

Task Abstraction, Design Rules of Thumb — Continued

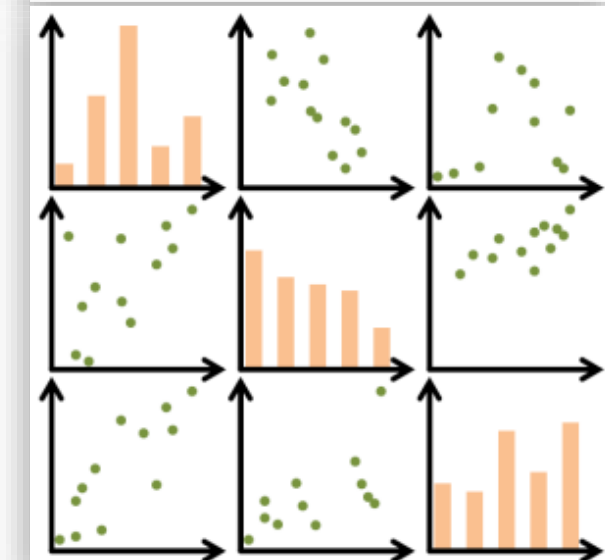
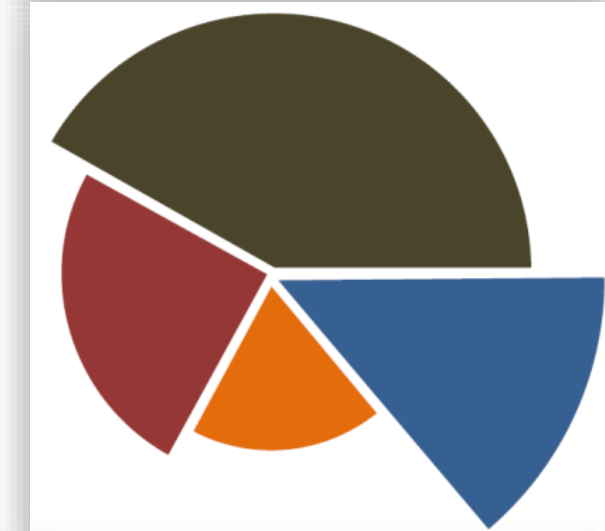
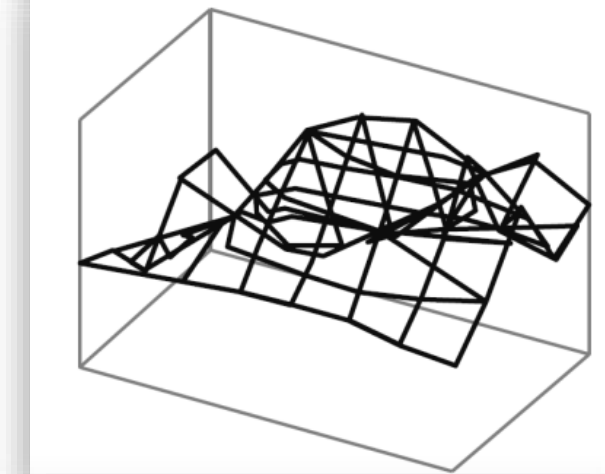
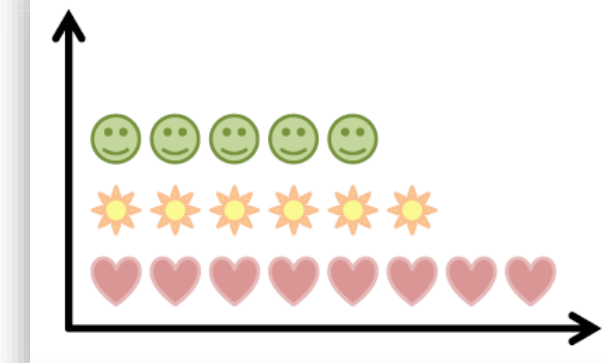
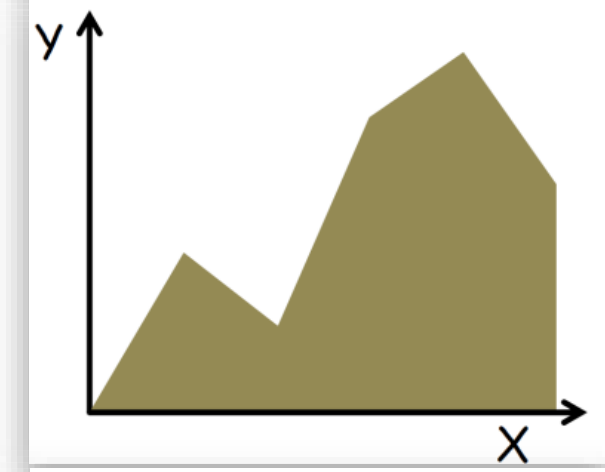
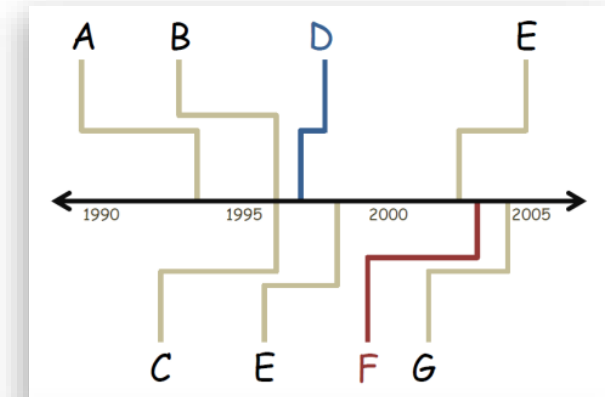
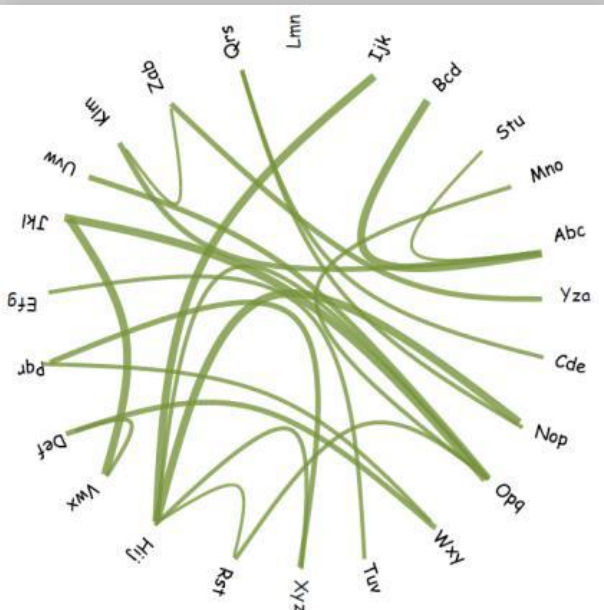
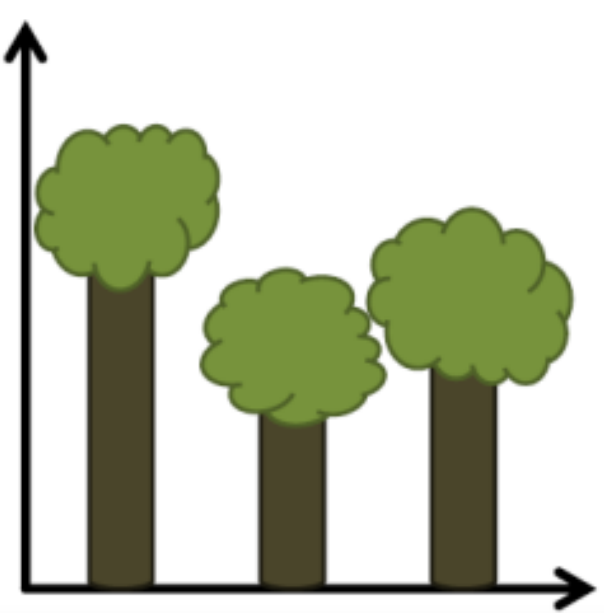
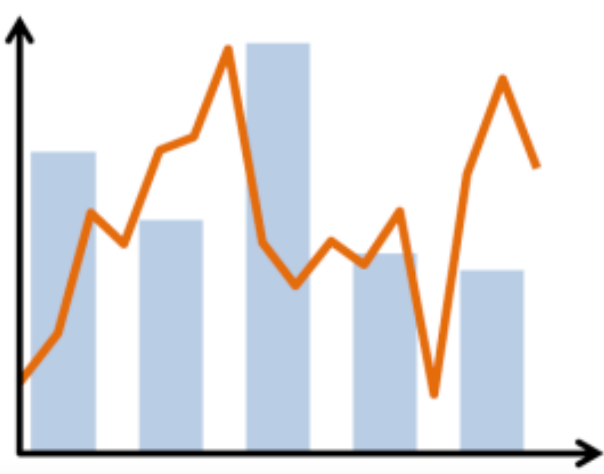
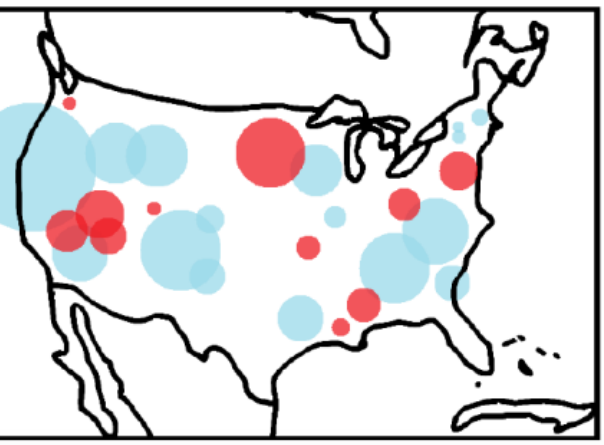
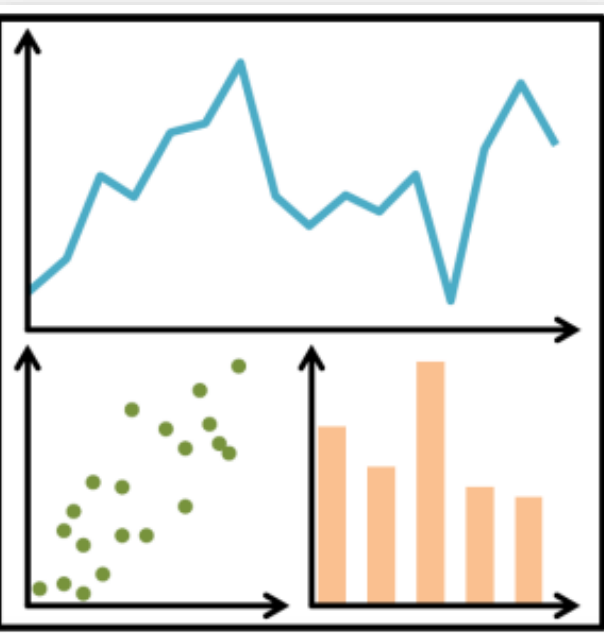
DS 4200

FALL 2020

Prof. Cody Dunne

NORTHEASTERN UNIVERSITY

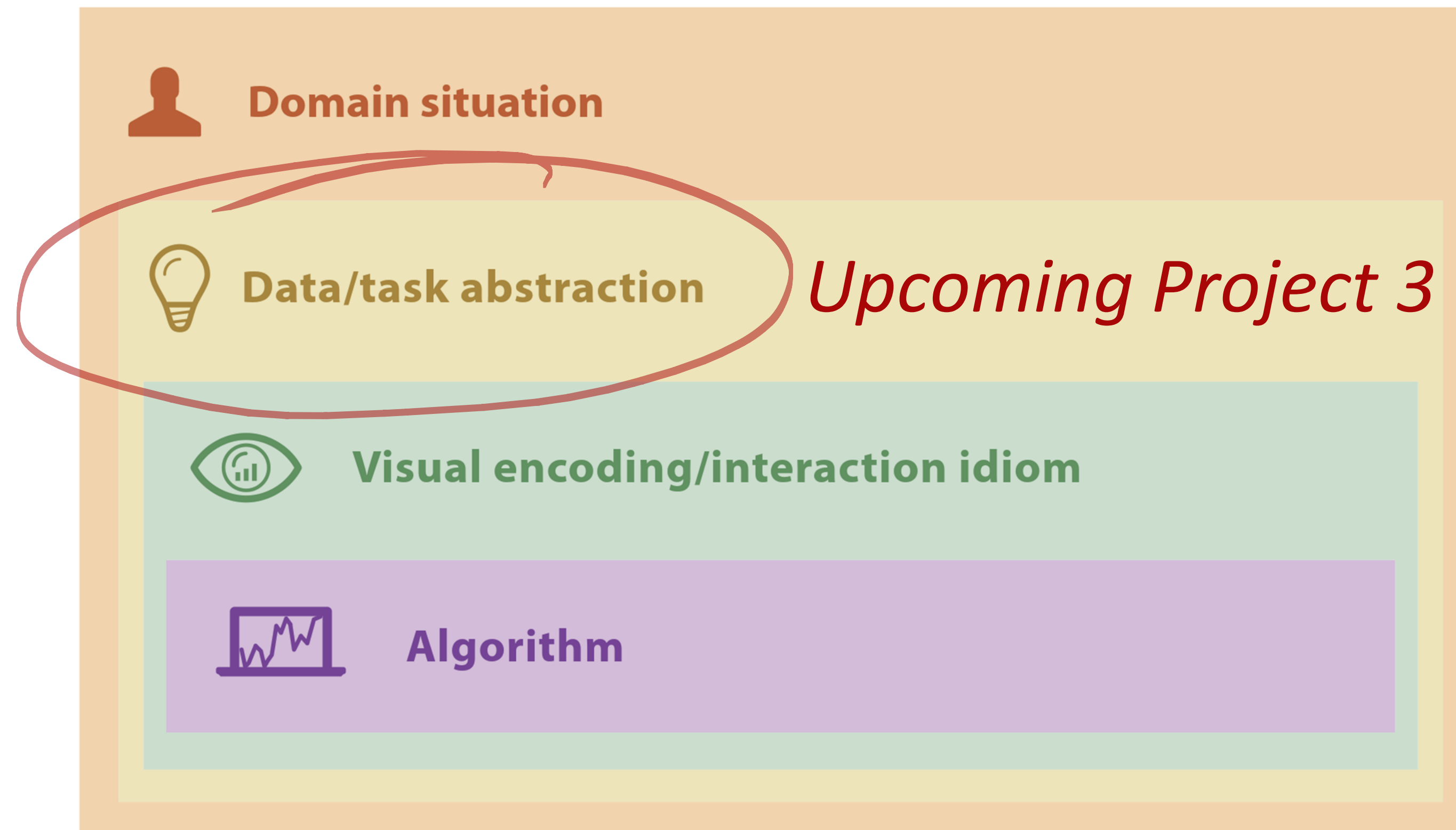
*Slides and inspiration from Michelle Borkin, Krzysztof Gajos, Hanspeter Pfister,
Miriah Meyer, Jonathan Schwabish, and David Sprague*







CHECK-IN





PREVIOUSLY, ON DS 4200...

Nested Model



Threats to Validity *✓ Final Project validation*

-  Domain situation
-  Data/task abstraction
-  Visual encoding/interaction idiom
-  Algorithm

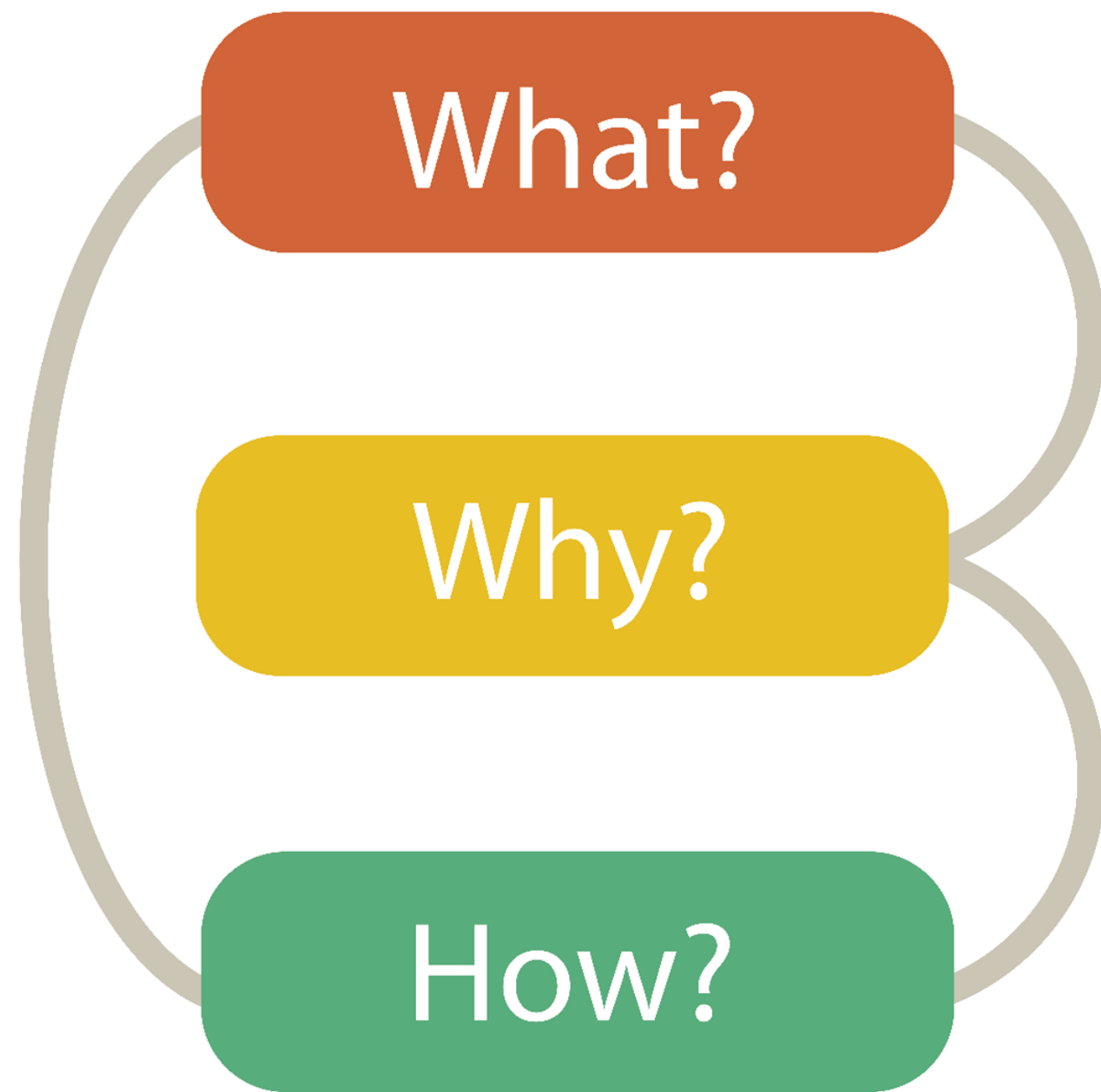
-  **Threat** Wrong problem ✓
 - Validate** Observe and interview target users
-  **Threat** Wrong task/data abstraction
-  **Threat** Ineffective encoding/interaction idiom ✓
 - Validate** Justify encoding/interaction design
-  **Threat** Slow algorithm
 - Validate** Analyze computational complexity
 - Implement system**
 - Validate** Measure system time/memory
- Validate** Qualitative/quantitative result image analysis
Test on any users, informal usability study ✓
- Validate** Lab study, measure human time/errors for task
- Validate** Test on target users, collect anecdotal evidence of utility
- Validate** Field study, document human usage of deployed system
- Validate** Observe adoption rates

Final project follow-up

Now, ON DS 4200...

