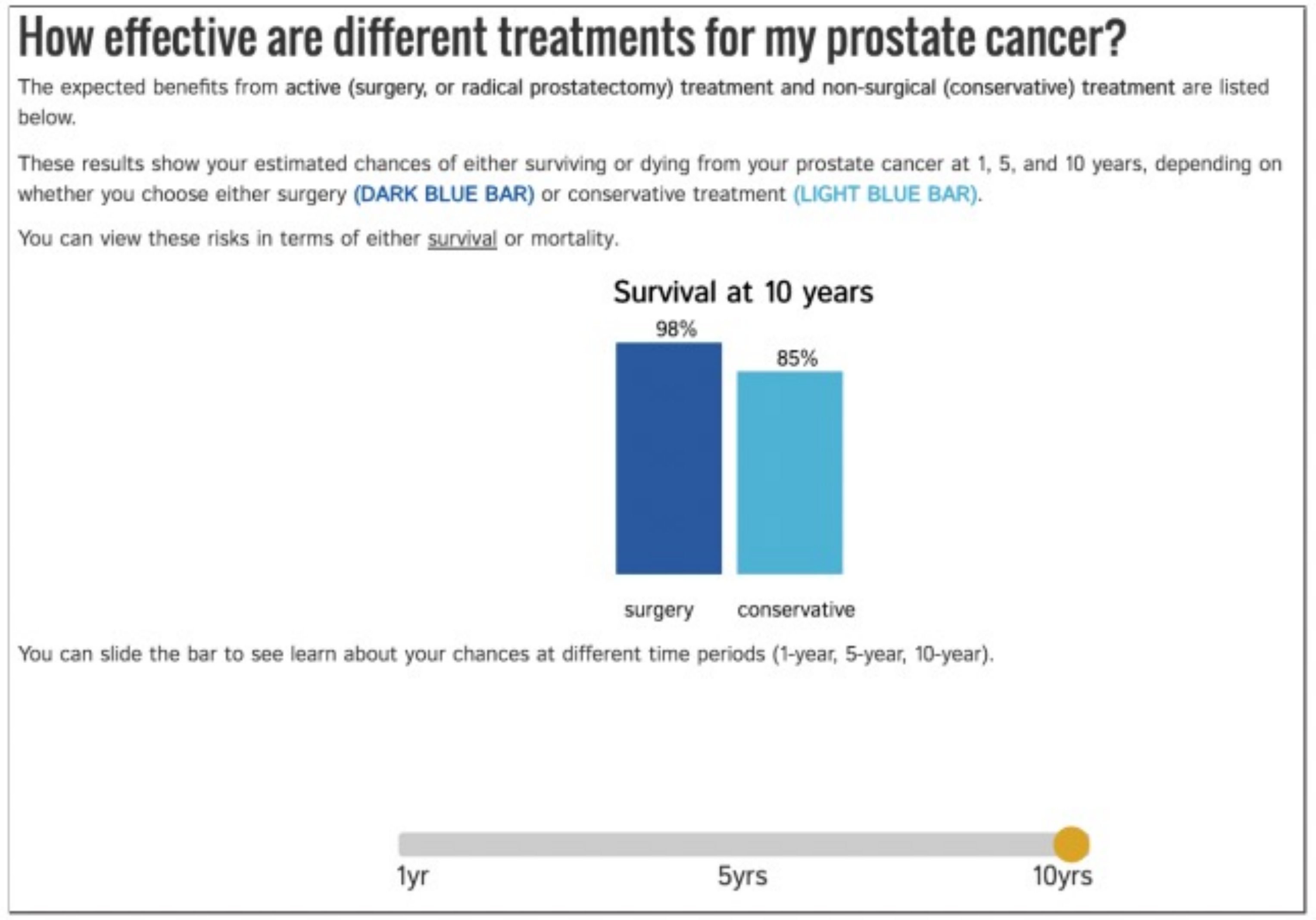
Data / Tasks: (1) Help patients understand personal risk at various years. (2) Help patients compare survival rates at various years given the two treatment options.

Example: PROACT

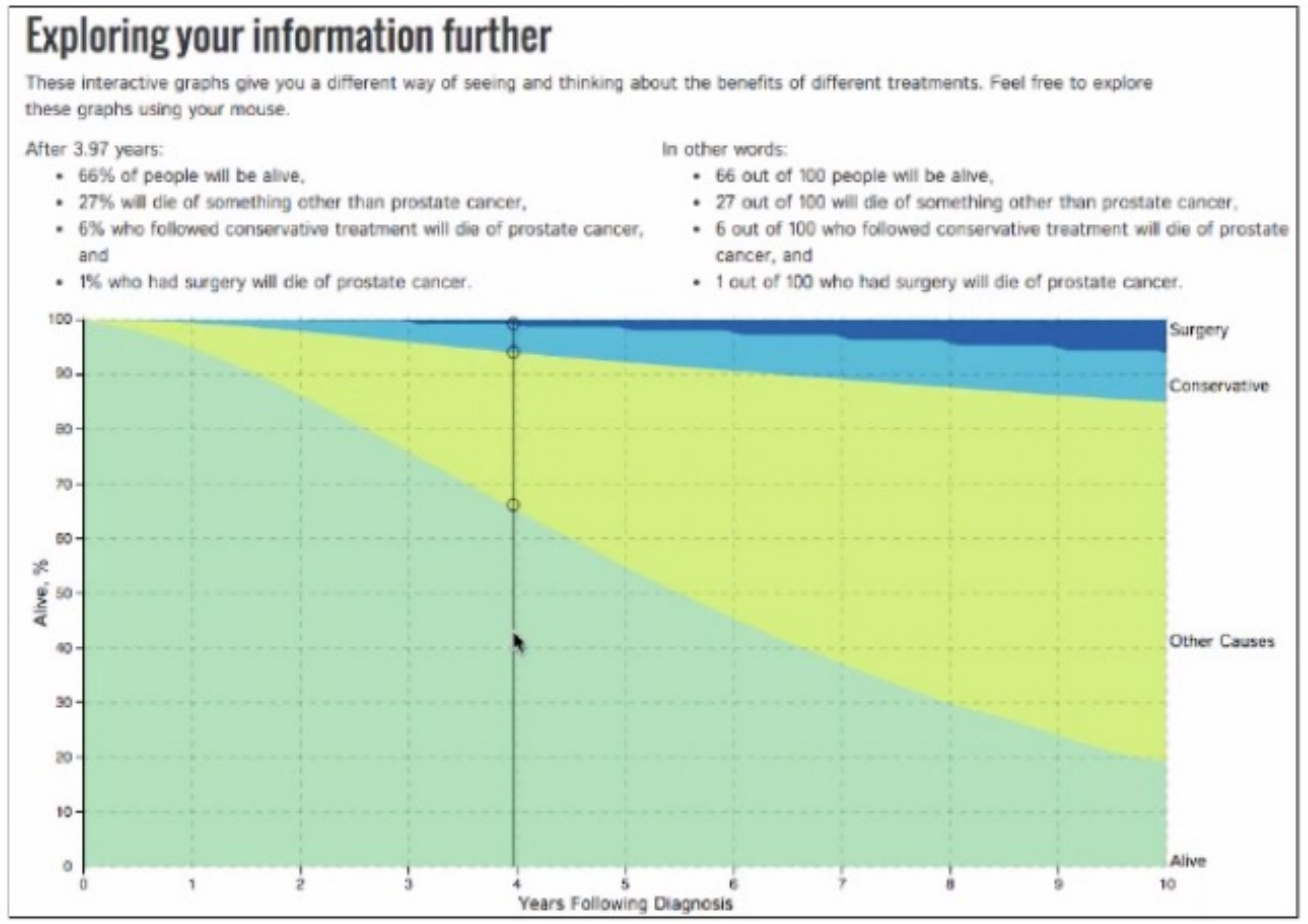
Visual Encoding / Interaction Paradigm:



Pie chart showing morbidity rates with slider for years.



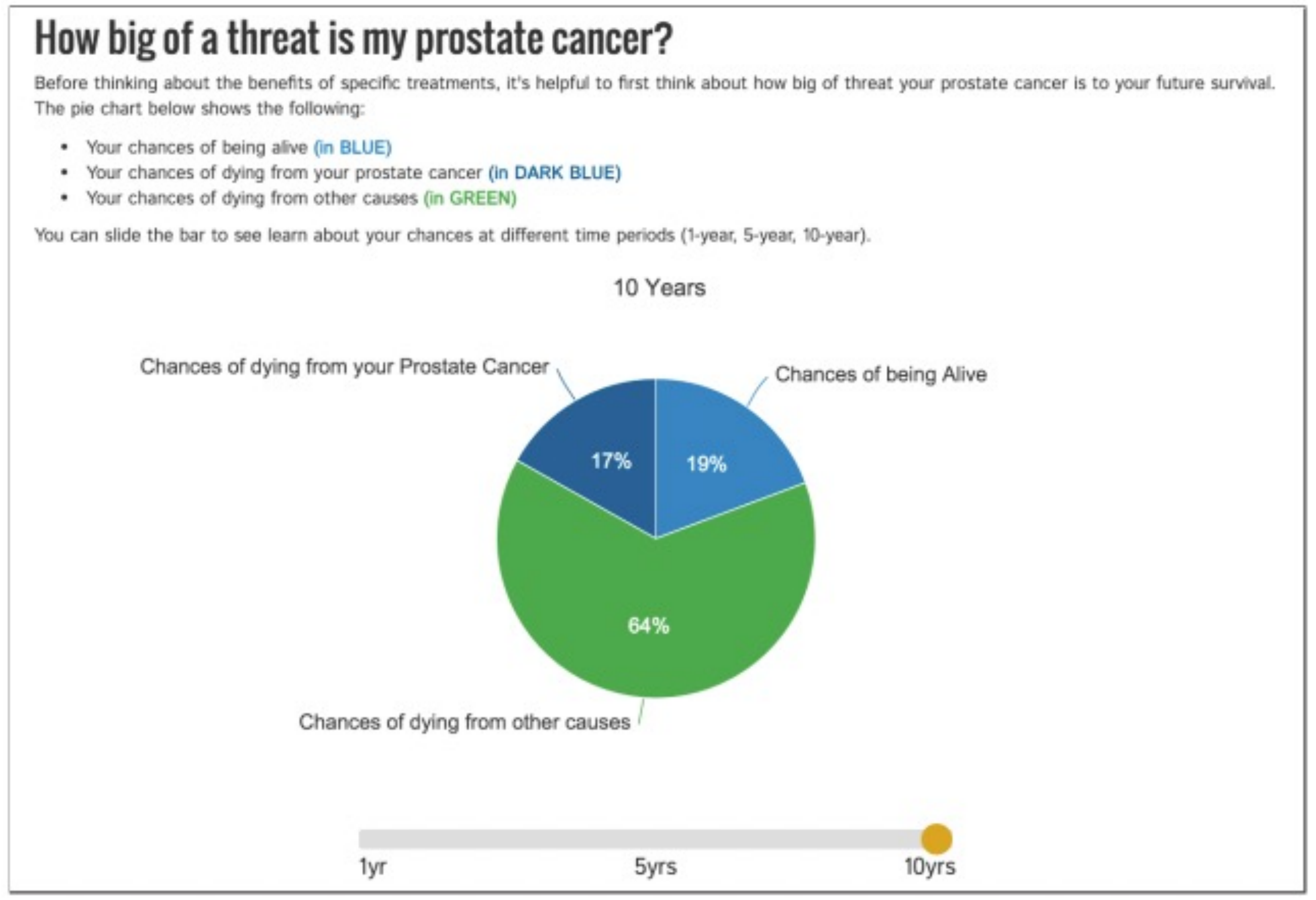
Bar chart showing survival rates with slider for years.



