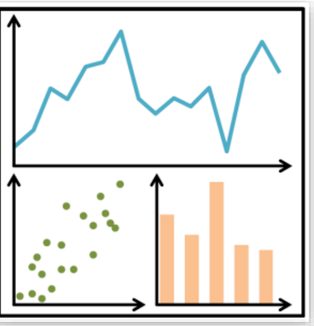
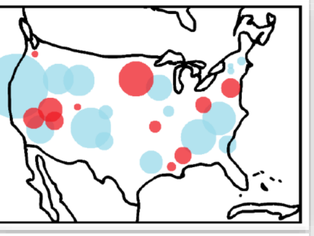
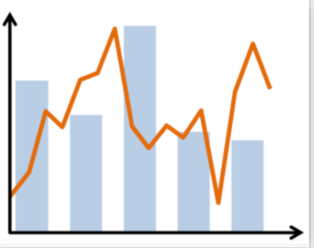
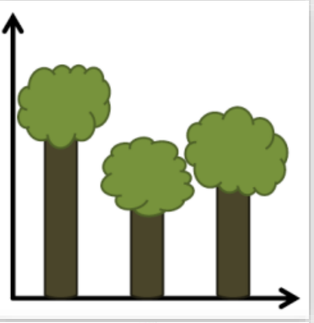
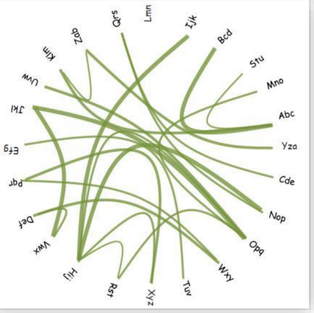






Lecture 7: D3, Data, Tasks

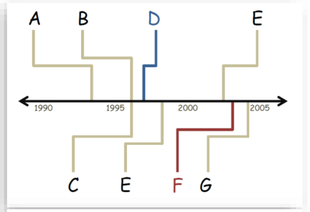
CS 7250

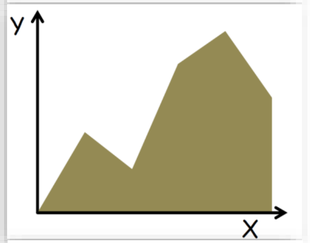
SPRING 2021

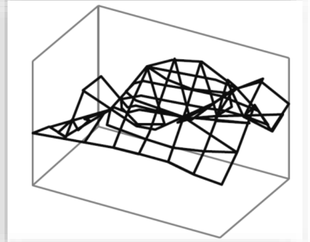
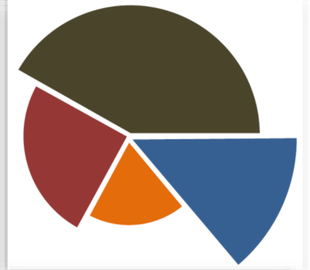
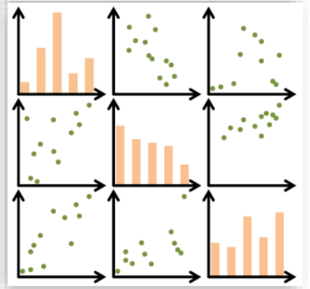
Prof. Cody Dunne

NORTHEASTERN UNIVERSITY

Slides and inspiration from Michelle Borkin, Krzysztof Gajos, Hanspeter Pfister, Miriah Meyer, Jonathan Schwabish, and David Sprague





Checking In

Previously, on CS 7250…

D3 Tutorial

Examples:

https://github.com/NEU-CS-7250-S21-Staff/D3_Examples

Now, on CS 7250…

In-Class Programming — D3 Line Chart

~45 min

Questions? Troubleshooting help?

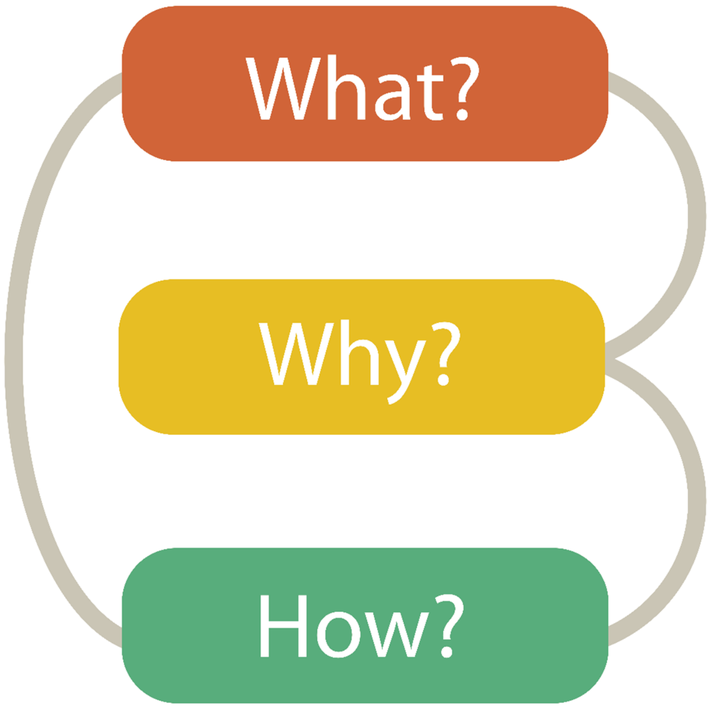
- Here
- khouryofficehours.com
- Teams

Data Types

Goals for Today

- Learn more about the attribute types
- Learn how to pick appropriate visual representations based on attribute type and perceptual properties

Analysis

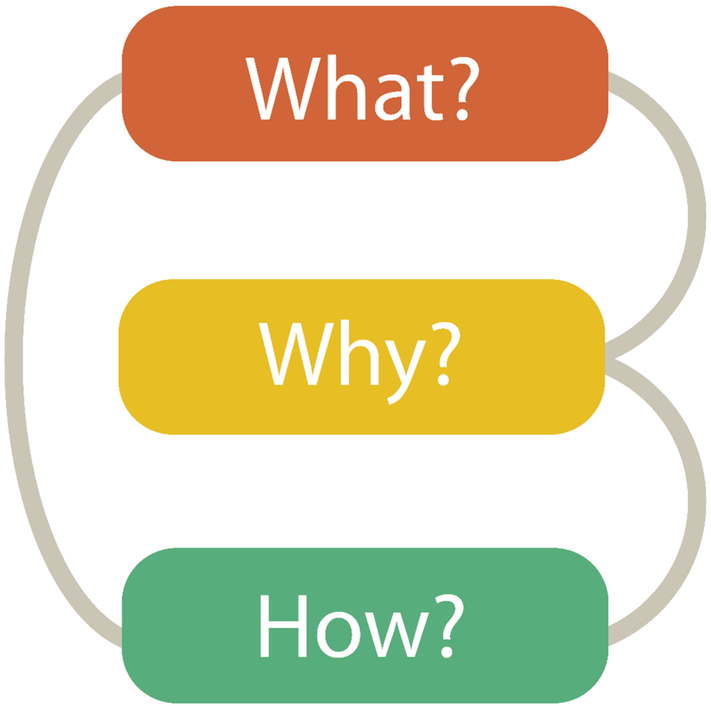


What data is shown?

Why is the user analyzing / viewing it?

How is the data presented?

Analysis

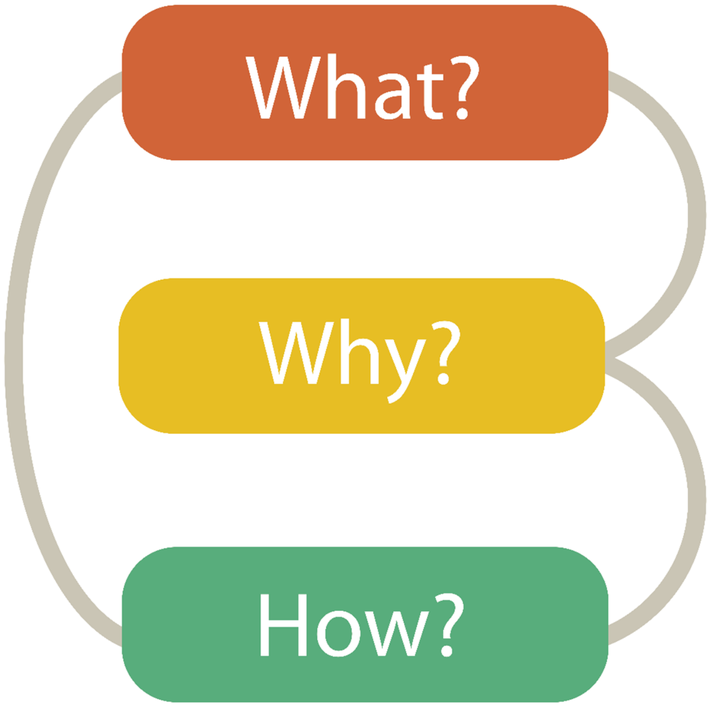


DATA ABSTRACTION

TASK ABSTRACTION

VISUAL ENCODING

Analysis



DATA ABSTRACTION

TASK ABSTRACTION

VISUAL ENCODING

Attribute Types

➔ Categorical

e.g.,
fruit (apple, pear, grape),
colleges (CAMD, Khoury, COE)

➔ Ordered

➔ *Ordinal*

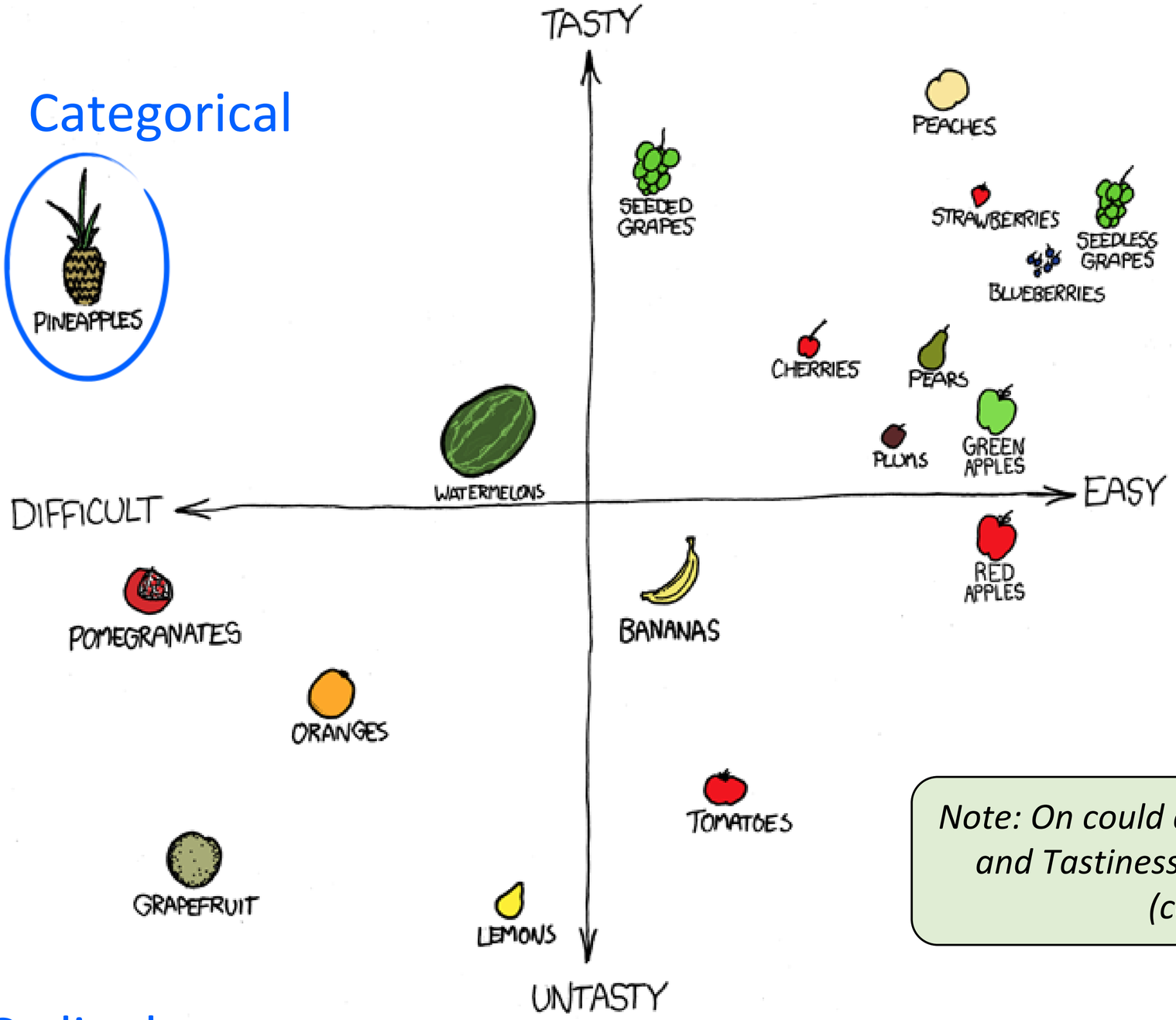
e.g.,
sizes (xs, s, m, l, xl),
months (J, F, M)

➔ *Quantitative* *(continuous)*

e.g.,
lengths (1’, 2.5’, 5’),
population

Categorical

Ordinal



Ordinal

Note: On could also argue that Difficulty and Tastiness could be quantitative (continuous)

