Cody Dunne
Northeastern University

Projects, Marks, and Channels

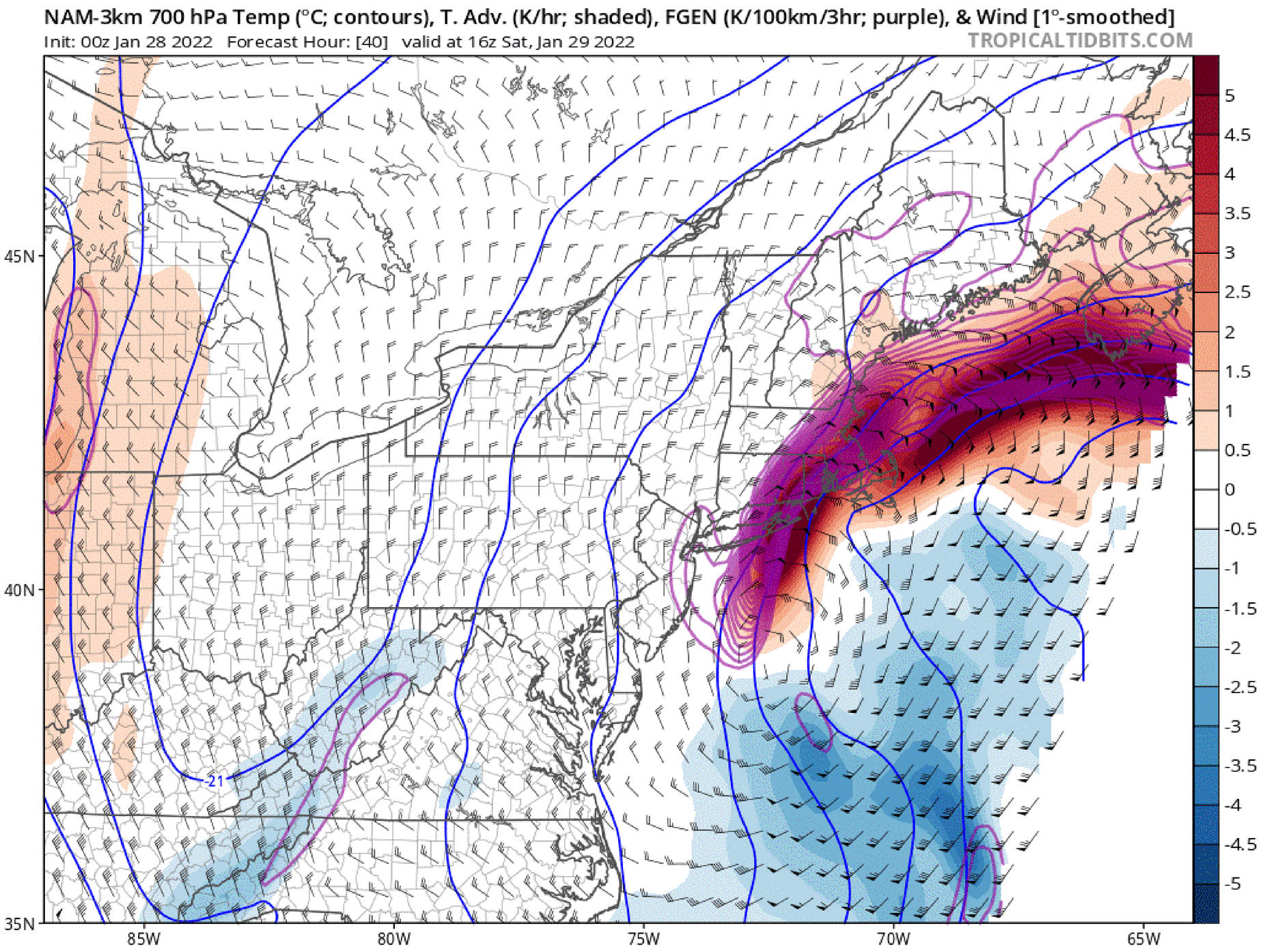
Feel free to interrupt with questions!

Checking In

S-L Assessments

Hall of Fame or Hall of Shame

Prof. Krzystzof Gajos


NAM-3km 700 hPa Temp (°C; contours), T. Adv. (K/hr; shaded), FGEN (K/100km/3hr; purple), & Wind [1°-smoothed]
Init: 00z Jan 28 2022 Forecast Hour: [40] valid at 16z Sat, Jan 29 2022
TROPICALTIDBITS.COM
45N
40N
35N
85W
80W
75W
70W
65W
-21
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
-4.5
-5

Matthew Kolosick - https://twitter.com/blkahn/status/905141924939649024/photo/1

In-class critique: wind maps

10m

Take 5 minutes to talk in your breakout room and critique one of these visualizations.

Room 1: http://hint.fm/wind/ (the original)

Room 2: https://www.accuweather.com/en/us/cambridge/02139/winter-weather-forecast/329319
(click the gray “wind” then “wind flow map”

Room 3: https://www.windy.com/?2022013003,43.676,-68.807,6

Room 4: https://earth.nullschool.net/#current/wind/surface/level/orthographic=-65.91,40.65,3843

Room 5: https://www.windfinder.com/#6/41.1456/-66.3873

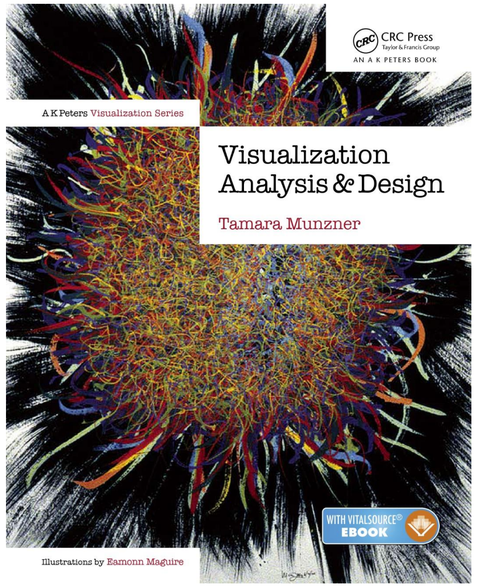
When we return, have someone from your room present findings to the class.

Plan for Today

- Discuss our expectations for projects
- Learn about the building blocks of visualizations

The Nested Model for Visualization Development

Used for your Projects



"Nested Model"

Domain situation
Observe target users using existing tools

Data/task abstraction

Visual encoding/interaction idiom
Justify design with respect to alternatives

Algorithm
Measure system time/memory
Analyze computational complexity

Analyze results qualitatively

Measure human time with lab experiment (*lab study*)

Observe target users after deployment (*field study*)

Measure adoption

Example

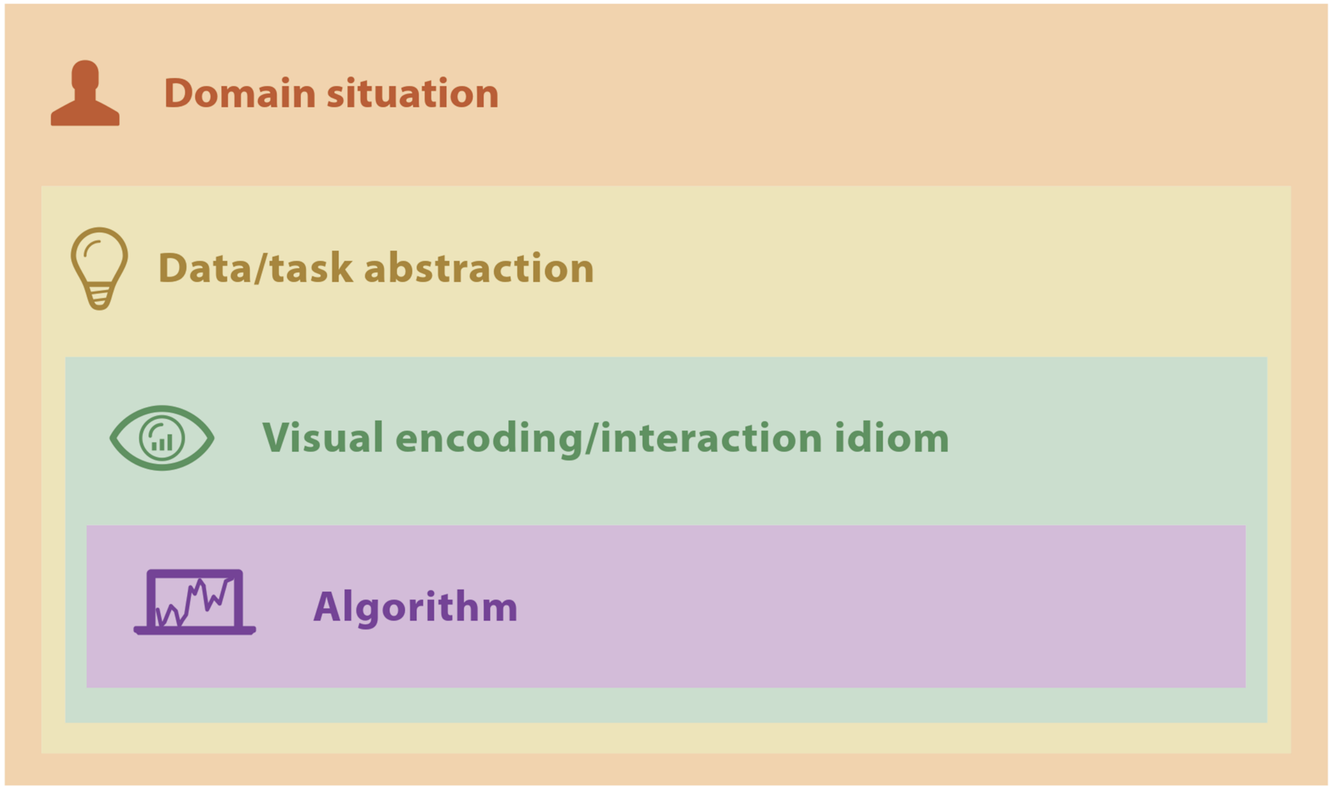
FAA (aviation)

What is the busiest time of day at Logan Airport?

Map vs. Scatter Plot vs. Bar

[Tamara Munzner]