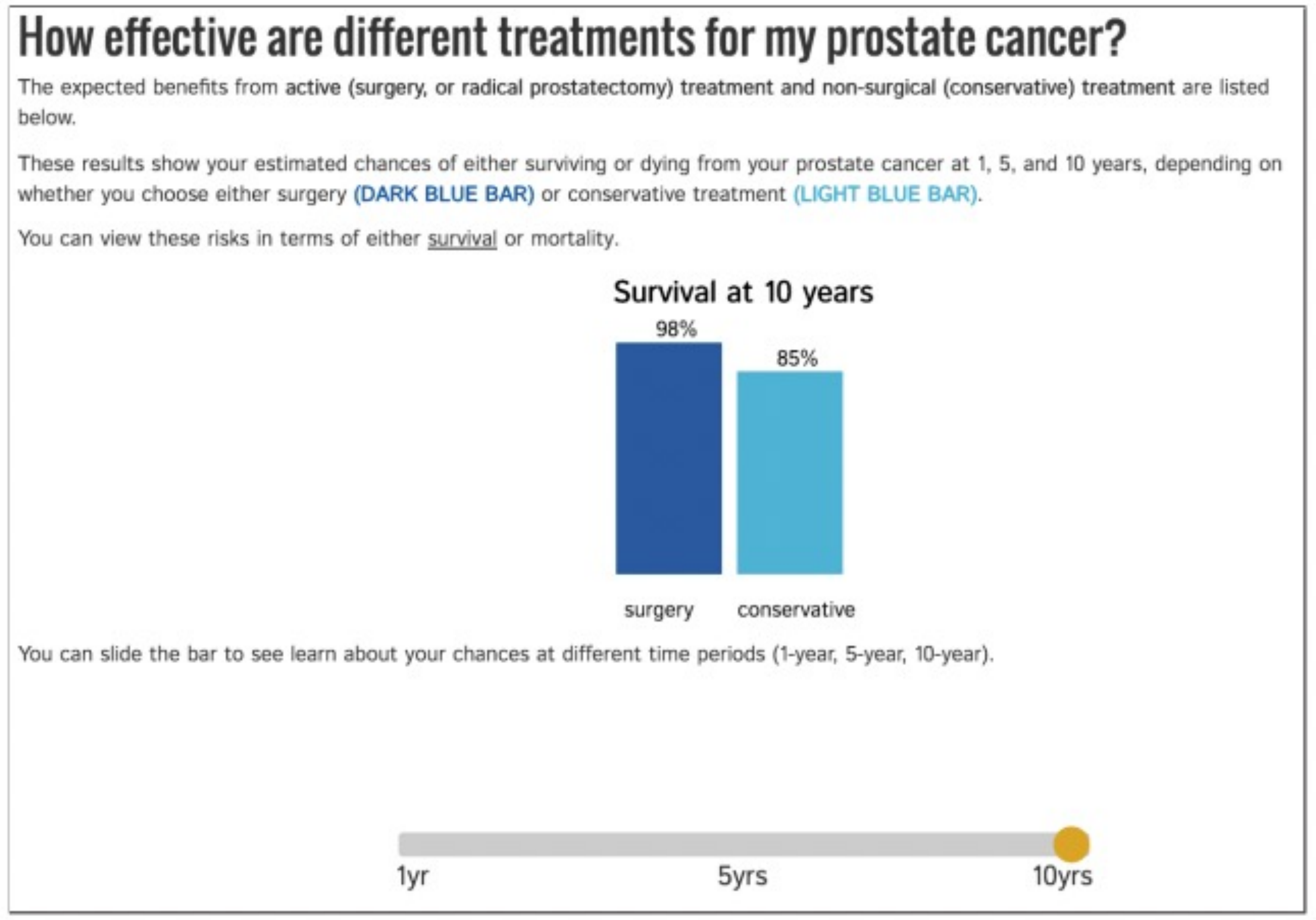
Temporal area chart of survival rates.

Example: PROACT

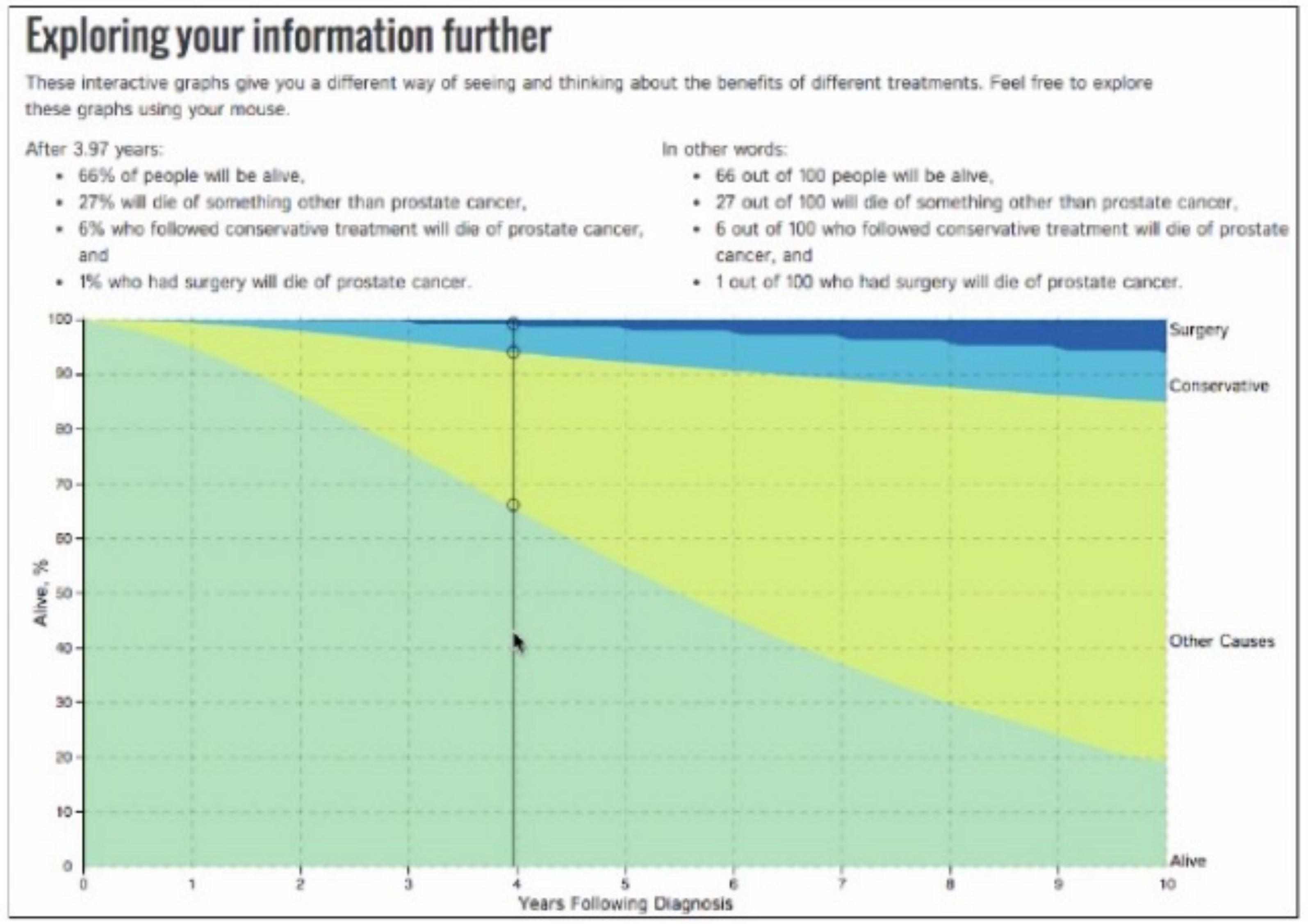
Evaluation showed…



Pie chart showing morbidity rates with ~~slider~~ for years.



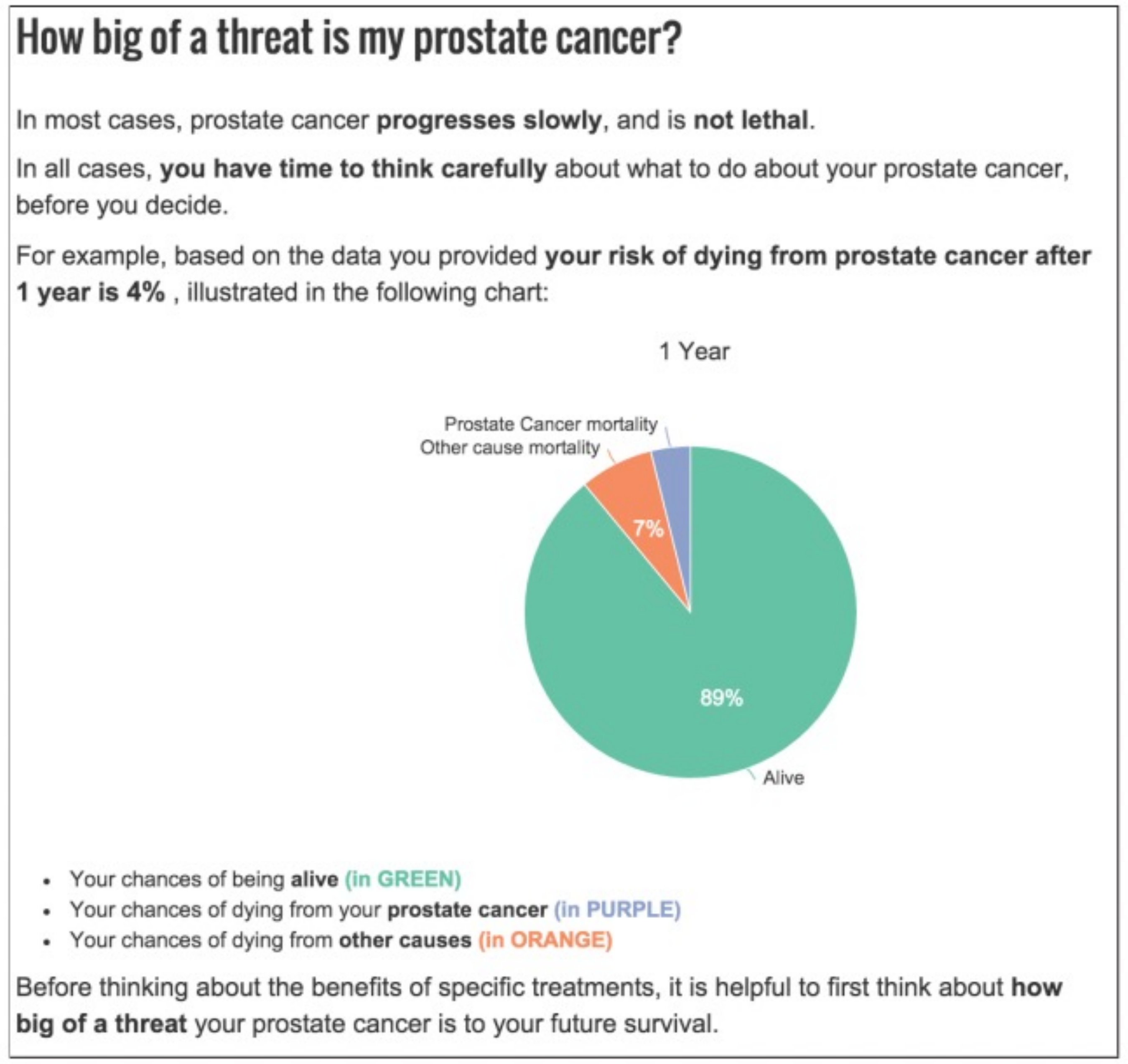
Bar chart showing survival rates with ~~slider~~ for years.