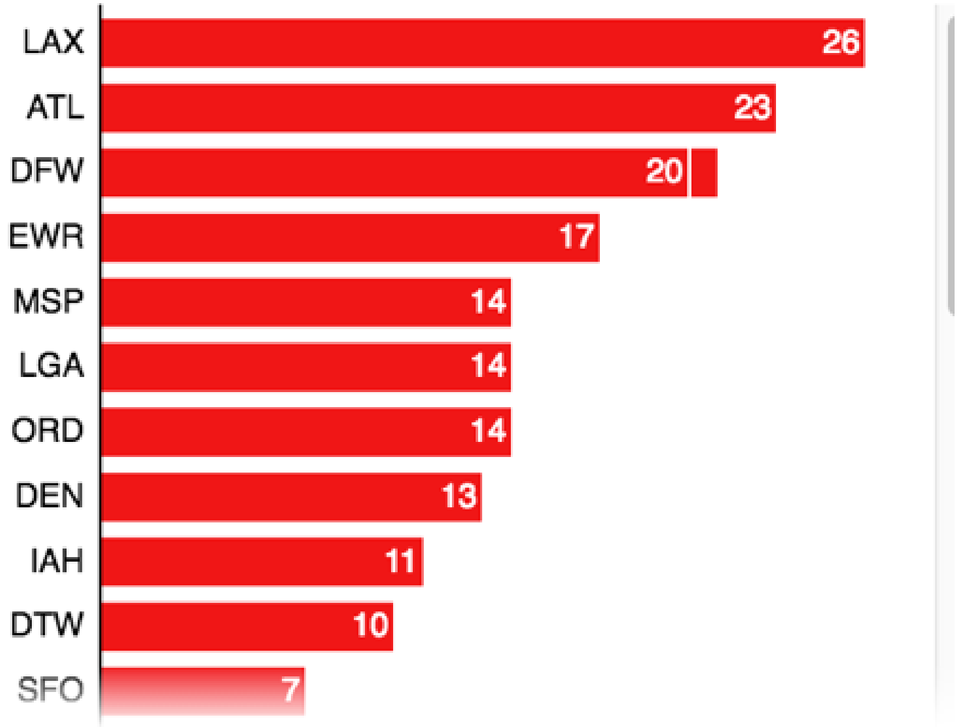
https://xkcd.com/388/

FlightAware MiseryMap
Back to main site

236
DELAYS

1
CANCELLATIONS

between 3 PM and 7 PM (all cancellations today) (all delays today)

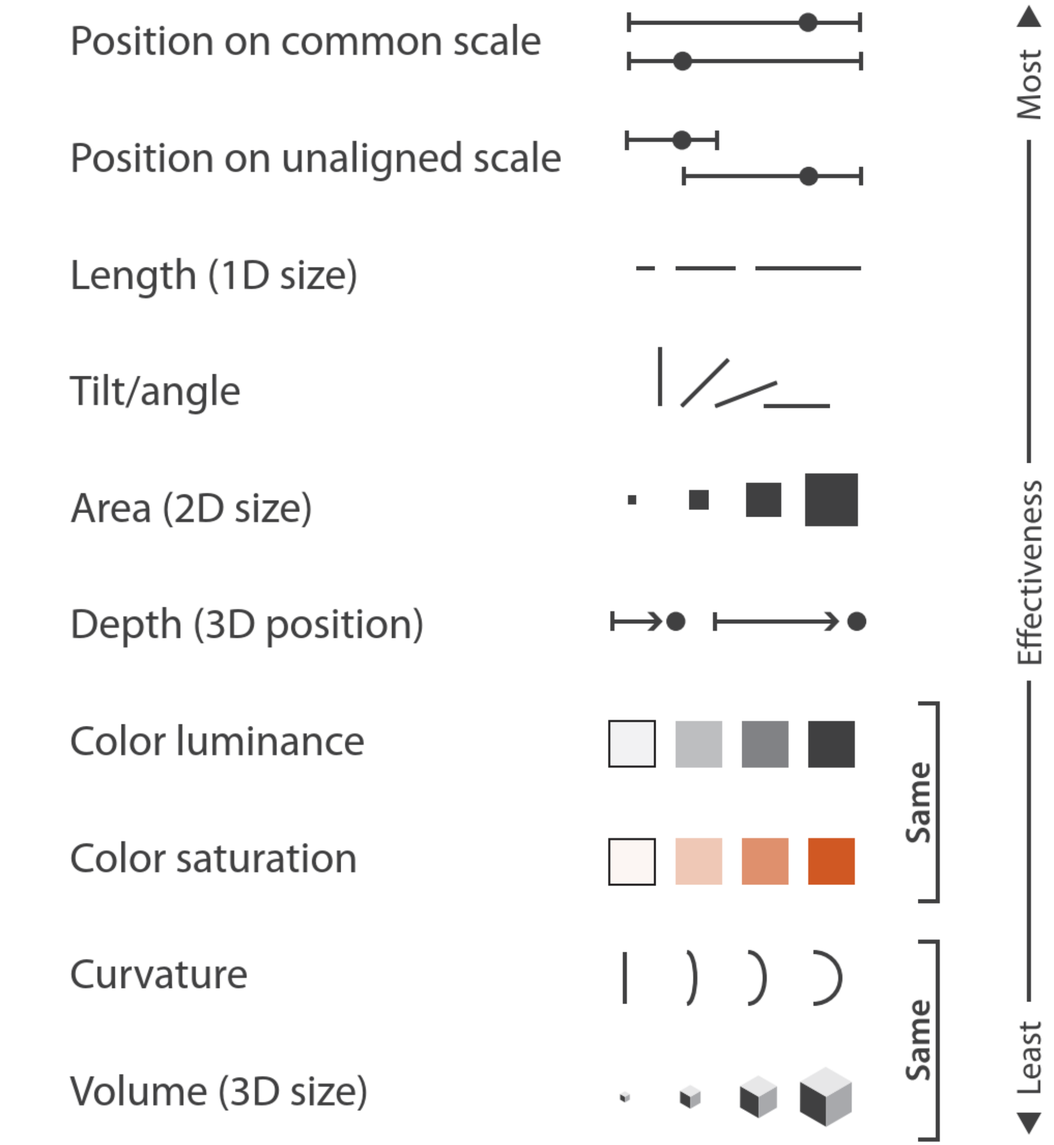


Categorical

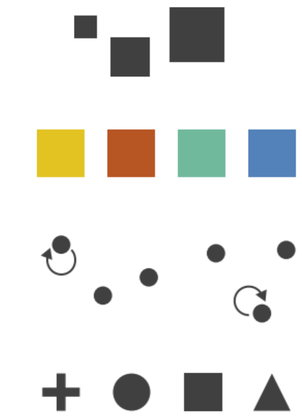
Quantitative

Channels: Expressiveness Types and Effectiveness Ranks

Magnitude Channels: Ordered Attributes



Identity Channels: Categorical Attributes



Summarizes results from [Cleveland & McGill (1984)](), [Heer & Bostock (2010)]()

Channel Ranking by Data Type

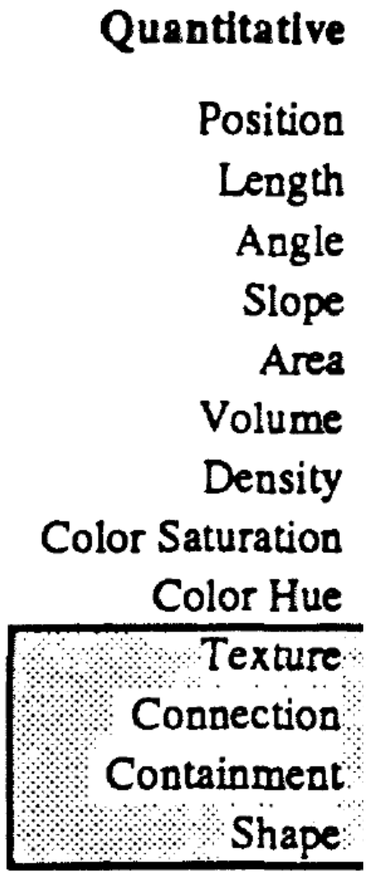


Figure 15: Ranking of Perceptual Tasks. *The tasks shown in the gray boxes are not relevant to that type of data.*

Mackinlay (1986)

Channel Ranking by Data Type

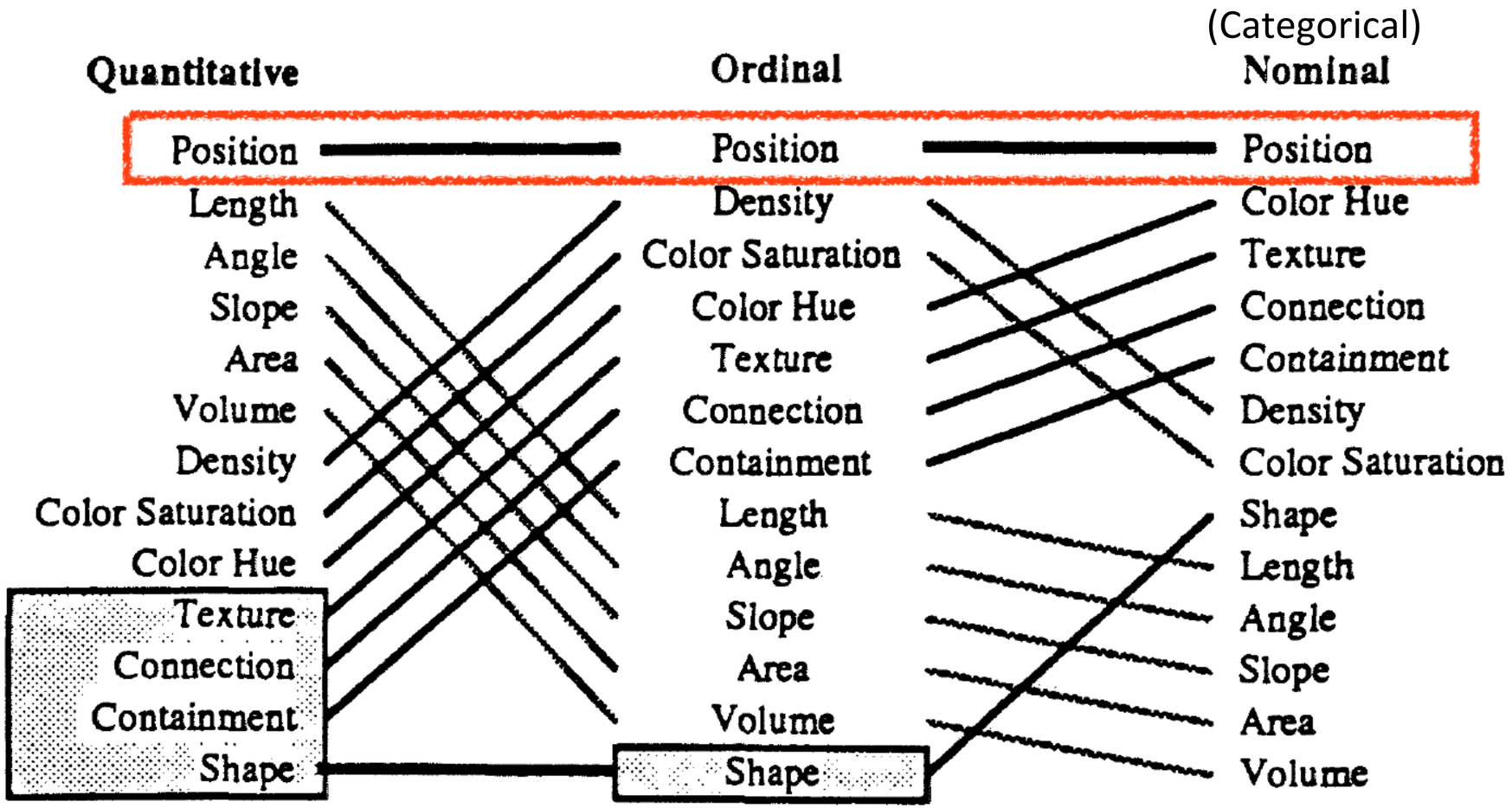


Figure 15: Ranking of Perceptual Tasks. *The tasks shown in the gray boxes are not relevant to that type of data.*

[Mackinlay (1986)](#)

Channel Ranking by Data Type

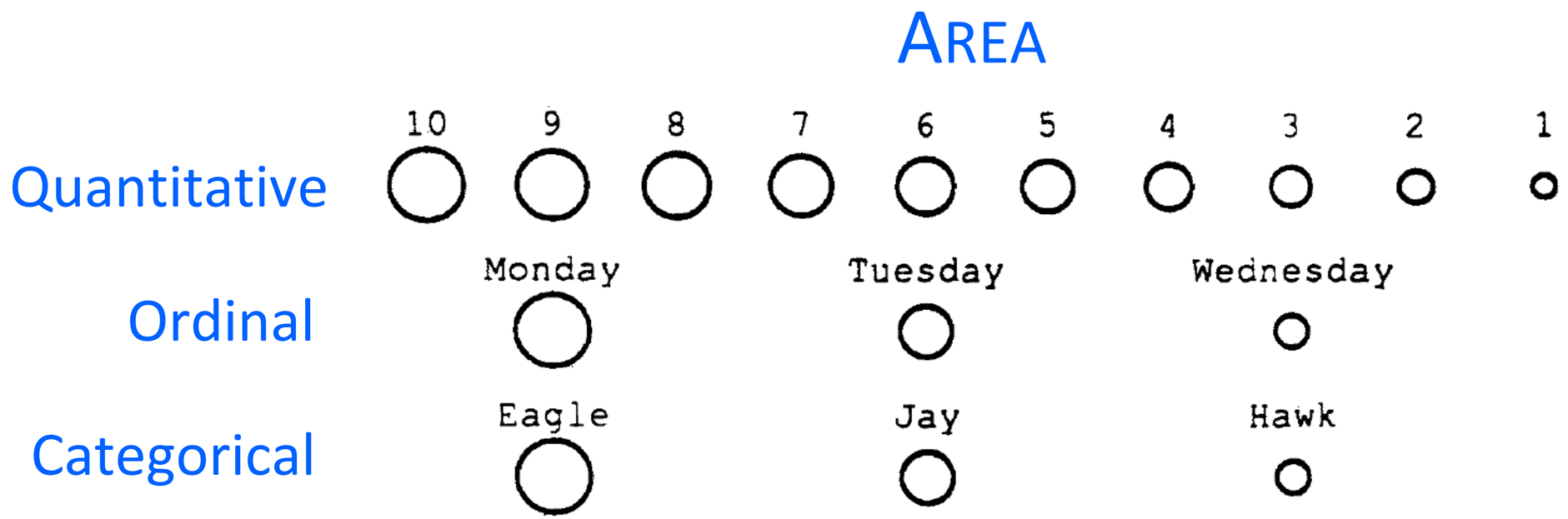


Figure 16: Analysis of the Area Task.