Nested Model



Nested Model

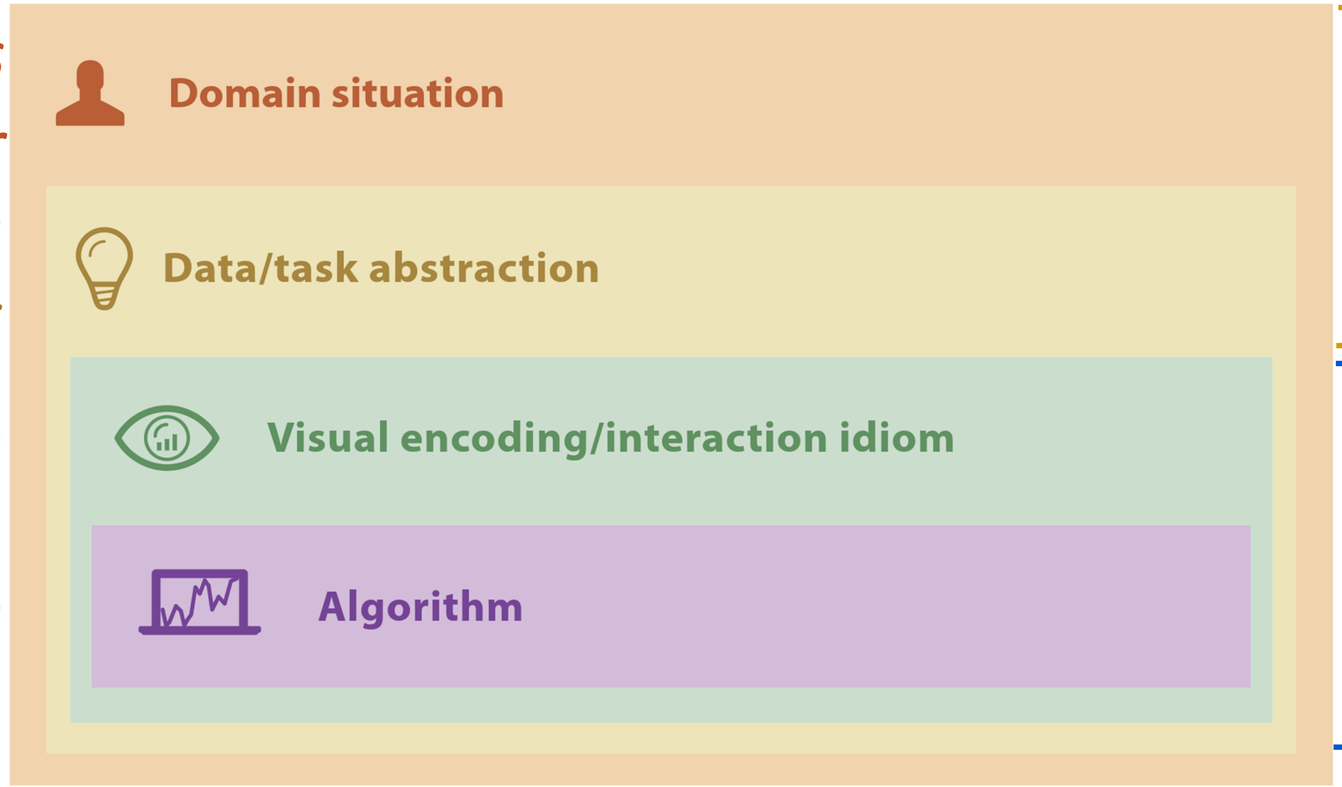
Human-centered design

Designer understands user

Abstract domain tasks

Visualization design

Implementation



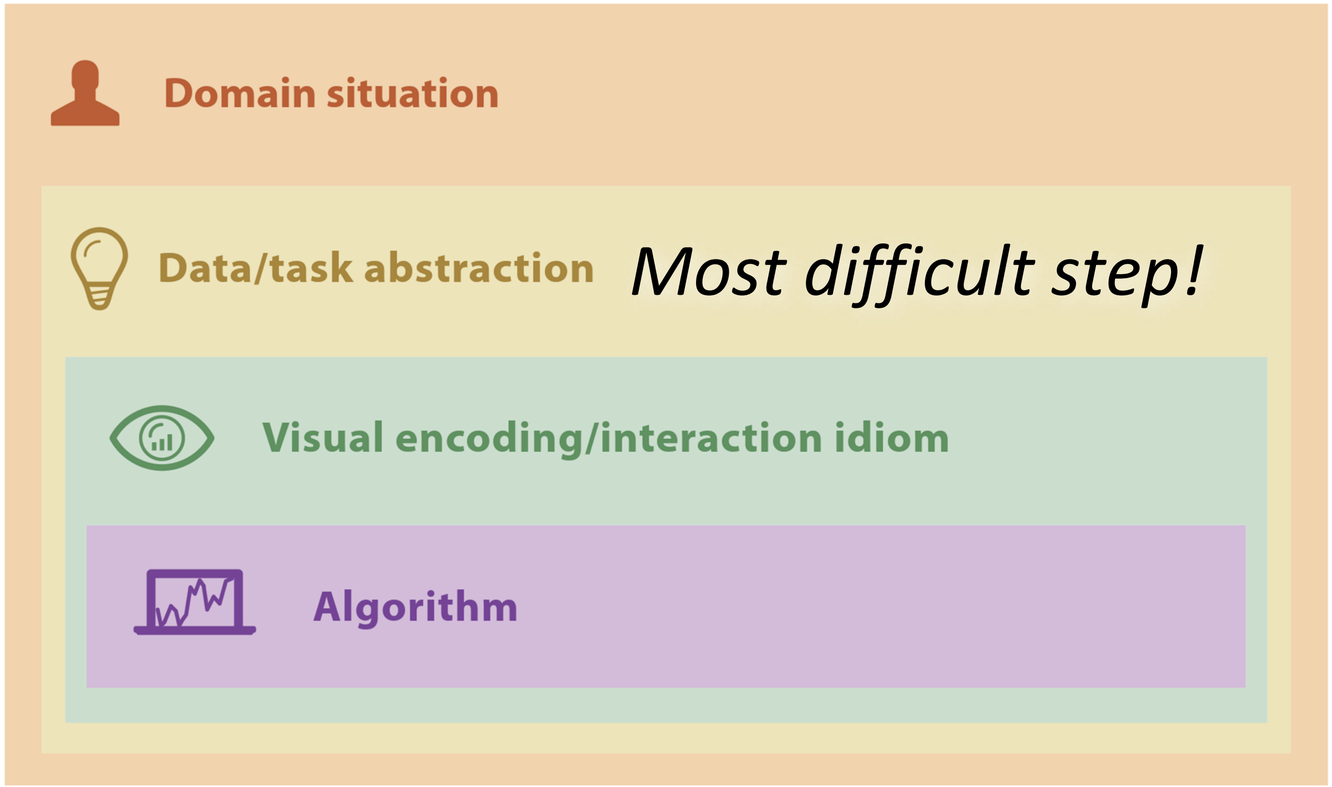
Identified

Designed

Nested Model

Design Study

TOP-DOWN
"problem-driven"

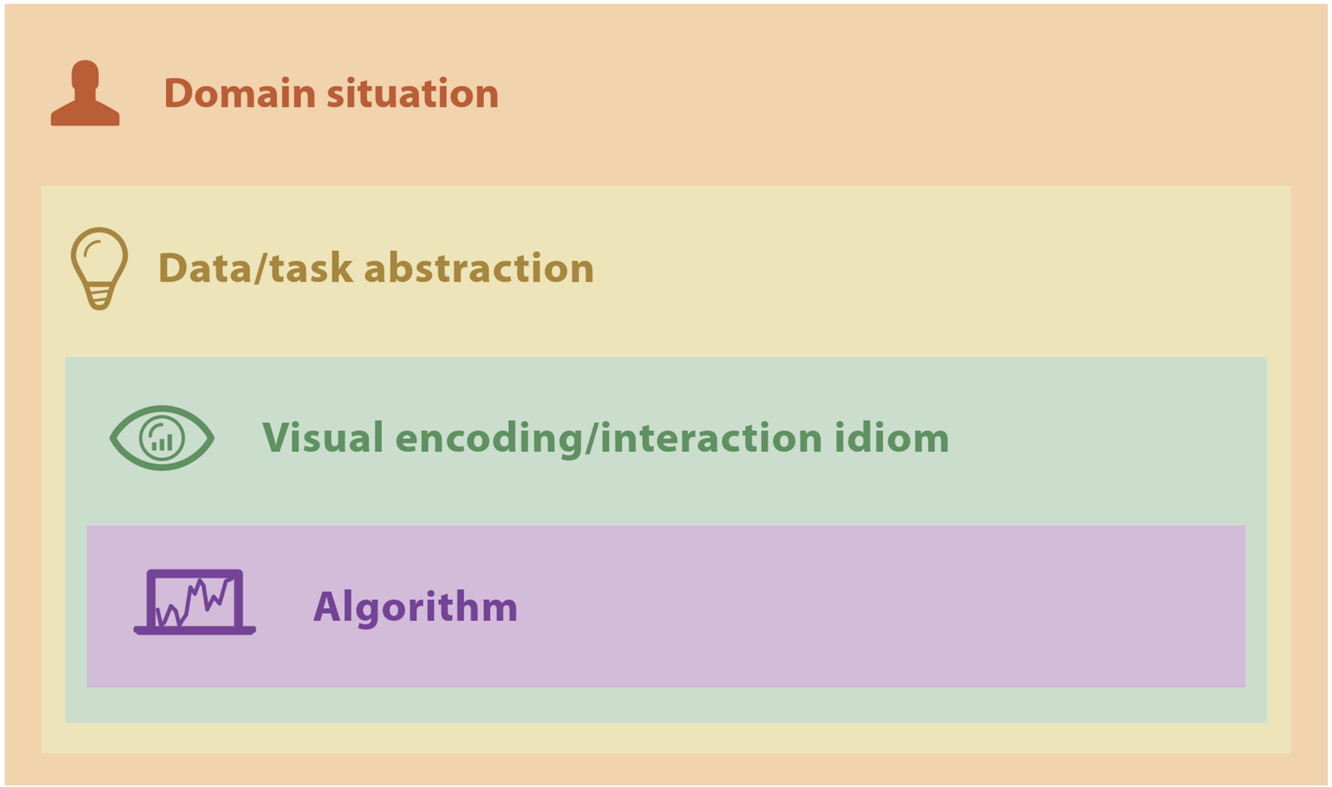


Technique

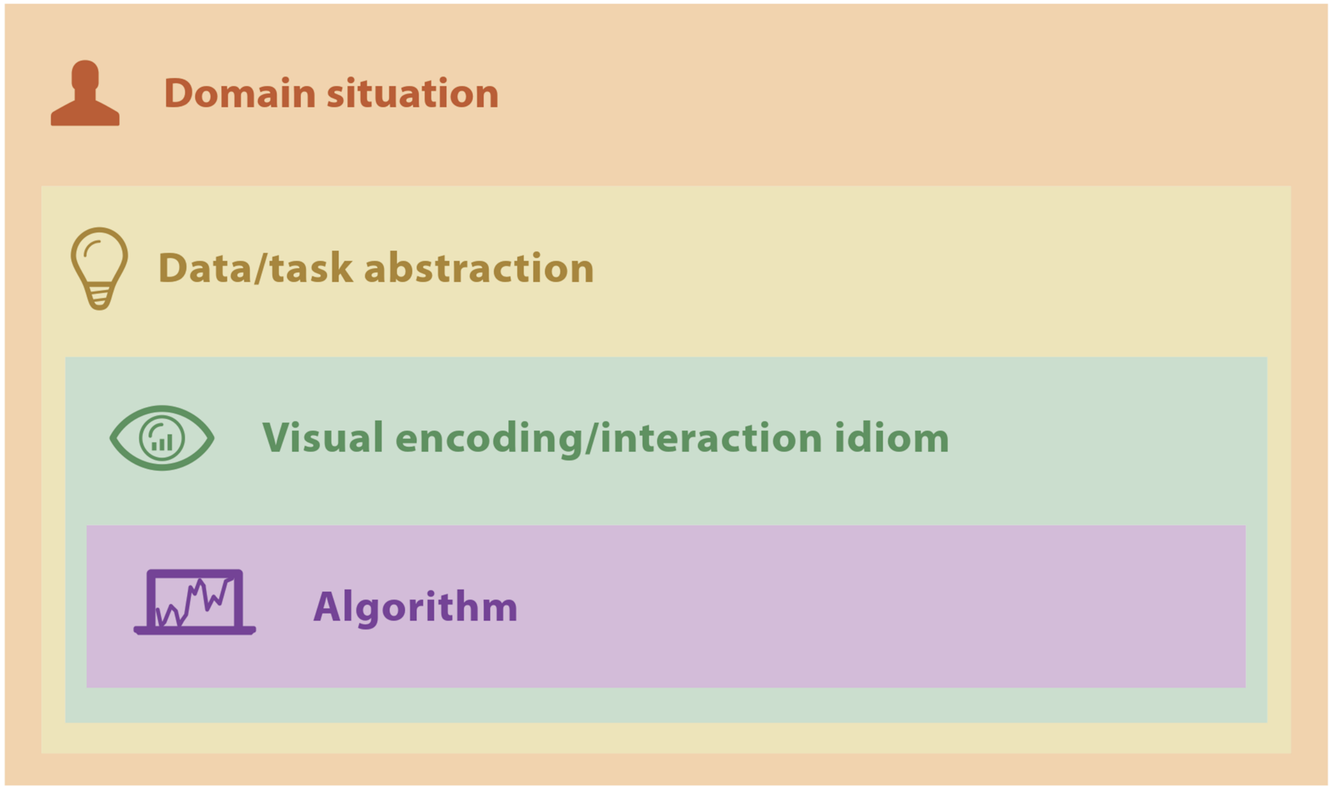
BOTTOM-UP
"technique-driven"

Nested Model

Mistakes propagate through model!

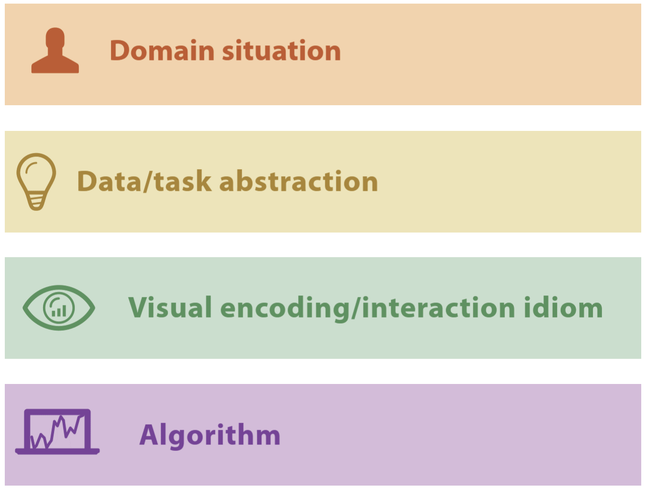


Threats to Validity



Threats to Validity

✓ *Final Project validation*



Final project follow-up

PROJECTS

(Using the nested model via *design study "lite" methodology*)
https://neu-ds-4200-s22.github.io/projects/overview

Experiential Learning Projects

Why are we doing experiential learning?

Design Study "Lite" Methodology (Borkin et al. 2017, Syeda et al. (2020))

- Design studies are a growing and valuable research area.
- Real-world data visualization experience.
- Visualization for exploration and communication.
- A more realistic experience of creating visualizations, and doing work in general.
- Teaches design, interview, evaluation, communication, and feedback techniques difficult to replicate in a classroom.
- Higher-stakes deliverables.
- Professional development.
- Make a positive impact in the community.
- Publication?

Experiential Learning Projects

What are the challenges?

- Real-world data is messy and difficult to gather and process.
- Partners may not have clear goals and expectations.
- There is communication and scheduling overhead, inc. for teaching staff to differentiate assignment grading if necessary.
- Project areas may be too predefined.
- Project areas may be too ambiguous.
- May not actually make a meaningful impact.
- Reduces time for white-room technical education.
- More ambiguous expectations and grading challenges.
- Possible variation in student workload.
- Students may not know they are signing up for Service-Learning in advance (common problem with our registrar).

Examples of Successful Course Projects

(Albeit with different requirements per course)

PROJECT EXAMPLE — WWOVIS

Close and Distant Reading via Named Entity Network Visualization: A Case Study of Women Writers Online

Sarah Campbell* Zheng-yan Yu† Sarah Connell‡ Cody Dunne§

Northeastern University

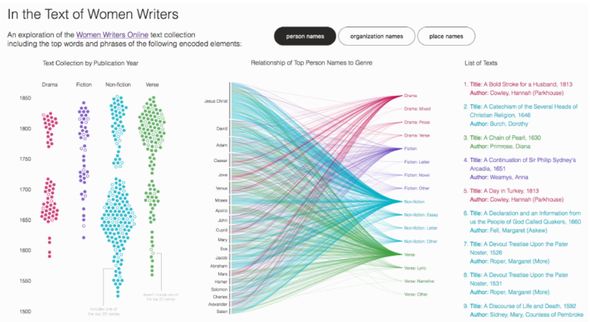


Figure 1: Three linked visualizations showing a named entity network queried from the Women Writers Online textbase. Left: a beeswarm visualization of the genre of each publication by year. Middle: a bipartite node-link visualization of the top 20 named entities connected to the genres of texts they reside in. Right: a list of texts that include at least one top 20 named entity, ordered alphabetically and linked to the full text. Marks are colored categorically by genre: drama is pink ●, fiction is purple ●, non-fiction is blue ●, and verse is green ●. Empty circles show texts that do not include any of the top 20 named entities, e.g. ○.