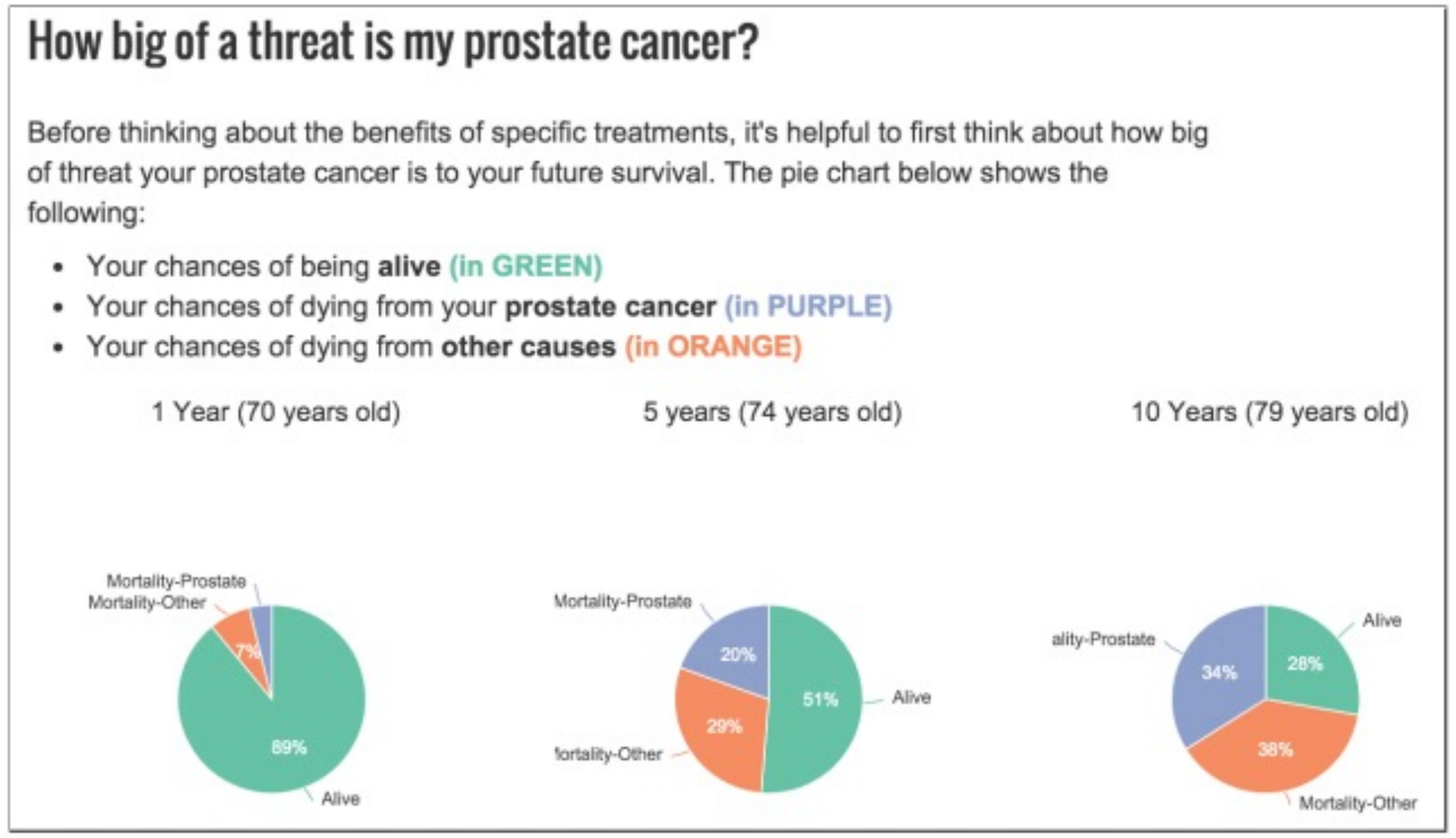
~~Temporal area chart~~ of survival rates.

Example: PROACT

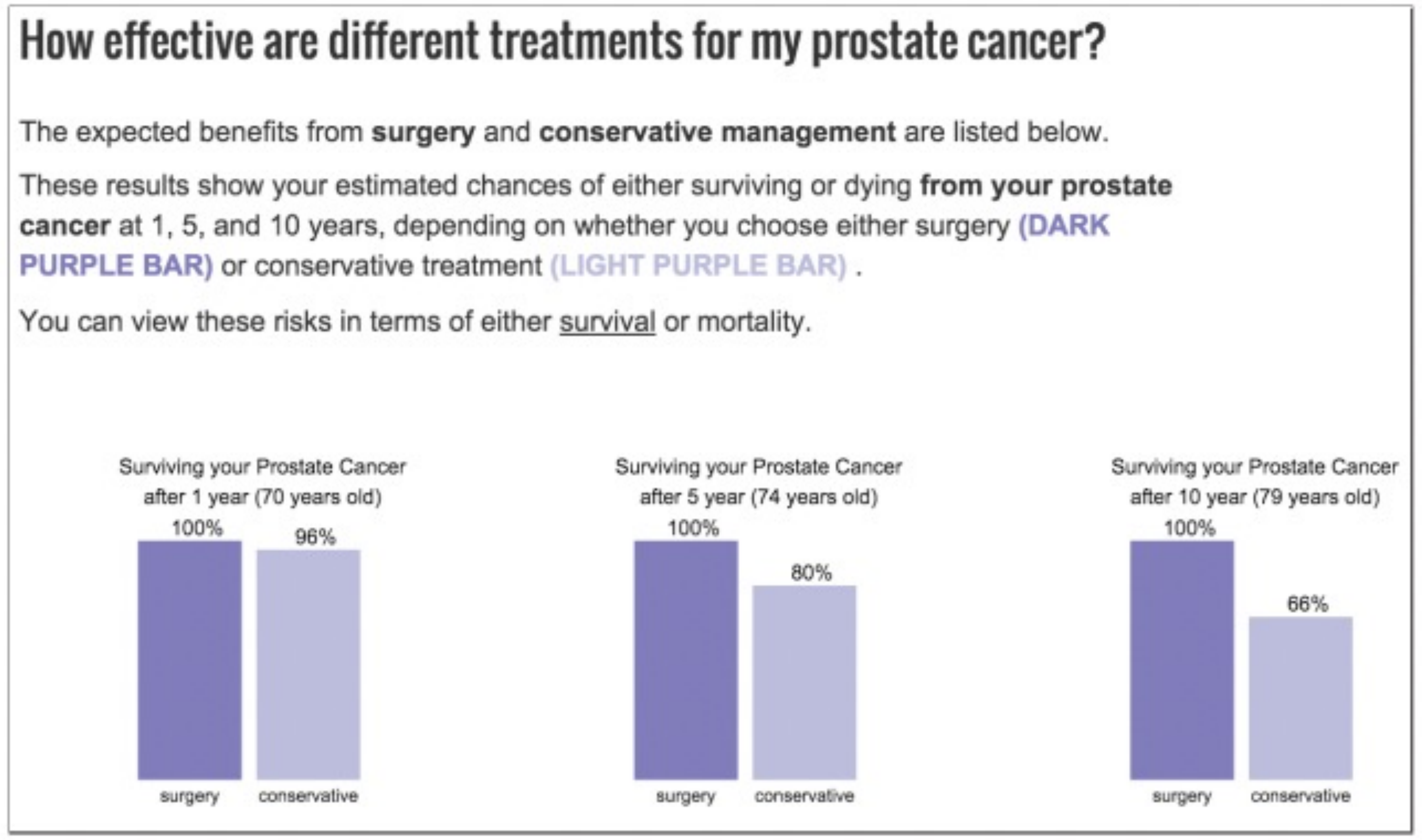
REVISED Visual Encoding / Interaction Paradigm:



Pie chart with 1 year survival.



Pie charts showing morbidity rates over years.



Bar charts showing survival rates over years.

Example: Jigsaw

Reflections on the evolution of the Jigsaw visual analytics system

Information Visualization
2014, Vol. 13(4) 336–345

DOI: 10.1177/1473871613495674
ivi.sagepub.com
SAGE

Carsten Görg[1], Zhicheng Liu[2] and John Stasko[3]

Example: Jigsaw

Reflections on the evolution of the Jigsaw visual analytics system

Information Visualization
2014, Vol. 13(4) 336–345
© The Author(s) 2013
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1473871613495674
ivi.sagepub.com
SAGE

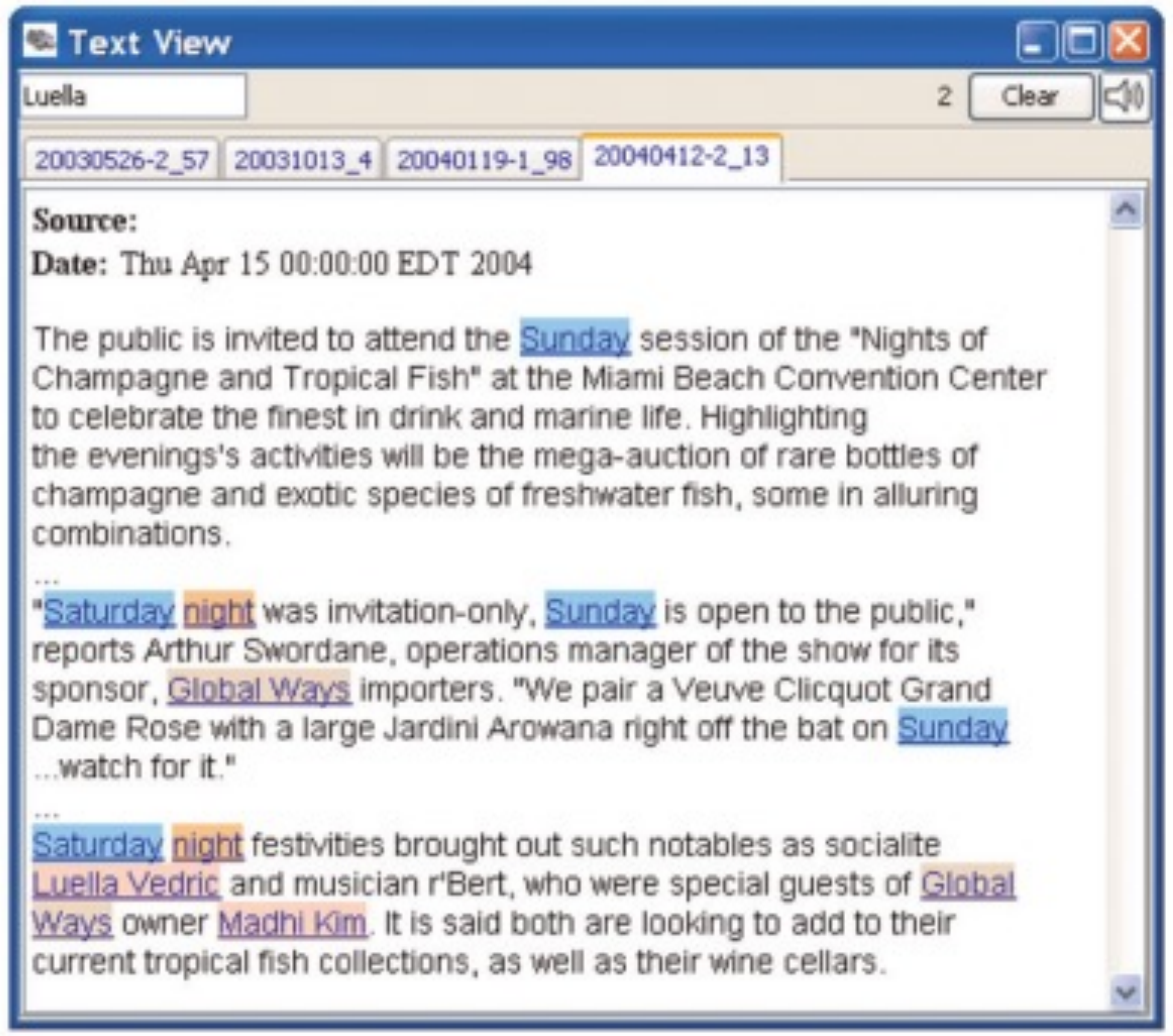
Carsten Görg[1], Zhicheng Liu[2] and John Stasko[3]

Domain Situation: Intelligence analysts need to analyze collections of text for connections.