Mackinlay (1986)

Channel Ranking by Data Type

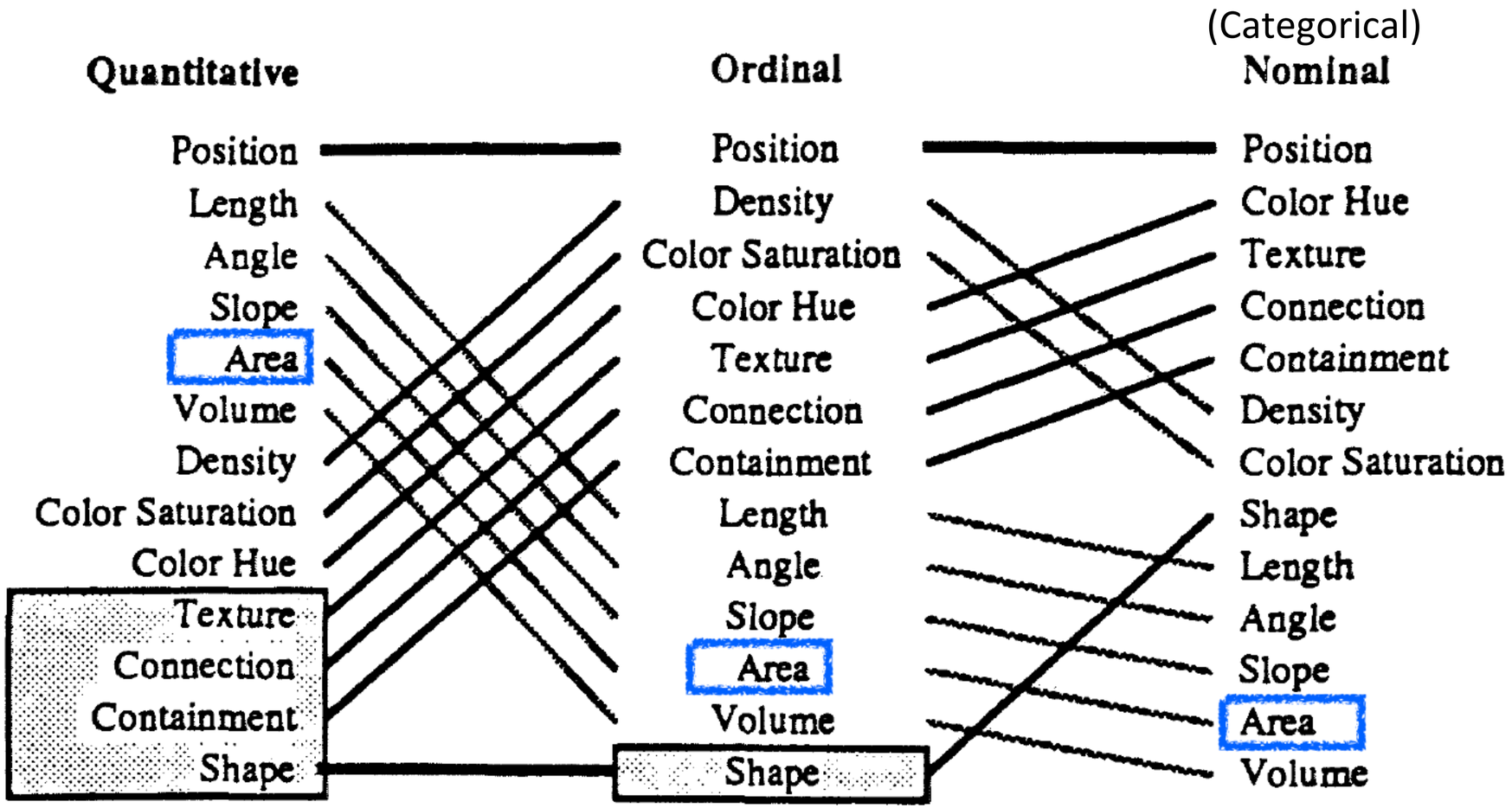


Figure 15: Ranking of Perceptual Tasks. *The tasks shown in the gray boxes are not relevant to that type of data.*

[Mackinlay (1986)](#)

Data Abstraction

What?

Datasets

Data Types

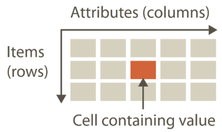
- Items
- Attributes
- Links
- Positions
- Grids

Data and Dataset Types

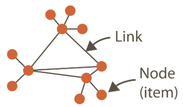
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

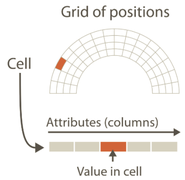
- Tables



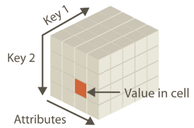
- Networks



- Fields (Continuous)



- *Multidimensional Table*



- *Trees*

- Geometry (Spatial)

Dataset Availability

- Static

- Dynamic

Attributes

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

Ordering Direction

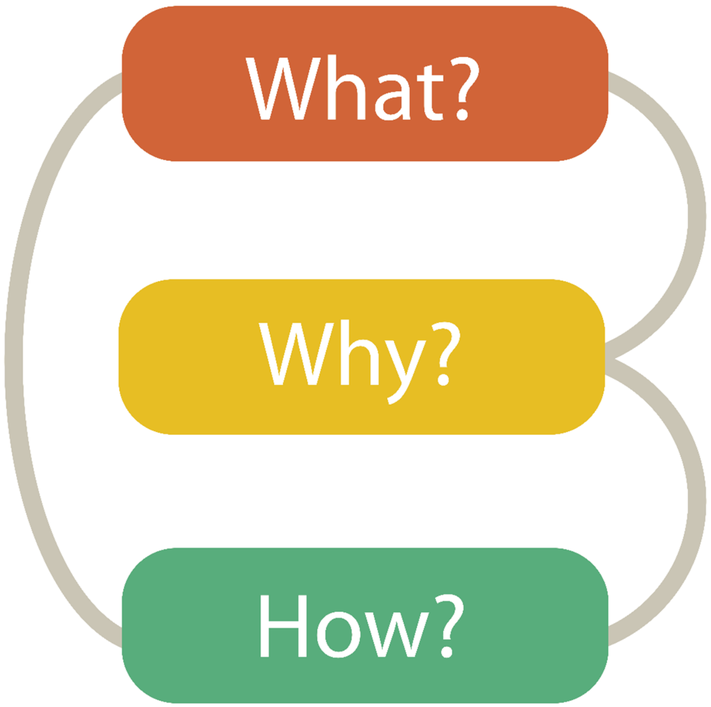
- Sequential

- Diverging

- Cyclic

TASK ABSTRACTION

Analysis

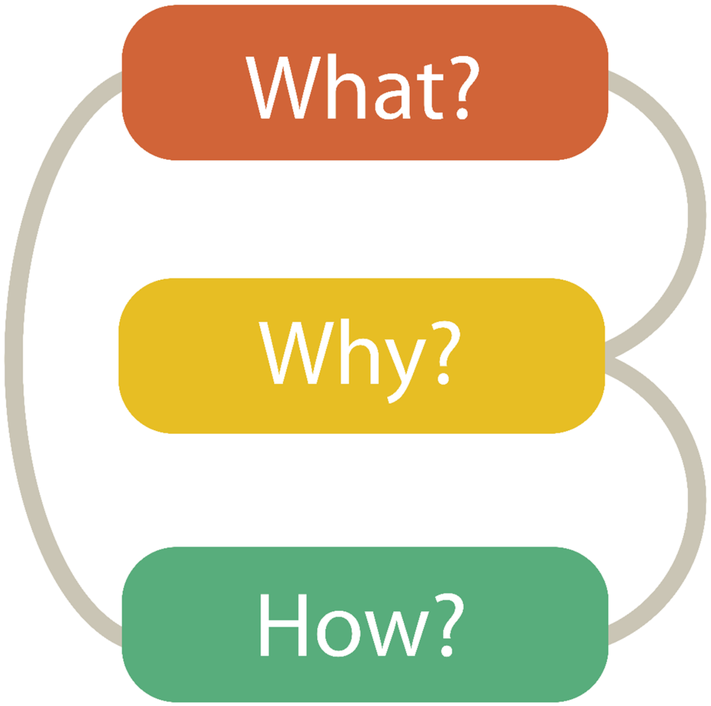


What data is shown?

Why is the user analyzing / viewing it?

How is the data presented?

Analysis

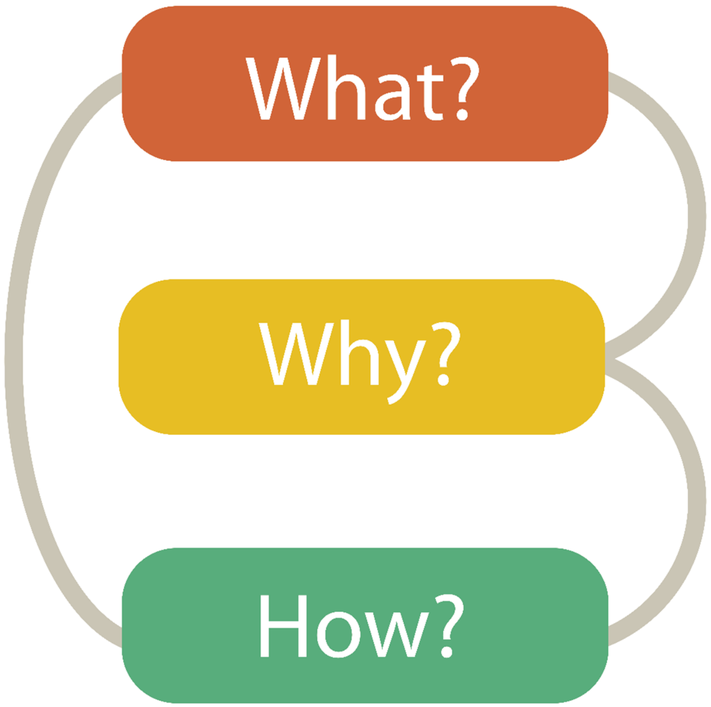


DATA ABSTRACTION

TASK ABSTRACTION

VISUAL ENCODING

Analysis



DATA ABSTRACTION

TASK ABSTRACTION

VISUAL ENCODING

Goals for Today

- Learn what “Tasks” are and why they are so important.
- Learn the differences between high, mid, and low level task classifications.
- Begin practicing how to classify tasks (key step in visualization design process!).

Task Abstraction

Why abstract?

Avoids domain specific terms thus easier to apply to other cases (broadly applicable results).

Why?

Actions

Analyze

- Consume
  - *Discover*
  - *Present*
  - *Enjoy*
- Produce
  - *Annotate*
  - *Record*
  - *Derive*

Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

Query

- Identify
- Compare
- Summarize

Targets

All Data

- Trends
- Outliers
- Features

Attributes

- One
- Many
  - *Distribution*
  - *Dependency*
  - *Correlation*
  - *Similarity*
  - *Extremes*

Network Data

- Topology
- *Paths*

Spatial Data

- Shape

What? Why? How?

Task Abstraction

Why abstract?

Avoids domain specific terms thus easier to apply to other cases (broadly applicable results).

Visualization Tools

Specific ⟷ *General*

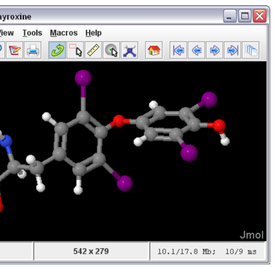
Gephi

Altair

D3

ArcGIS ESRI

tableau



Task Abstraction

Why abstract?

Avoids domain specific terms thus easier to apply to other cases (broadly applicable results).

Why?