TASK ABSTRACTION

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

Targets

→ Analyze

→ Consume

→ Discover



→ Present



→ Enjoy



→ Produce

→ Annotate



→ Record



→ Derive

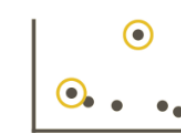


→ Search

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

→ Query

→ Identify



→ Compare

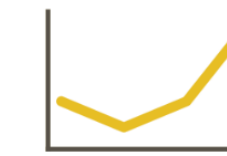


→ Summarize



→ All Data

→ Trends



→ Outliers



→ Features



→ Attributes

→ One

→ Distribution



→ Extremes

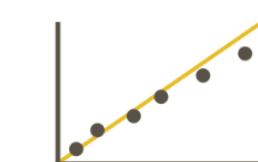


→ Many

→ Dependency



→ Correlation



→ Similarity



→ Network Data

→ Topology

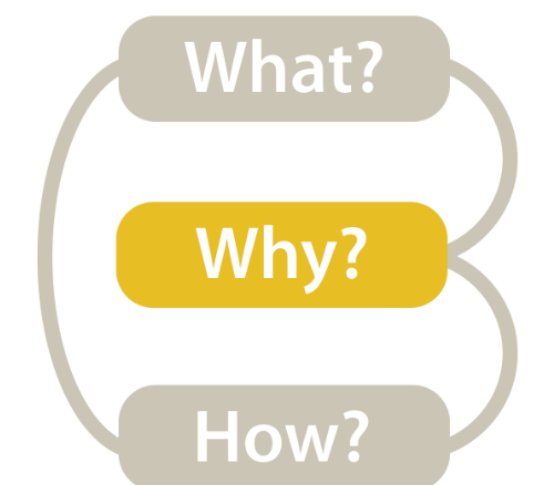
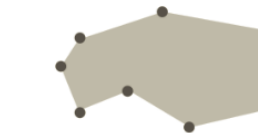


→ Paths



→ Spatial Data

→ Shape



TASK ABSTRACTION

Why abstract?

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

Visualization Tools

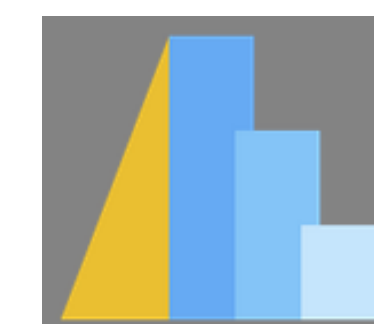
Specific

General



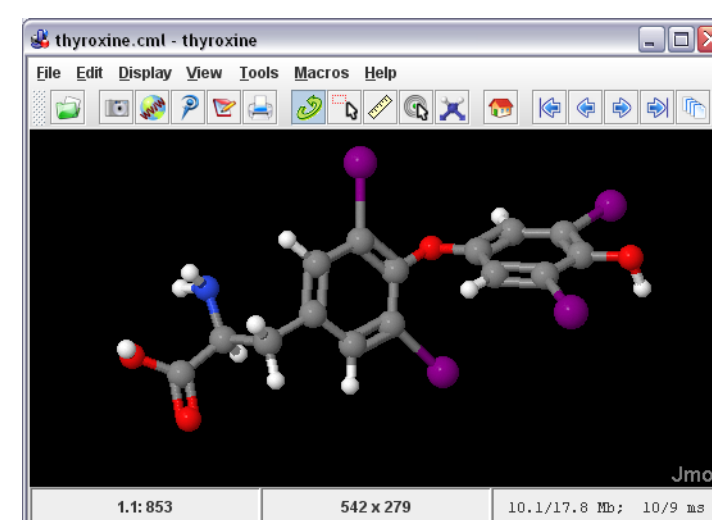
 Gephi

Altair



 ArcGIS™
ESRI

 + a b l e a u®



TASK ABSTRACTION

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

Why?

Actions

Targets

→ Analyze

→ Consume

→ Discover



→ Present



→ Enjoy



→ Produce

→ Annotate



→ Record



→ Derive

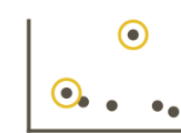


→ Search

	Target known	Target unknown
Location known	<i>Lookup</i>	<i>Browse</i>
Location unknown	<i>Locate</i>	<i>Explore</i>

→ Query

→ Identify



→ Compare



→ Summarize



→ All Data

→ Trends



→ Outliers



→ Features



→ Attributes

→ One

→ Distribution



→ Extremes

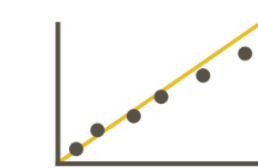


→ Many

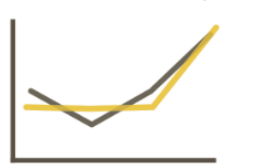
→ Dependency



→ Correlation



→ Similarity



→ Network Data

→ Topology

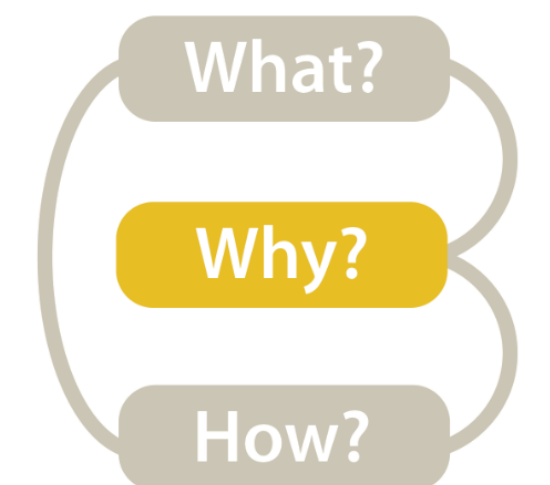


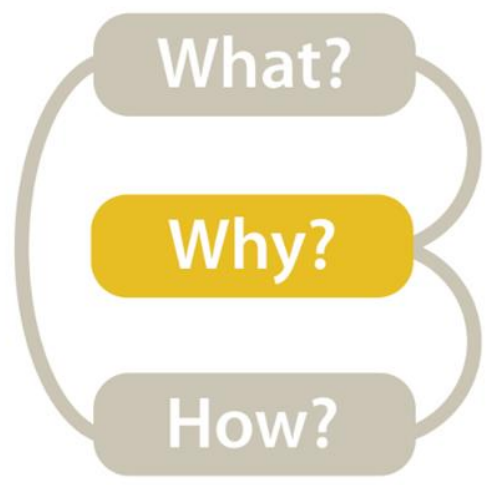
→ Paths



→ Spatial Data

→ Shape





High-level

ACTIONS define user goals.

➔ Analyze

➔ Consume

➔ *Discover*



➔ *Present*



➔ *Enjoy*



➔ Produce

➔ *Annotate*

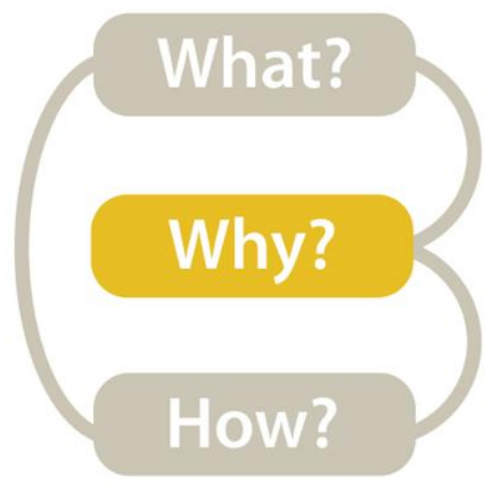


➔ *Record*



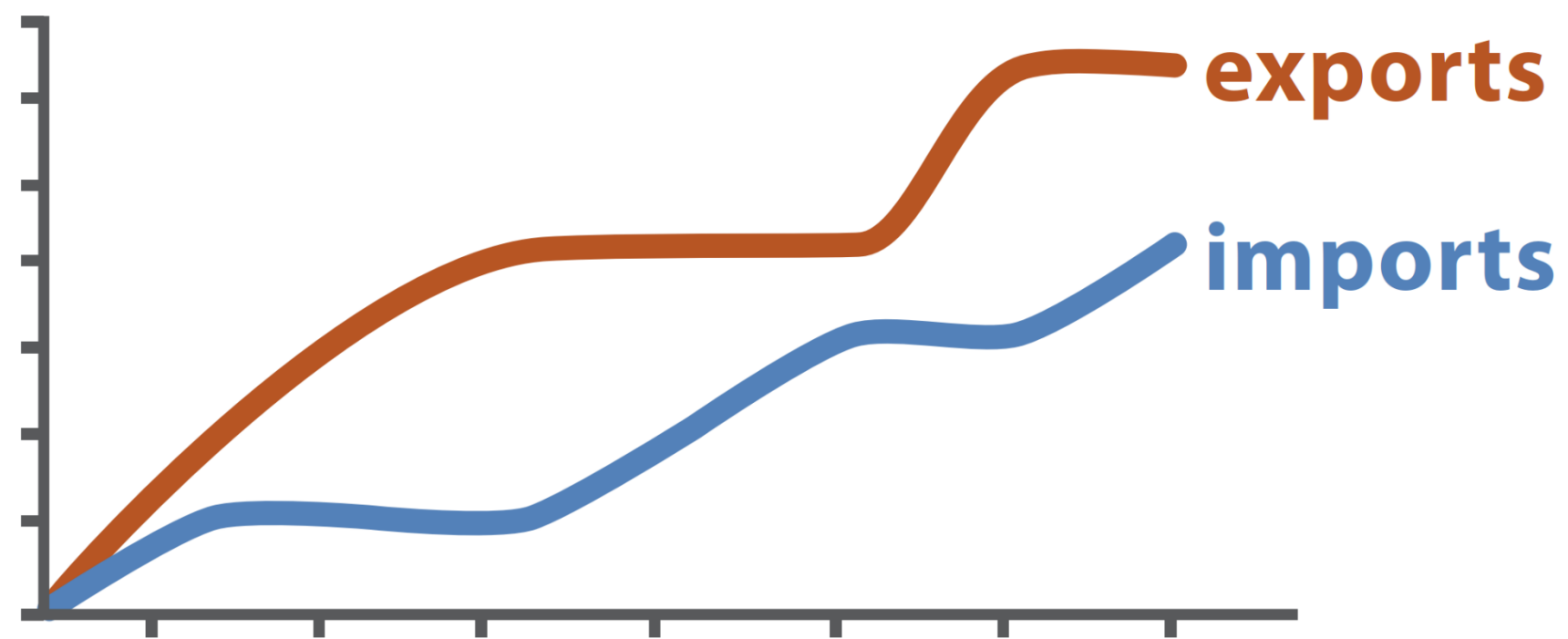
➔ *Derive*



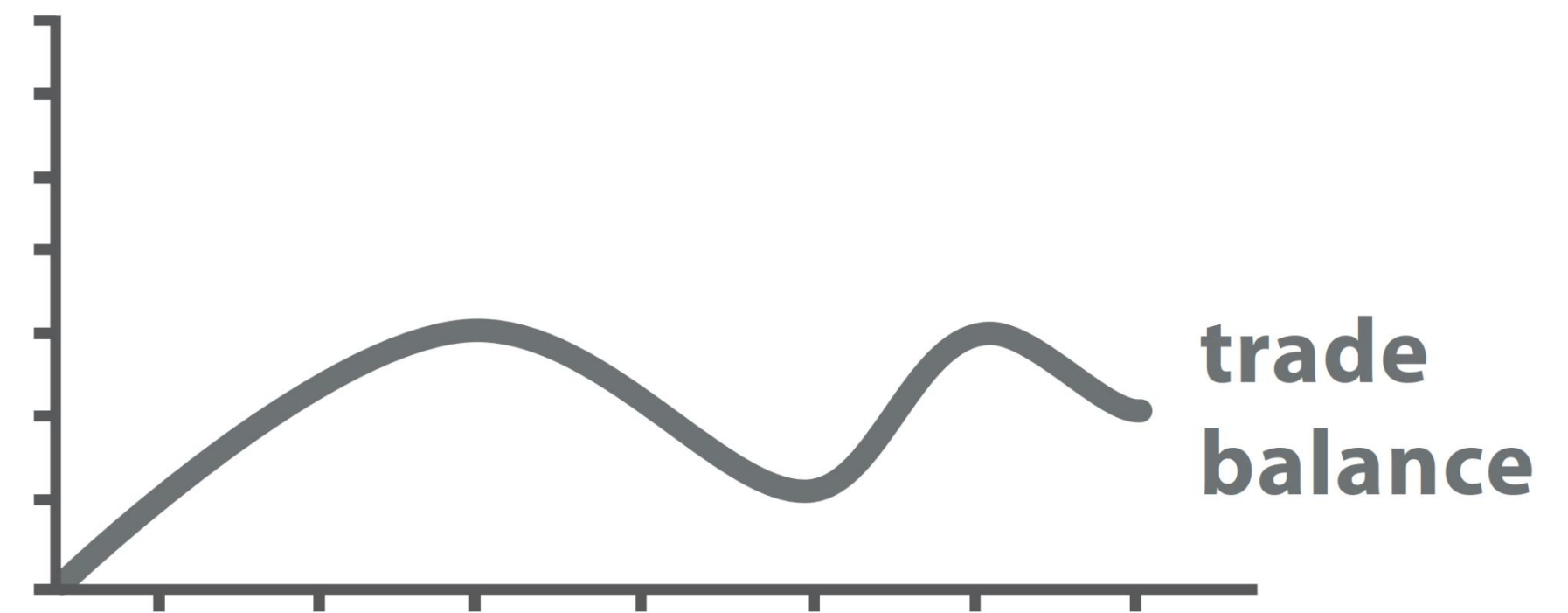


ACTIONS define user goals.

→ *Derive*

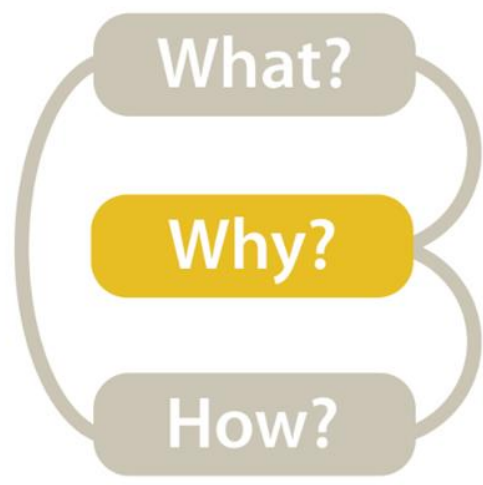


Original Data







$$\text{trade balance} = \text{exports} - \text{imports}$$

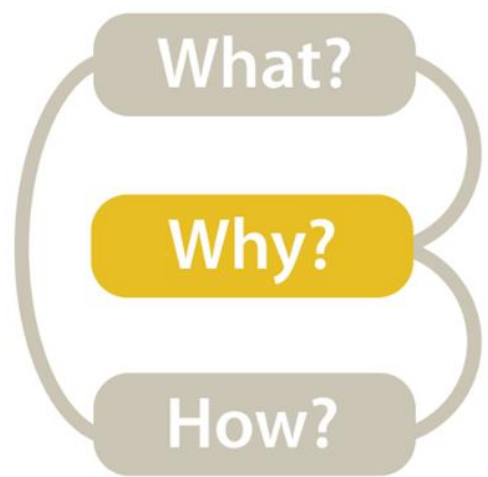
Derived Data



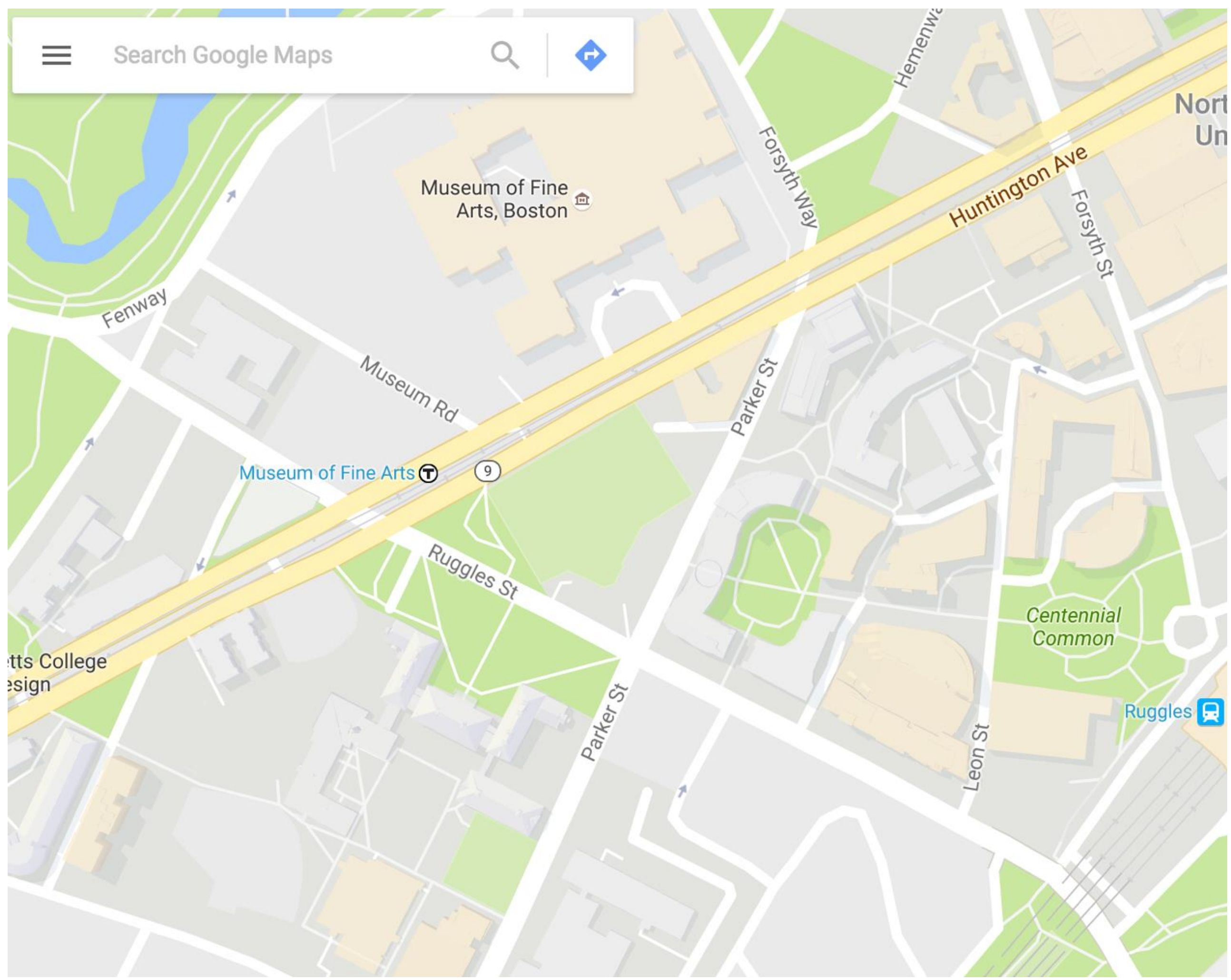
ACTIONS define user goals. *Mid-level*





➔ **Search**

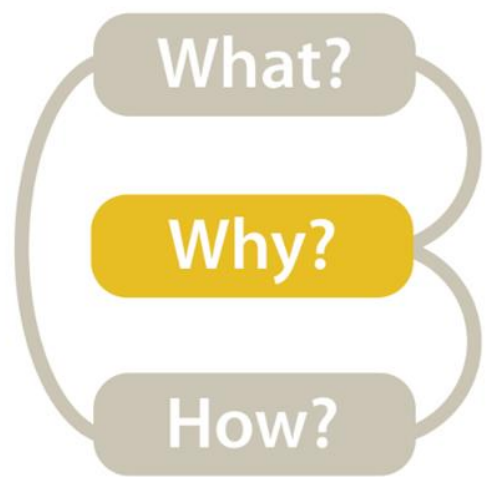
	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>