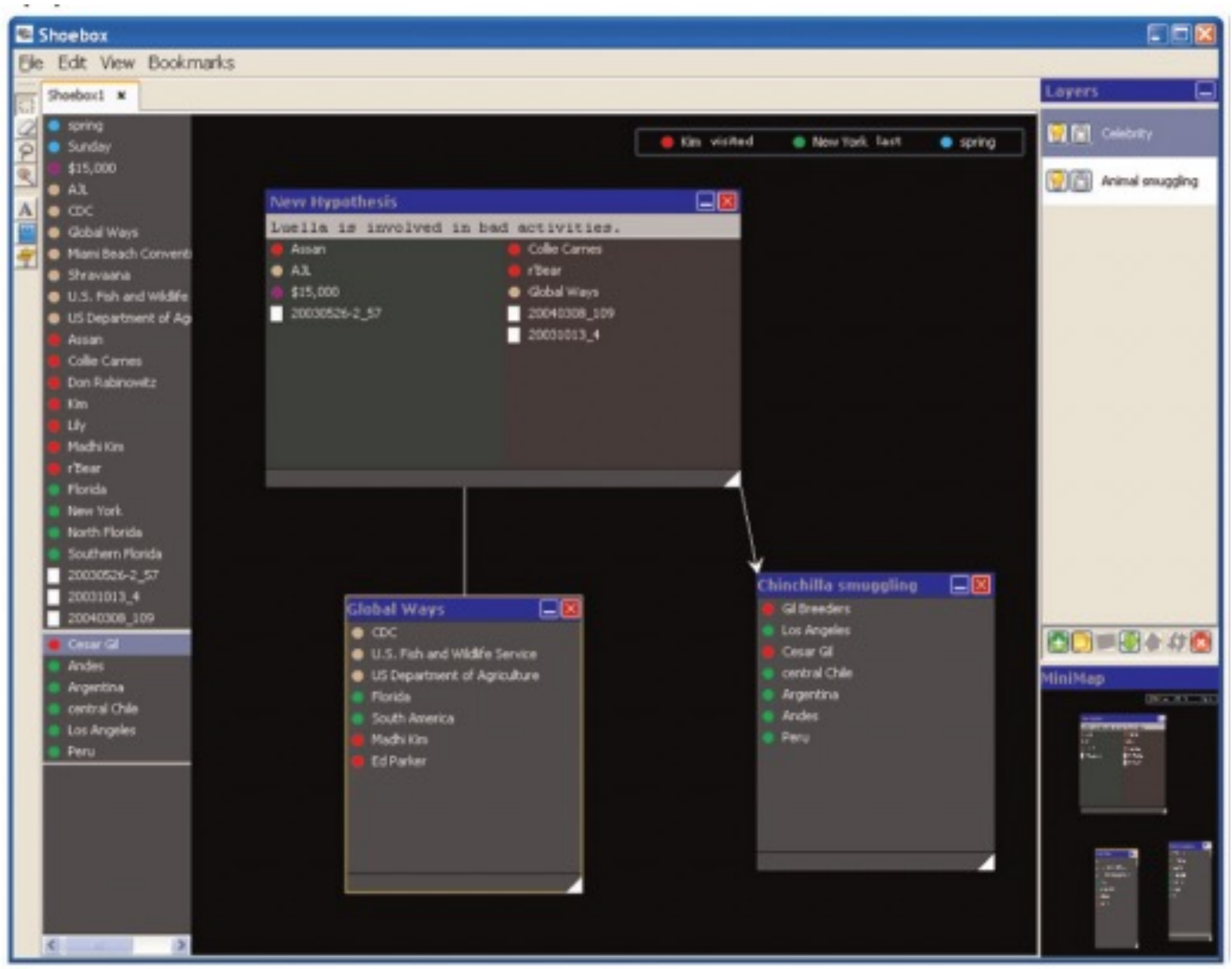
Data / Tasks: Group documents based on shared people, places, organizations, etc..

Example: Jigsaw

Visual Encoding / Interaction Paradigm:



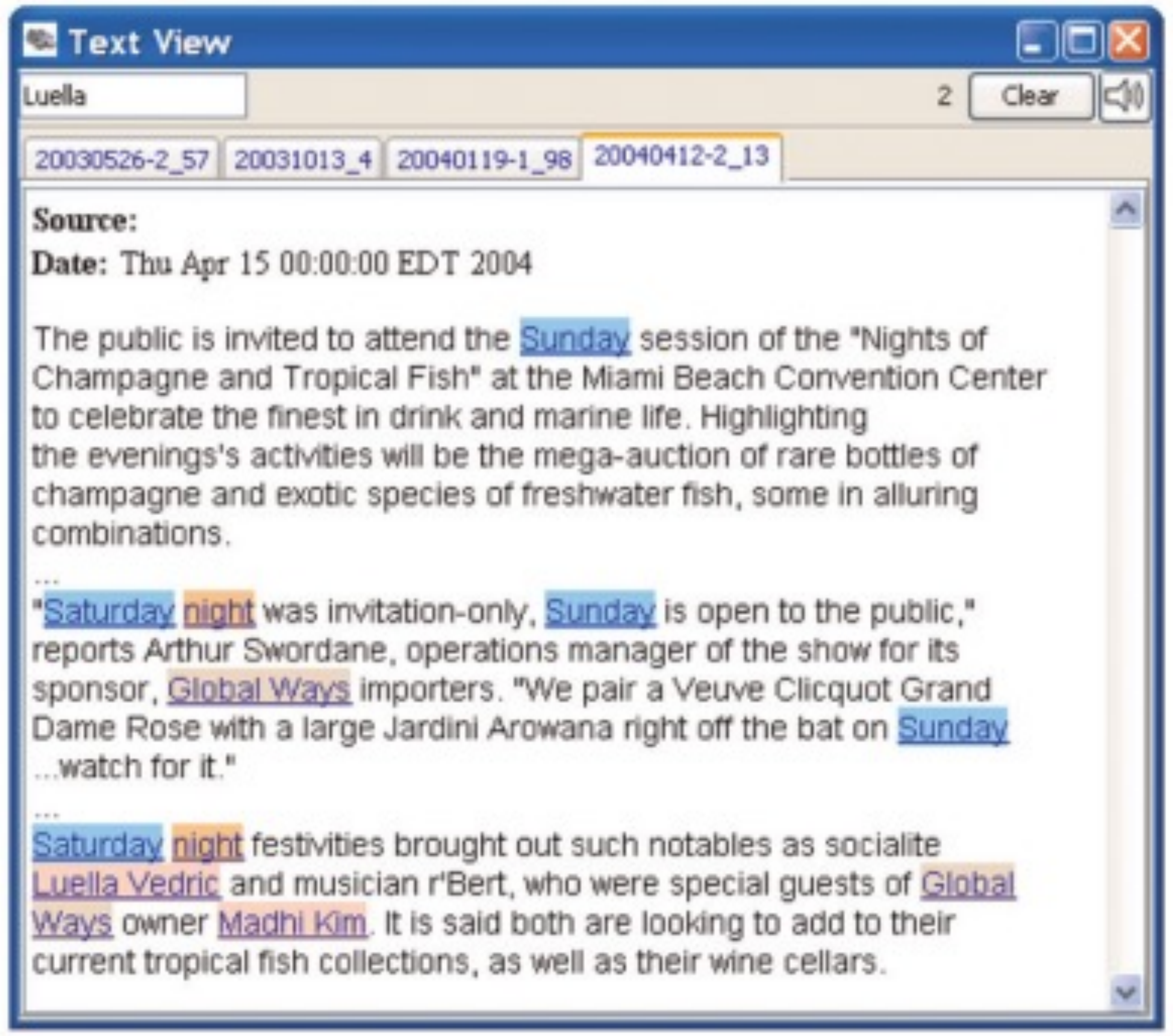
Documents shown in tabs with entities highlighted.



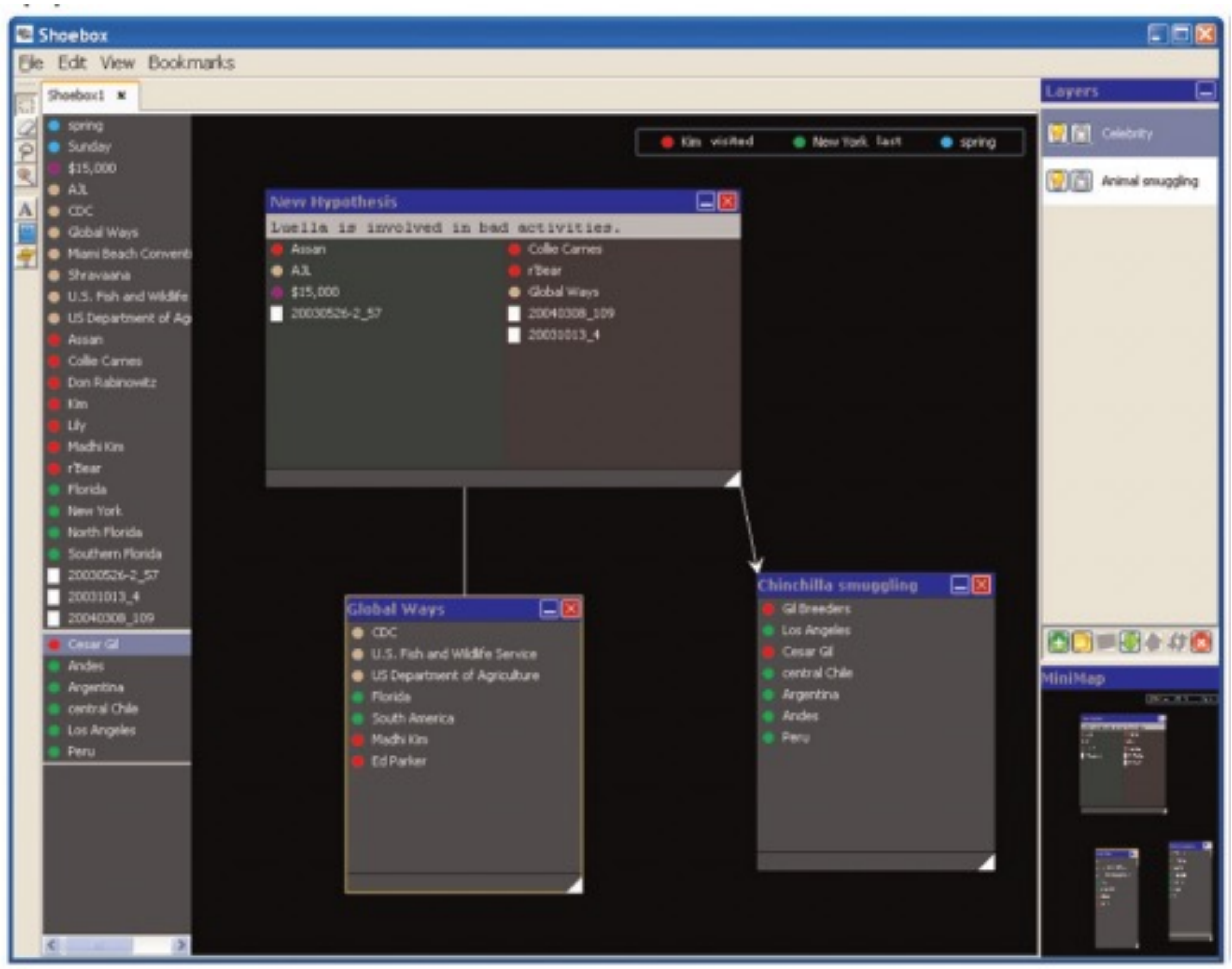
Manual note taking with automated layout.

Example: Jigsaw

Visual Encoding / Interaction Paradigm:



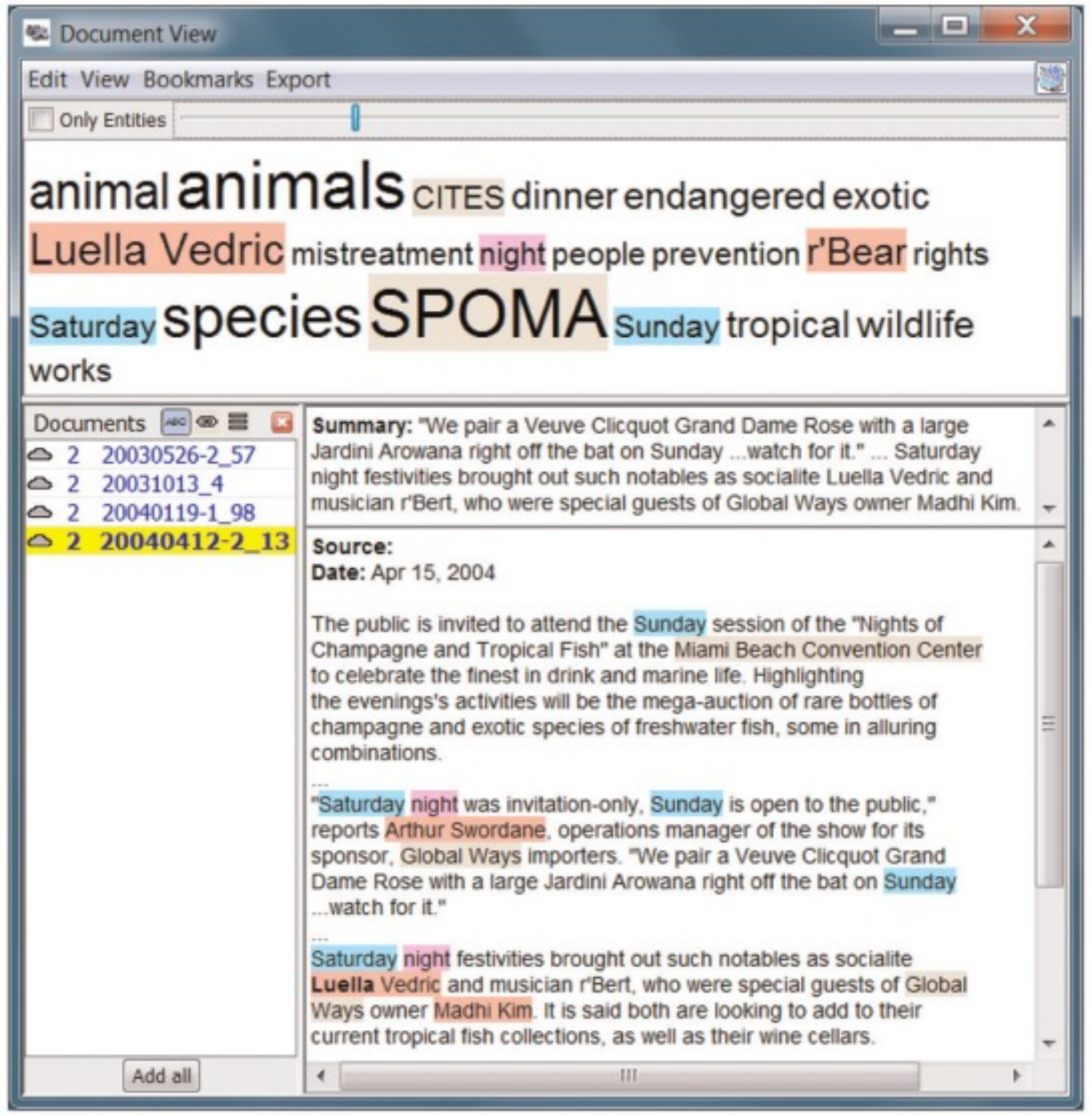
Documents shown in ~~tabs~~ with entities highlighted. Not enough information.



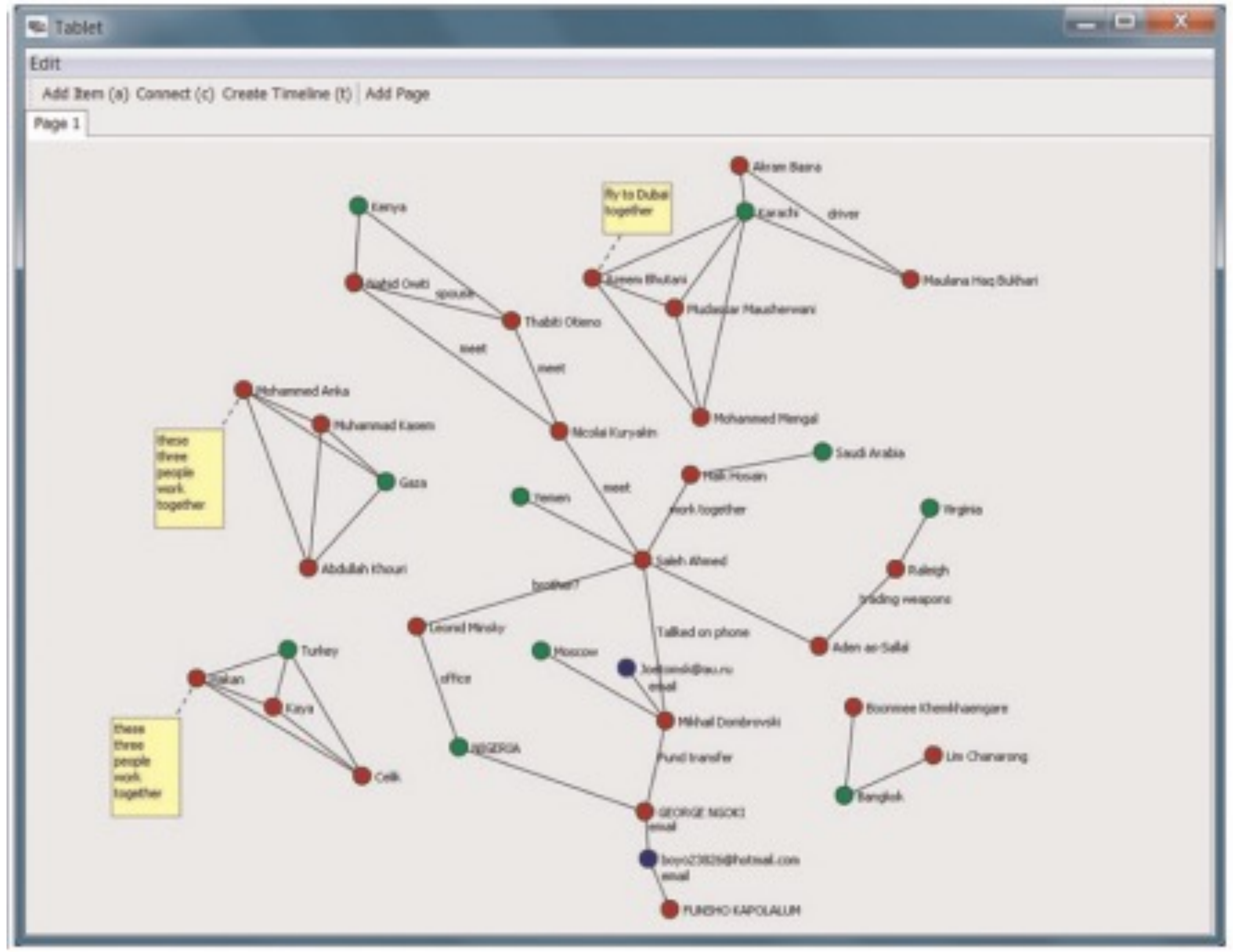
Manual note taking with ~~automated layout~~. Not enough flexibility.

Example: Jigsaw

REVISED Visual Encoding / Interaction Paradigm:



Documents shown in scrollable list with tag cloud and 1 sentence summary.



Freestyle organization (drag & drop), with ability to draw links and add post-it notes.

Example: DIVA

Drug-Drug Interaction Visual Analytics (DIVA)

Stakeholder - U.S. Food and Drug Administration (FDA) ▪ *2017-2018*

Tabassum Kakar

https://tkakar.github.io/Website/DIVA.html

Example: DIVA

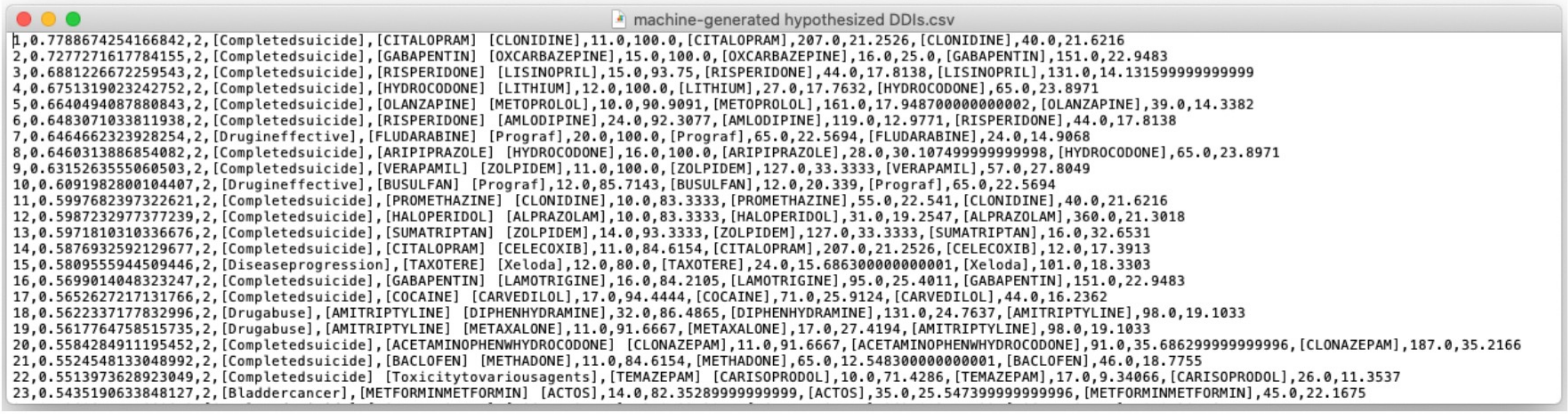
Drug-Drug Interaction Visual Analytics (DIVA)

Stakeholder - U.S. Food and Drug Administration (FDA) ▪ *2017-2018*

Tabassum Kakar

Domain Situation: FDA Analysts need to detect adverse reactions caused by drug-drug interactions (DDI).

Data / Tasks: Validate potential DDI signals in ML output.


machine-generated hypothesized DDIs.csv

```
1,0.7788674254166842,2,[Completedsuicide],[CITALOPRAM] [CLONIDINE],11.0,100.0,[CITALOPRAM],207.0,21.2526,[CLONIDINE],40.0,21.6216
2,0.7277271617784155,2,[Completedsuicide],[GABAPENTIN] [OXCARBAZEPINE],15.0,100.0,[OXCARBAZEPINE],16.0,25.0,[GABAPENTIN],151.0,22.9483
3,0.6881226672259543,2,[Completedsuicide],[RISPERIDONE] [LISINOPRIL],15.0,93.75,[RISPERIDONE],44.0,17.8138,[LISINOPRIL],131.0,14.131599999999999
4,0.6751319023242752,2,[Completedsuicide],[HYDROCODONE] [LITHIUM],12.0,100.0,[LITHIUM],27.0,17.7632,[HYDROCODONE],65.0,23.8971
5,0.6640494087880843,2,[Completedsuicide],[OLANZAPINE] [METOPROLOL],10.0,90.9091,[METOPROLOL],161.0,17.948700000000002,[OLANZAPINE],39.0,14.3382
6,0.6483071033811938,2,[Completedsuicide],[RISPERIDONE] [AMLODIPINE],24.0,92.3077,[AMLODIPINE],119.0,12.9771,[RISPERIDONE],44.0,17.8138
7,0.6464662323928254,2,[Drugineffective],[FLUDARABINE] [Prograf],20.0,100.0,[Prograf],65.0,22.5694,[FLUDARABINE],24.0,14.9068
8,0.6460313886854082,2,[Completedsuicide],[ARIPIPRAZOLE] [HYDROCODONE],16.0,100.0,[ARIPIPRAZOLE],28.0,30.107499999999998,[HYDROCODONE],65.0,23.8971
9,0.6315263555060503,2,[Completedsuicide],[VERAPAMIL] [ZOLPIDEM],11.0,100.0,[ZOLPIDEM],127.0,33.3333,[VERAPAMIL],57.0,27.8049
10,0.6091982800104407,2,[Drugineffective],[BUSULFAN] [Prograf],12.0,85.7143,[BUSULFAN],12.0,20.339,[Prograf],65.0,22.5694
11,0.5997682397322621,2,[Completedsuicide],[PROMETHAZINE] [CLONIDINE],10.0,83.3333,[PROMETHAZINE],55.0,22.541,[CLONIDINE],40.0,21.6216
12,0.5987232977377239,2,[Completedsuicide],[HALOPERIDOL] [ALPRAZOLAM],10.0,83.3333,[HALOPERIDOL],31.0,19.2547,[ALPRAZOLAM],360.0,21.3018
13,0.5971810310336676,2,[Completedsuicide],[SUMATRIPTAN] [ZOLPIDEM],14.0,93.3333,[ZOLPIDEM],127.0,33.3333,[SUMATRIPTAN],16.0,32.6531
14,0.5876932592129677,2,[Completedsuicide],[CITALOPRAM] [CELECOXIB],11.0,84.6154,[CITALOPRAM],207.0,21.2526,[CELECOXIB],12.0,17.3913
15,0.5809555944509446,2,[Diseaseprogression],[TAXOTERE] [Xeloda],12.0,80.0,[TAXOTERE],24.0,15.686300000000001,[Xeloda],101.0,18.3303
16,0.5699014048323247,2,[Completedsuicide],[GABAPENTIN] [LAMOTRIGINE],16.0,84.2105,[LAMOTRIGINE],95.0,25.4011,[GABAPENTIN],151.0,22.9483
17,0.5652627217131766,2,[Completedsuicide],[COCAINE] [CARVEDILOL],17.0,94.4444,[COCAINE],71.0,25.9124,[CARVEDILOL],44.0,16.2362
18,0.5622337177832996,2,[Drugabuse],[AMITRIPTYLINE] [DIPHENHYDRAMINE],32.0,86.4865,[DIPHENHYDRAMINE],131.0,24.7637,[AMITRIPTYLINE],98.0,19.1033
19,0.5617764758515735,2,[Drugabuse],[AMITRIPTYLINE] [METAXALONE],11.0,91.6667,[METAXALONE],17.0,27.4194,[AMITRIPTYLINE],98.0,19.1033
20,0.5584284911195452,2,[Completedsuicide],[ACETAMINOPHENWHYDROCODONE] [CLONAZEPAM],11.0,91.6667,[ACETAMINOPHENWHYDROCODONE],91.0,35.686299999999996,[CLONAZEPAM],187.0,35.2166
21,0.5524548133048992,2,[Completedsuicide],[BACLOFEN] [METHADONE],11.0,84.6154,[METHADONE],65.0,12.548300000000001,[BACLOFEN],46.0,18.7755
22,0.5513973628923049,2,[Completedsuicide] [Toxicitytovariousagents],[TEMAZEPAM] [CARISOPRODOL],10.0,71.4286,[TEMAZEPAM],17.0,9.34066,[CARISOPRODOL],26.0,11.3537
23,0.5435190633848127,2,[Bladdercancer],[METFORMINMETFORMIN] [ACTOS],14.0,82.35289999999999,[ACTOS],35.0,25.547399999999996,[METFORMINMETFORMIN],45.0,22.1675
```