Actions

Analyze

- Consume
  - *Discover*
  - *Present*
  - *Enjoy*
- Produce
  - *Annotate*
  - *Record*
  - *Derive*

Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

Query

- Identify
- Compare
- Summarize

Targets

All Data

- Trends
- Outliers
- Features

Attributes

- One
- Many
  - *Distribution*
  - *Dependency*
  - *Correlation*
  - *Similarity*
  - *Extremes*

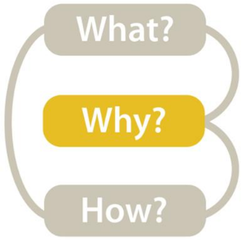
Network Data

- Topology
- *Paths*

Spatial Data

- Shape

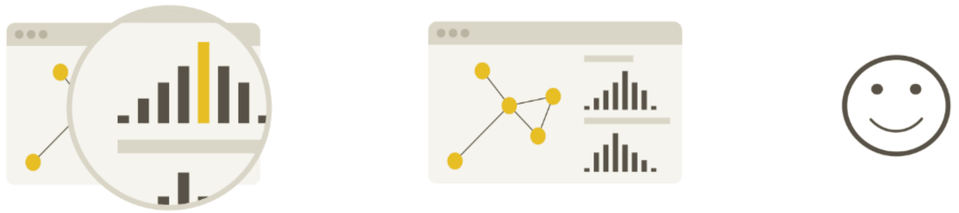
What? Why? How?



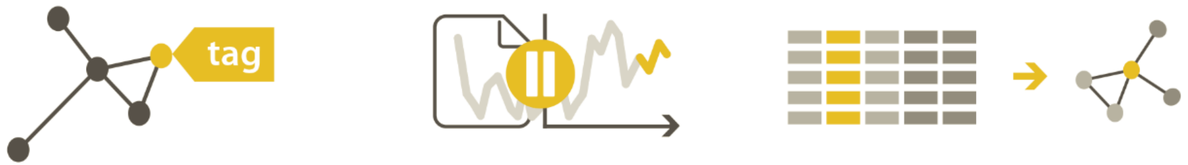
High-level

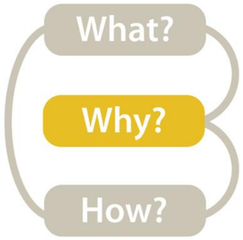
ACTIONS define user goals.

- Analyze
  - Consume
    - *Discover*
    - *Present*
    - *Enjoy*



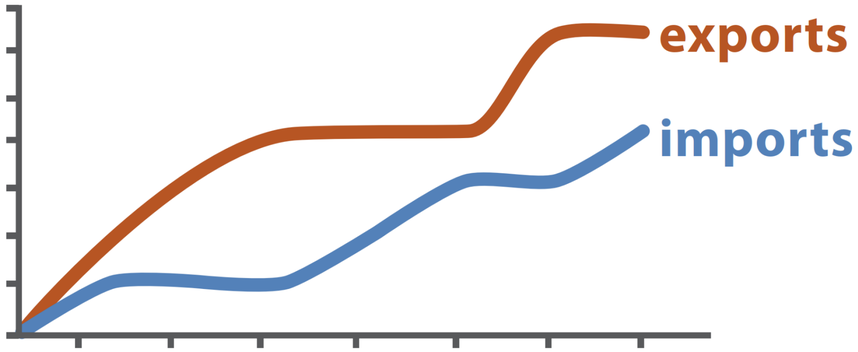
  - Produce
    - *Annotate*
    - *Record*
    - *Derive*



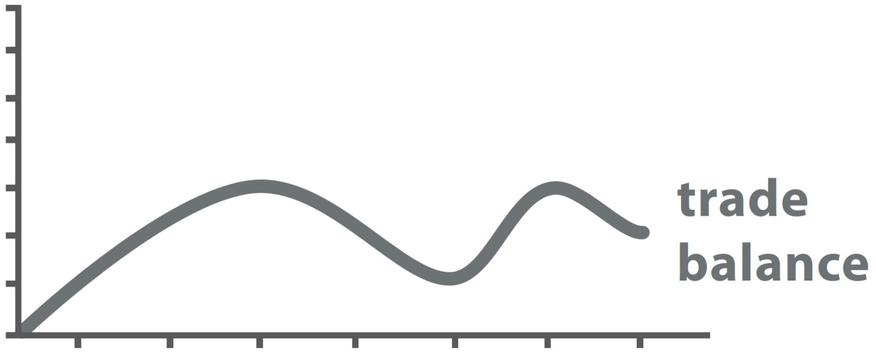


→ *Derive*

ACTIONS define user goals.

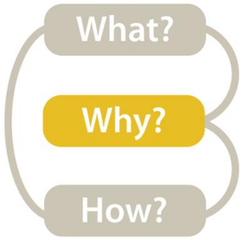


Original Data



trade balance = exports – imports

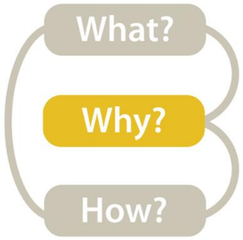
Derived Data



ACTIONS define user goals. *Mid-level*

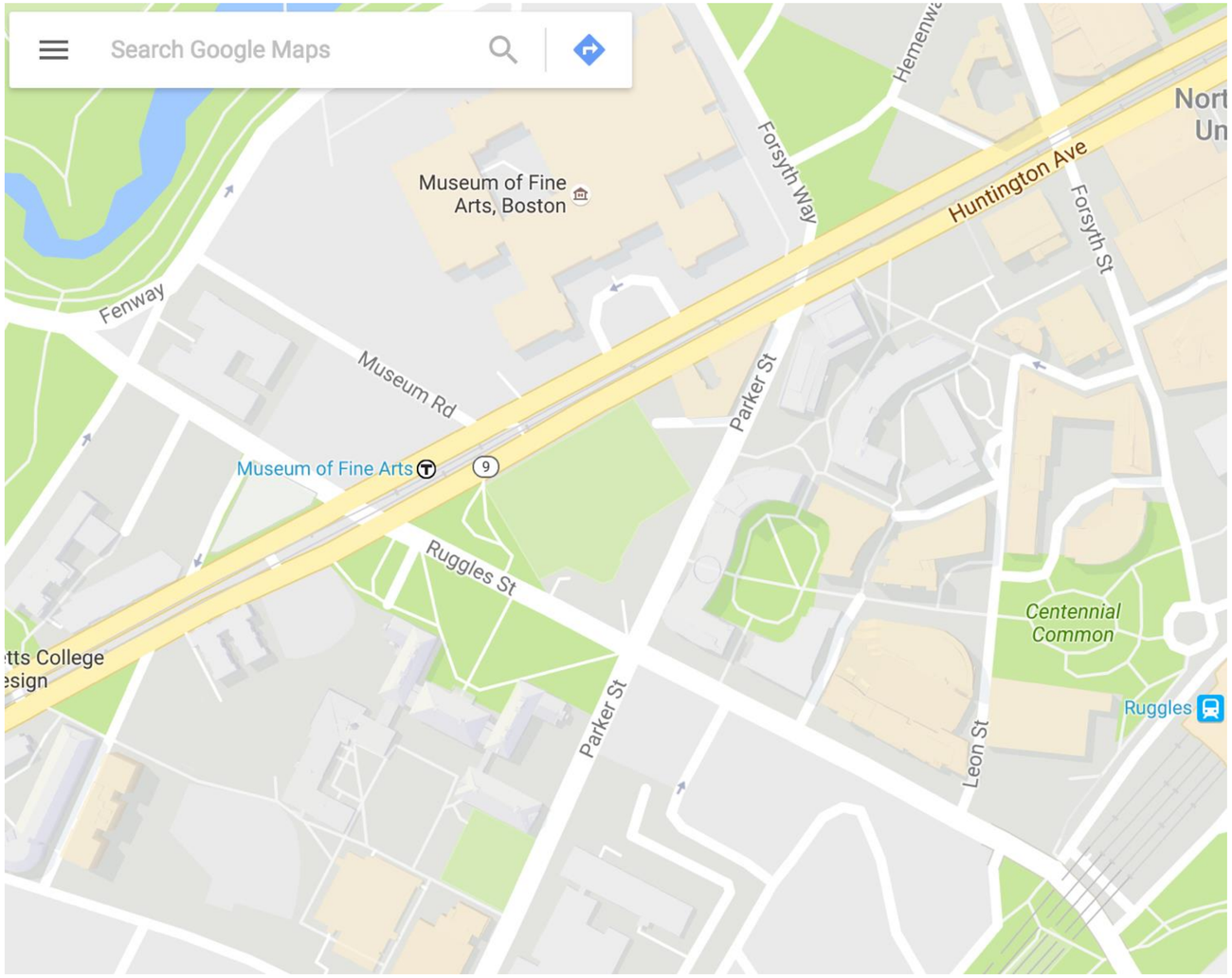
Search

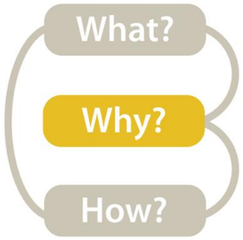
	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*



Search

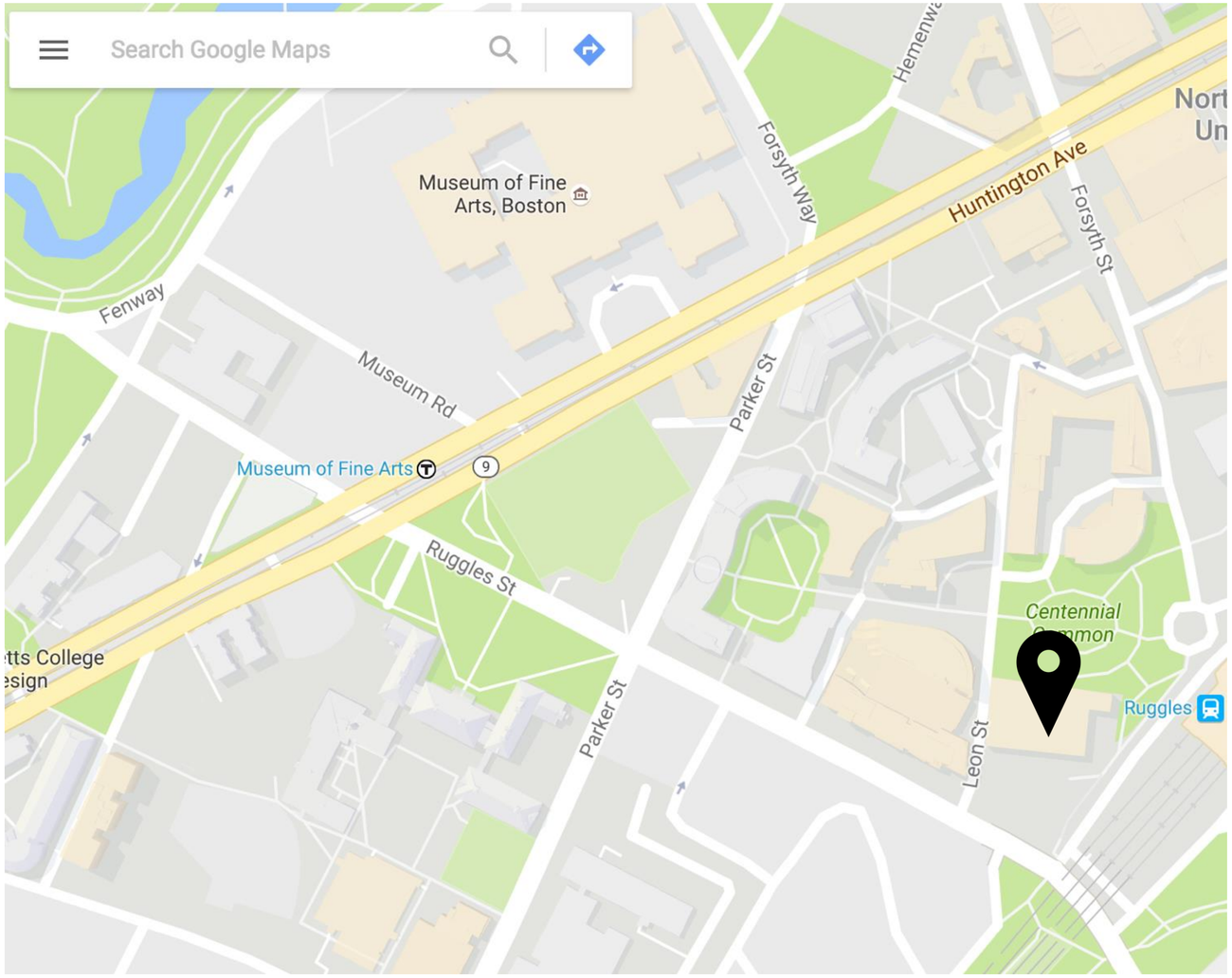
	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*





Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*



What is the address of Ryder hall?