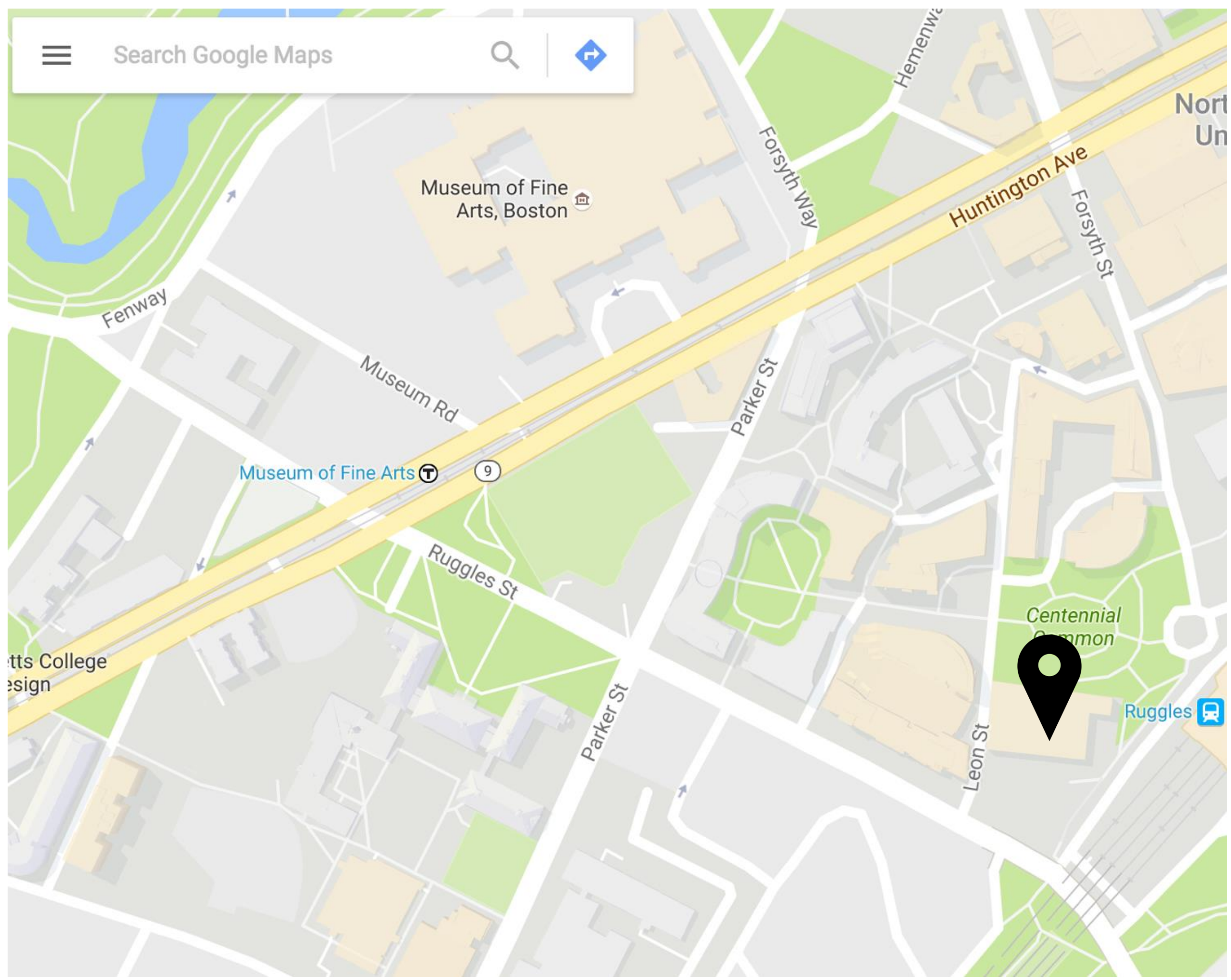
➔ Search







	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

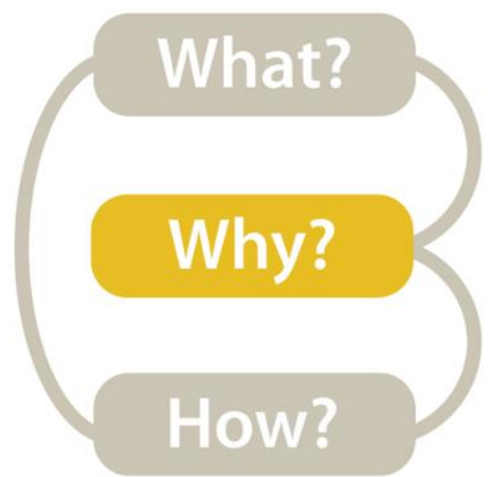


➔ Search

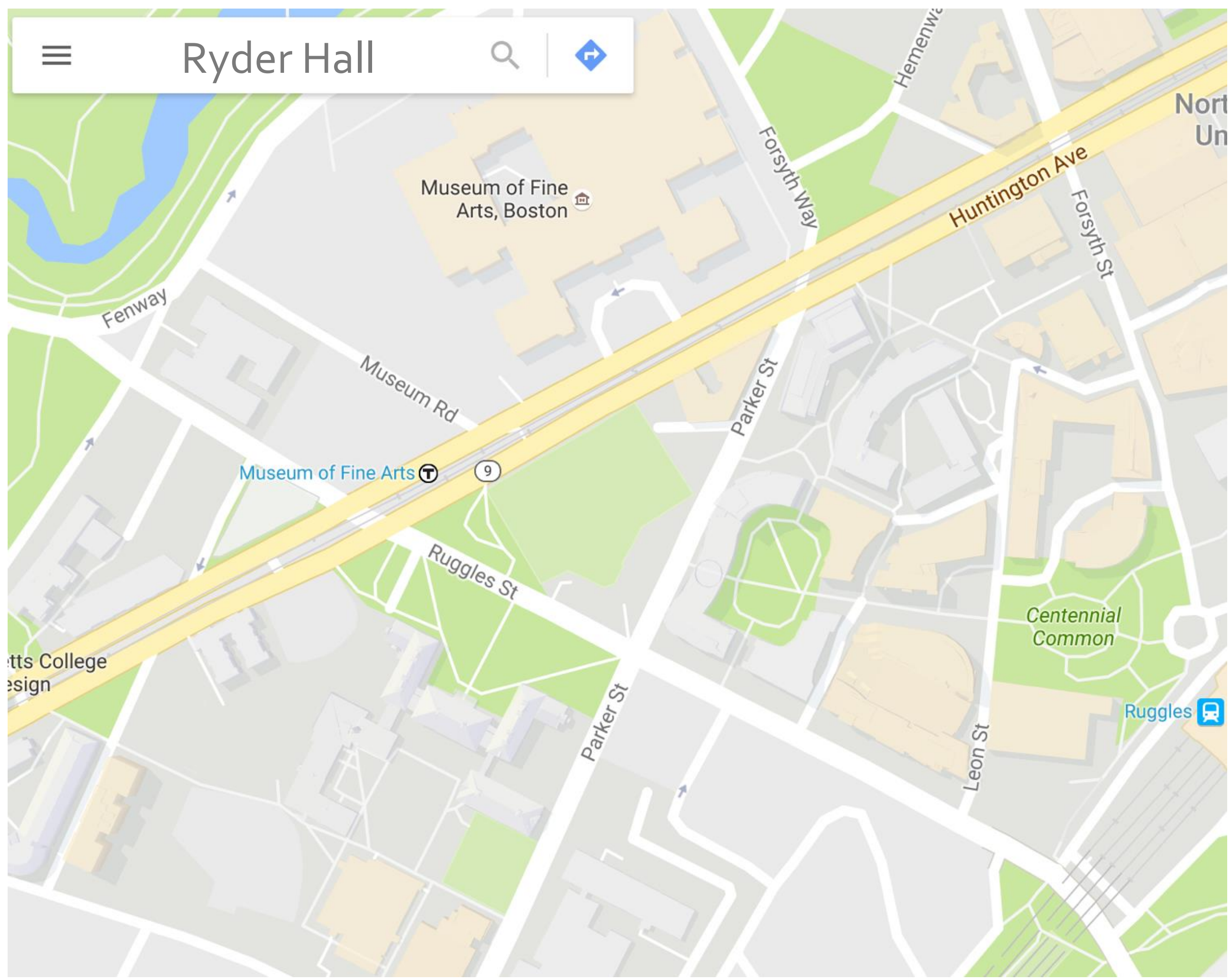


	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

What is the address of Ryder hall?

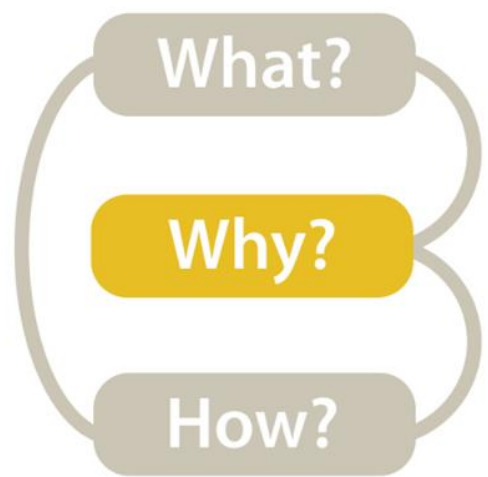


➔ Search



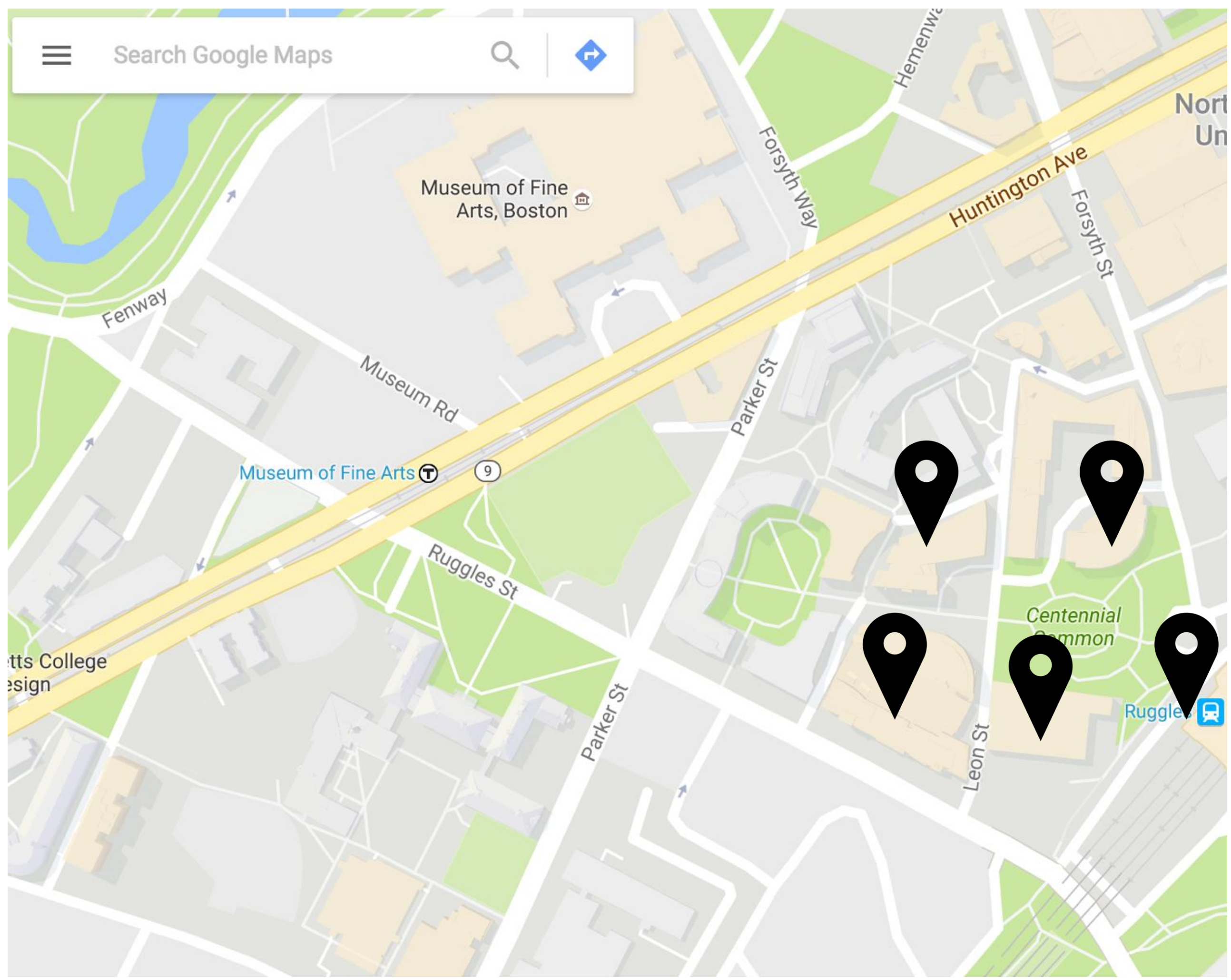
	Target known	Target unknown
Location known	<i>Lookup</i>	<i>Browse</i>
Location unknown	<i>Locate</i>	<i>Explore</i>

Where is Ryder Hall?



➔ Search

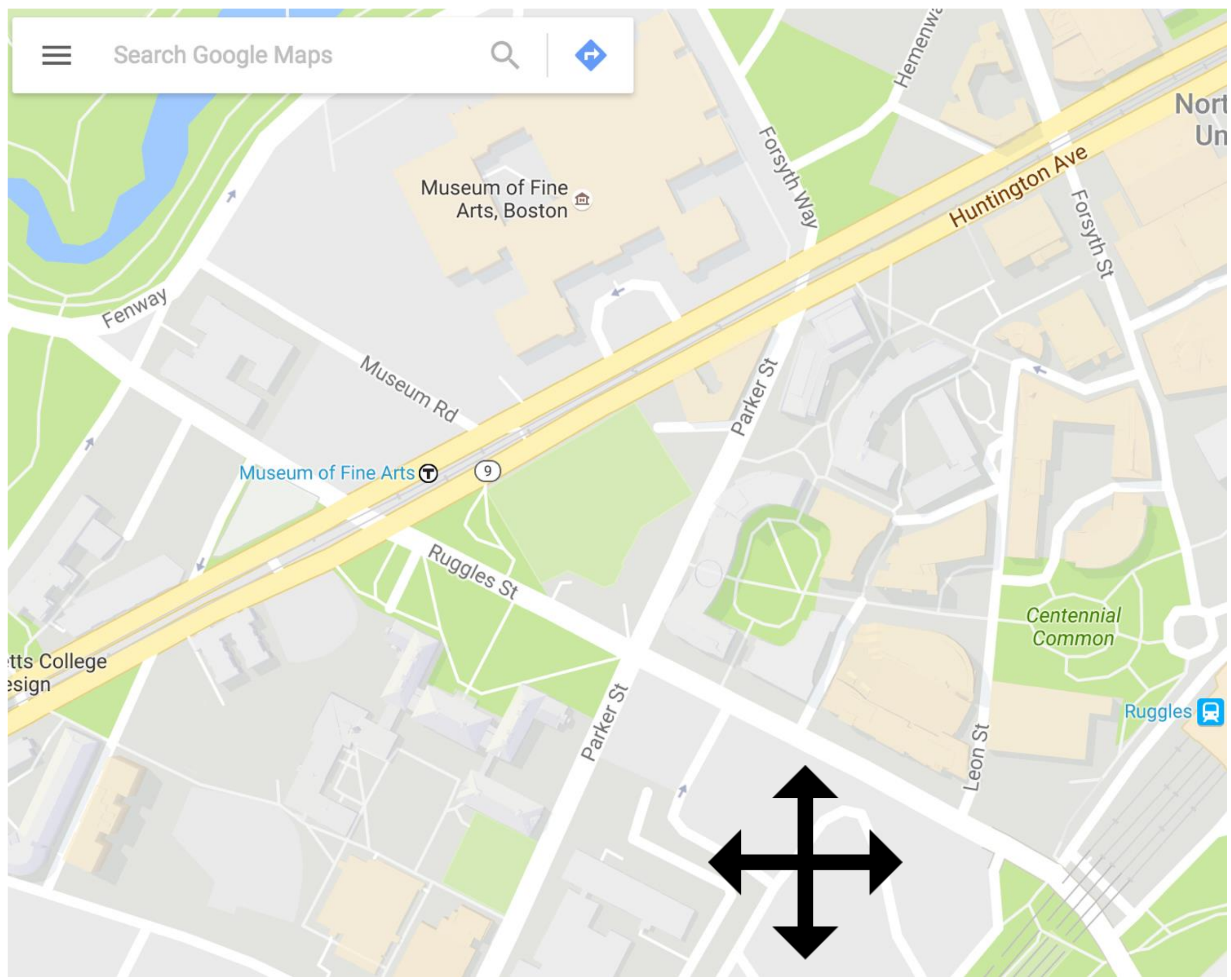
	Target known	Target unknown
Location known	<i>Lookup</i>	<i>Browse</i>
Location unknown	<i>Locate</i>	<i>Explore</i>







What buildings are near Ryder Hall?