Example of a few Drug-drug Interaction (DDI) signals generated by machine-learning

https://tkakar.github.io/Website/DIVA.html

Example: DIVA

User Research

Target Users

Our users for DIVA are drug safety evaluators at the Division of Pharmacovigilance at the FDA who are responsible for the detection adverse drug reactions (signals).

Key Research Questions and Areas

- How drug safety reports are analyzed?
- What are the key challenges users currently face?
- Observe and understand users' current review process.

Methodology

The user research took a total of two weeks and included the following activities with five users.

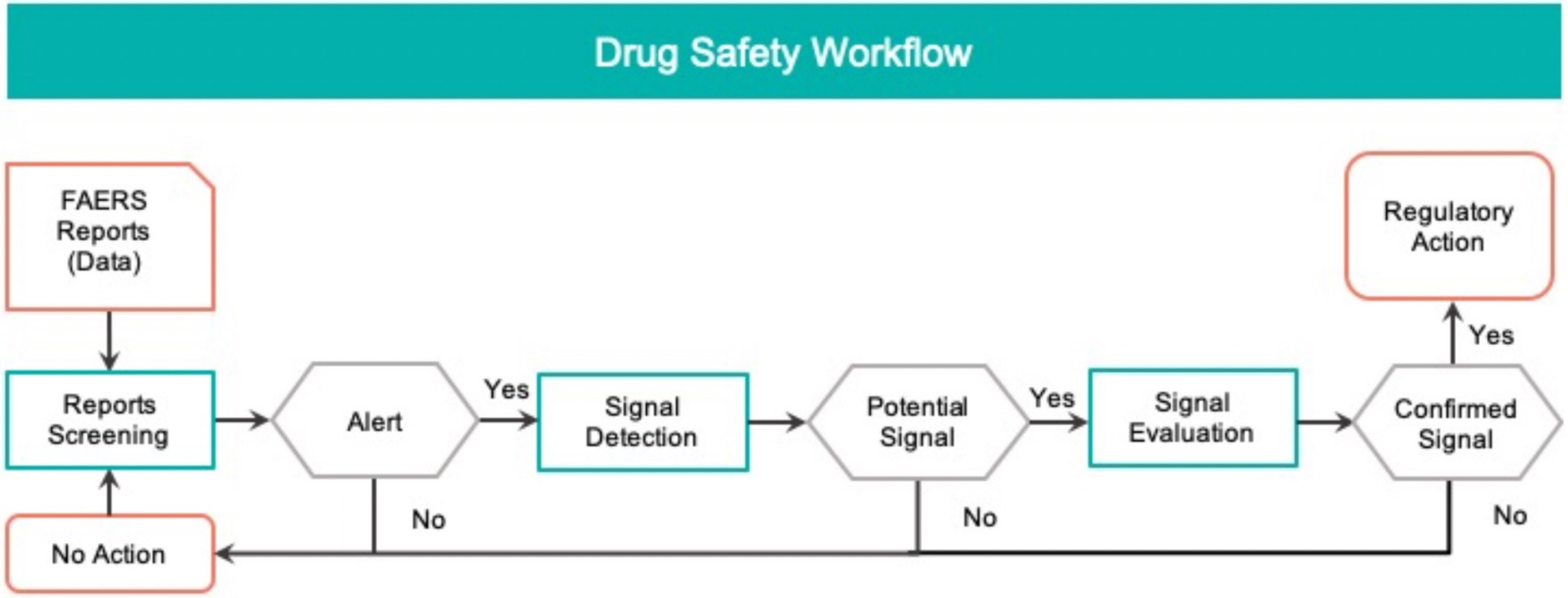
- **Week 1 (Phase 1).** 45 minutes remote interviews to get an overview of their workflow.
- **Week 2 (Phase 2a).** 30 minutes contextual inquiry to observe users at work.
- **Week 2 (Phase 2b).** 30 minutes in person interviews for follow-up discussion and clarification.

https://tkakar.github.io/Website/DIVA.html

Example: DIVA

User Workflow for Signal Detection

Based on our interviews in Phase 1, the drug safety review workflow is summarized below.

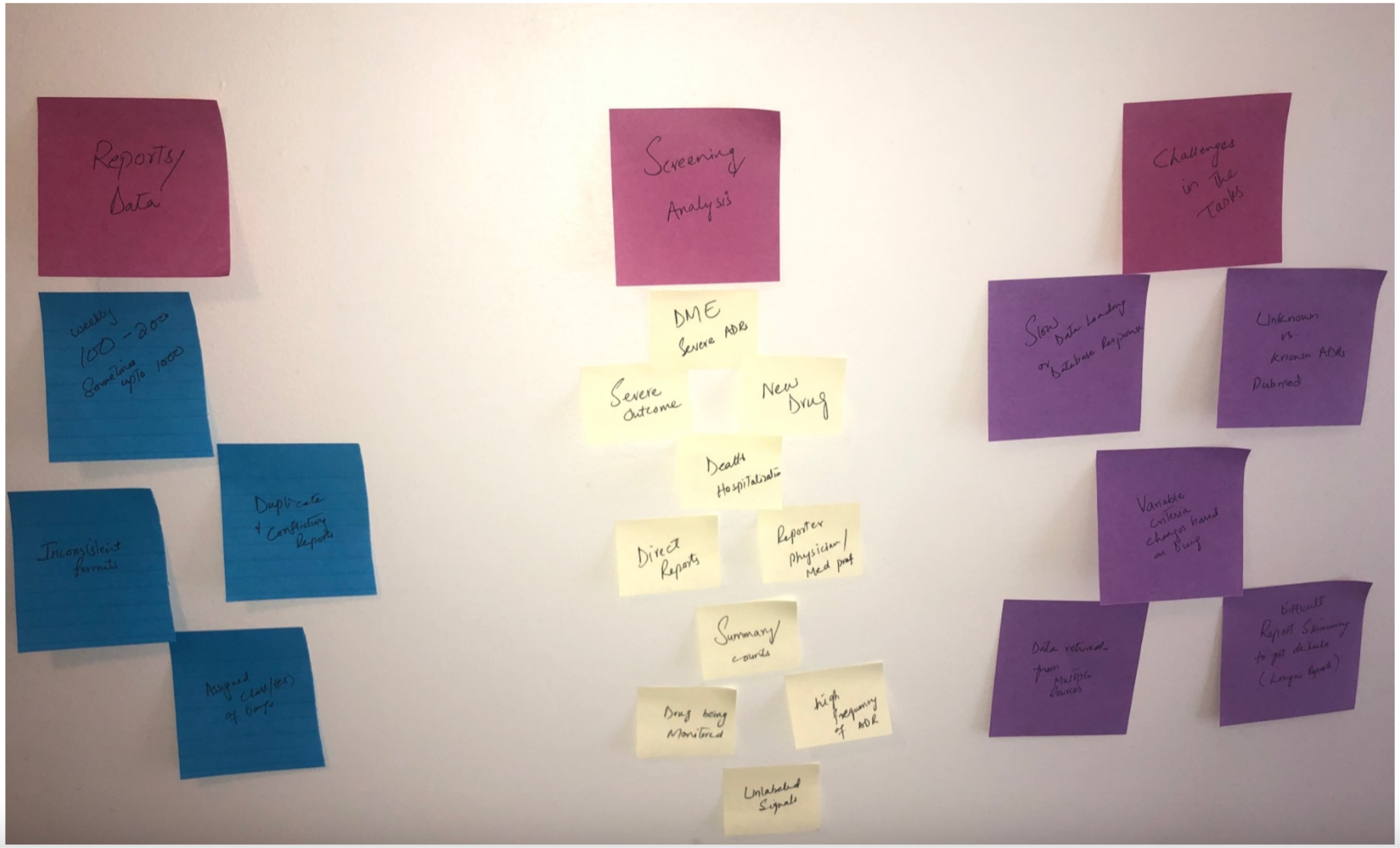


https://tkakar.github.io/Website/DIVA.html

Example: DIVA

Affinity Diagram to Categorize Users' Needs

Based on the user research, an affinity diagram was created to identify and group users' needs and concerns.



https://tkakar.github.io/Website/DIVA.html

Example: DIVA

Key Findings from User Research

Based on the user research, the following features were identified as high priority for the visual analytics to help in the analysis of the DDI signals.

- Provide an overview of all DDI signals.
- Allow analysts to segment and prioritize signals.
- Integrate information about previously known (discovered) signals.
- Facilitate identification of unknown (undiscovered) signals.
- Facilitate identification of severe adverse reactions.
- Access to raw reports supporting the signals.

Design Process

DDI Signals Data to be Visualized

Based on our user research, we integrated domain knowledge into an already complex DDI signal generated by machine learning. The final data model to be visualized consists of the following elements.

- The two interacting drugs (DDI).
- A set of adverse reactions associated with the DDI.
- A numeric score (machine learning metric) that represents the significance of the signal.
- Domain knowledge about the signal status being known or known.
- Domain knowledge about the signal having a severe reaction (DME).

Literature Review

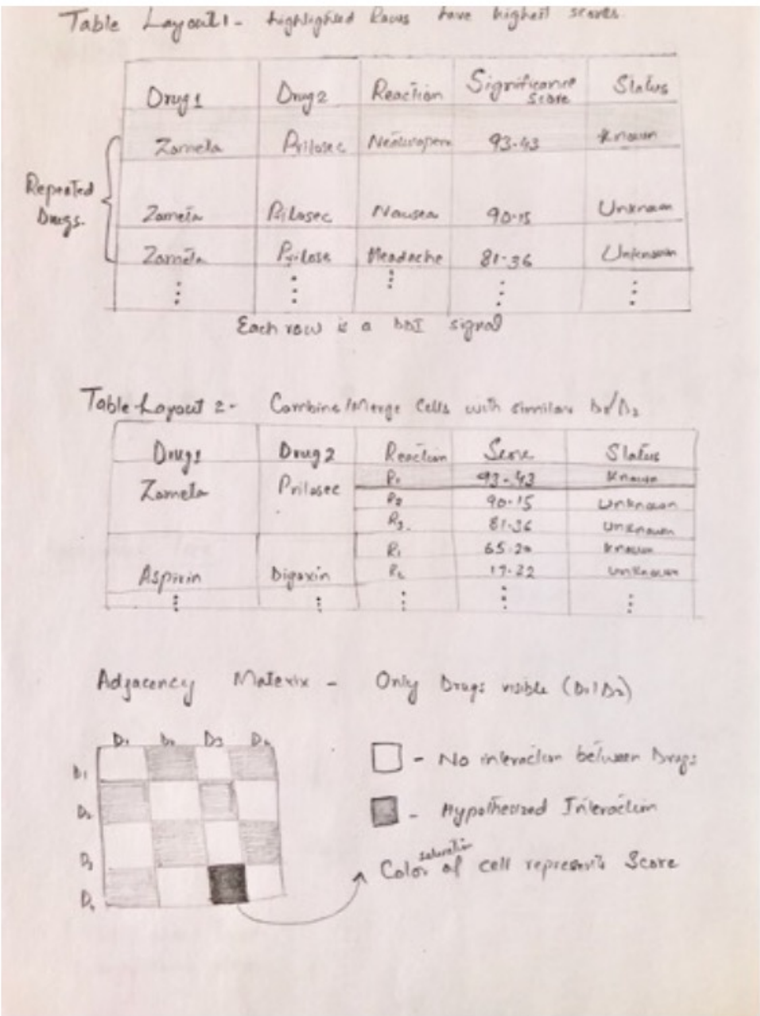
To start, we reviewed academic research papers and other published best practices for designing visualizations for machine-learning output and drug related information. In addition to the required features, we learned some best practices and guidelines for designing multiple coordinated views - such as details on demand, selecting chart types based on the data and task, and direct manipulation of data.