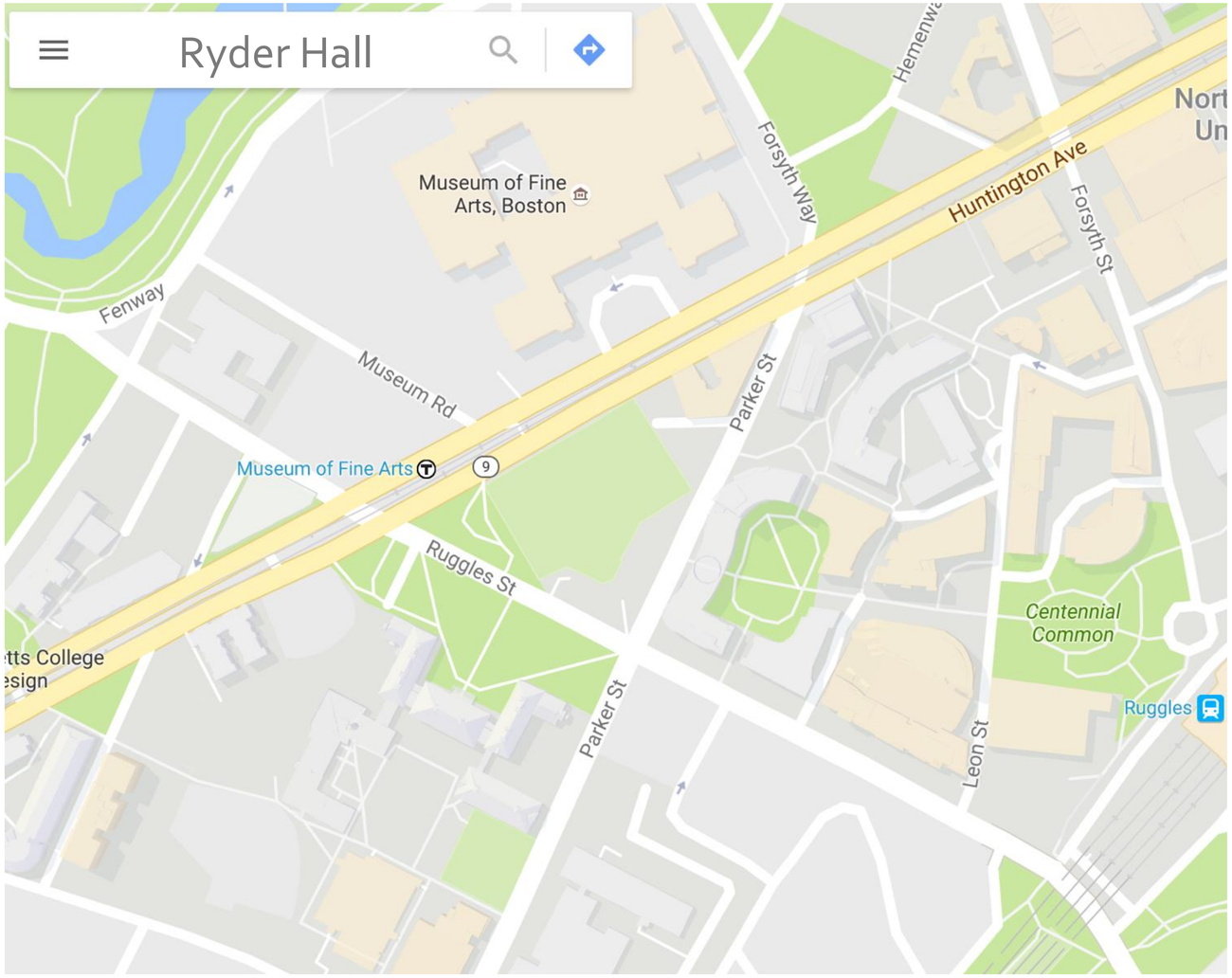
What?

Why?

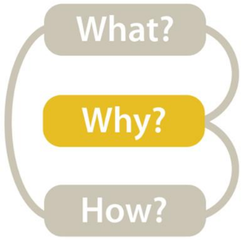
How?

Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

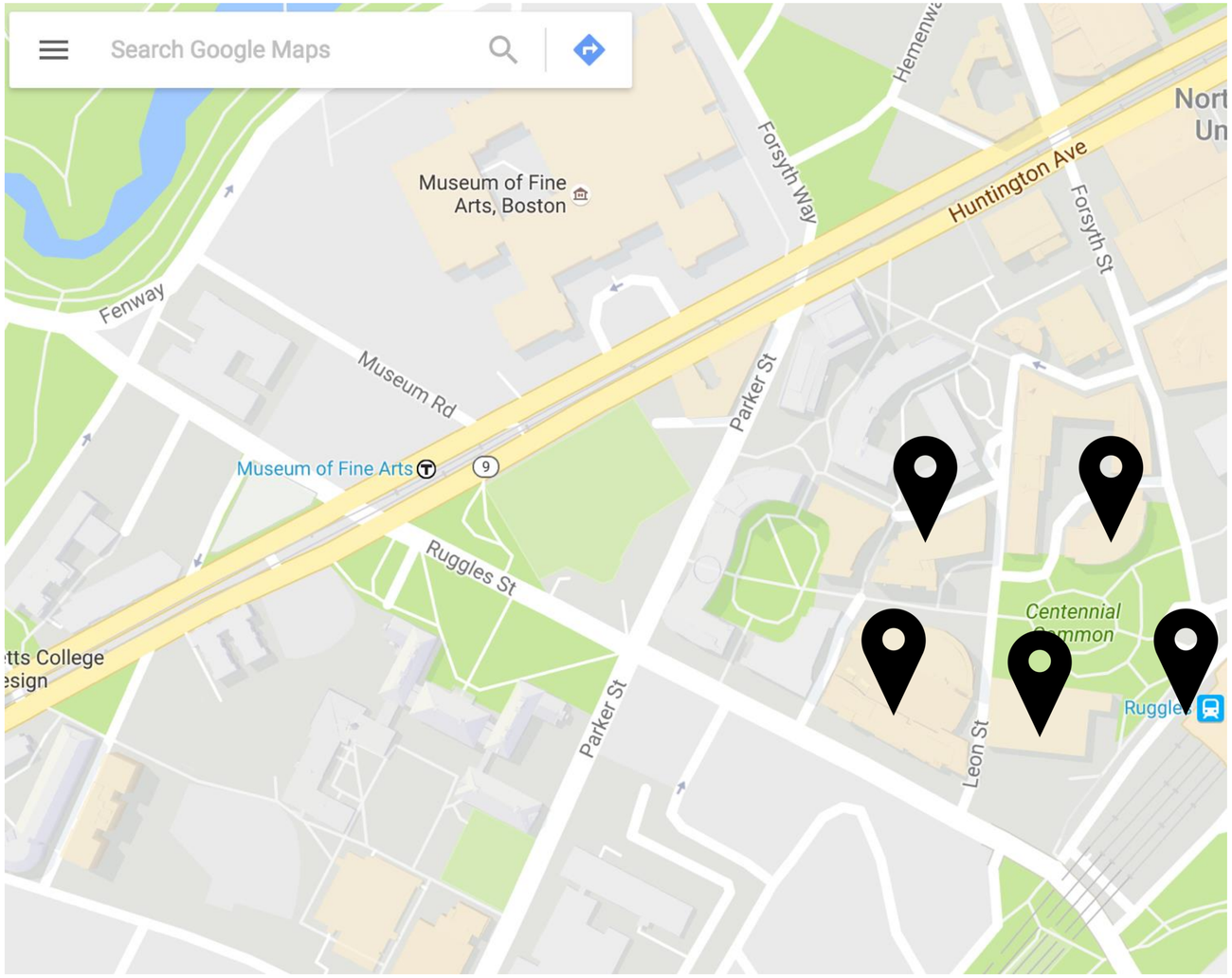


Where is Ryder Hall?

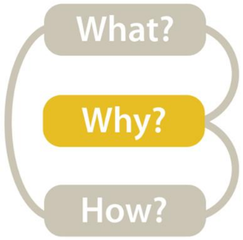


Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

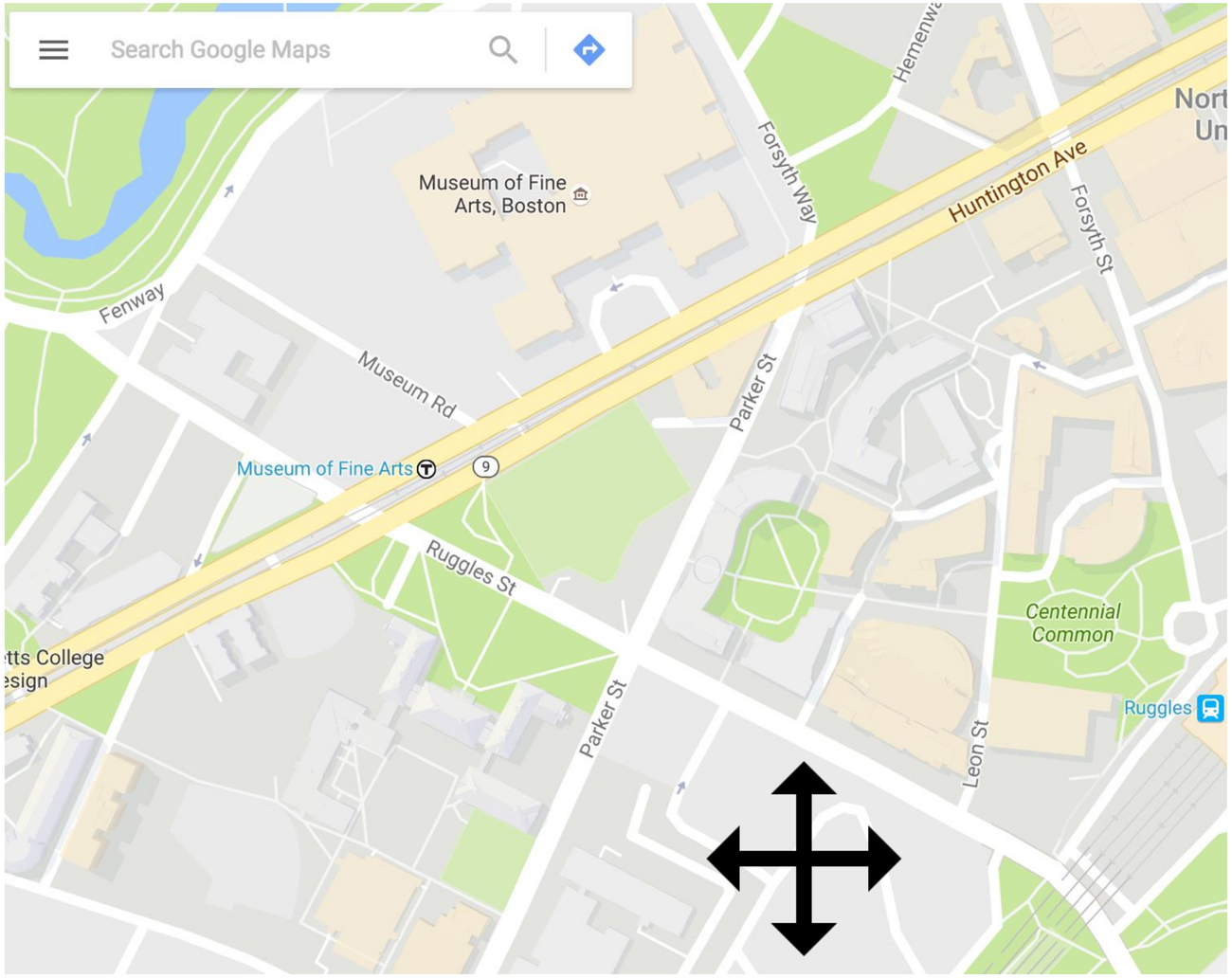


What buildings are near Ryder Hall?

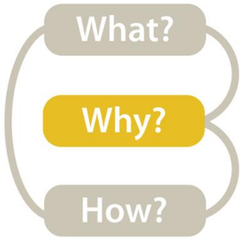


Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*



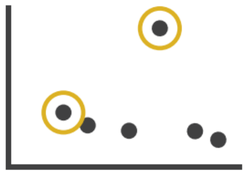
What is south of Huntington Ave?



ACTIONS define user goals. *Low-level*

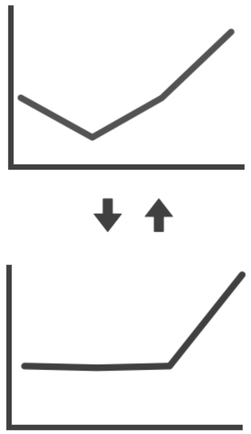
Query

→ Identify



single target

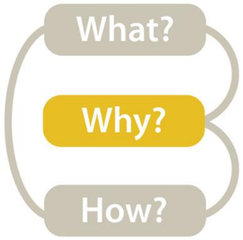
→ Compare



multiple targets

→ Summarize

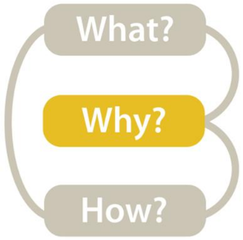
all targets



TARGETS are aspects of the data interest that are interest to the user.

Targets

- **All Data**
  - Trends
  - Outliers
  - Features
- **Attributes**
  - One
    - *Distribution*
    - *Extremes*
  - Many
    - *Dependency*
    - *Correlation*
    - *Similarity*
- **Network Data**
  - Topology
  - *Paths*
- **Spatial Data**
  - Shape



ACTIONS define user goals.

Lots of other task taxonomies...!