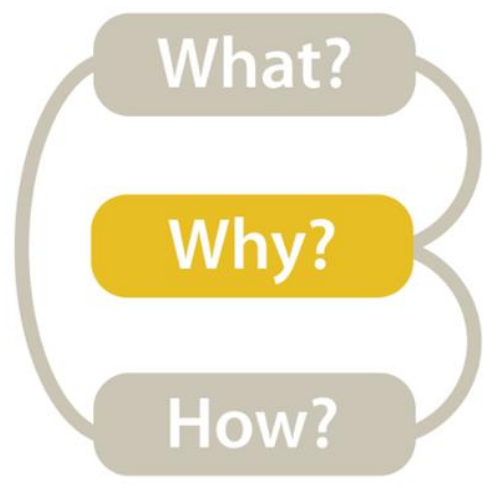
- What?
- Why?
- How?

➔ Search



	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

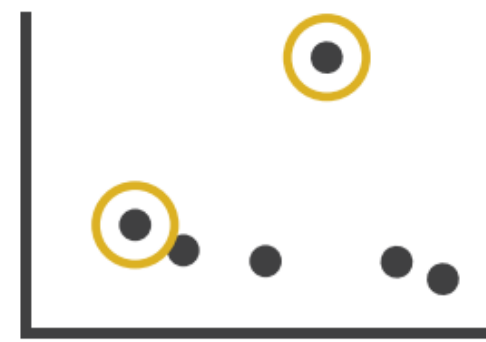
What is south of Huntington Ave?



ACTIONS define user goals. *Low-level*

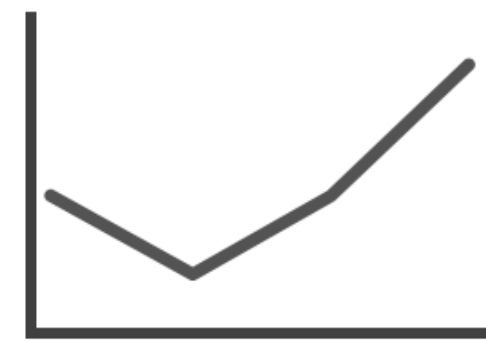
➔ Query

➔ Identify



single target

➔ Compare

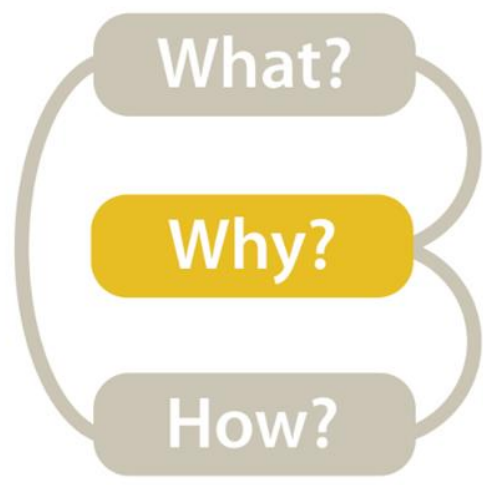


multiple targets

➔ 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*

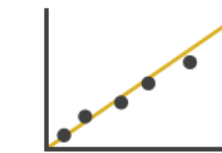


➔ Many

➔ *Dependency*



➔ *Correlation*

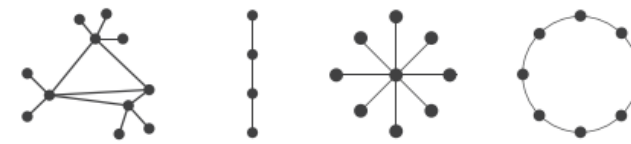


➔ *Similarity*



➔ Network Data

➔ Topology



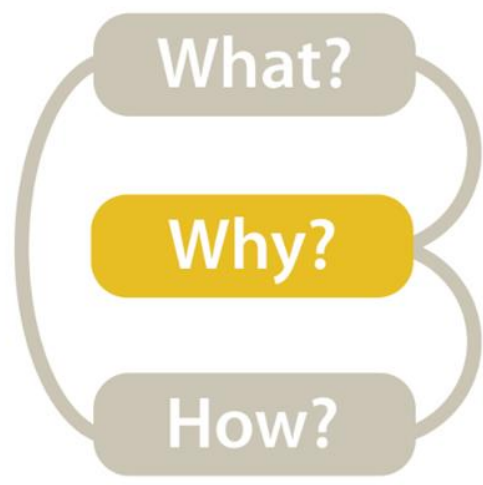
➔ *Paths*



➔ Spatial Data

➔ Shape

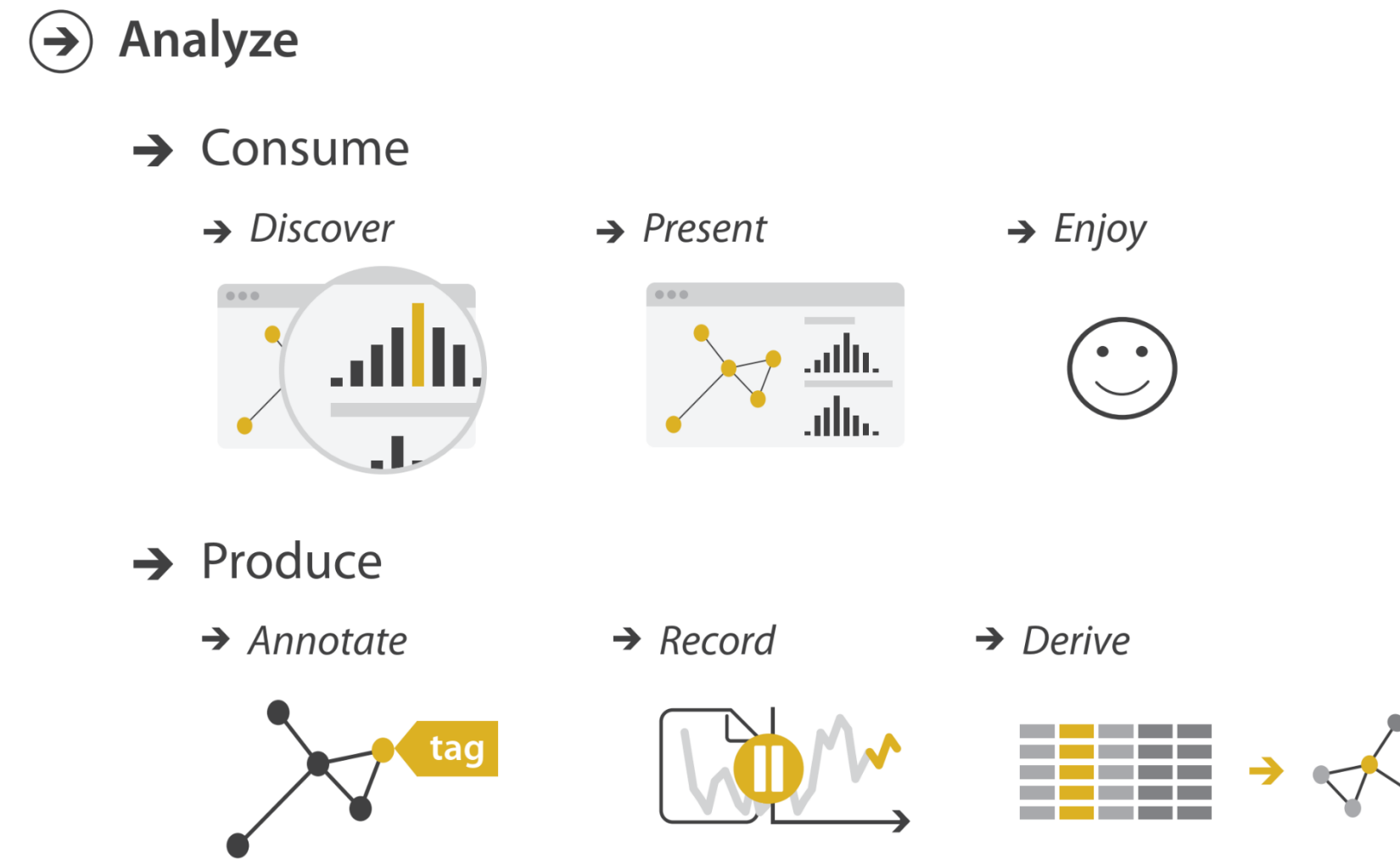




ACTIONS define user goals.

Lots of other task taxonomies...!

High-level

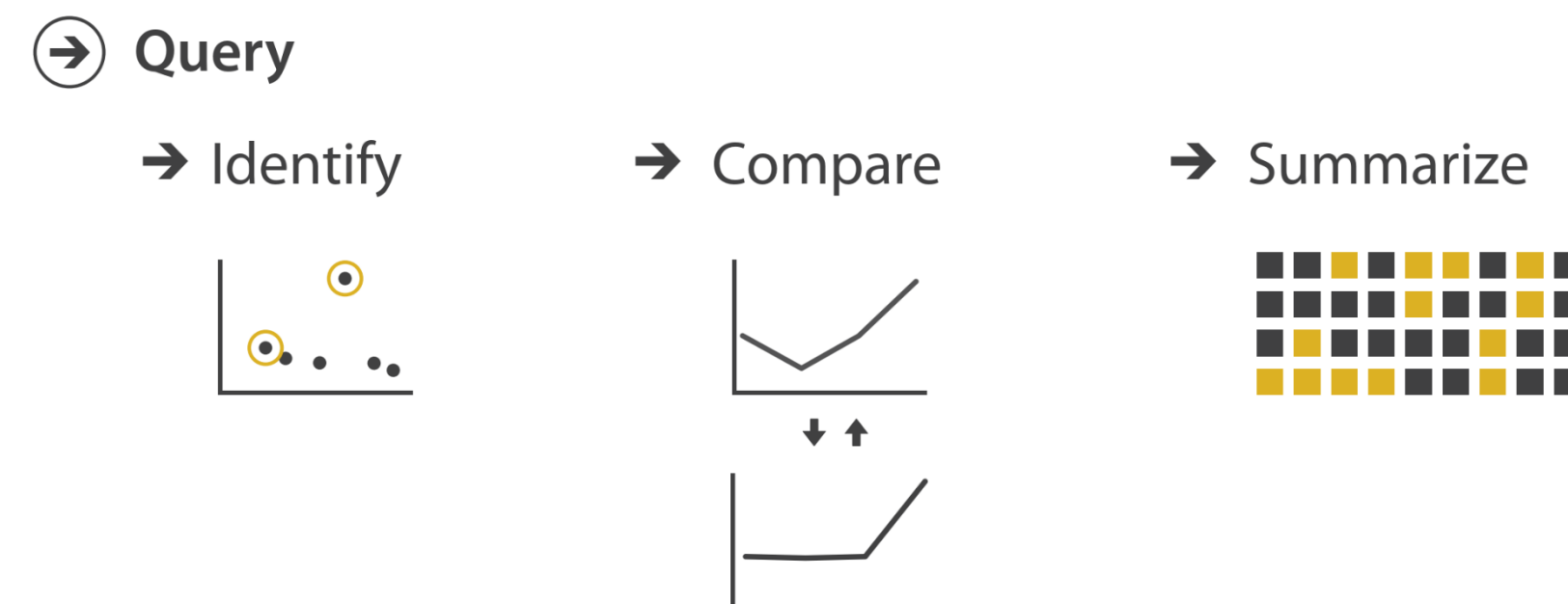


Mid-level

→ Search

	Target known	Target unknown
Location known	<i>Lookup</i>	<i>Browse</i>
Location unknown	<i>Locate</i>	<i>Explore</i>

Low-level



Analytic Task Taxonomy *Low-level*

Retrieve Value *How long is the movie Gone with the Wind?*

Filter *What comedies have won awards?*

Compute Derived Value *How many awards have MGM studio won in total?*

Find Extremum *What director/film has won the most awards?*

Sort *Rank movies by most number of awards.*

Determine Range *What is the range of film lengths?*

Characterize Distribution *What is the age distribution of actors?*

Find Anomalies *Are there exceptions to the relationship between number of awards won and total movies made by an actor?*

Cluster *Is there a cluster of typical film lengths?*

Correlate *Is there a trend of increasing film length over the years?*

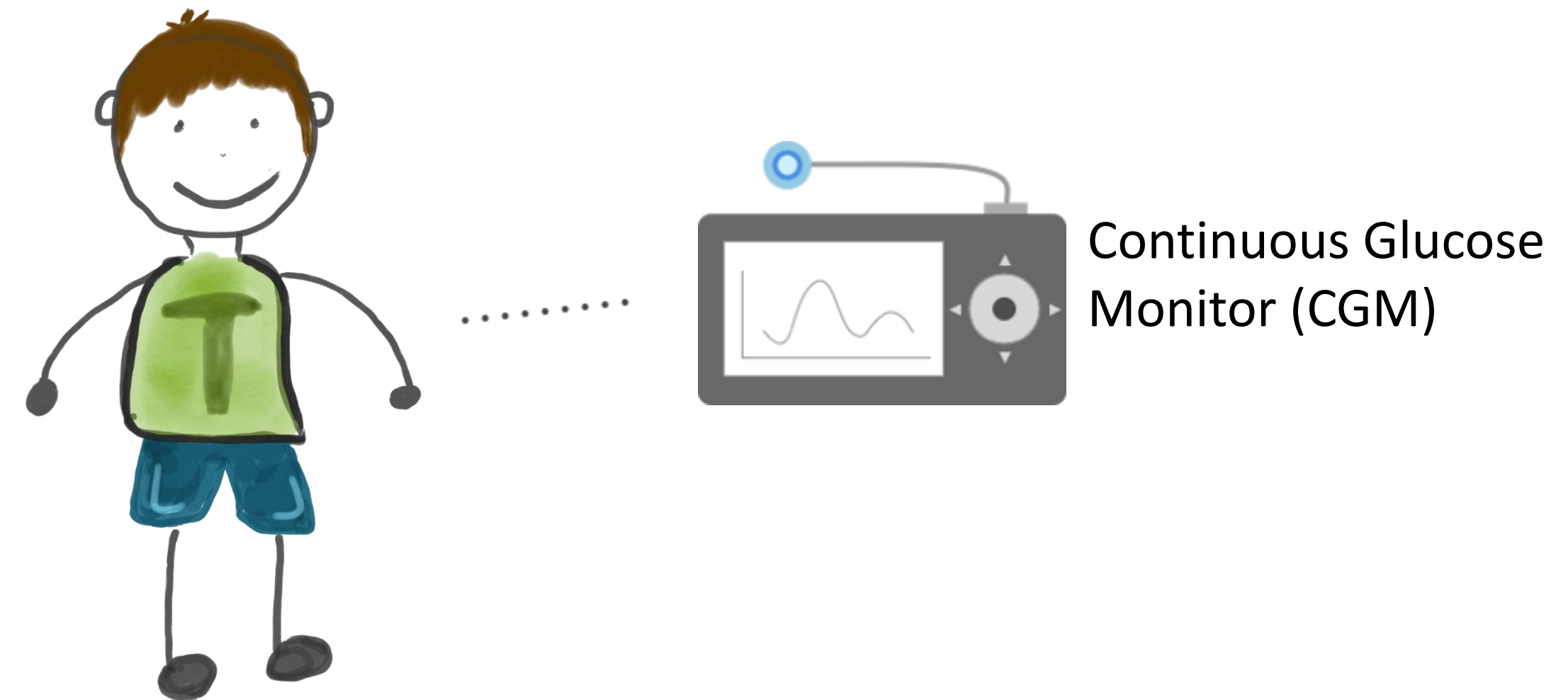
AN EXAMPLE OF TASK ANALYSIS

→ VISUALIZATION DESIGN

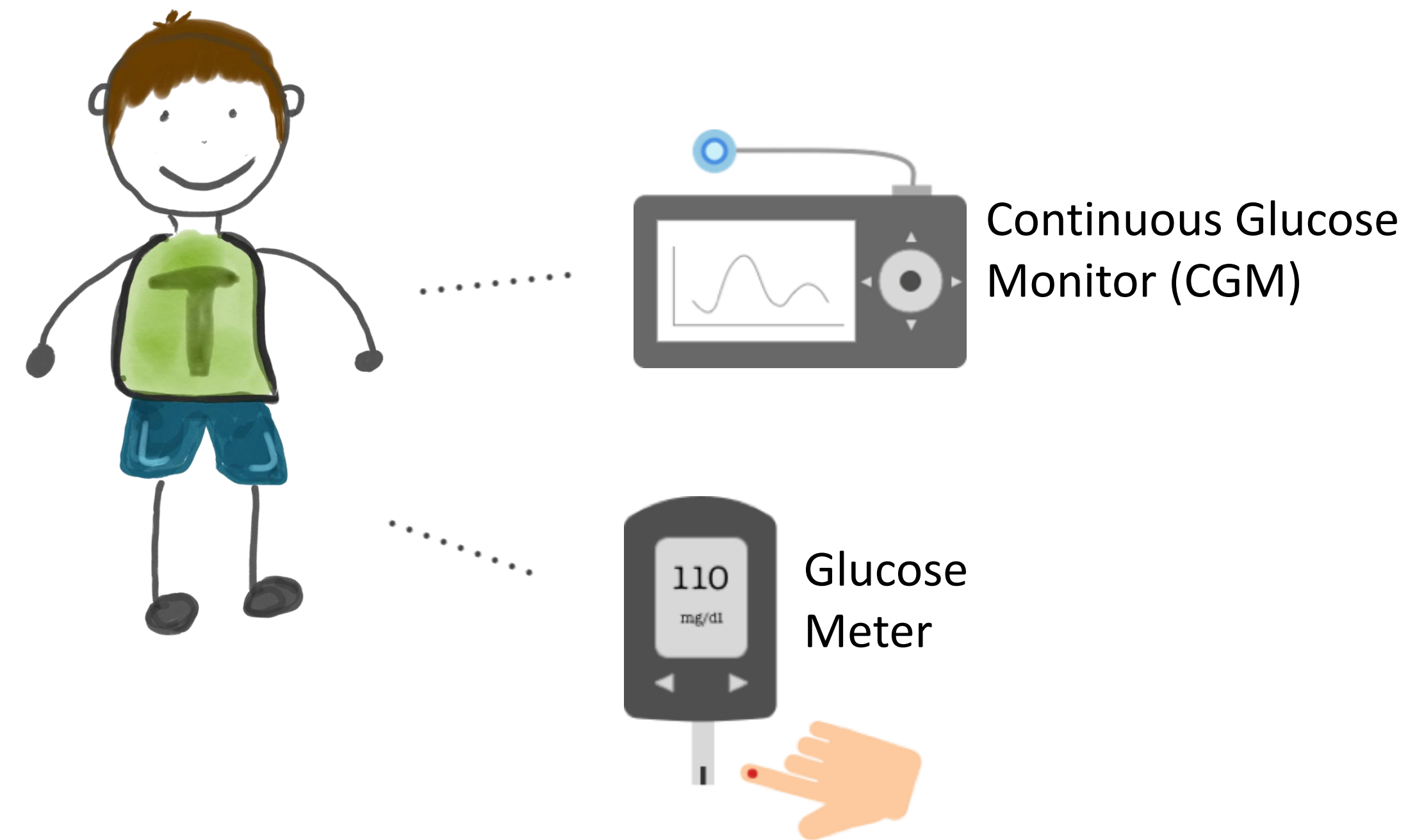
Imagine a 10-year-old kid, who has been diagnosed with type 1 diabetes...