ABSTRACT

Close reading and distant reading are widely used in digital humanities and can benefit from information visualizations. Digital humanities scholars have curated numerous TEI-encoded textual collections which provide the data necessary for blending both close and distant reading – however we do not have tools to support general users in conducting these blended analyses. In this paper we focus on one such collection: Women Writers Online (WWO). We contribute the design and implementation of a multiple coordinated view network visualization to facilitate close and distant reading in WWO and a transparent view into our iterative design process to help guide future designers and humanists in applying our approach to other textual collections.

Index Terms: Applied computing—Education—Digital libraries and archives; Human-centered computing—Visualization—Visualization application domains—Information visualization; Human-centered computing—Visualization—Visualization techniques—Graph Drawings;

*e-mail: campbell.sar@husky.neu.edu
†e-mail: yu.zheng@husky.neu.edu
‡e-mail: sa.connell@northeastern.edu
§e-mail: c.dunne@northeastern.edu

1 INTRODUCTION

Close and distant reading are two important tools in the digital humanities toolbox which can both benefit from information visualization [6]. **Close reading** is the traditional method for literary criticism. Several visualizations have been developed to support close reading, but existing approaches can benefit from adding supplementary **named entity** information, especially acting persons and their relationships [6]. **Distant reading**, introduced by Moretti [11], alternatively focuses on an abstract view of global features of one or more texts. Network/graph visualizations can be particularly useful for examining relationships between these features and texts for corpus analysis [6]. We designed an interactive visualization to support a blend of close and distant reading – both explorations at scale and text-level investigation.

This paper focuses on the application of our visualization approach in service to the Women Writers Project (WWP). The WWP is a long-term digital humanities research project at Northeastern University, devoted to early modern women's writing and electronic text encoding. The goal of the project is to bring texts from pre-Victorian women writers out of the archive and make them more accessible to a wide audience of teachers, students, scholars, and the general user. We focus on the WWP's major textual collection, Women Writers Online (WWO). WWO is a full-text collection of early women's writing in English. It currently includes full transcriptions, encoded following the standards of the Text Encoding Initiative (TEI), of 407 texts published between 1526 and 1850. In addition to the collection's broad chronological framing, the texts in WWO also represent a very diverse set of genres, ranging from prophecies, religious meditations, petitions, and recipe books to

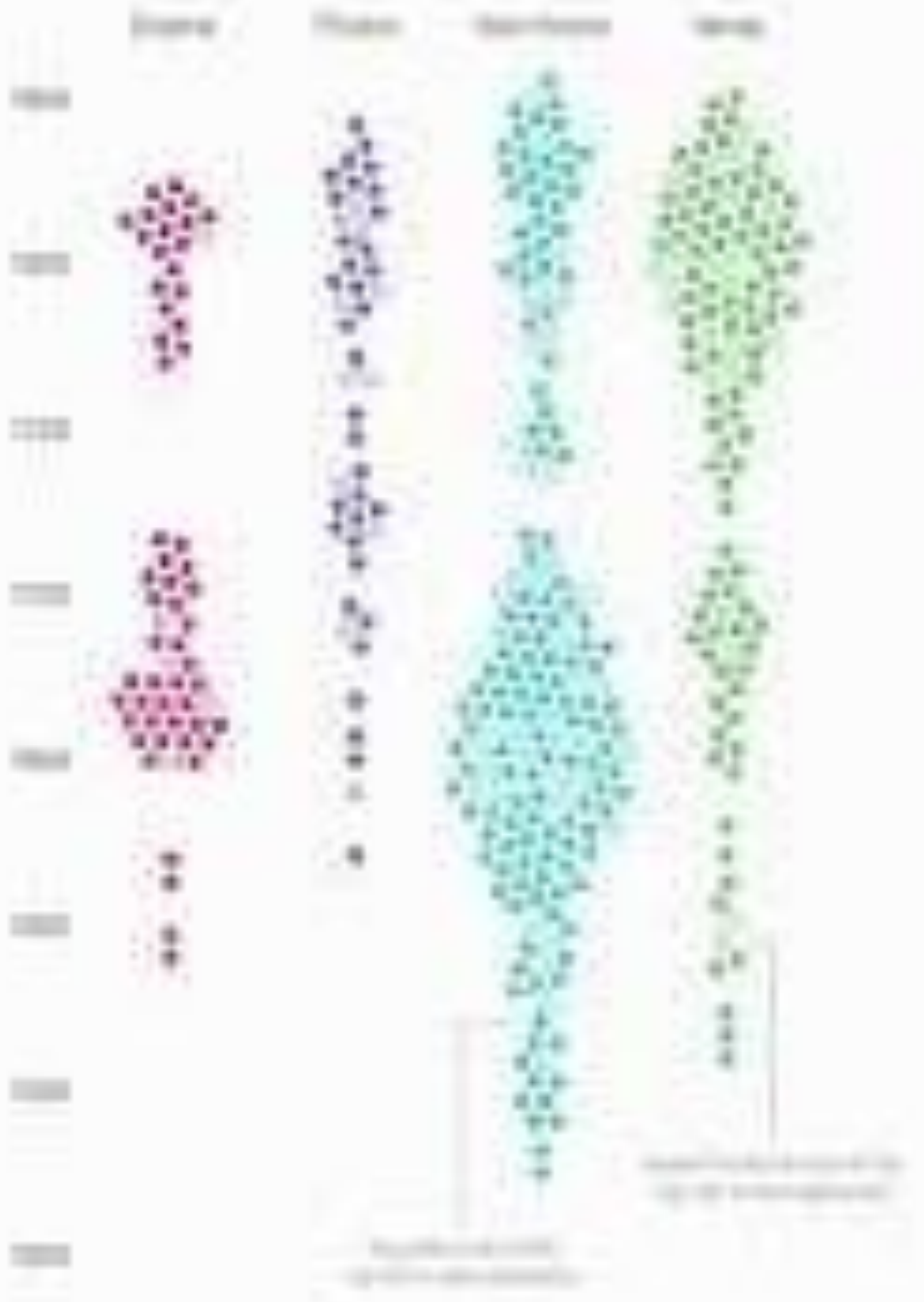
CS 7260 FALL 2017: VISUALIZATION FOR NETWORK SCIENCE

Cambpell et al. VIS4DH 2018

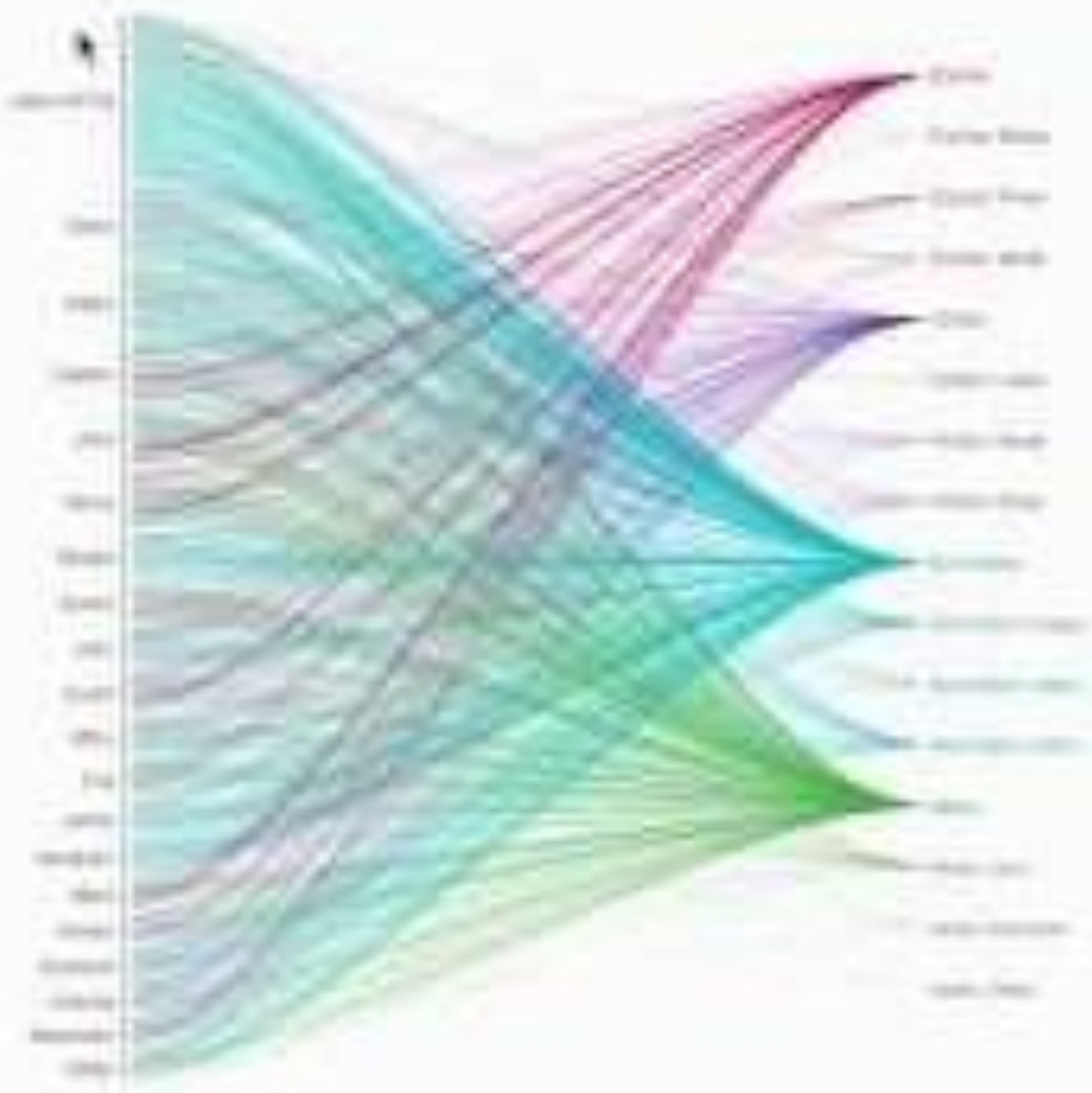
including the top words and phrases of the following encoded elements:

person names | organization names | place names

Text Collection by Publication Year



Relationship of Top Person Names to Genre



List of Texts

PROJECT IDEAS:
VIS + X

Where X = a CS subfield (ML | SEC | NLP | HCC | GAM | NS | SYS | …)

OR

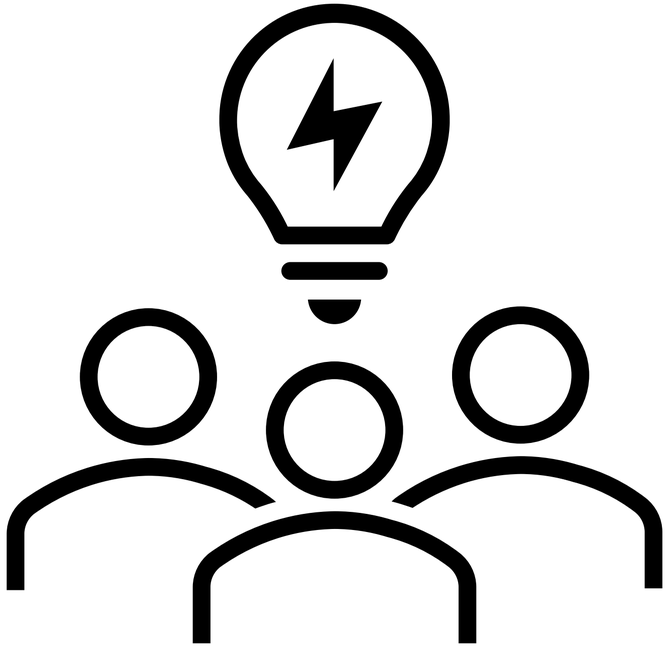
Where X = a domain application (health, energy, transportation, astronomy, crime…)

PROJECTS

What questions do you have for me?

MARKS AND CHANNELS

IN-CLASS EXERCISE



In-class brainstorming: building blocks

9m

1. Take 5 minutes to talk in your breakout room and brainstorm what you think the building blocks to a visualization are.