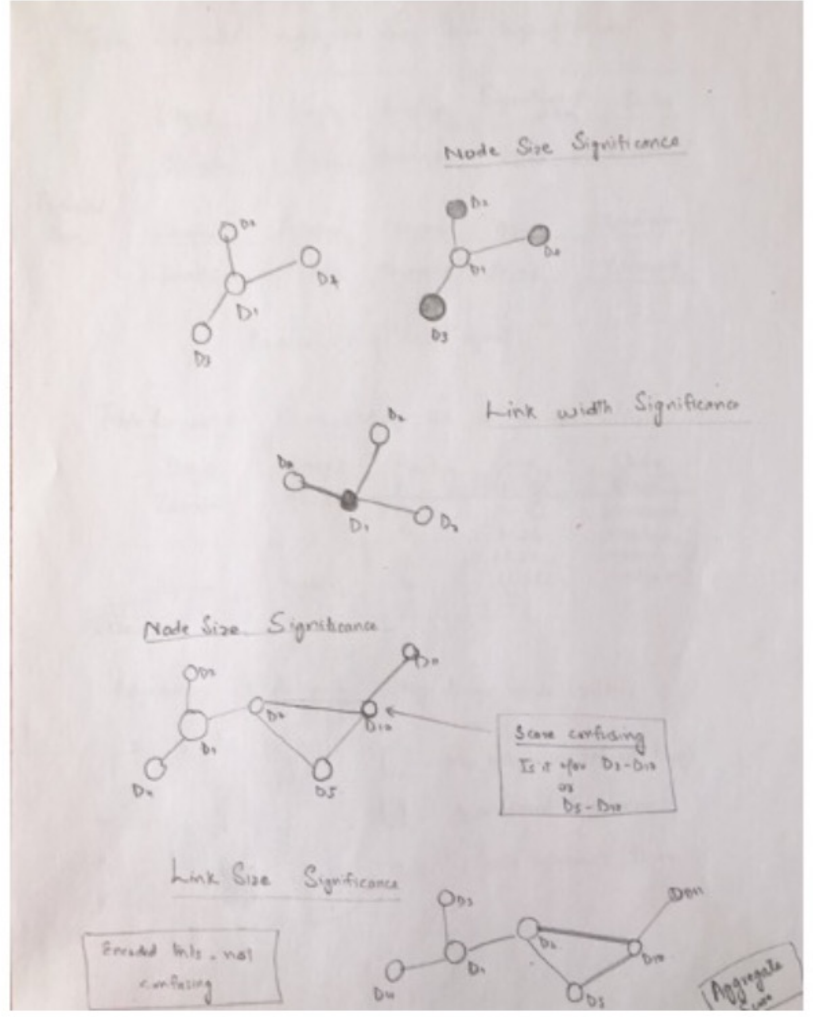
https://tkakar.github.io/Website/DIVA.html

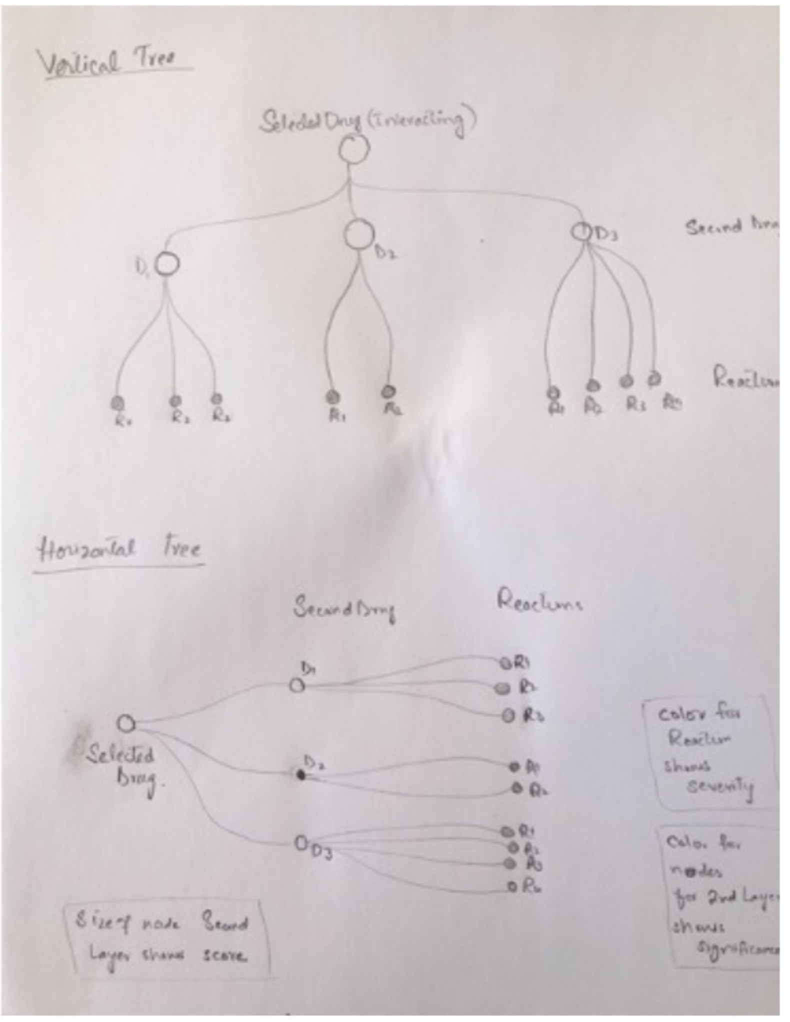
Example: DIVA

Visual Design and Prototypes to Get User Feedback

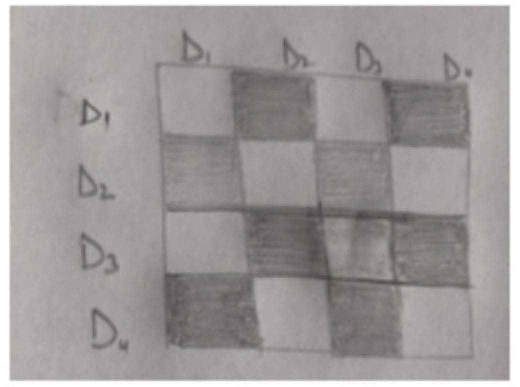
Given the complexity of the DDI signals and user tasks, designing a single visualization is not feasible. Based on literature review and heuristic analysis of different visualization techniques for the underlying data, we considered various visualizations.



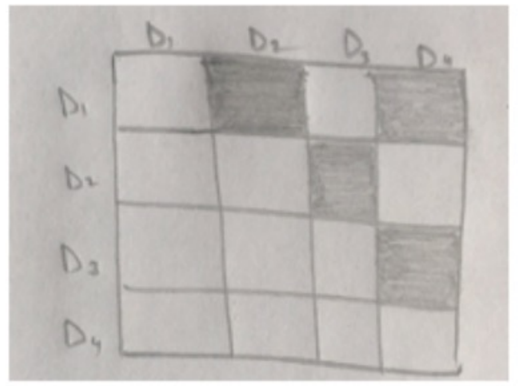




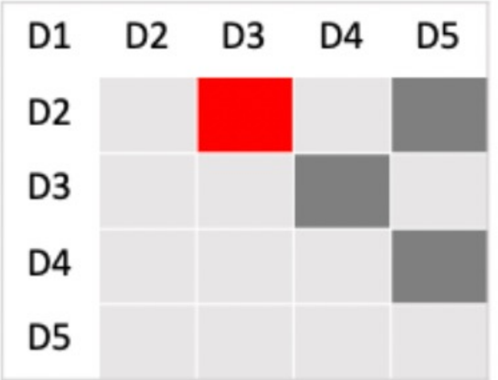
To help users explore the set of all the machine-generated drug-drug interactions (DDIs) the first suitable candidate is an Adjacency Matrix (below figures). Due to sparsity of the matrix for real data, alternatively, a node-link diagram is designed.



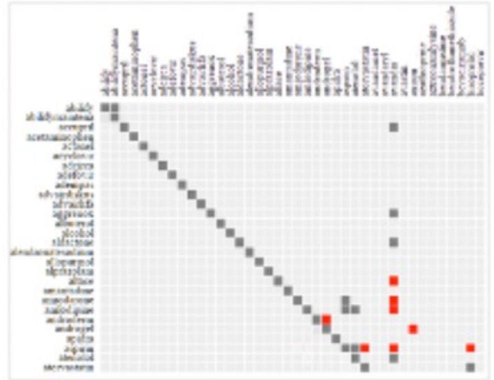
Each axis represents four drugs (D_1,D_2, etc.). Shaded cells represent an interaction between two drugs exist in the machine generated signals.



No order between DDIs. Interaction between D_1-D_2 is same as D_2-D_1. Only used the only the upper triangular matrix.

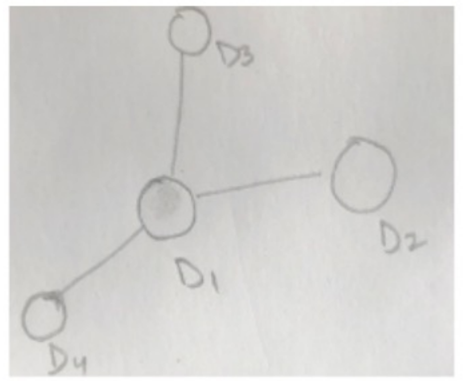


Color of the cell is mapped to the significance score of the DDI interaction used by the machine learning algorithm to rank signals.

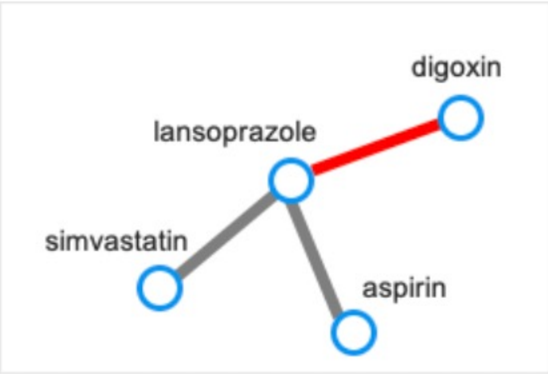


Adjacency matrix using D3.js with real DDI data has high sparsity, due to hundreds of drugs having unequal number of interactions with each other

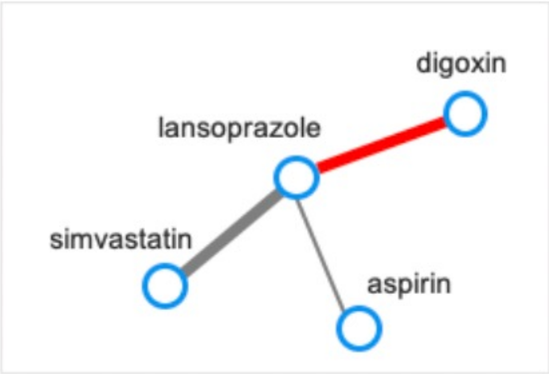
In the node-link diagrams (following figure), the nodes are mapped to the drugs and the links depict the interaction between drugs. The visual encodings, i.e., color and size of the links (Screening view) and nodes (Triage view) are mapped to the significance score of the signals and status of the reactions, respectively.



Nodes represent drugs, and links between nodes represent a machine-generated drug-drug interaction.



Links are encoded with score using color hue. Red color shows a high score, grey means a low score.