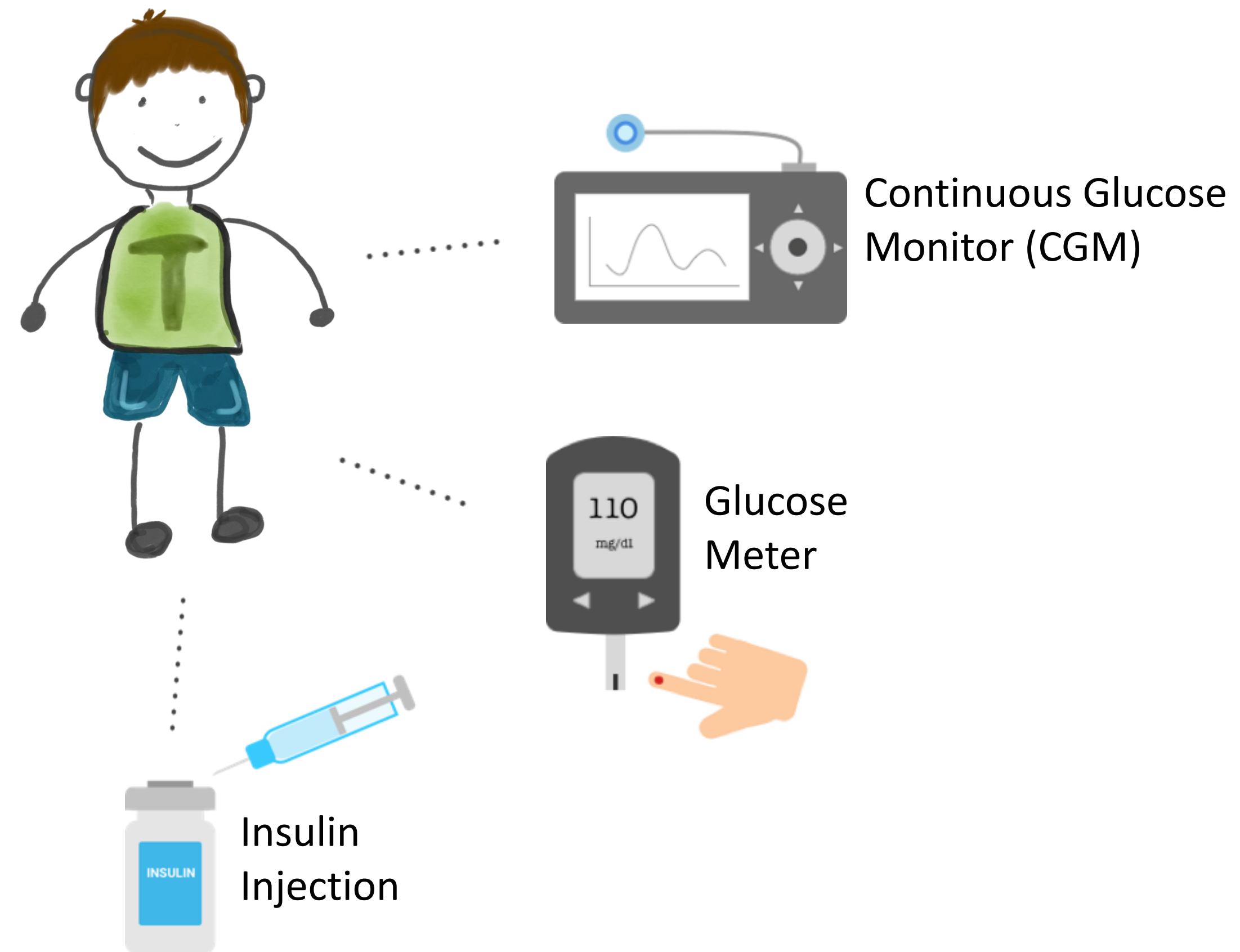
Imagine a 10-year-old kid, who has been diagnosed with type 1 diabetes...



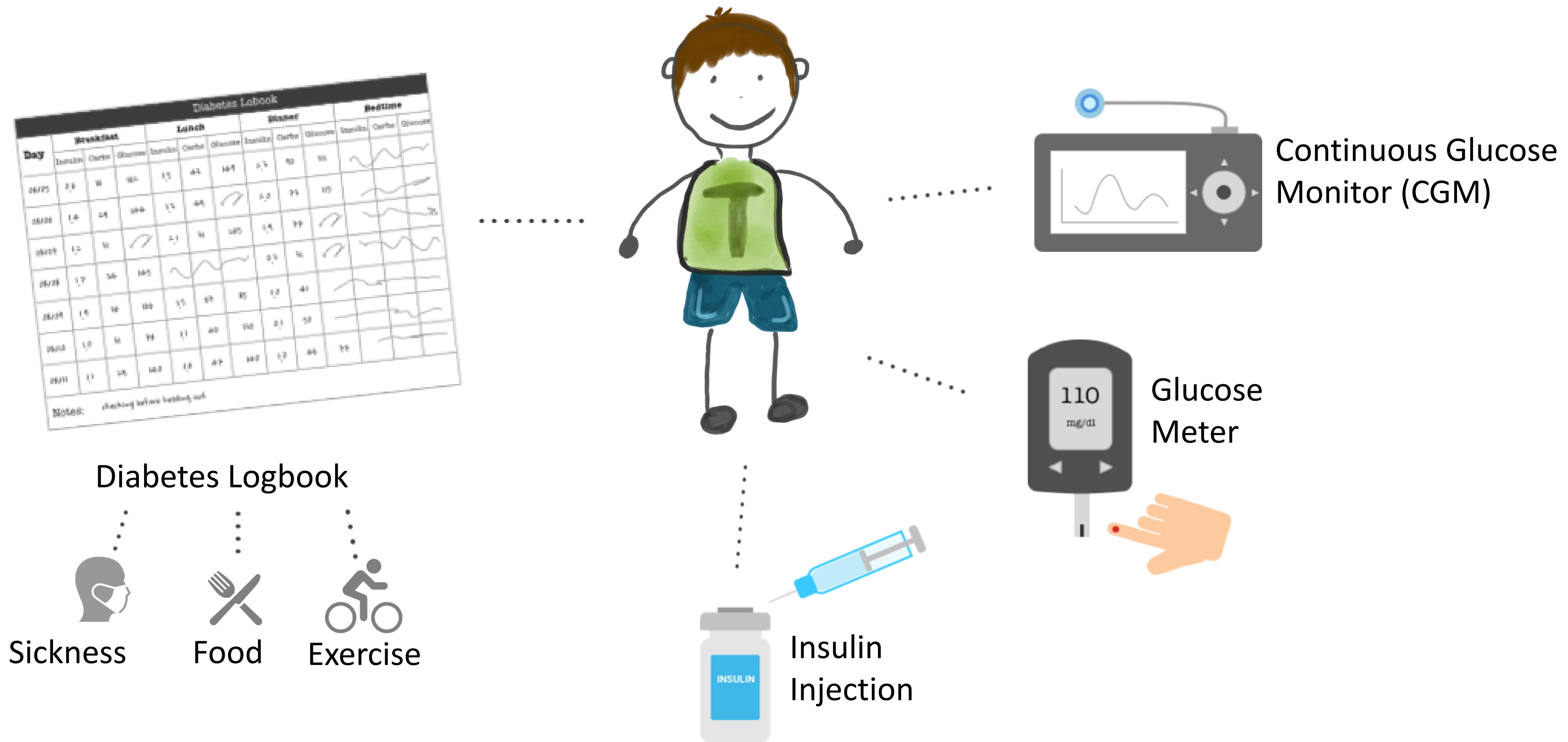
Imagine a 10-year-old kid, who has been diagnosed with type 1 diabetes...



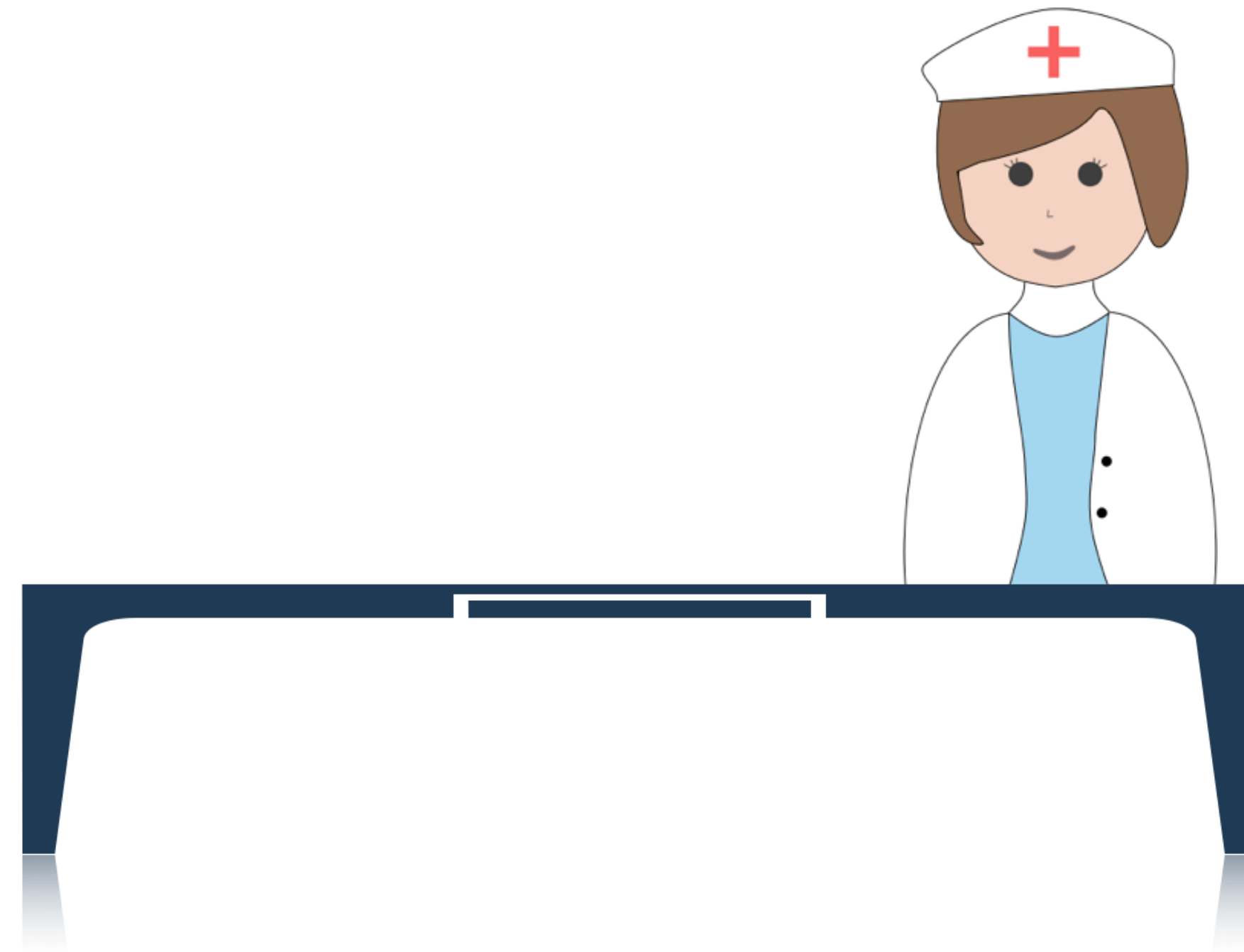
Imagine a 10-year-old kid, who has been diagnosed with type 1 diabetes...



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



During a clinical visit ...



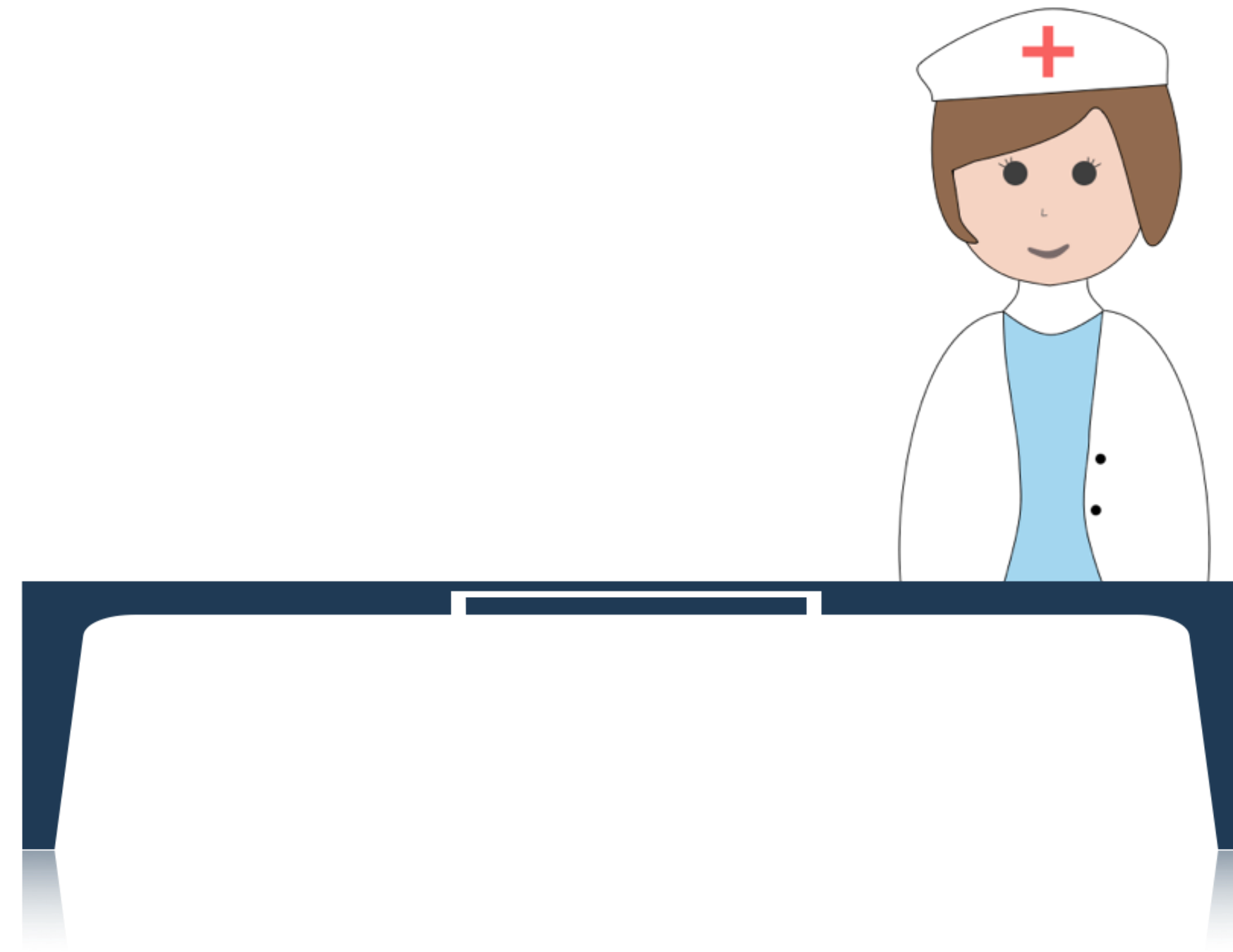
During a clinical visit ...



Diabetes Logbook

Day	Breakfast			Lunch			Dinner			Bedtime		
	Insulin	Carbs	Glucose	Insulin	Carbs	Glucose	Insulin	Carbs	Glucose	Insulin	Carbs	Glucose
08/25	2.0	30	165	1.5	45	180	1.5	30	150	1.0	20	140
08/26	1.4	25	155	1.2	35	165	1.0	25	150	0.8	15	135
08/27	1.5	30	160	1.3	40	170	1.1	28	155	0.9	18	140
08/28	1.7	35	165	1.4	45	175	1.2	32	160	1.0	22	145
08/29	1.6	32	162	1.3	42	172	1.1	30	158	0.9	20	142
08/30	1.8	38	170	1.5	50	180	1.3	38	168	1.1	25	150
08/31	1.9	40	175	1.6	52	185	1.4	40	170	1.2	28	155

Note: checking before testing out.