Actions

High-level

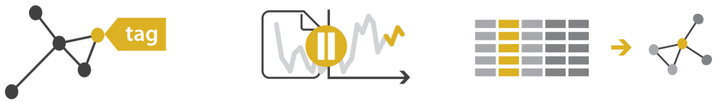
Analyze

→ Consume

→ *Discover* → *Present* → *Enjoy*

→ Produce

→ *Annotate* → *Record* → *Derive*



Mid-level

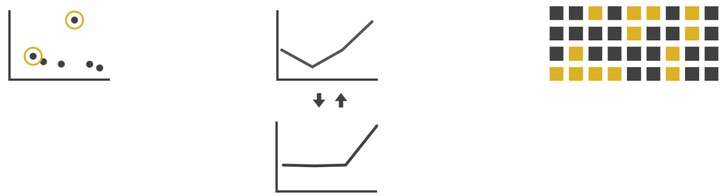
Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

Low-level

Query

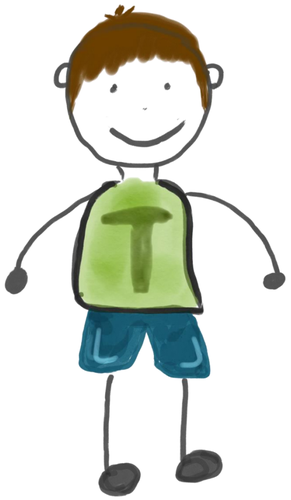
→ Identify → Compare → Summarize



An Example of Task Analysis → Visualization Design

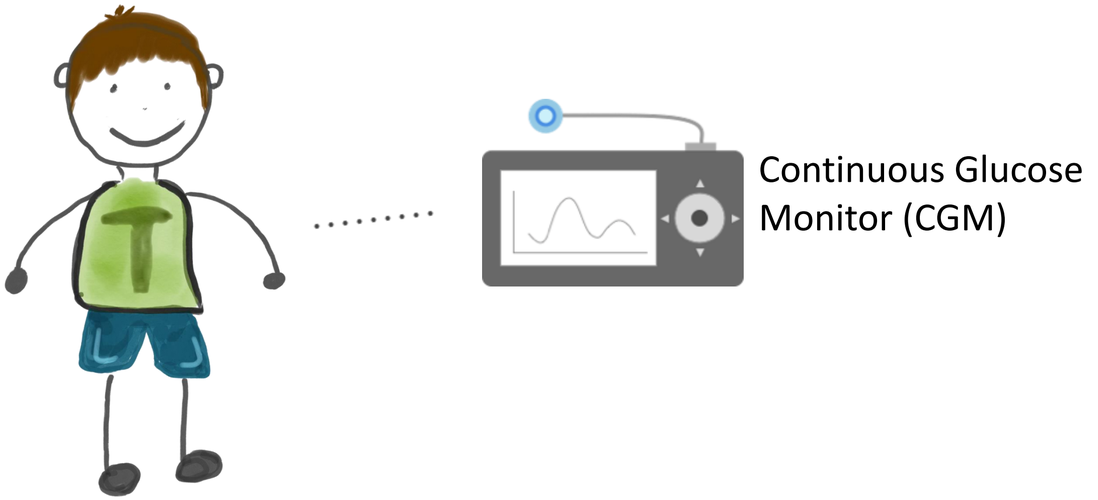
Zhang et al., 2018

Imagine a 10-year-old kid, who has been diagnosed with type 1 diabetes…



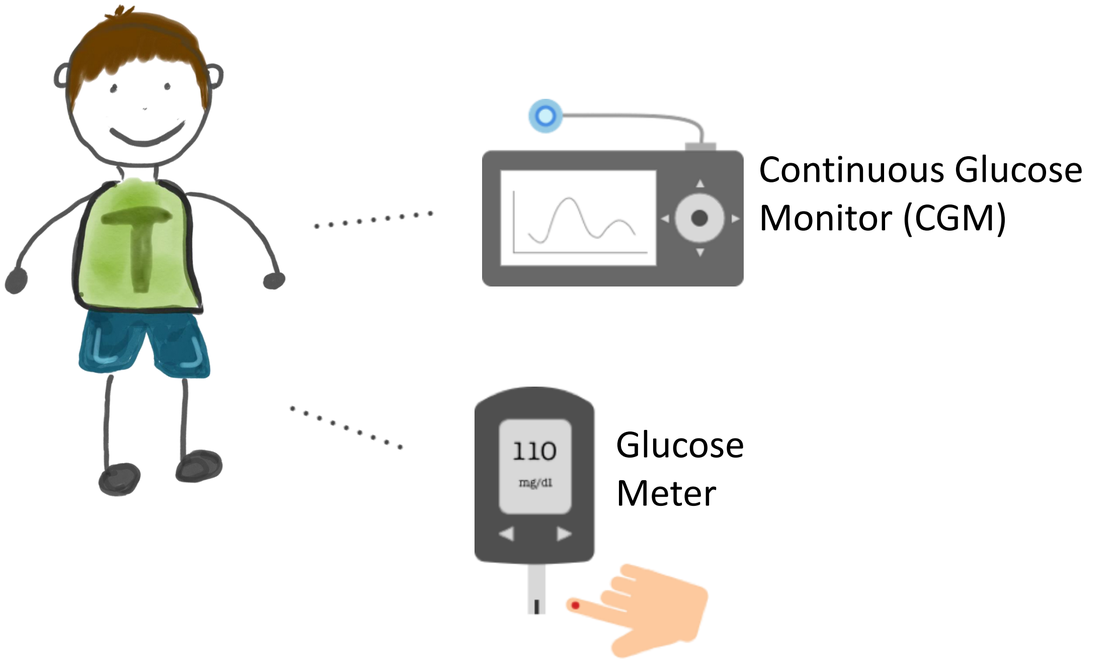
Zhang et al., 2018

Imagine a 10-year-old kid, who has been diagnosed with type 1 diabetes…



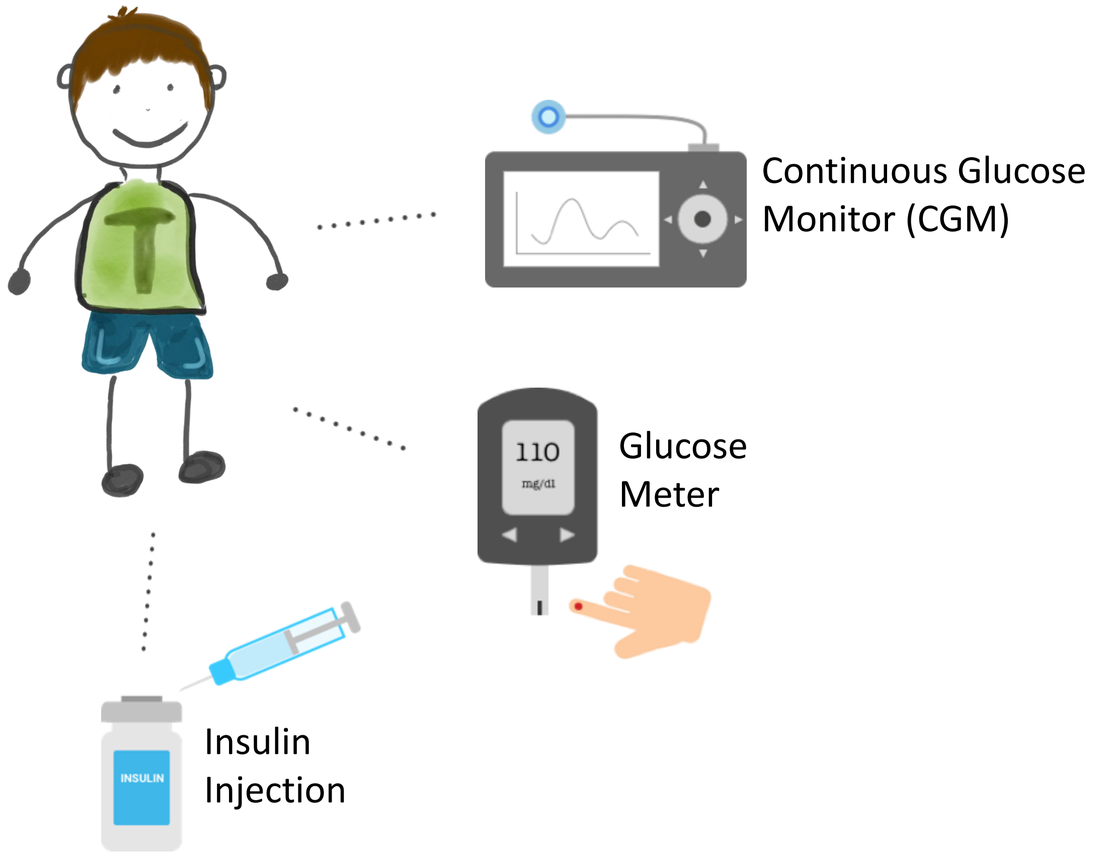
Zhang et al., 2018

Imagine a 10-year-old kid, who has been diagnosed with type 1 diabetes…



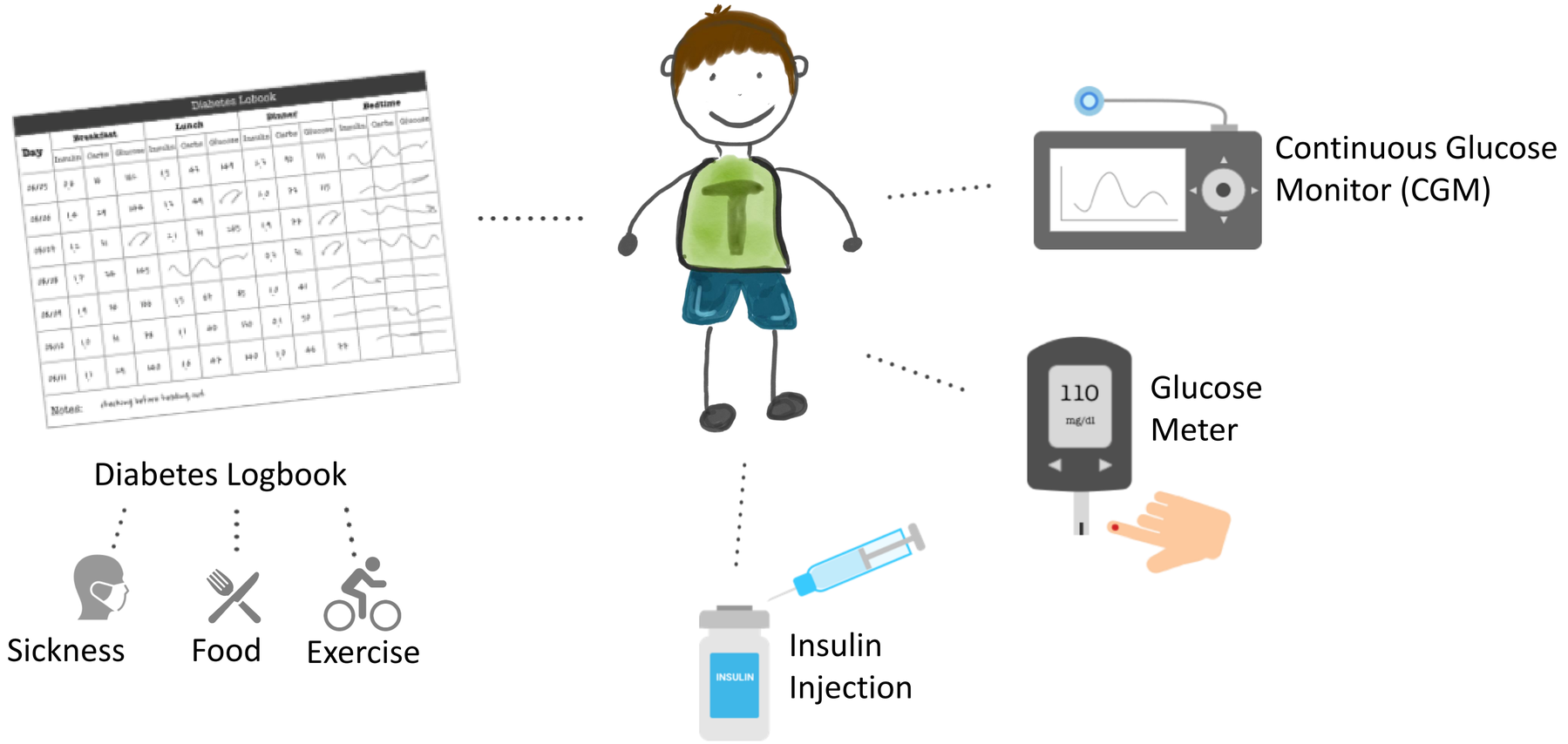
Zhang et al., 2018

Imagine a 10-year-old kid, who has been diagnosed with type 1 diabetes…



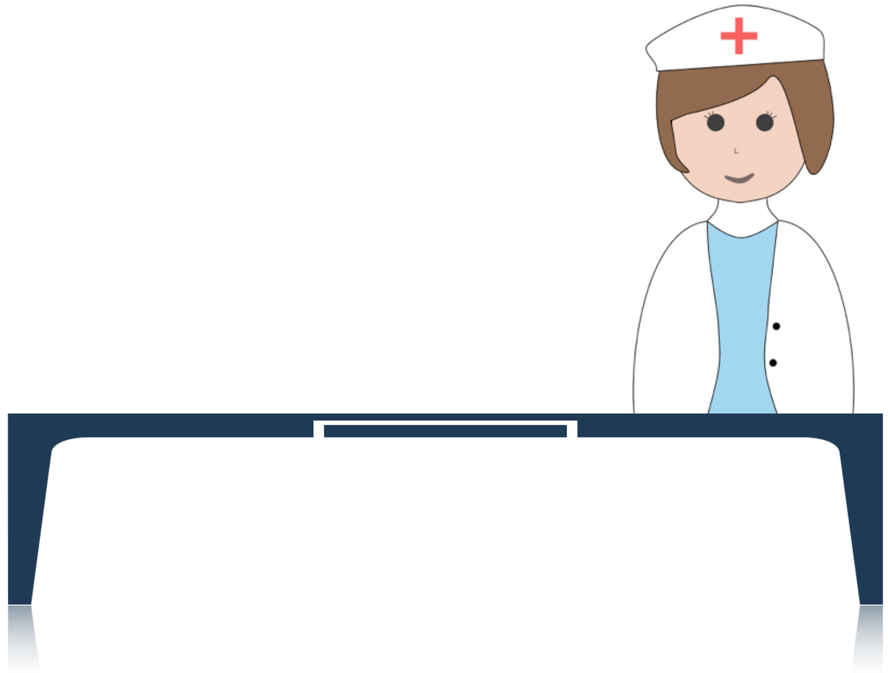
Zhang et al., 2018

Imagine a 10-year-old kid, who has been diagnosed with type 1 diabetes…



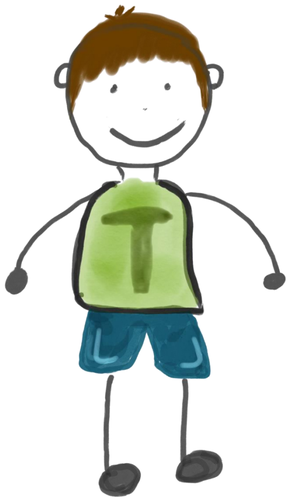
[Zhang et al., 2018]