2. Be prepared to share with the class.

Goals for Today

- Learn the basic visual primitives of visualizations (marks and channels)
- Understand how marks and channels are assembled to make visualizations
- Learn which marks and channels are most effective for a given task (“perceptual ordering”)

Visualization Building Blocks

MARK = basic graphical element in an image

Visualization Building Blocks

CHANNEL = way to control the appearance of marks, independent of the dimensionality of the geometric primitive

Visualization Building Blocks

Marks as Items/Nodes

- Points
- Lines
- Areas

Marks as Links

- Containment
- Connection

Channels :

- Position
  - Horizontal
  - Vertical
  - Both

- Color

- Shape

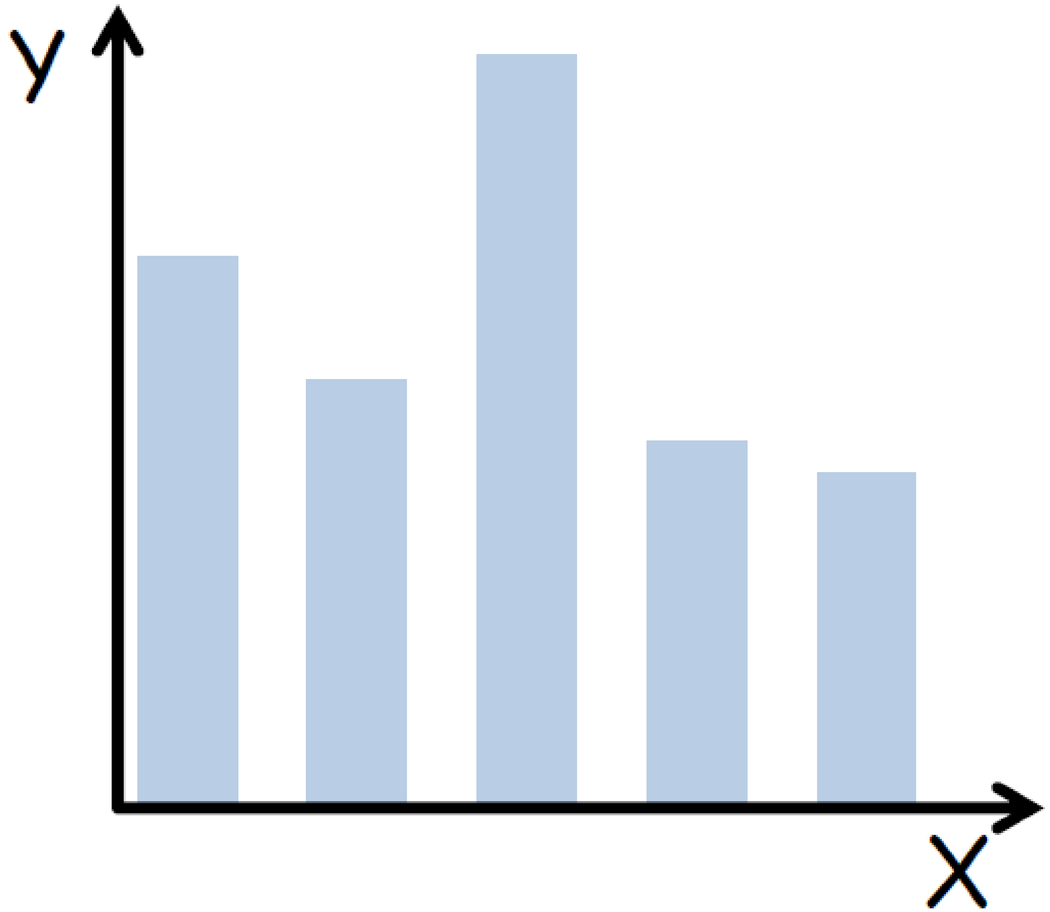
- Tilt

- Size
  - Length
  - Area
  - Volume

Note: these are all really important concepts when it comes time to coding your visualizations...!

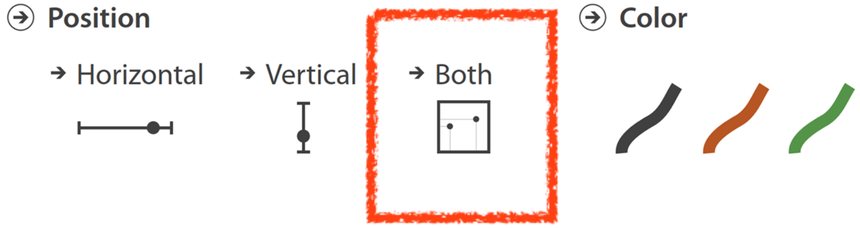
Visualization Building Blocks

of attributes encoded: 2



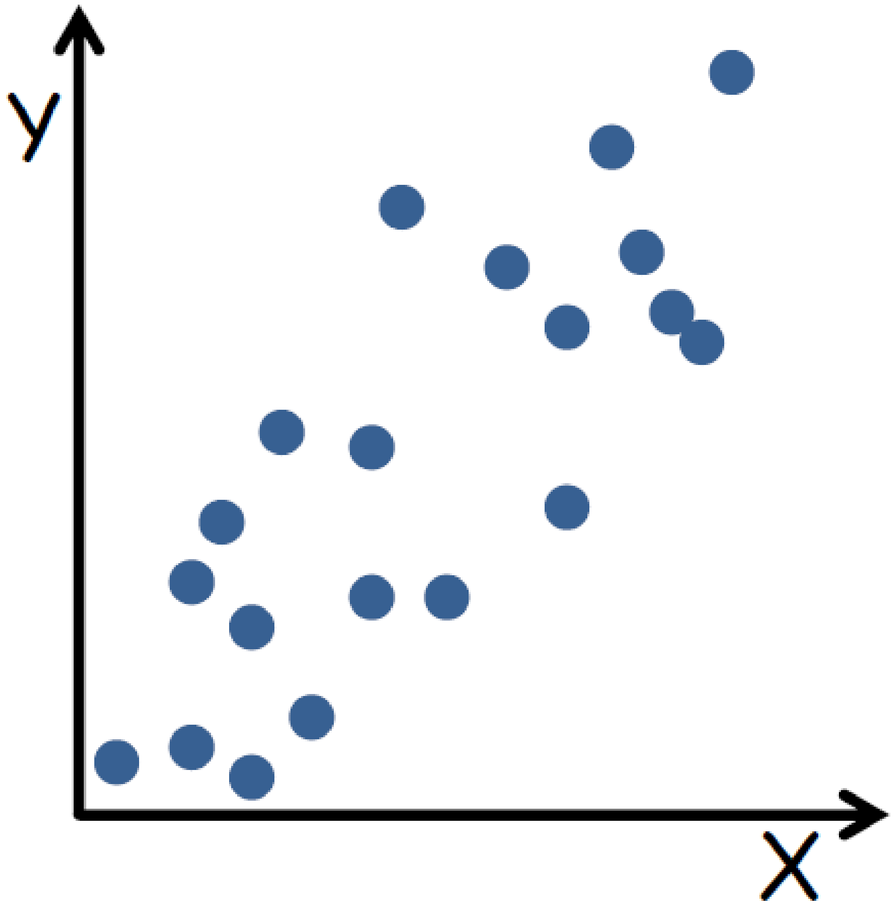
MARK:

CHANNEL :

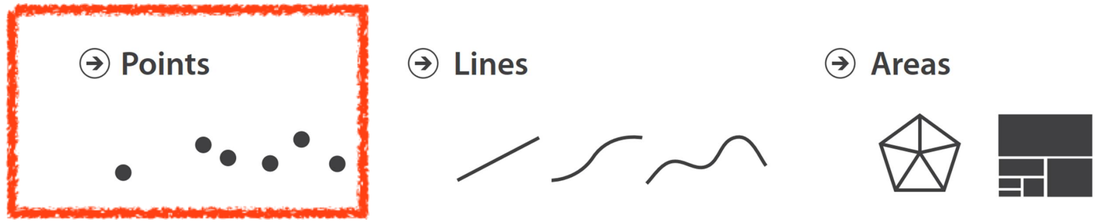


Visualization Building Blocks

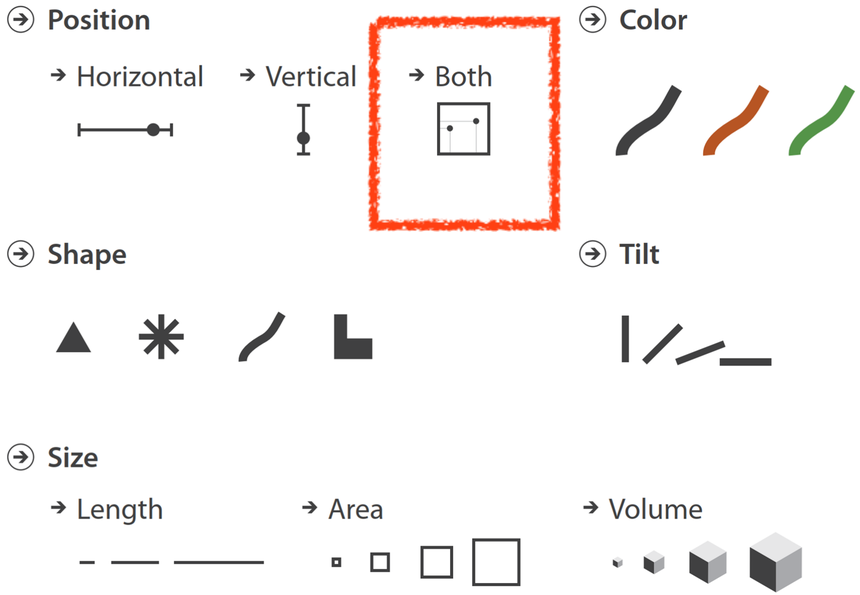
of attributes encoded: 2



MARK:

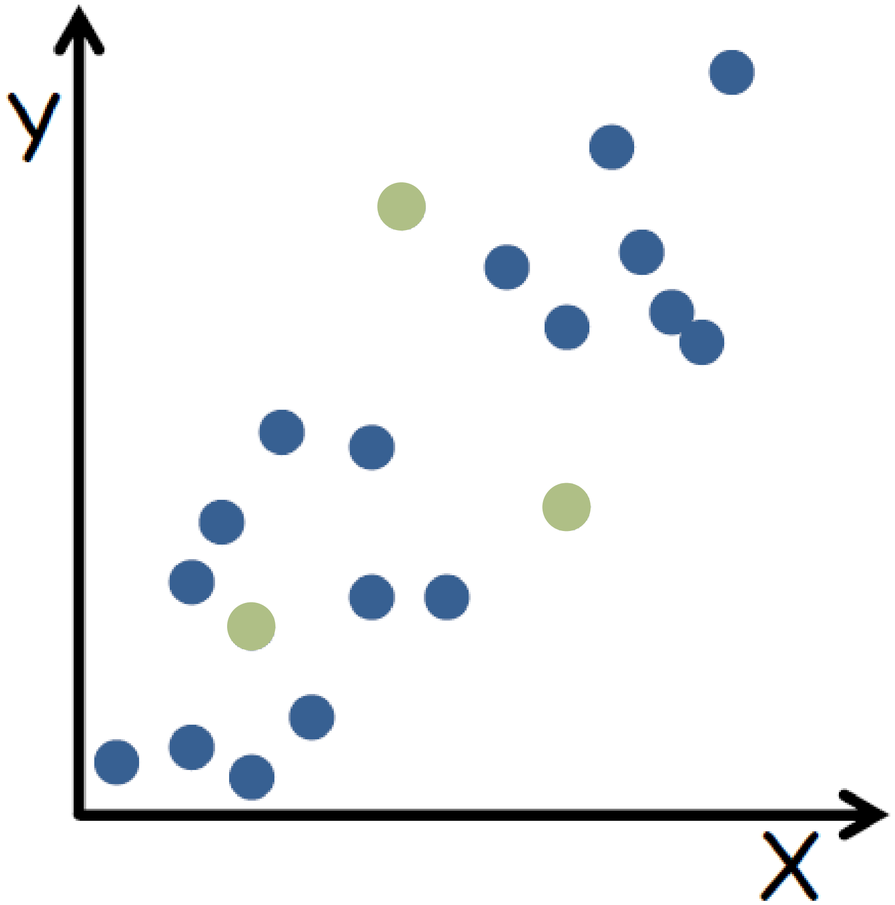


CHANNEL :

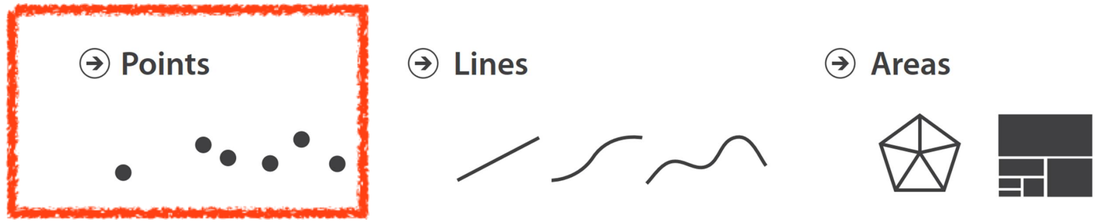


Visualization Building Blocks

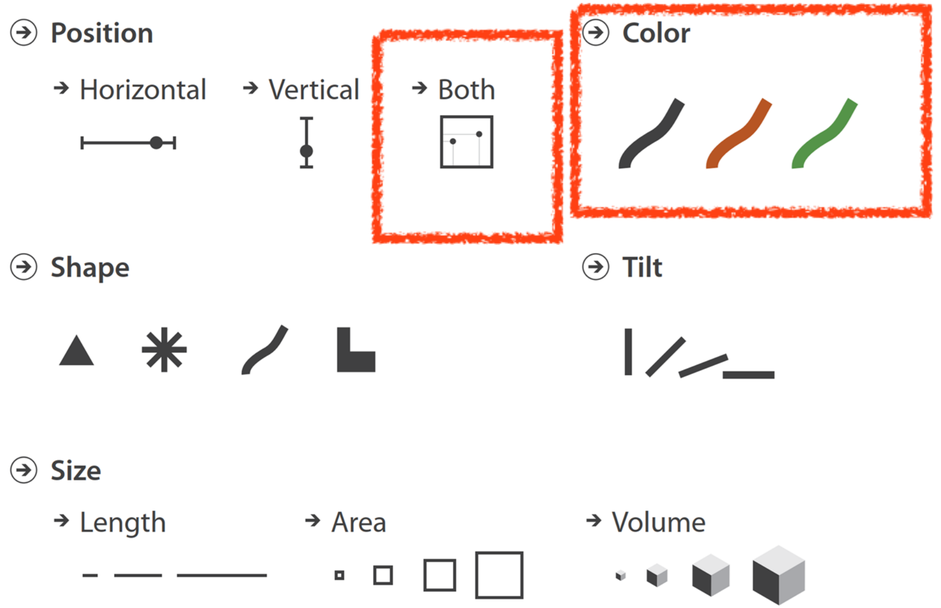
of attributes encoded: 3



MARK:

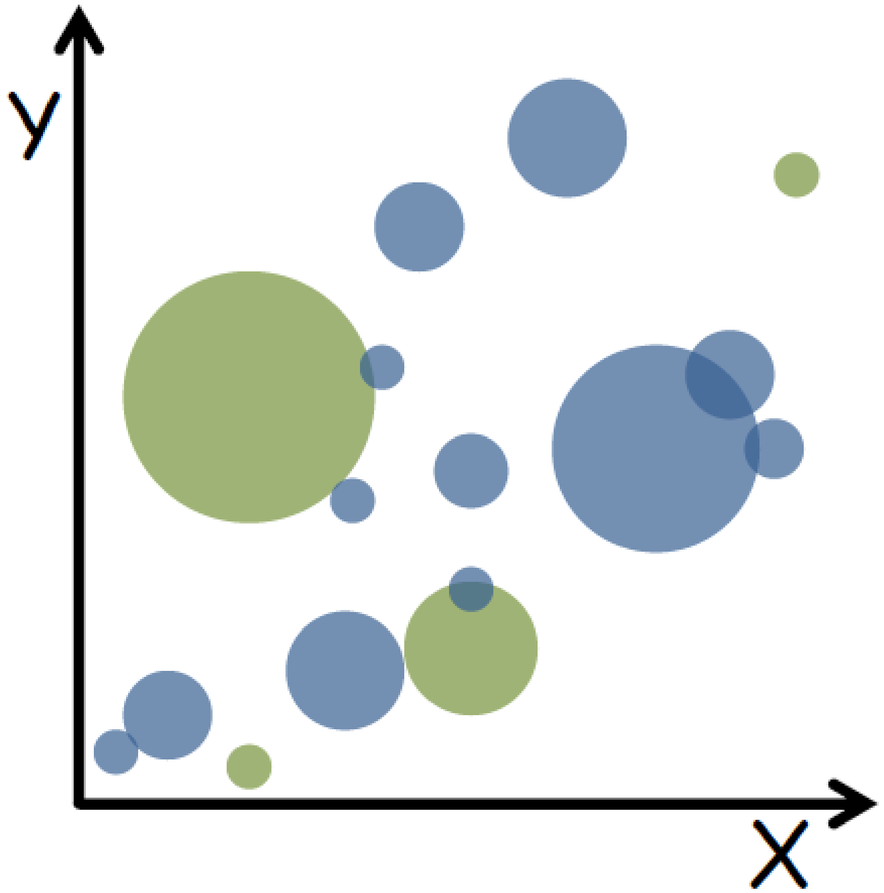


CHANNEL :

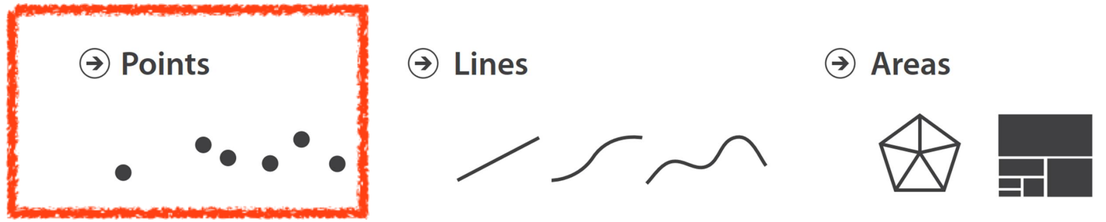


Visualization Building Blocks

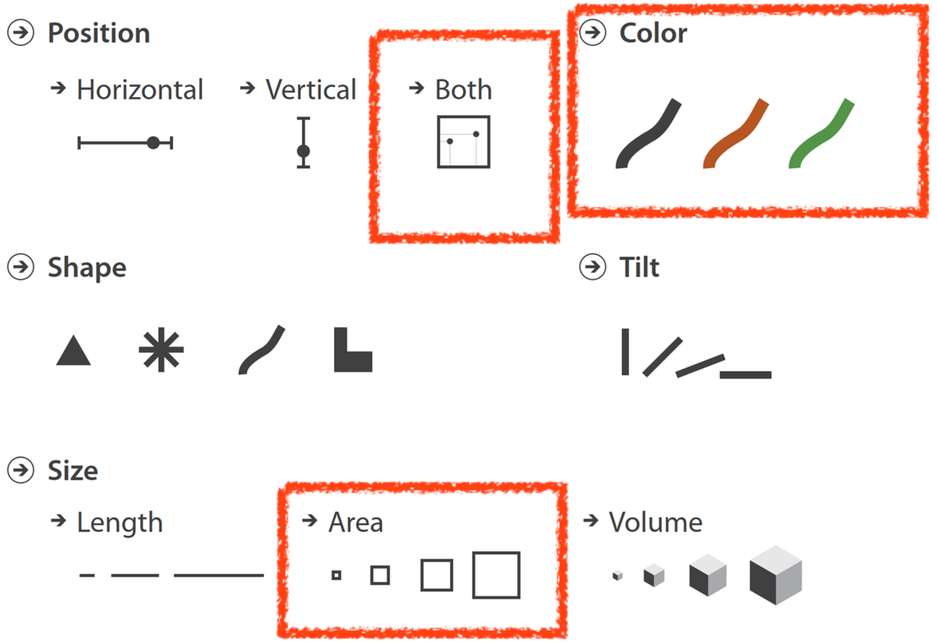
of attributes encoded: 4



MARK:

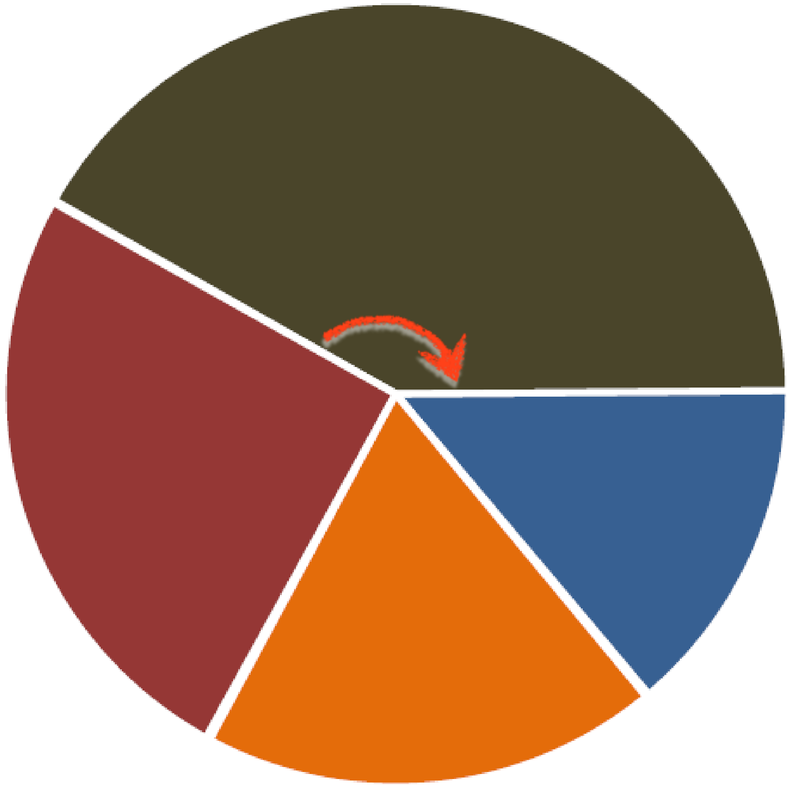


CHANNEL :

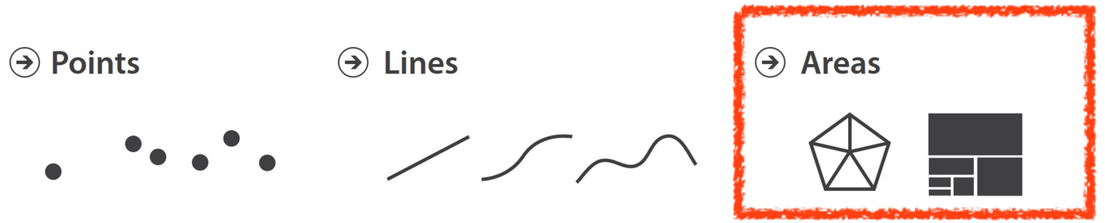


Visualization Building Blocks

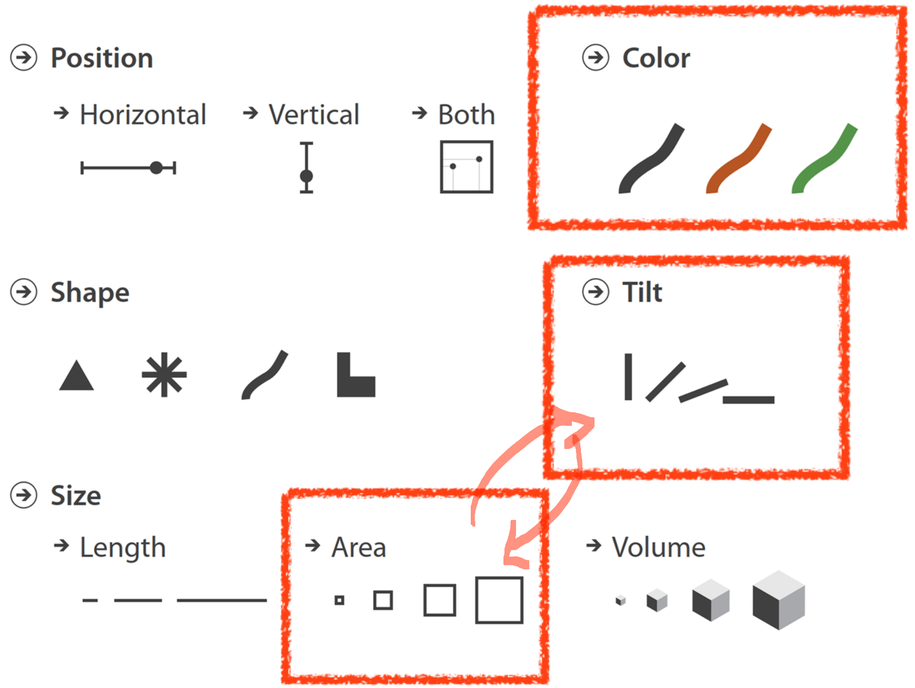
of attributes encoded: 2



MARK:



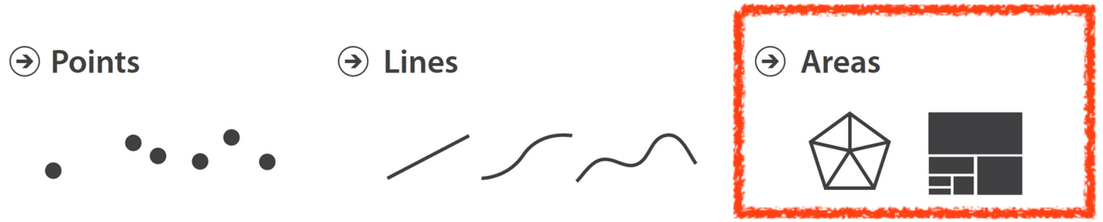
CHANNEL :



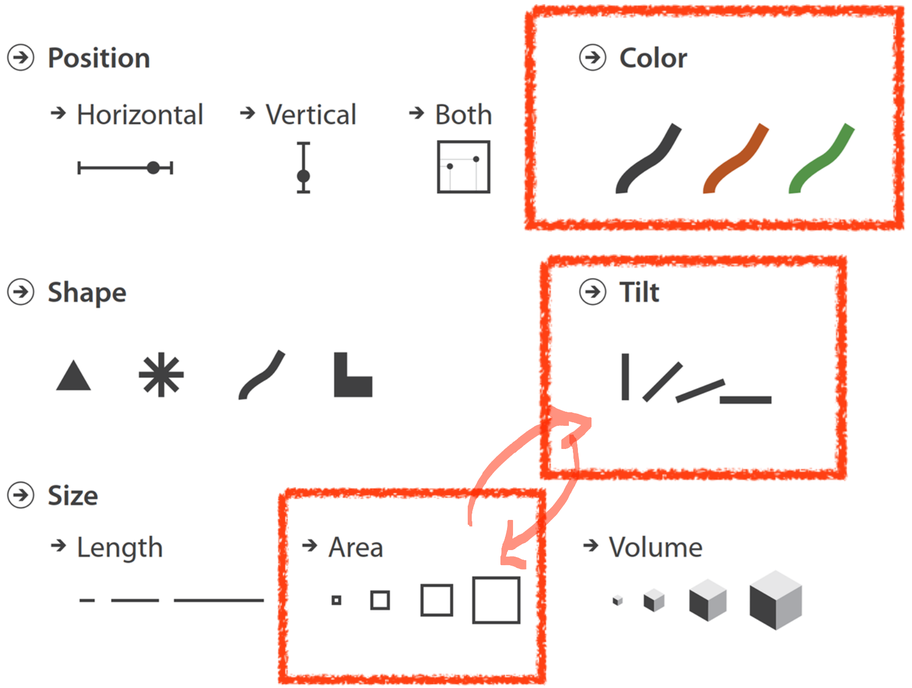
Visualization Building Blocks

of attributes encoded: 2

MARK:

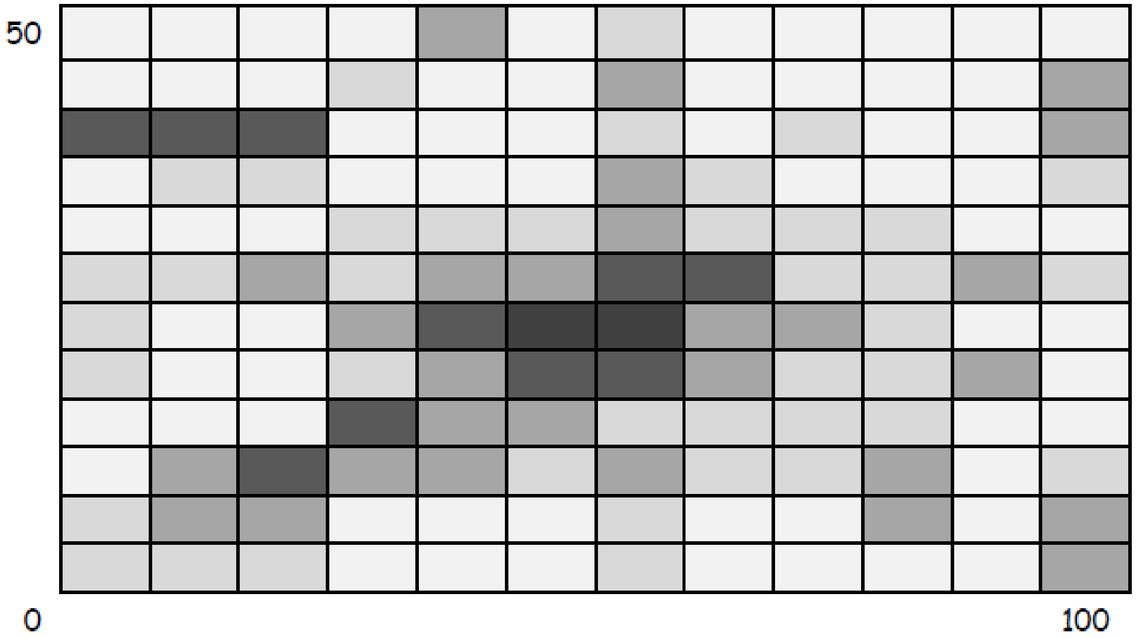


CHANNEL :

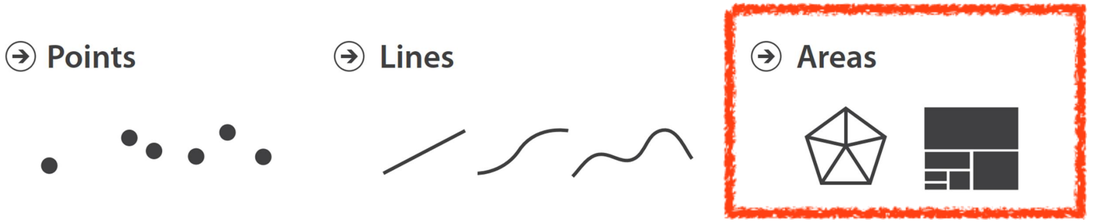


Visualization Building Blocks

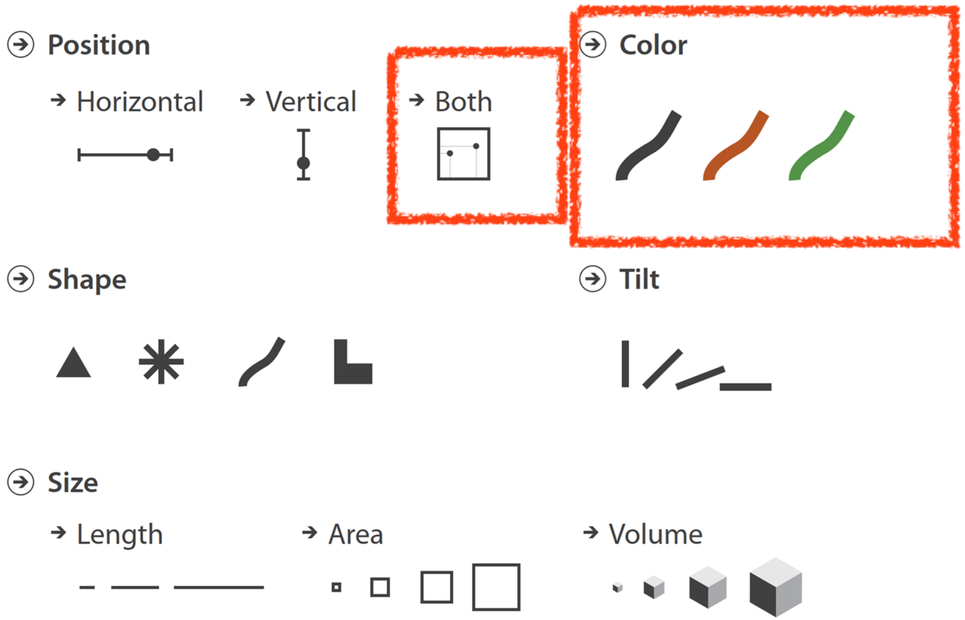
of attributes encoded:



MARK:



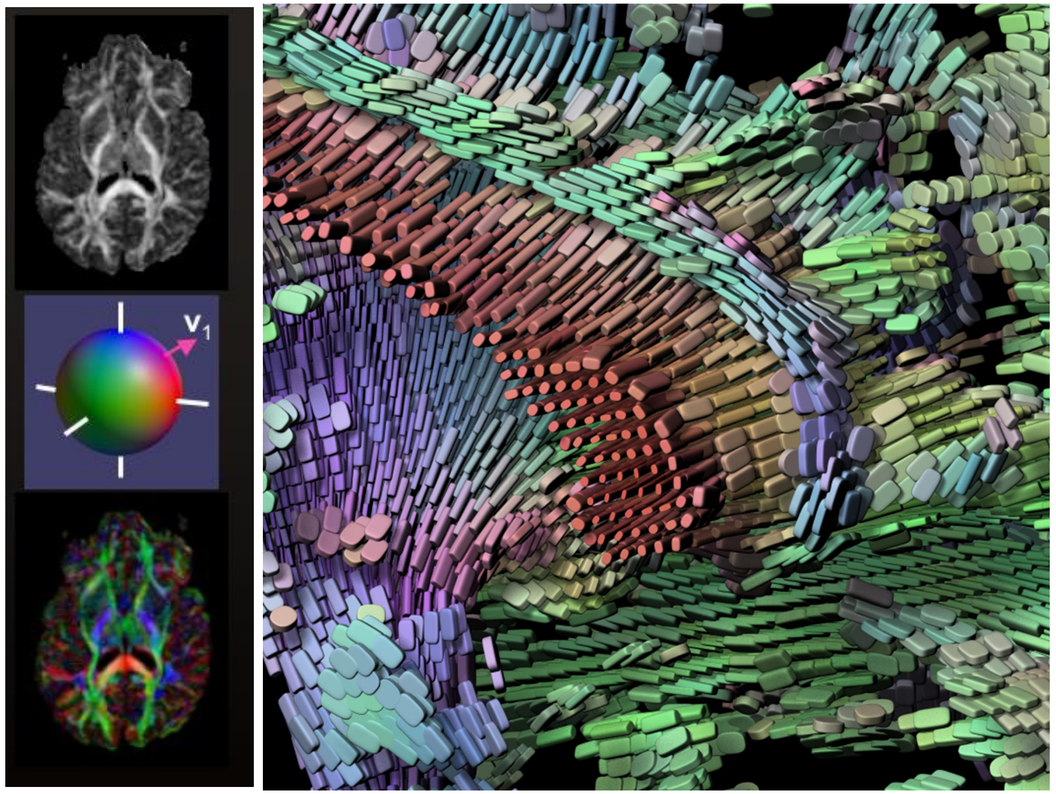
CHANNEL :



Visualization Building Blo

Don't overload the user with encodings!

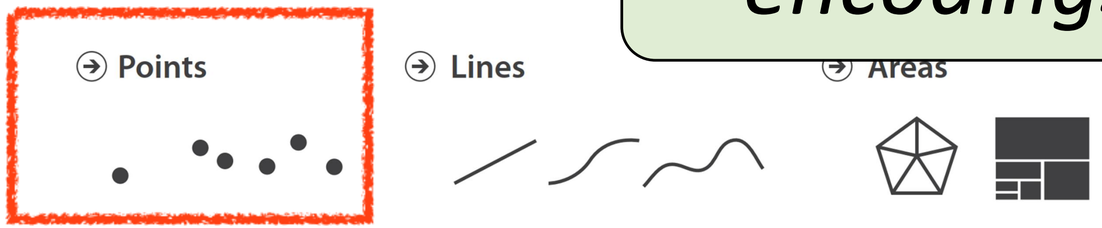
# of attributes encoded: ?



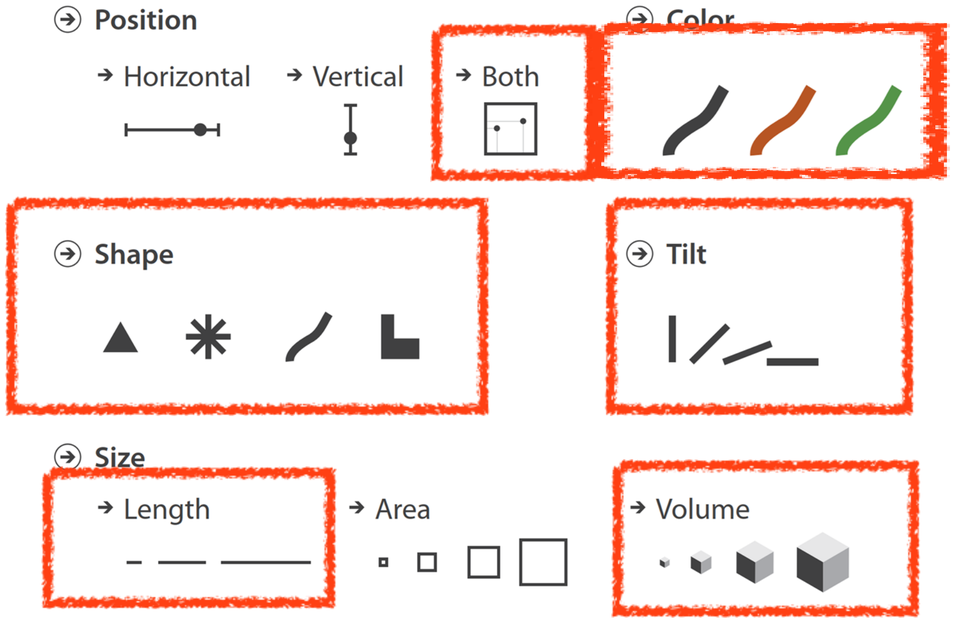
Kindlmann (2004)

+ position in 3D space

MARK:



CHANNEL :