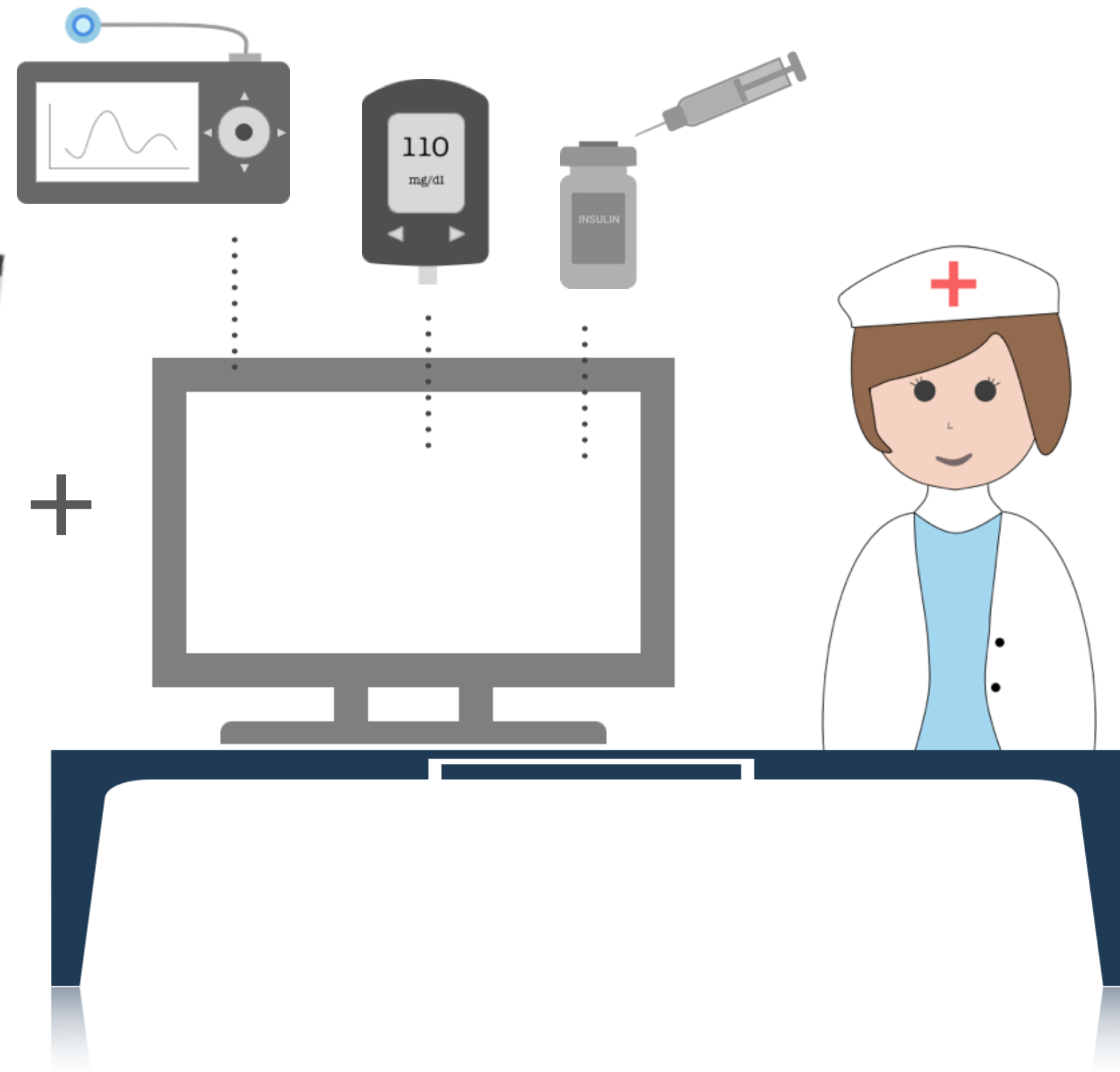
During a clinical visit ...



Diabetes Logbook

Day	Breakfast				Lunch				Dinner				Medicine
	Insulin	Carbs	Glucose	Carbs	Glucose	Carbs	Glucose	Carbs	Glucose	Carbs	Glucose		
08/25	2.0	30	100	1.5	45	105	1.5	30	100	1.5	30	100	
08/26	1.4	25	100	1.5	45	105	1.5	30	100	1.5	30	100	
08/27	1.5	30	100	2.1	50	105	1.5	30	100	1.5	30	100	
08/28	1.7	30	100			105	1.5	30	100	1.5	30	100	
08/29	1.8	30	100	1.5	45	105	1.5	30	100	1.5	30	100	
08/30	1.8	30	100	1.1	40	100	1.5	30	100	1.5	30	100	
08/31	1.1	25	100	1.8	45	105	1.5	30	100	1.5	30	100	

Note: checking before testing out.



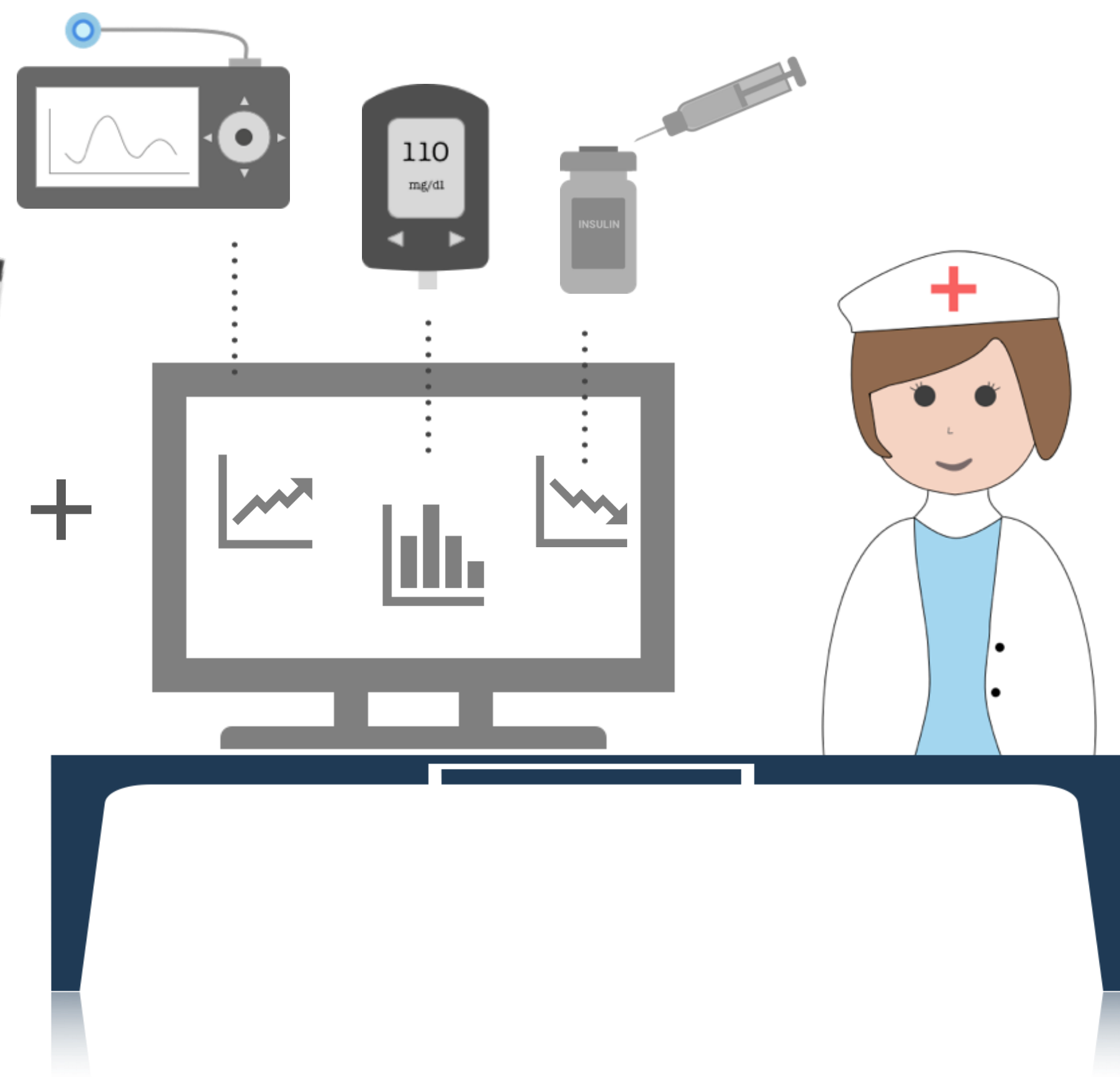
During a clinical visit ...



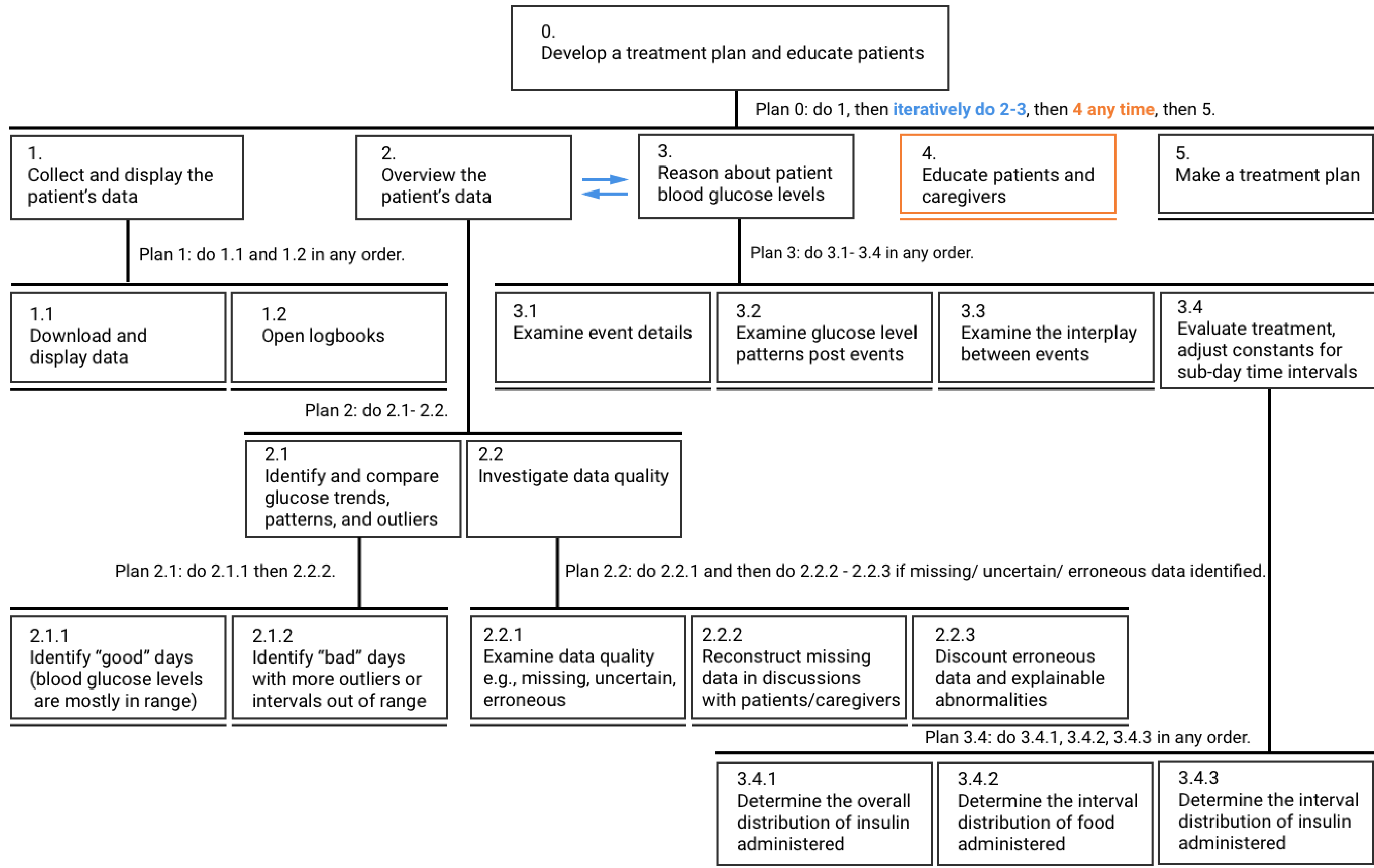
Diabetes Logbook

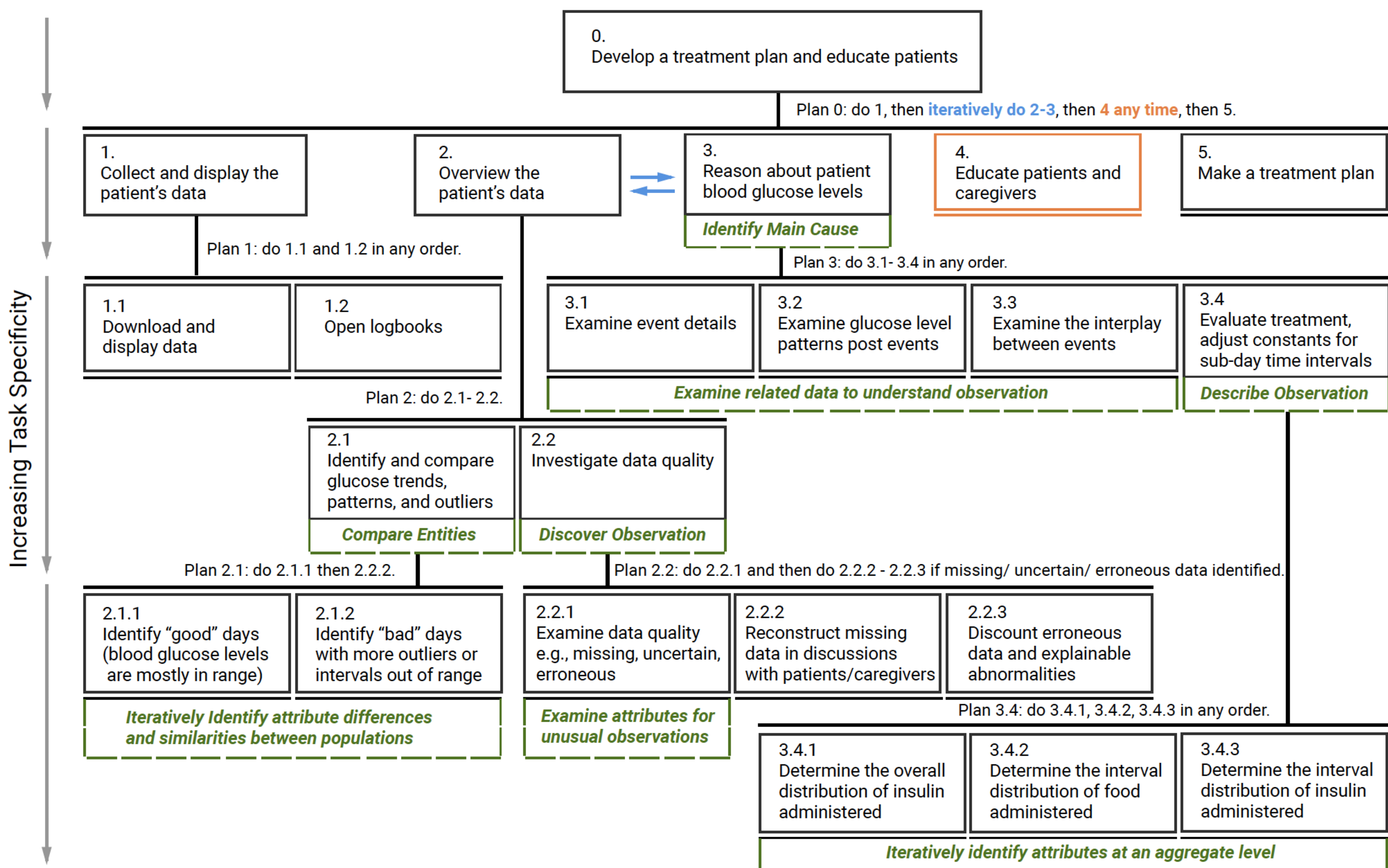
Day	Breakfast			Lunch			Dinner			Bedtime		
	Insulin	Carbs	Glucose	Insulin	Carbs	Glucose	Insulin	Carbs	Glucose	Insulin	Carbs	Glucose
08/25	2.0	30	160	1.5	45	145	1.5	30	150	1.0	20	120
08/26	1.4	25	155	1.2	40	140	1.2	25	135	0.8	15	115
08/27	1.5	30	160	1.3	45	145	1.3	30	140	1.0	20	125
08/28	1.7	35	165	1.4	50	150	1.4	35	145	1.1	25	130
08/29	1.6	30	160	1.3	45	145	1.3	30	140	1.0	20	125
08/30	1.5	30	155	1.2	40	140	1.2	30	135	0.9	15	120
08/31	1.3	25	150	1.1	35	135	1.1	25	130	0.8	15	115

Note: checking before heading out.