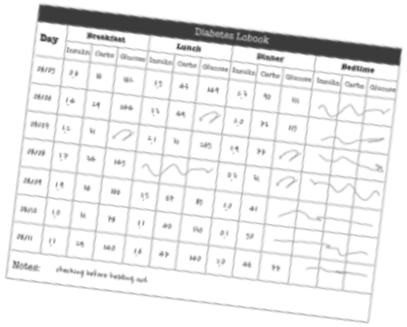
During a clinical visit …

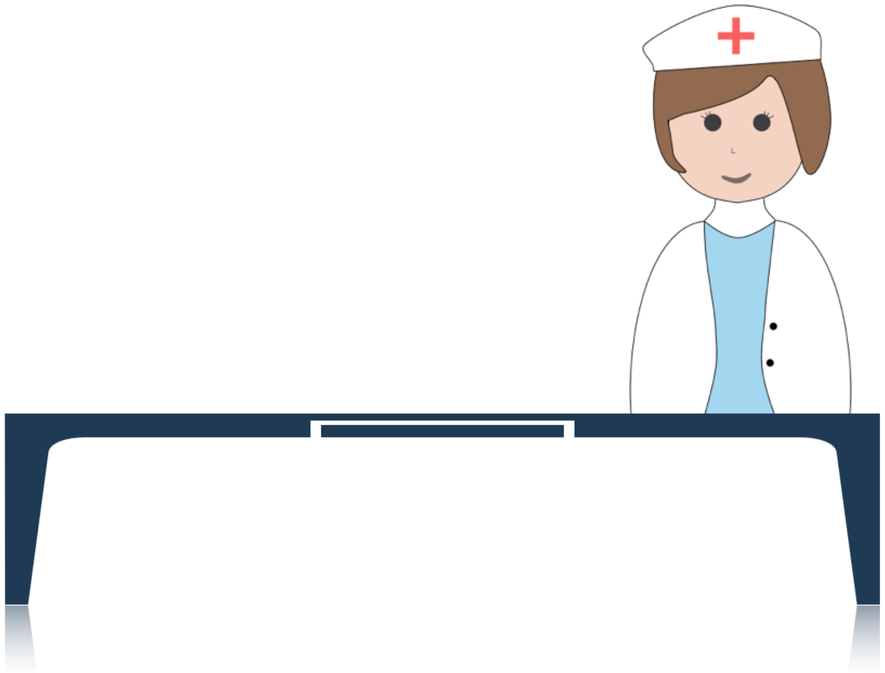


Zhang et al., 2018

During a clinical visit …

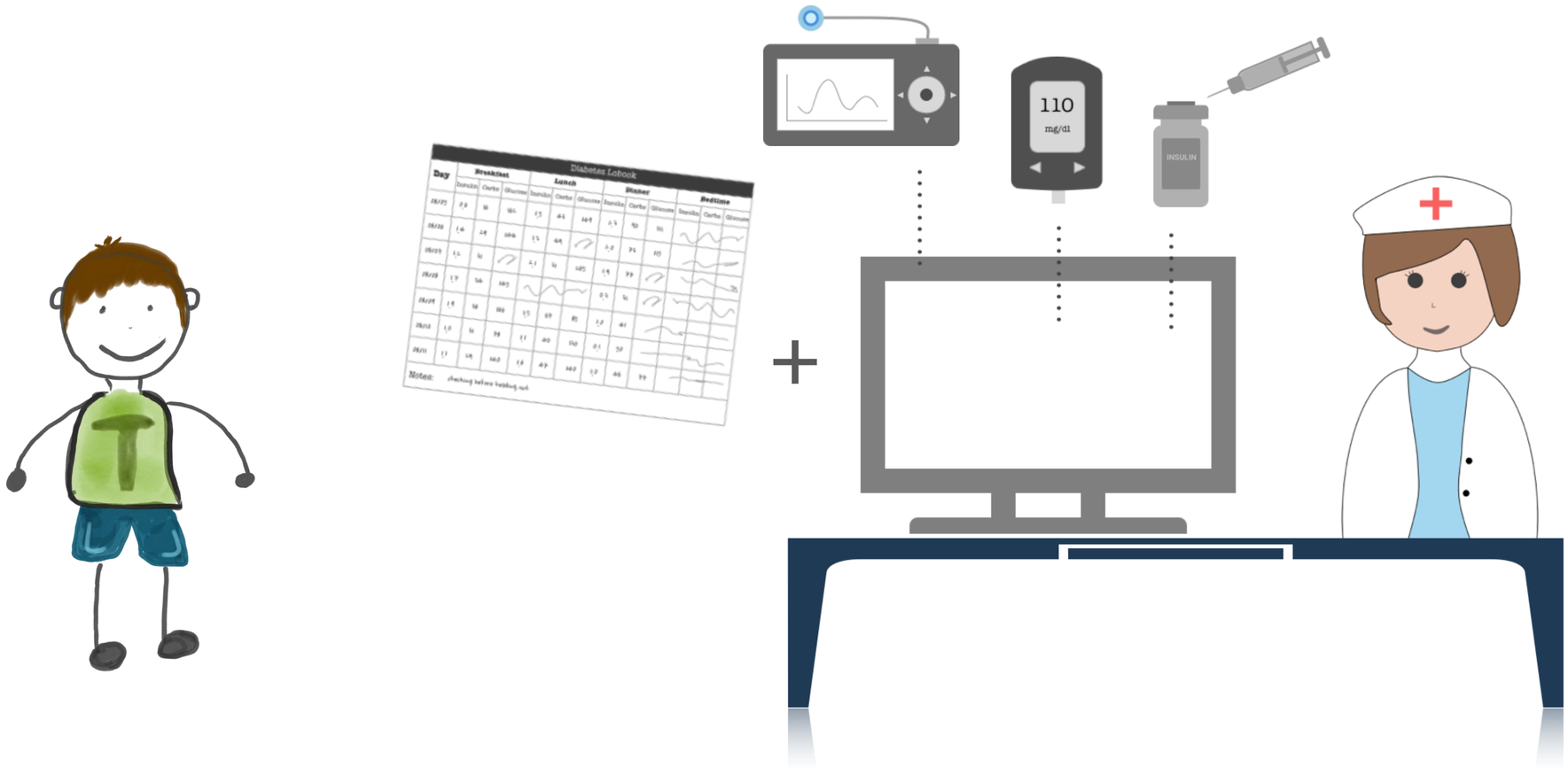






[Zhang et al., 2018]

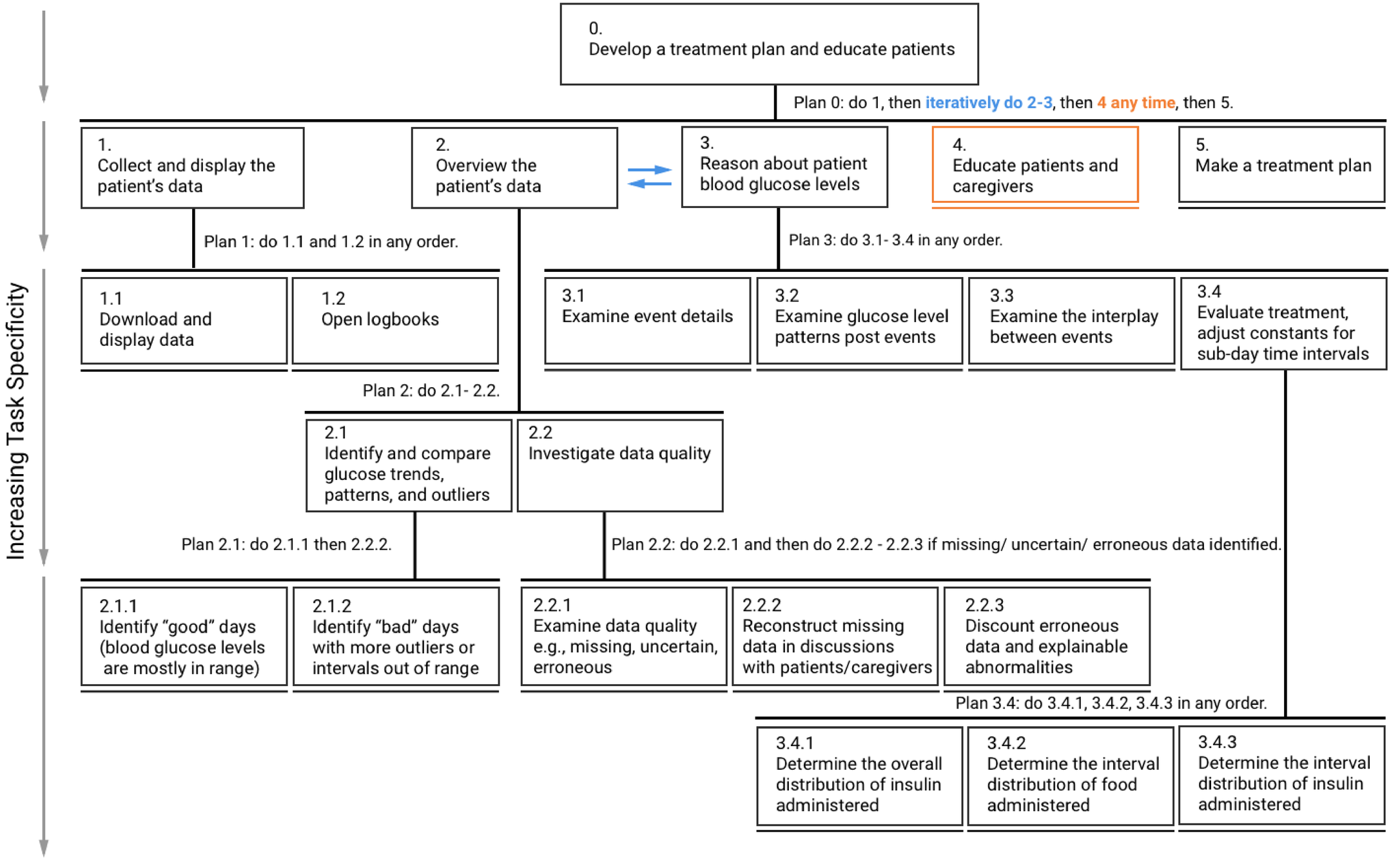
During a clinical visit …



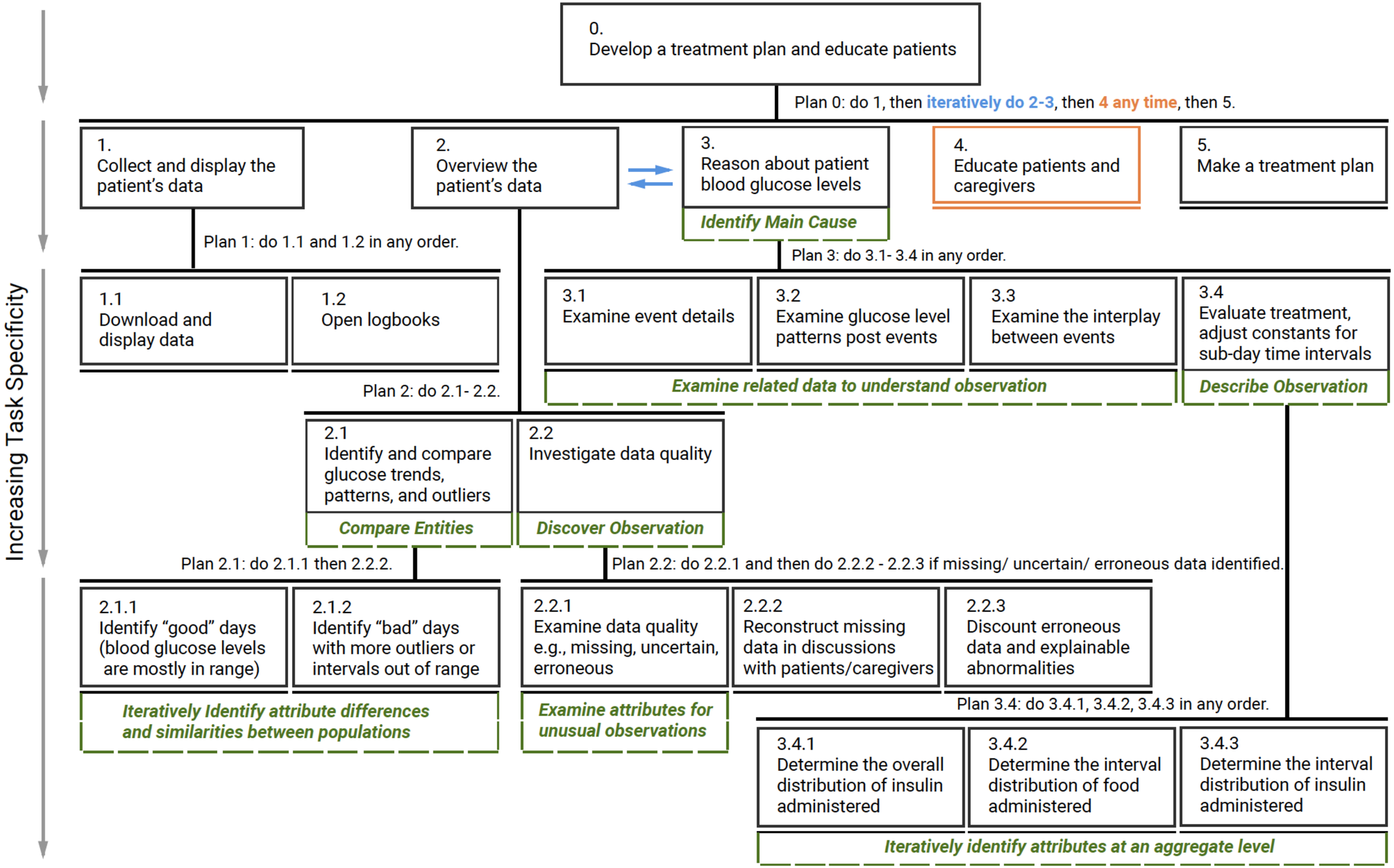
Zhang et al., 2018

During a clinical visit …

Zhang et al., 2018



Zhang et al., 2018



Zhang et al., 2018

Design Requirements

- DR1. Composite Visualization of **Integrated** Data
- DR2. Visualization of **Folded Temporal** Data
- DR3. **Align and Scale** Temporal Data
- DR4. **Summary** Statistics