Visualization Building Blocks

Marks as Items/Nodes

- Points

- Lines

- Areas

Marks as Links

- Containment

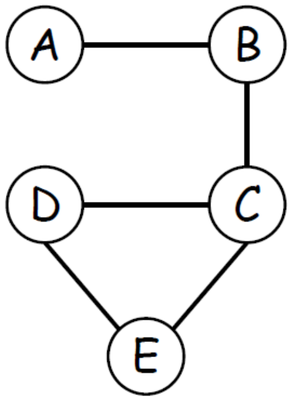
- Connection

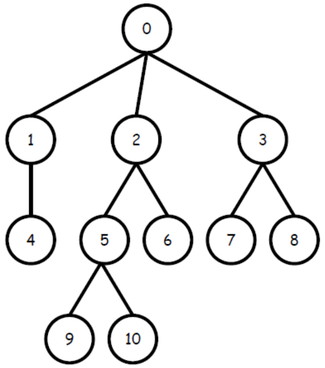
Visualization Building Blocks

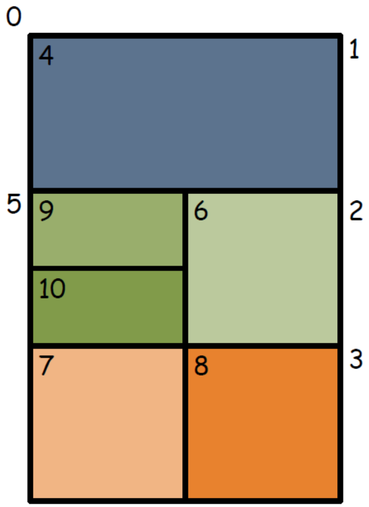
Marks as Links

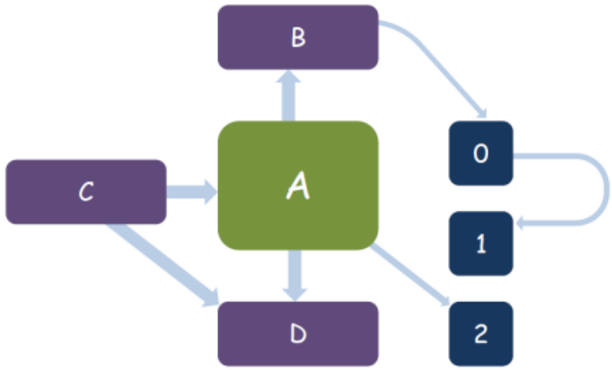
- Containment

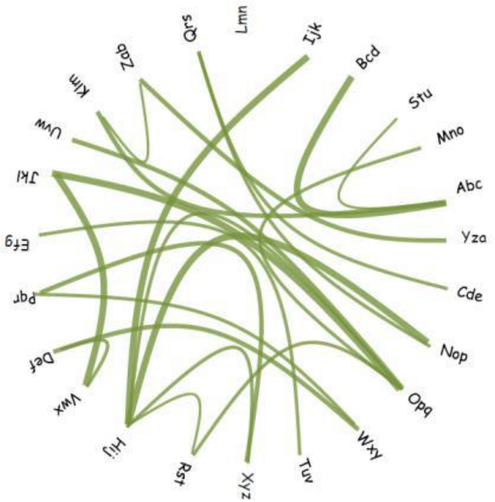
- Connection











Visualization Building Blocks

Marks as Items/Nodes

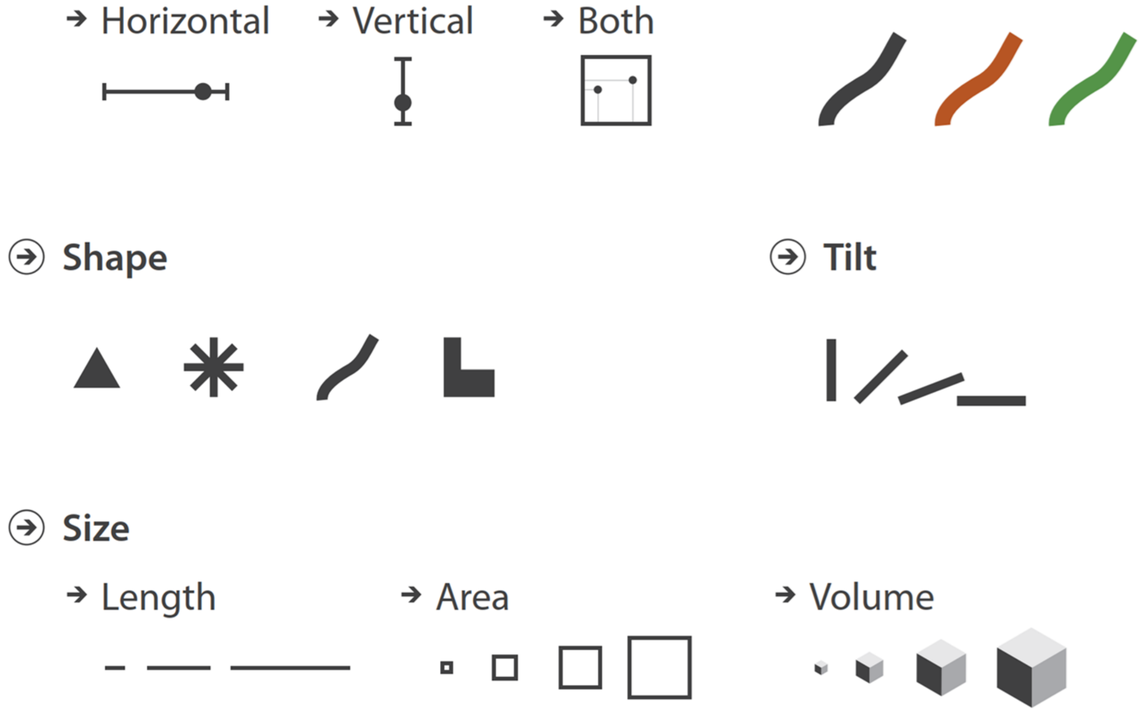
- Points
- Lines
- Areas

Marks as Links

- Containment
- Connection

Channels :

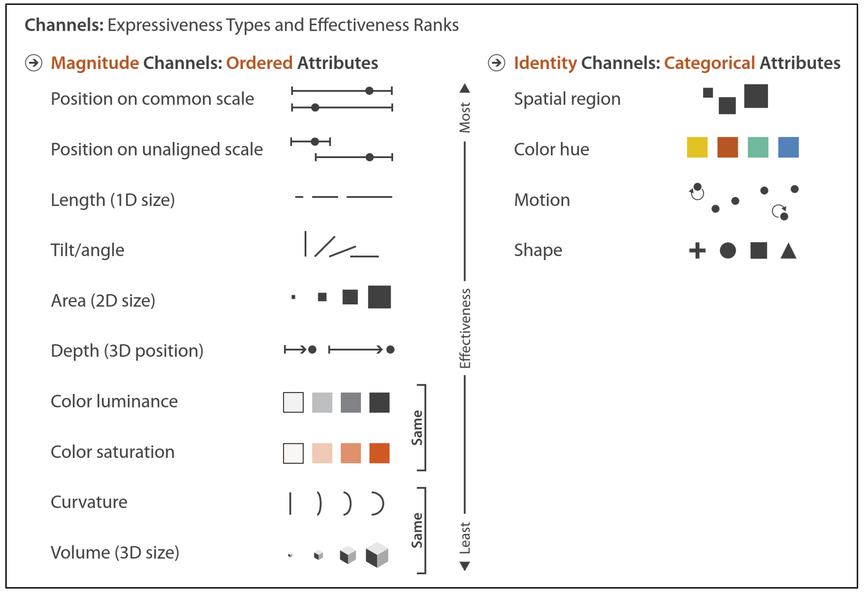
- Position
  - Horizontal
  - Vertical
  - Both
- Color
- Shape
- Tilt
- Size
  - Length
  - Area
  - Volume

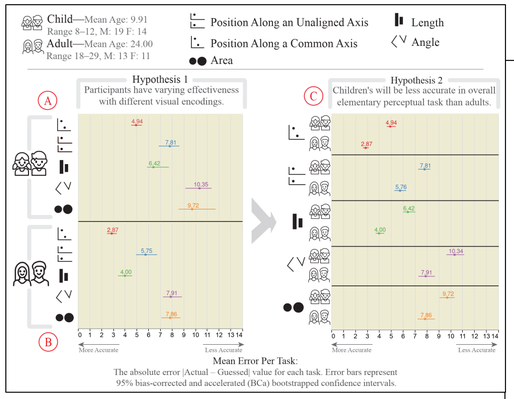


Note: these are all really important concepts when it comes time to coding your visualizations...!

How do I pick *which* marks or channels to use?

How to pick? User study results!





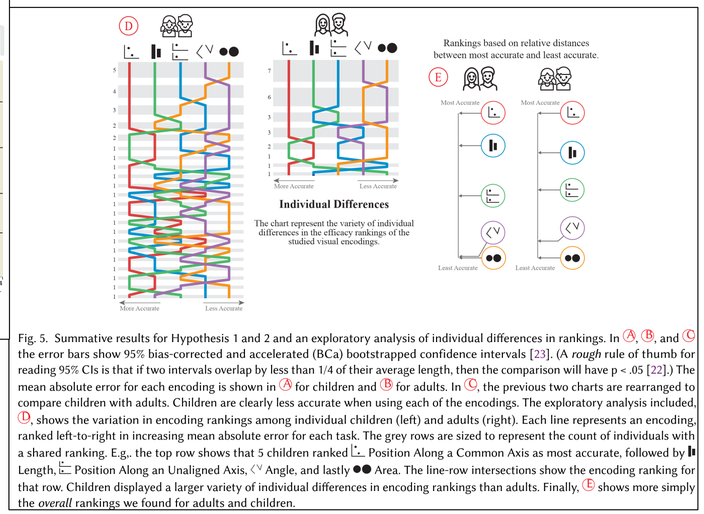


Fig. 5. Summative results for Hypothesis 1 and 2 and an exploratory analysis of individual differences in rankings. In Ⓐ, Ⓑ, and Ⓒ the error bars show 95% bias-corrected and accelerated (BCa) bootstrapped confidence intervals [23]. (A *rough* rule of thumb for reading 95% CIs is that if two intervals overlap by less than 1/4 of their average length, then the comparison will have $p < .05$ [22].) The mean absolute error for each encoding is shown in Ⓐ for children and Ⓑ for adults. In Ⓒ, the previous two charts are rearranged to compare children with adults. Children are clearly less accurate when using each of the encodings. The exploratory analysis included, Ⓓ, shows the variation in encoding rankings among individual children (left) and adults (right). Each line represents an encoding, ranked left-to-right in increasing mean absolute error for each task. The grey rows are sized to represent the count of individuals with a shared ranking. E.g., the top row shows that 5 children ranked Position Along a Common Axis as most accurate, followed by Length, Position Along an Unaligned Axis, Angle, and lastly Area. The line-row intersections show the encoding ranking for that row. Children displayed a larger variety of individual differences in encoding rankings than adults. Finally, Ⓔ shows more simply the *overall* rankings we found for adults and children.

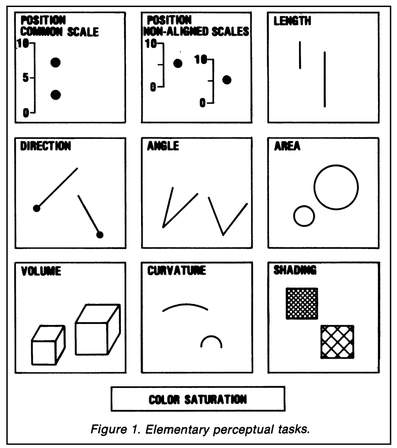


Figure 1. Elementary perceptual tasks.

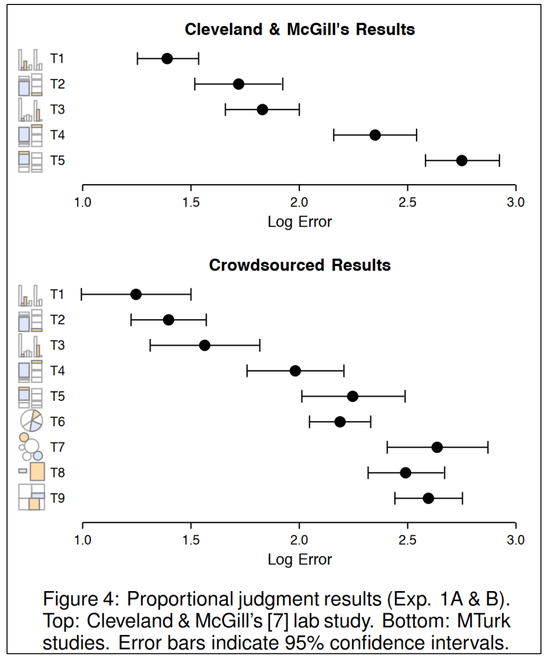


Figure 4: Proportional judgment results (Exp. 1A & B). Top: Cleveland & McGill's [7] lab study. Bottom: MTurk studies. Error bars indicate 95% confidence intervals.

[Munzer, 2014]
[Cleveland & McGill, 1984]
[Heer & Bostock, 2010]
[Mackinlay, 1986]
Panavas et al., 2021 (under submission)

“Ordering of Elemental Perceptual Tasks”

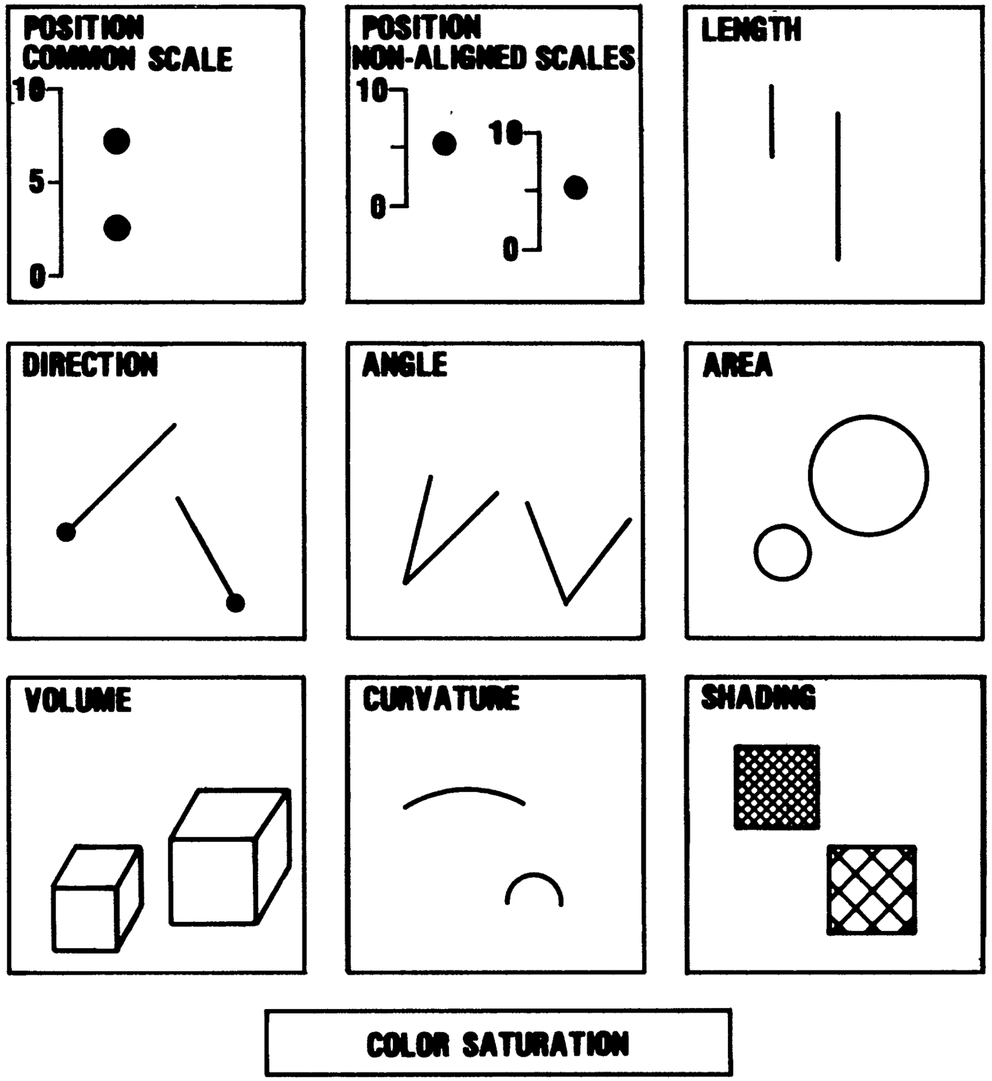


Figure 1. Elementary perceptual tasks.