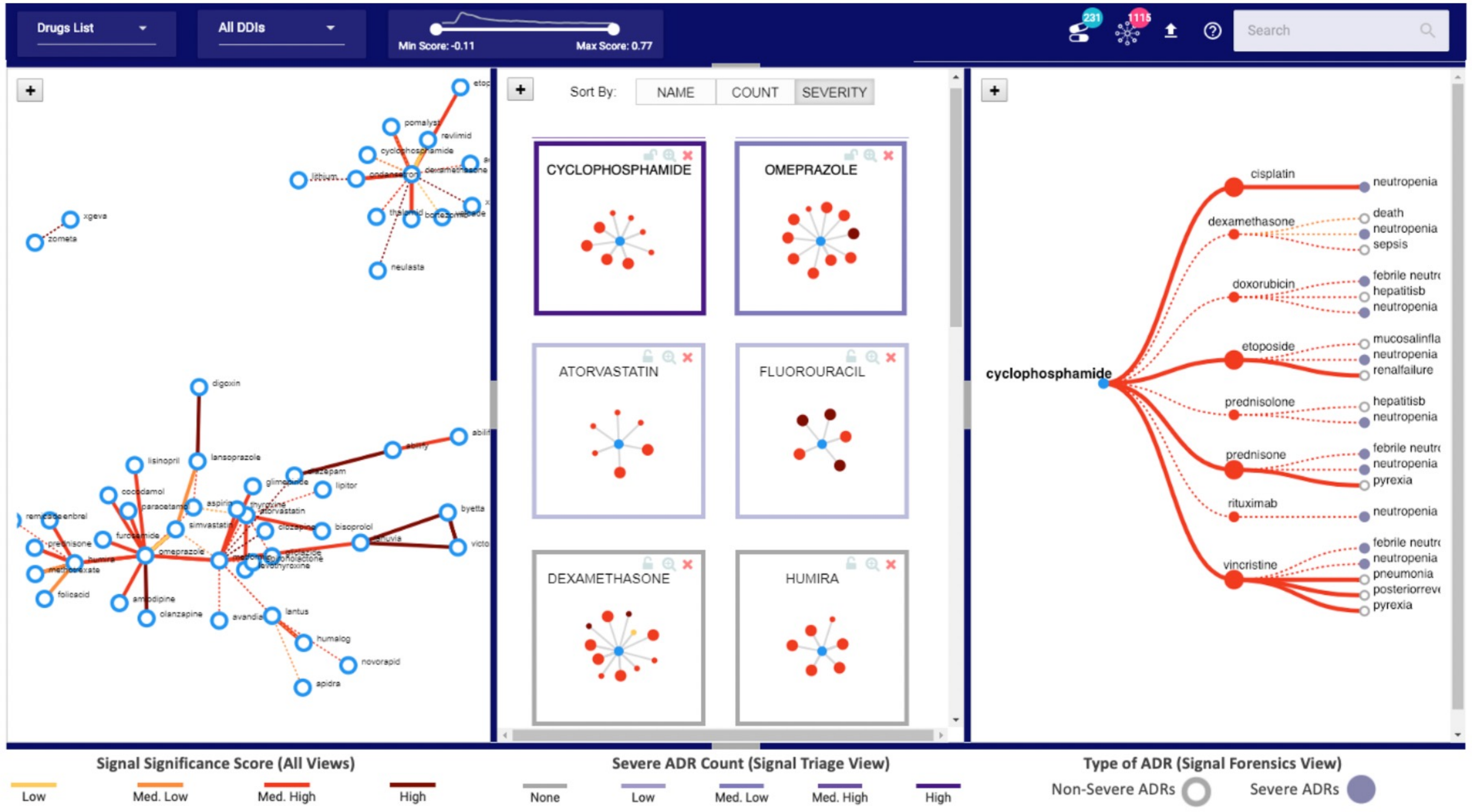
The signal status (known or unknown) is mapped to the link size. Thin = known signal and thick= unknown.

In the final design, color saturation is used to encode score. Both link shape and size are used to encode status.

https://tkakar.github.io/Website/DIVA.html

The node-link design for all the views are iteratively refined to choose the most suitable visual encodings for mapping the underlying signals data.

Example: DIVA

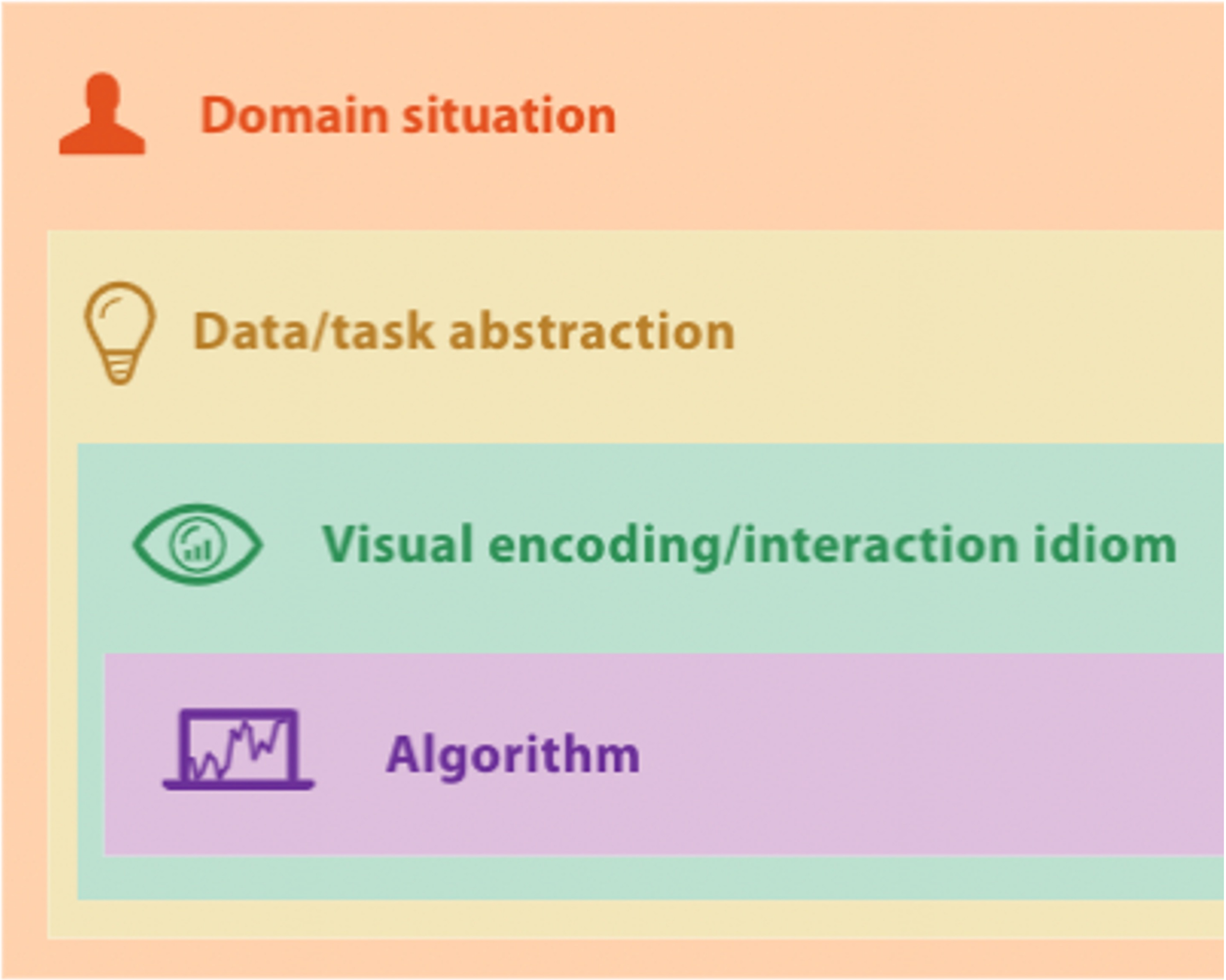


DIVA Interface. **Left:** Screening view to explore the DDI signals. **Middle:** Triage view to prioritize signals related to specific drugs. **Right:** Forensics view to get the adverse reaction details of a drug related signals.

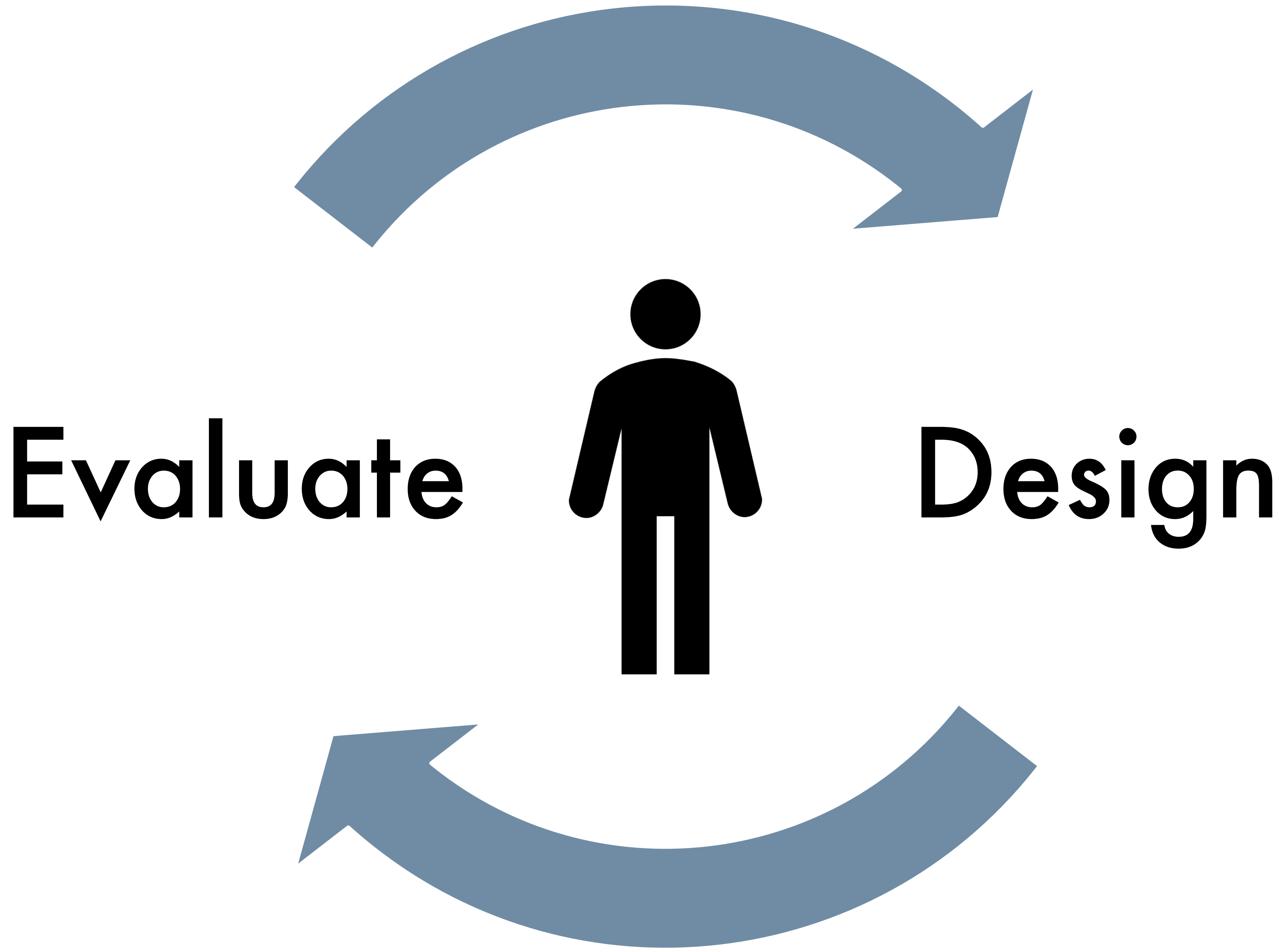
https://tkakar.github.io/Website/DIVA.html

Visualization Design Pipeline

Human-Centered, Iterative Design



Munzner's Nested Model



Let’s take a break! Stretch, go for a walk, be social ☺
Be back here in 10 mins.

Finish ic-01

Summary

Today we:

- Reviewed Visualization Design Pipeline
- Finished ic-01

ic-01 is DUE today.

hw-01 is OUT today.