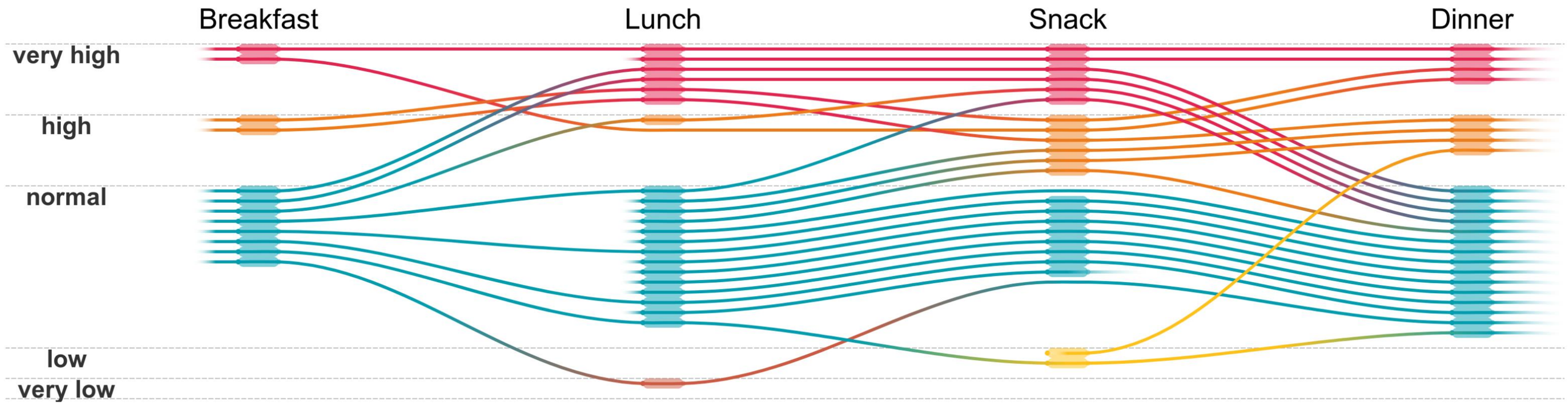
Zhang et al., 2018

14-Day Overview

Detail View

Summary Statistics Panel

Alternate



Di Bartolomeo et al., 2020

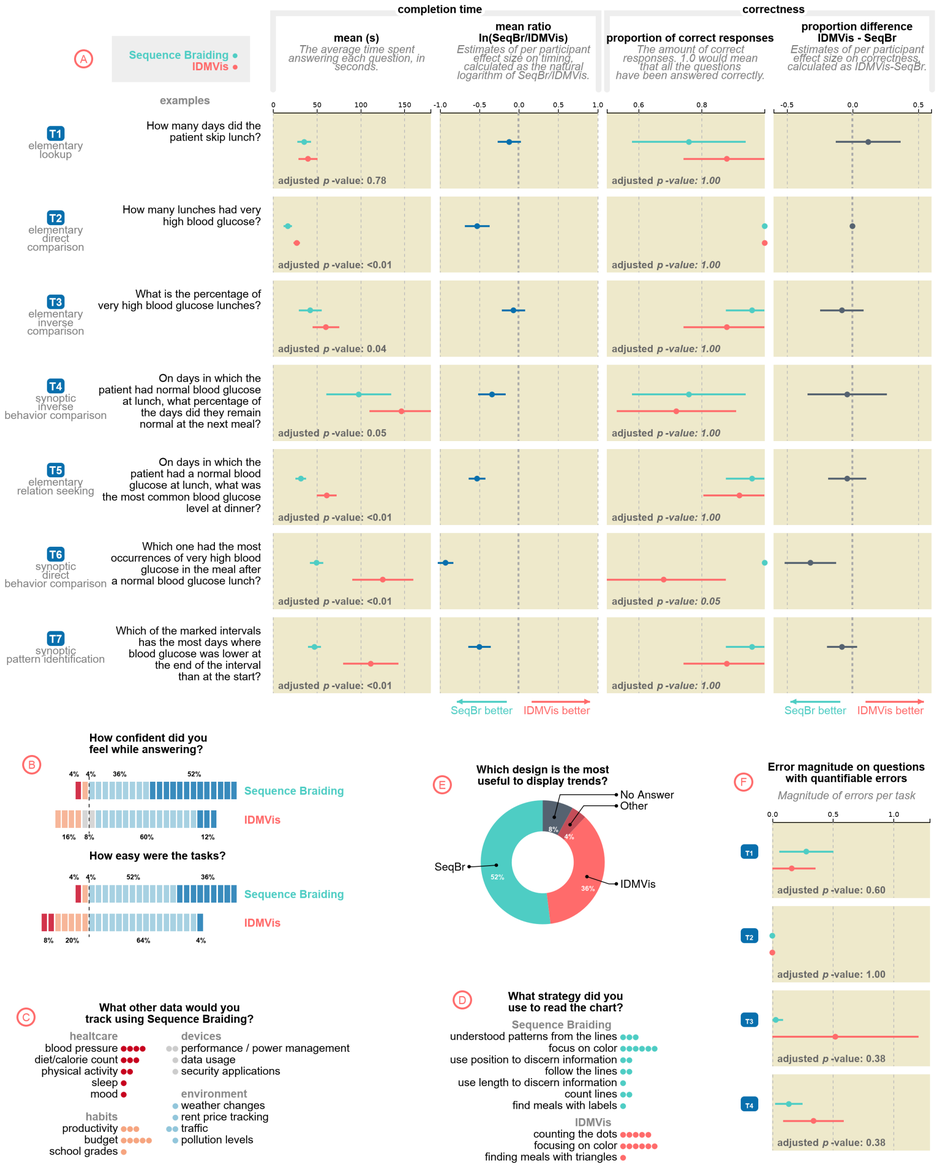
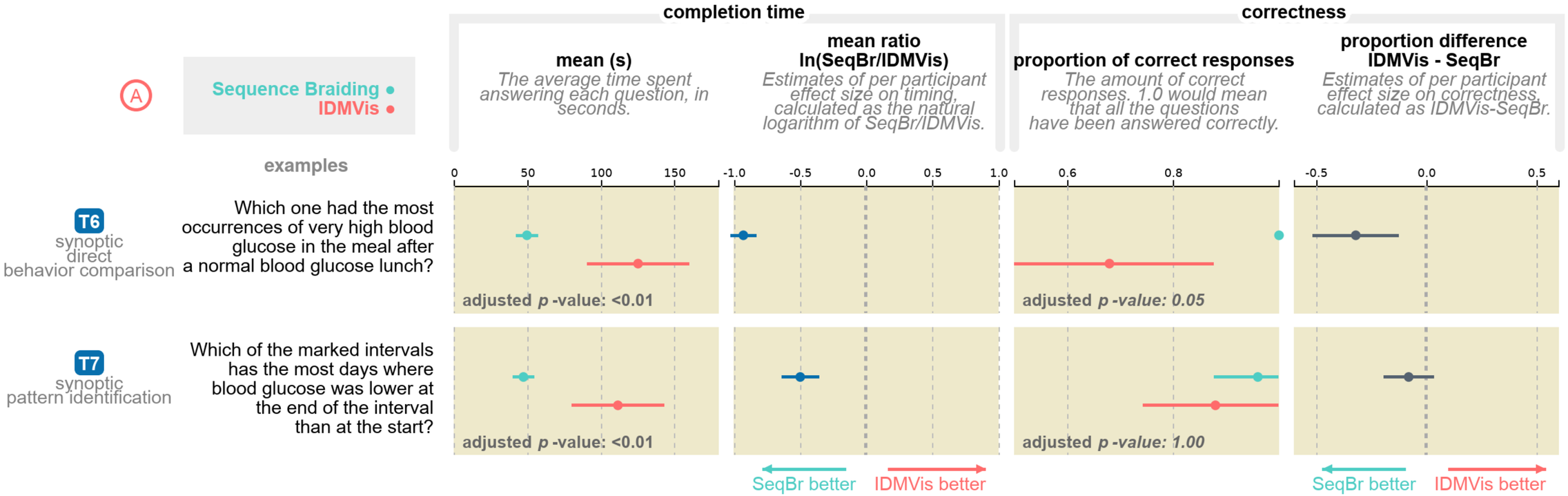
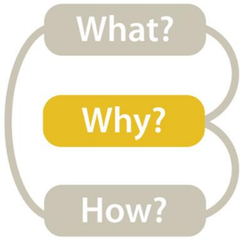


Fig. 9: Results of our evaluation comparing SEQUENCE BRAIDING vs. IDMVis [63]. (A) Completion time and correctness per task. Each row corresponds to the task at left, which is classified based on Andrienko & Andrienko [3]. The specific question instantiating that task for the study is in the second column. (B) Participants' Likert scale responses regarding confidence and ease of use. (C) Participants' answers when asked what other types of data would they use with SEQUENCE BRAIDING. (D) Participants' reported strategies used. (E) Participants' preference for which method was most useful for displaying trends. (F) Error magnitude per task, for those which are quantifiable.

[Di Bartolomeo et al., 2020]


completion time
correctness
mean (s)
The average time spent answering each question, in seconds.
mean ratio ln(SeqBr/IDMVis)
Estimates of per participant effect size on timing, calculated as the natural logarithm of SeqBr/IDMVis.
proportion of correct responses
The amount of correct responses. 1.0 would mean that all the questions have been answered correctly.
proportion difference IDMVis - SeqBr
Estimates of per participant effect size on correctness, calculated as IDMVis-SeqBr.
A
Sequence Braiding
IDMVis
examples
T6
synoptic
direct
behavior comparison
Which one had the most occurrences of very high blood glucose in the meal after a normal blood glucose lunch?
T7
synoptic
pattern identification
Which of the marked intervals has the most days where blood glucose was lower at the end of the interval than at the start?
0
50
100
150
-1.0
-0.5
0.0
0.5
1.0
0.6
0.8
-0.5
0.0
0.5
adjusted p -value: <0.01
adjusted p -value: 0.05
adjusted p -value: <0.01
adjusted p -value: 1.00
SeqBr better
IDMVis better
SeqBr better
IDMVis better



Interview Advice

- Have a designated note-taker and designated leader
- Be prepared. (Have some questions prepared in advance.)
- Start slow, safe, and personal.
- Coax, don’t hammer.
- Make some questions open ended.
- Ask what you don’t know.
- Let the interviewees wander a bit–but be careful.
- Listen, really listen.
- For software, look for “work arounds” and hacks.
- Make sure to write down your thoughts and impressions immediately after the interview.
- You are the visualization expert – don’t ask them what vis they want, don’t think too early about what vis to build.

Shel Israel, 2012

Upcoming Assignments & Communication

https://northeastern.instructure.com/courses/63405/assignments/syllabus

Look at the upcoming assignments and deadlines regularly!

- Textbook, Readings, & Reading Quizzes — Variable days
- In-Class Activities — 11:59pm same day as class
  - F: In-class project pitches
  - T: In-class project group finalizing & work
  - Next F: Lecture & in-class activity on Altair
- Assignments & Projects— Generally due R 11:59pm
  - R (2 days): Project 1 (pitches), Assignments 4a (critique) & 4b (Altair setup) due
  - Next R (9 days): Project 2 (proposals), Assignment 5 due (D3)
  - Next-next R (16 days): Assignments 6a (Altair) & 6b (critique) due.
  - Next-next-next R (23 days): Project 3 — Interview & Task Analysis

Everyday Required Supplies:

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

Use Canvas Discussions for general questions, email the TAs/S-LTA/instructor for questions specific to you: codydunne-and-tas@ccs.neu.edu. Include links!