Increasing Task Specificity





Design Requirements

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

14-Day Overview



Detail View

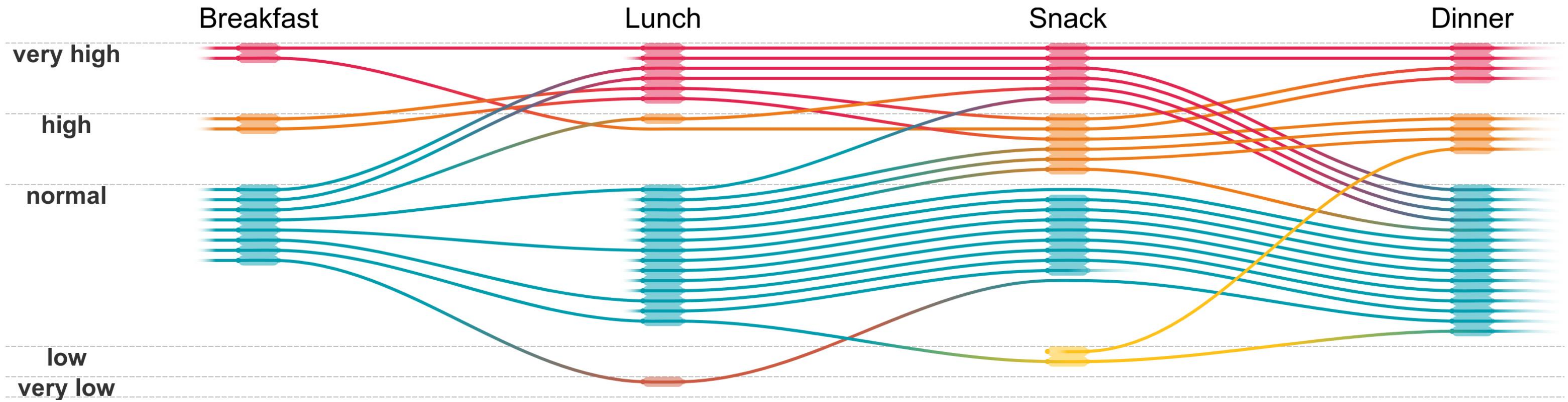
Summary Statistics Panel

Hierarchical Task Analysis

Task Abstraction

Design

Alternate



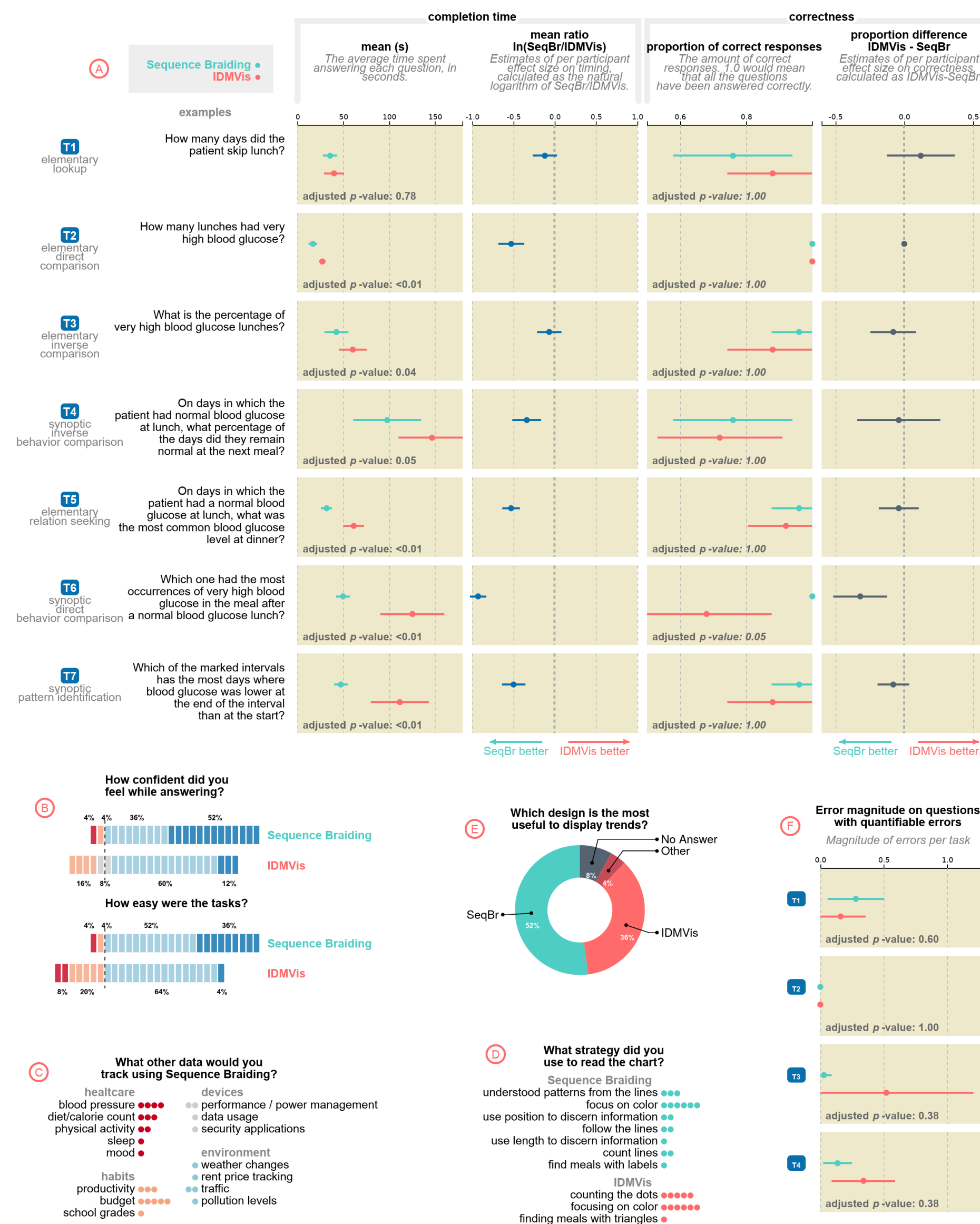


Fig. 9: Results of our evaluation comparing SEQUENCE BRAIDING vs. IDMVis [63]. **A** Completion time and correctness per task. Each row corresponds to the task at left, which is classified based on Andrienko & Andrienko [3]. The specific question instantiating that task for the study is in the second 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.

A

Sequence Braiding •
IDMVis •

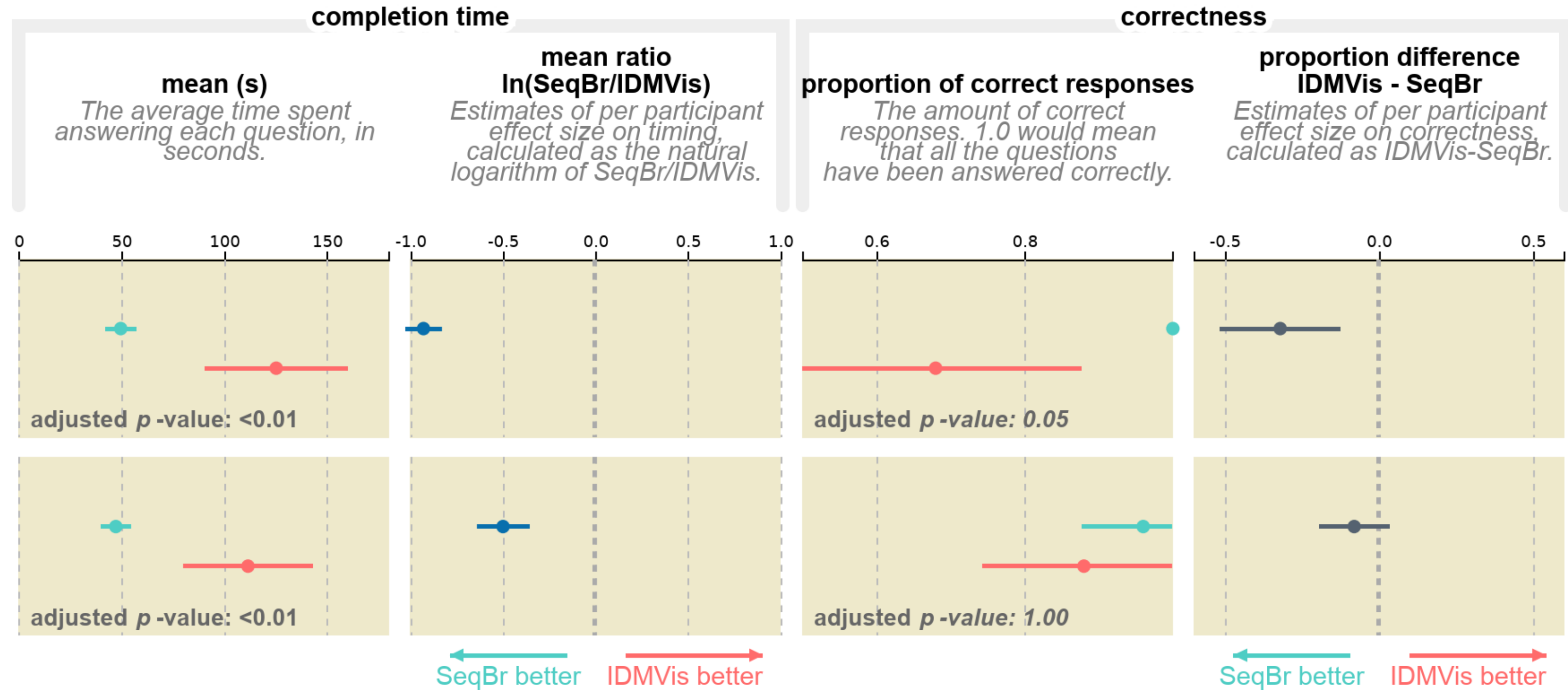
examples

T6
synoptic
direct
behavior comparison

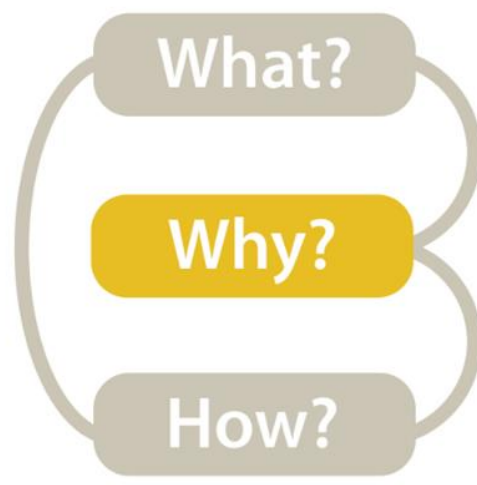
Which one had the most occurrences of very high blood glucose in the meal after a normal blood glucose lunch?

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?

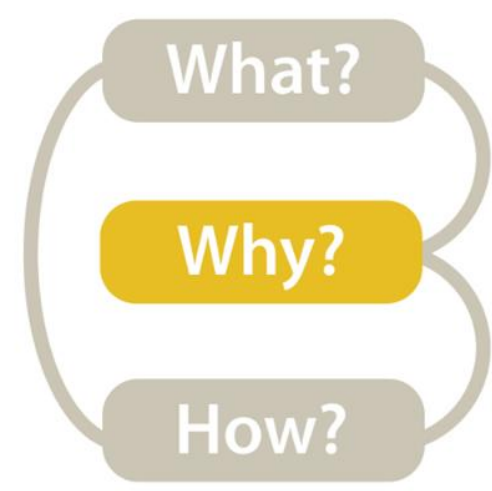


IN-CLASS EXERCISE:
MOCK INTERVIEW, TASK ANALYSIS



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.



Task Analysis

Visualization for Public Transit Development

15m

INSTRUCTIONS:

- Break-out into groups of ~3 people in Teams.
- Pretend you are transportation engineers, e.g., for the MBTA, City of Boston.
- Discuss the user tasks and goals and abstract them using one of these taskonomies.
- Save your notes for a later exercise!!!

Retrieve Value *How long is the movie Gone with the Wind?*

Filter *What comedies have won awards?*

Compute Derived Value *How many awards have MGM studio won in total?*

Find Extremum *What director/film has won the most awards?*

Sort *Rank movies by most number of awards.*

Determine Range *What is the range of film lengths?*

Characterize Distribution *What is the age distribution of actors?*

Find Anomalies *Are there exceptions to the relationship between number of awards won and total movies made by an actor?*

Cluster *Is there a cluster of typical film lengths?*

Correlate *Is there a trend of increasing film length over the years?*

Low-level

Analyze

High-level

→ Consume

→ Discover



→ Present

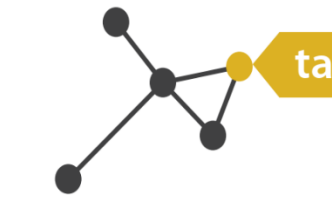


→ Enjoy

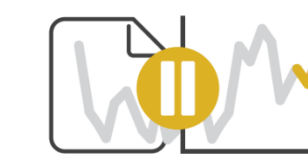


→ Produce

→ Annotate



→ Record



→ Derive



Search

Mid-level

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

Query

Low-level

→ Identify



→ Compare

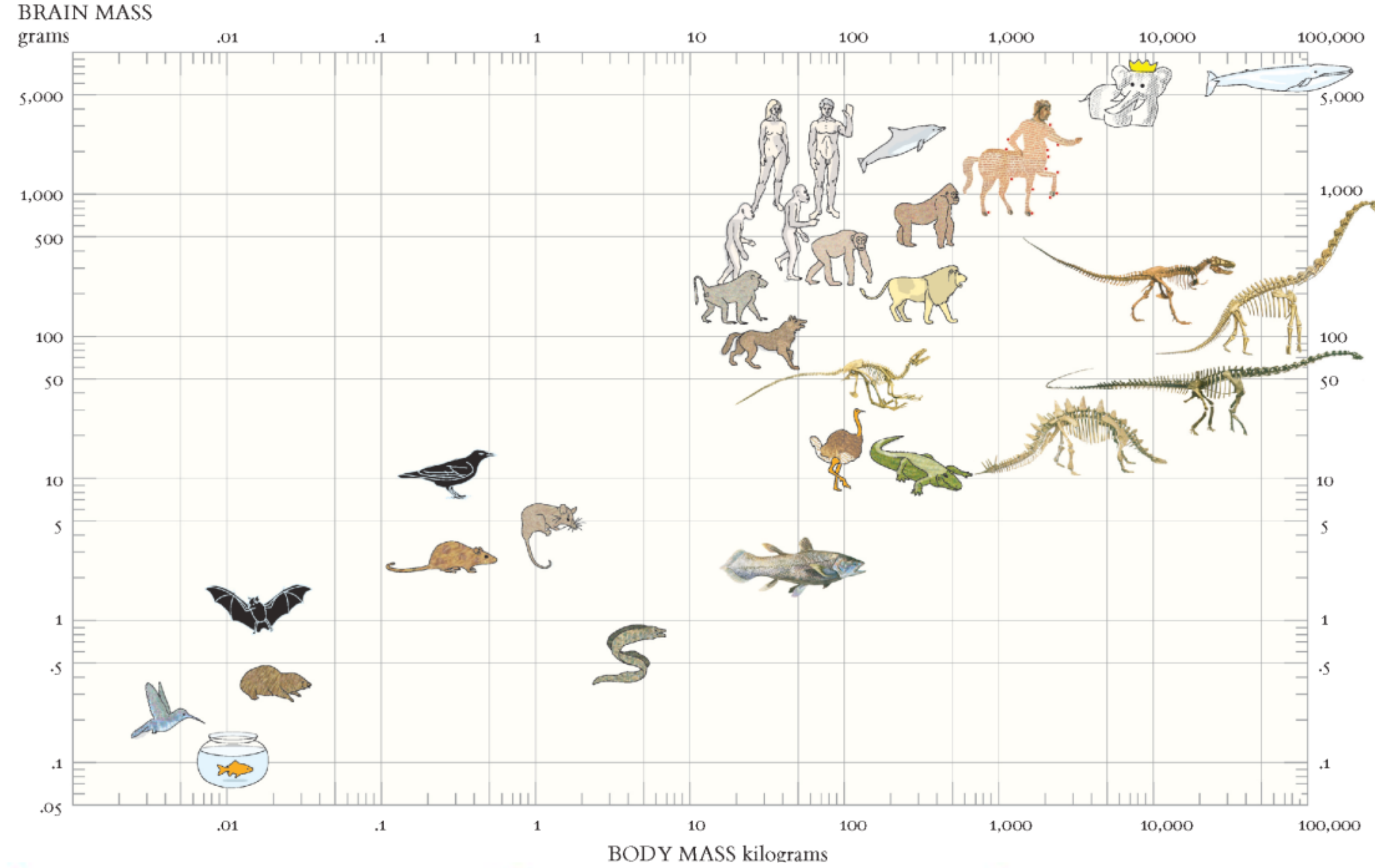


→ Summarize

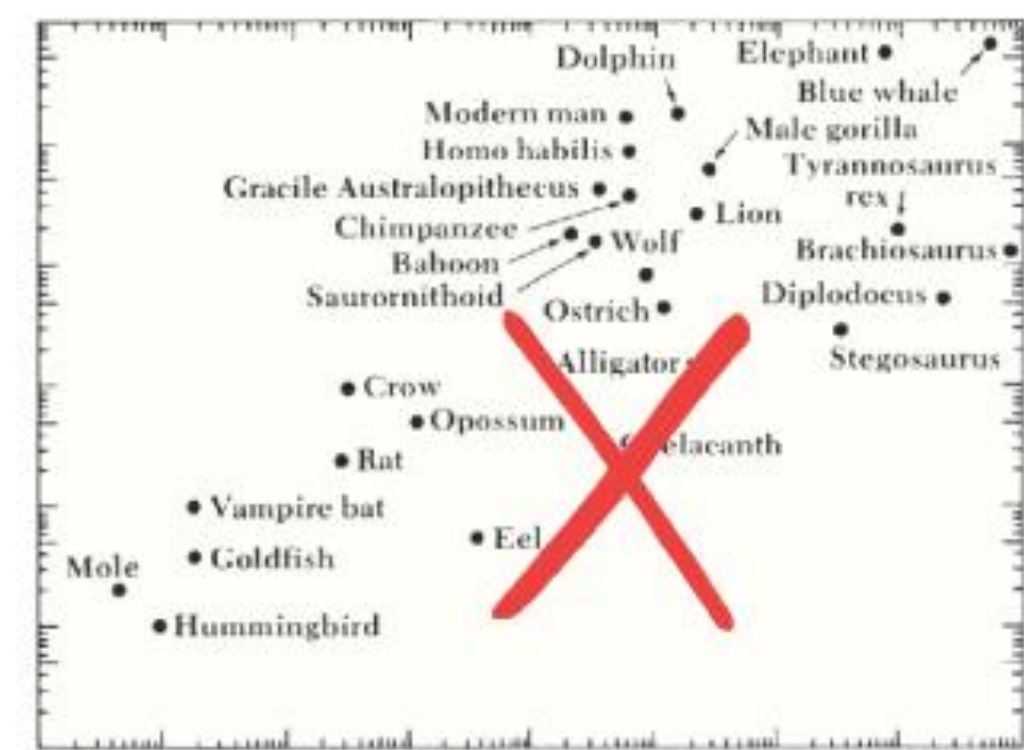


DESIGN RULES OF THUMB — CONTINUED

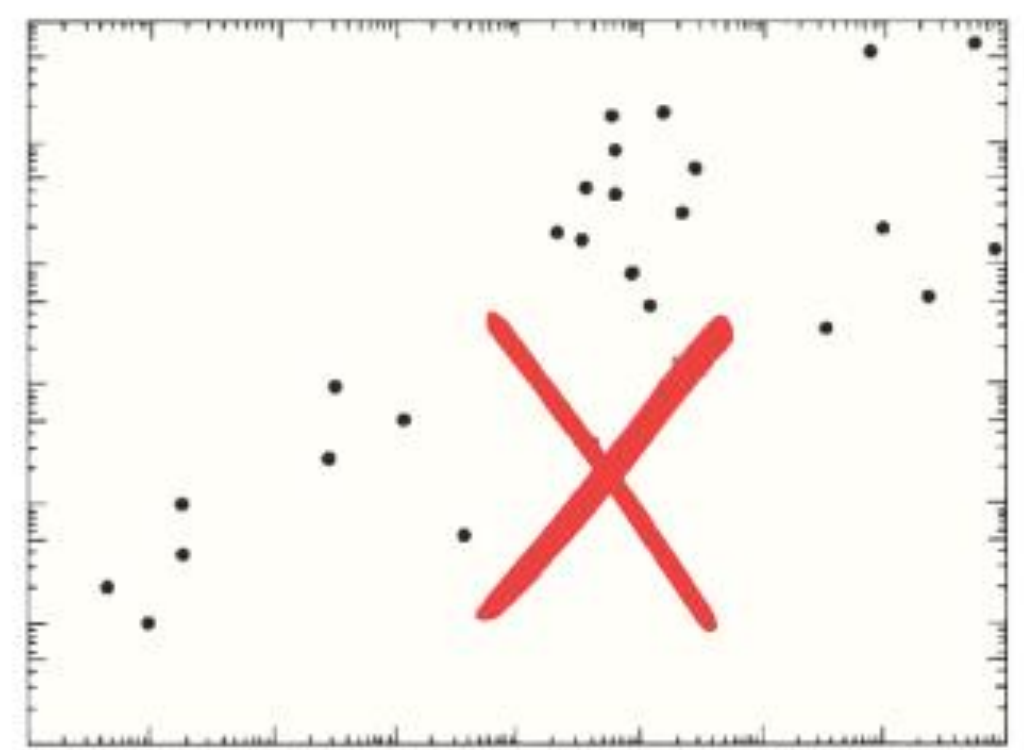
Not all “visual embellishments” are “chart junk”!