[Cleveland & McGill (1984)]

“Ordering of Elemental Perceptual Tasks”

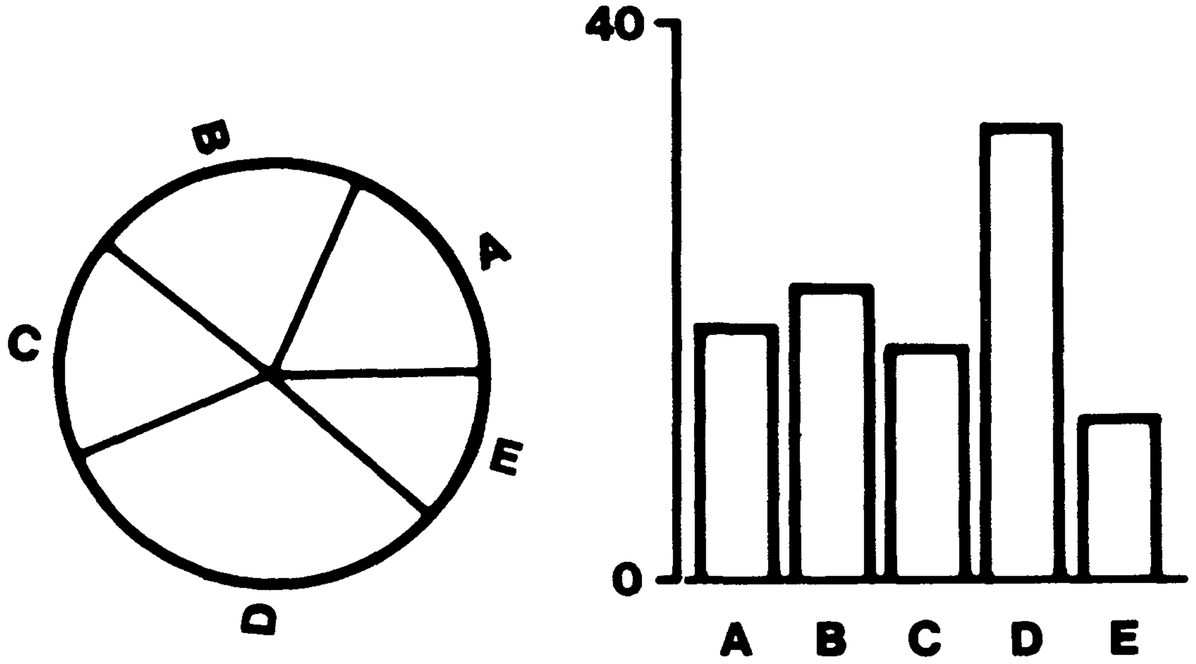
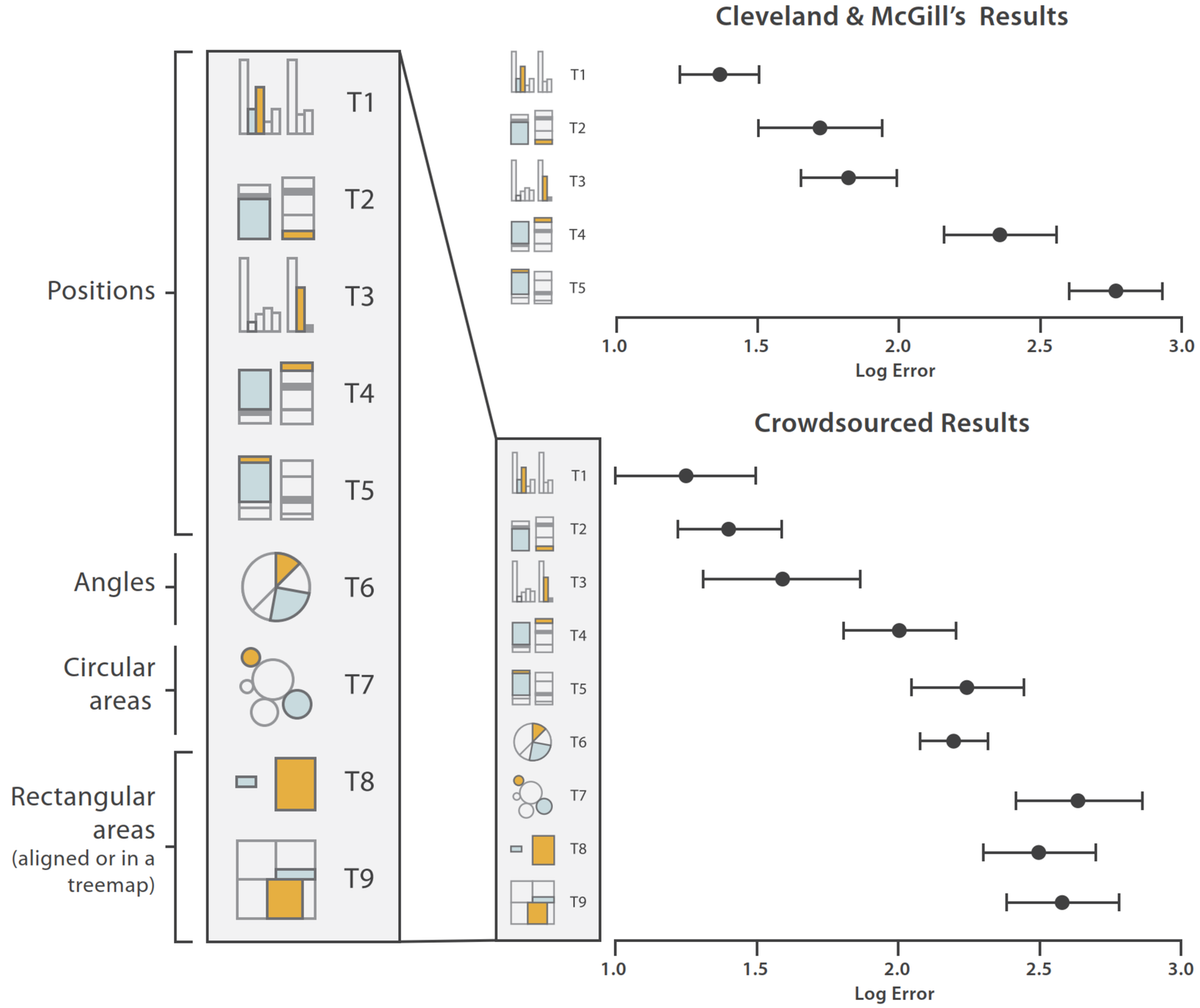


Figure 3. Graphs from position–angle experiment.

TASK: Which segment/bar is the maximum, and what is its percentage/value?

Cleveland & McGill (1984),
larger replication on AMT by Heer & Bostock (2010)

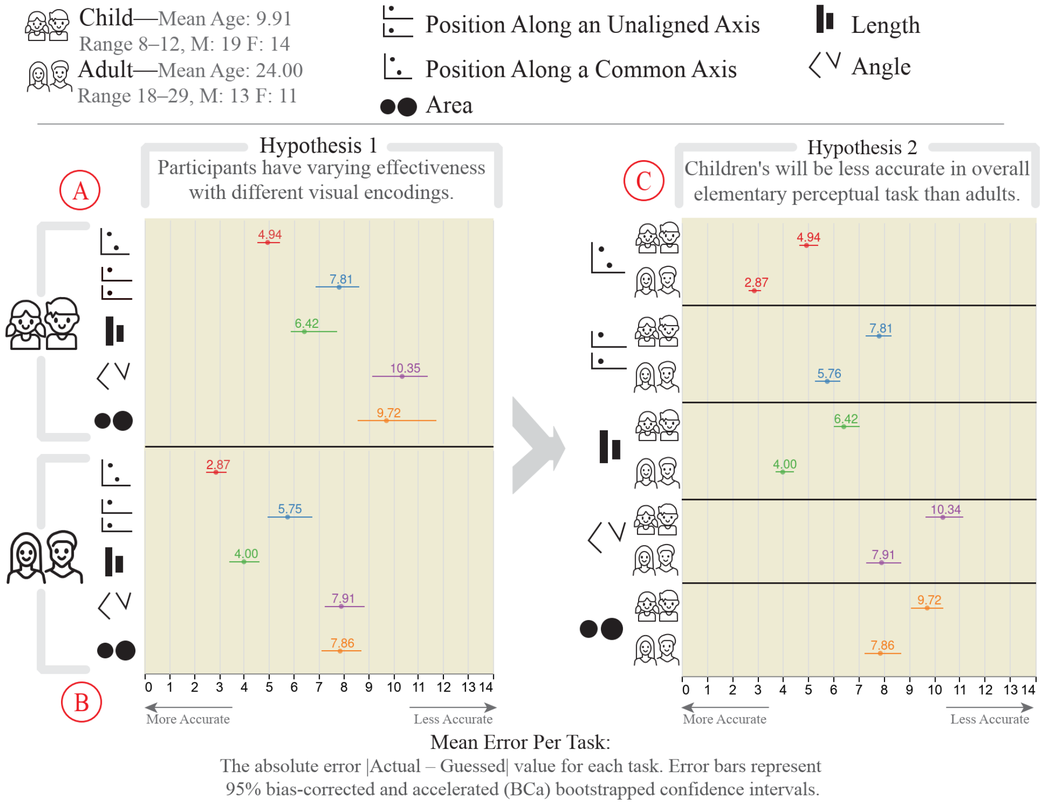


[Heer & Bostock (2010)](#)

Channels: Expressiveness Types and Effectiveness Ranks

- **Magnitude Channels: Ordered Attributes**

Caveats



Mean Error Per Task:
The absolute error |Actual – Guessed| value for each task. Error bars represent 95% bias-corrected and accelerated (BCa) bootstrapped confidence intervals.

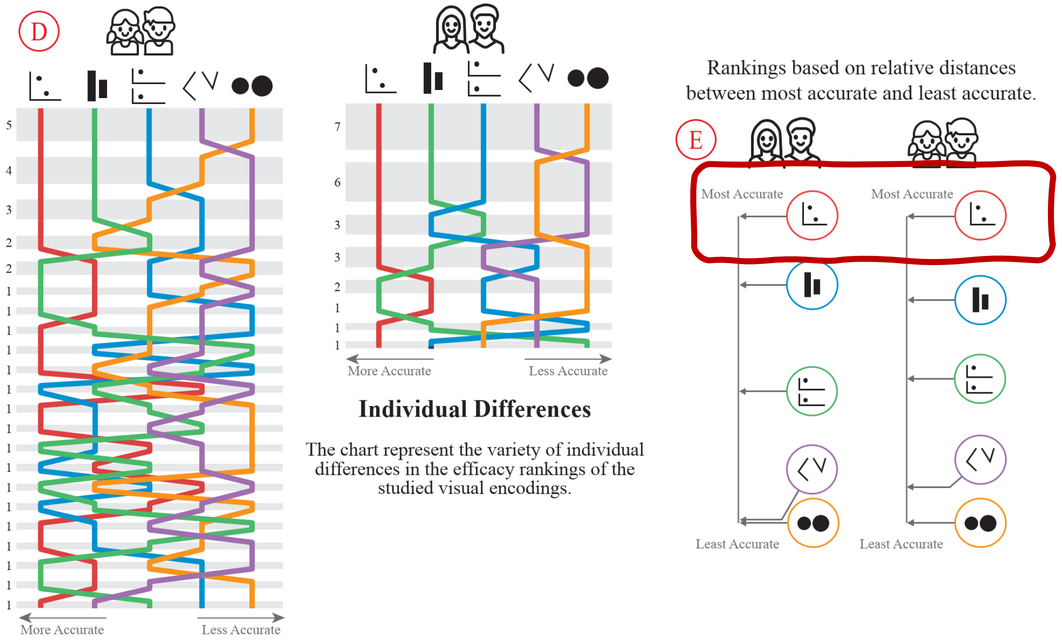


Fig. 5. Summative results for Hypothesis 1 and 2 and an exploratory analysis of individual differences in rankings. In Ⓐ, Ⓑ, and Ⓒ the error bars show 95% bias-corrected and accelerated (BCa) bootstrapped confidence intervals [23]. (A *rough* rule of thumb for reading 95% CIs is that if two intervals overlap by less than 1/4 of their average length, then the comparison will have $p < .05$ [22].) The mean absolute error for each encoding is shown in Ⓐ for children and Ⓑ for adults. In Ⓒ, the previous two charts are rearranged to compare children with adults. Children are clearly less accurate when using each of the encodings. The exploratory analysis included, Ⓓ, shows the variation in encoding rankings among individual children (left) and adults (right). Each line represents an encoding, ranked left-to-right in increasing mean absolute error for each task. The grey rows are sized to represent the count of individuals with a shared ranking. E.g,. the top row shows that 5 children ranked Position Along a Common Axis as most accurate, followed by Length, Position Along an Unaligned Axis, Angle, and lastly Area. The line-row intersections show the encoding ranking for that row. Children displayed a larger variety of individual differences in encoding rankings than adults. Finally, Ⓔ shows more simply the *overall* rankings we found for adults and children.

Panavas et al., 2022

Expressiveness and Effectiveness

Effectiveness principle: the importance of the attribute should match the salience of the channel; that is, its noticeability.
(i.e., encode most important attributes with highest ranked channels)

Expressiveness principle: the visual encoding should express all of, and only, the information in the dataset attributes.
(i.e., data characteristics should match the channel)

Mackinlay (1986)

My Summary: Prioritize choosing the most appropriate channel for each attribute

For Next Time

neu-ds-4200-s22.github.io/schedule

Look at the upcoming assignments and deadlines

- Textbook, Readings, & Reading Quizzes—Variable days
- In-Class Activities—If due, they are due 11:59pm the same day as class

Everyday Required Supplies:

- 5+ colors of pen/pencil
- White paper
- Laptop and charger

Use Canvas Discussions for general questions, email codydunne-and-tas@ccs.neu.edu for questions specific to you.

Week	Topics	Assignments
#1: Jan 17–21	What is visualization Design rules of thumb	A1—Setting up
#2: Jan 24–28	JS development, projects Marks & channels	A2—Encodings & xenographics
#3: Jan 31–Feb 04	Data types and tasks, Tableau D3 tutorial 1/2	P1—Pitches★
#4: Feb 07–11	In-class group formation D3 tutorial 2/2	A3—Tableau analysis P2—Proposal★
#5: Feb 14–18	Altair and JupyterLab Arrange tables	A4—D3 basic charts
#6: Feb 21–25	Color Pop-out, illusions	A5—Altair basic charts P3—Interview & tasks
#7: Feb 28–Mar 04	Interaction & animation (2)	A6—D3 event handling P4—Data, Initial sketches
#8: Mar 07–11	Trees & networks (2)	P5—Final sketches & plan★