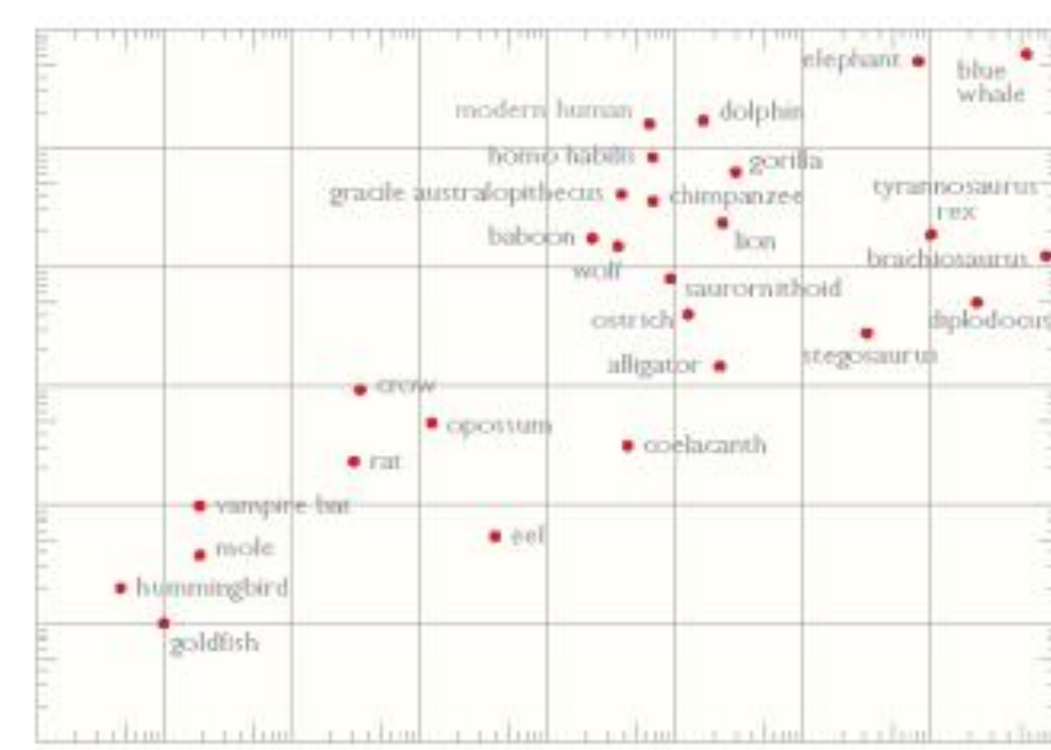
1



2



3



“Chart Junk”

Chart junk can... persuade, help with memorability, engage
... bias, limit data-ink ratio, clutter, lower trust

Take-away: *it depends on your audience, task, and context...*

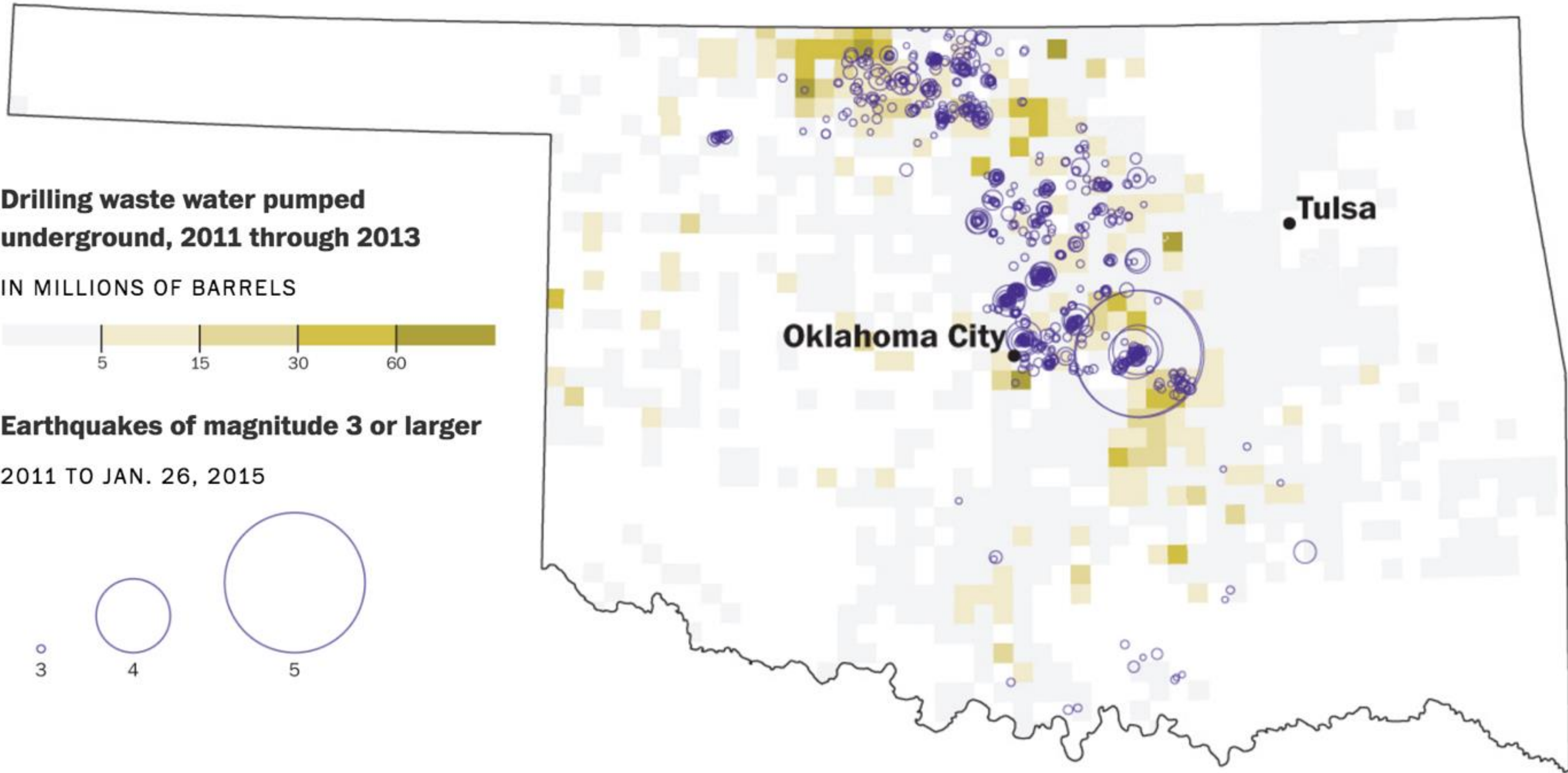
Hall of Fame or Hall of
Shame

Damaging quakes in Oklahoma

A lawsuit claims that Oklahoma's great increase in earthquake activity has been caused by pumping waste from drilling operations back underground. The suit involves the largest measured quake in the history of the state, a 5.6 tremor that happened in Prague, east of Oklahoma City in November 2011. The pace of quakes with magnitude 3 or higher has increased since then, with 567 in 2014, and 52 in less than four weeks this month. [Read related article.](#)

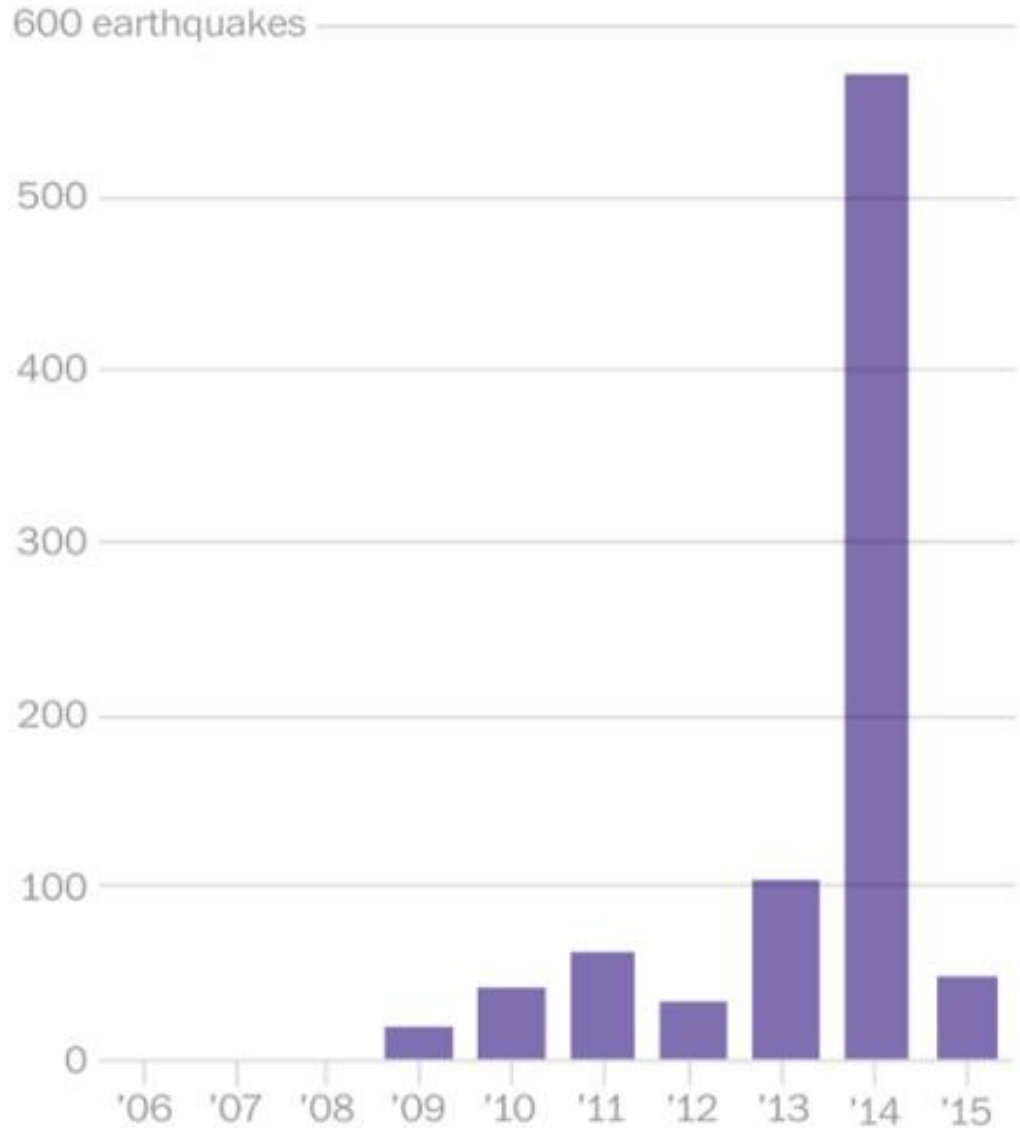
By Dan Keating and Darla Cameron

Published: Jan. 28, 2015

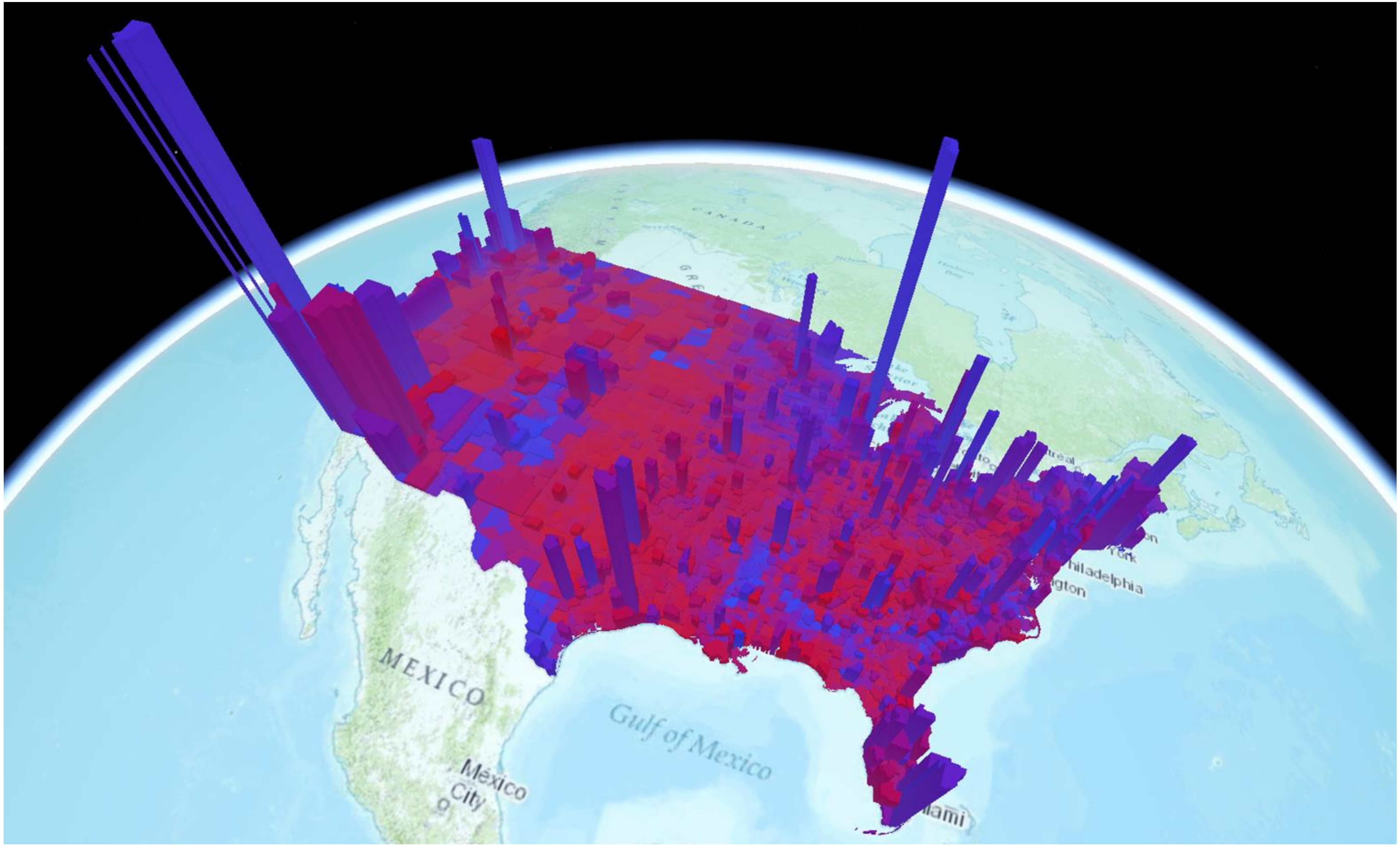


2014 was a record year

Oklahoma saw a record number of earthquakes with a magnitude of 3 or larger in 2014.

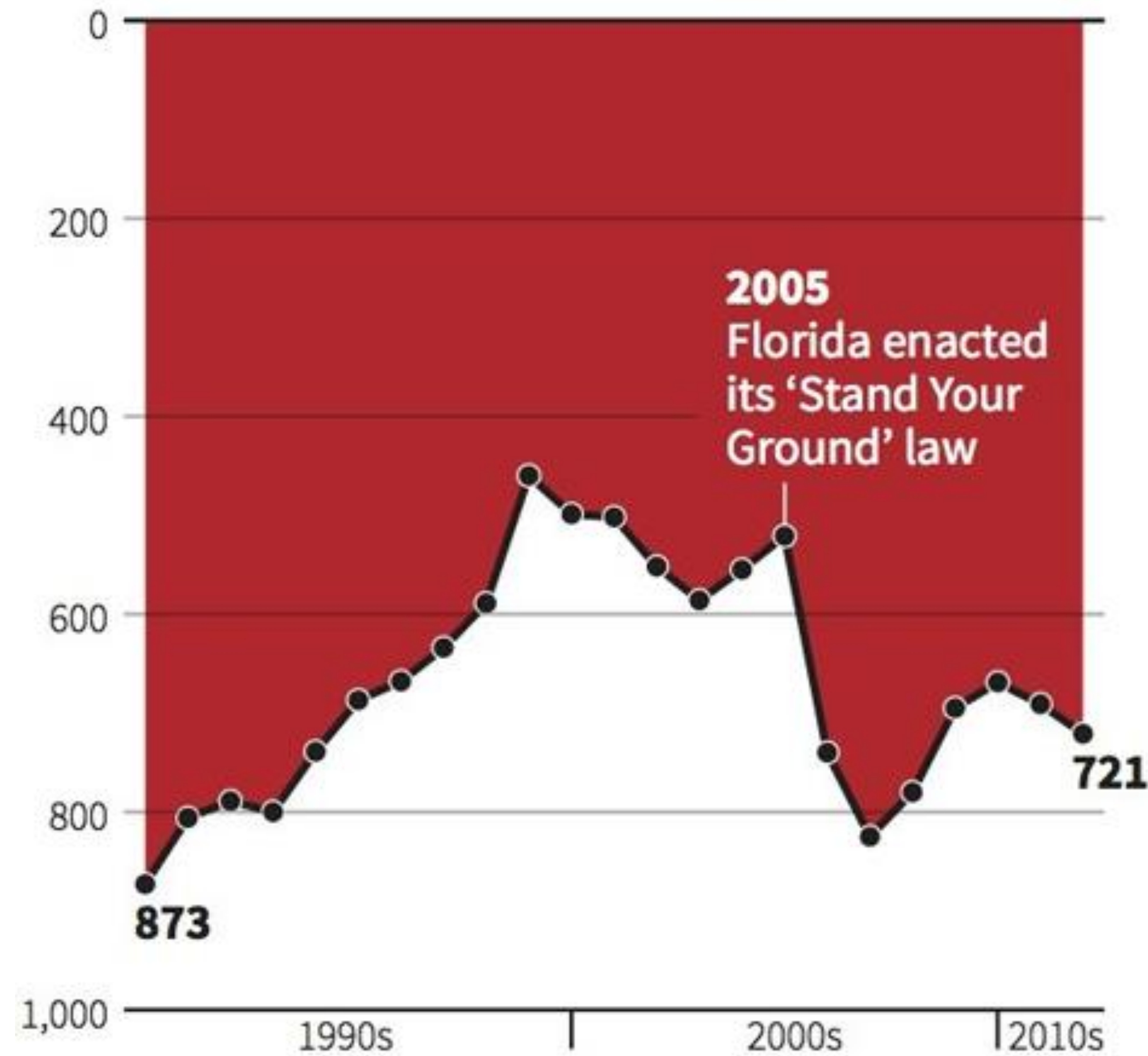


Note: 2015 data is through Jan. 26.

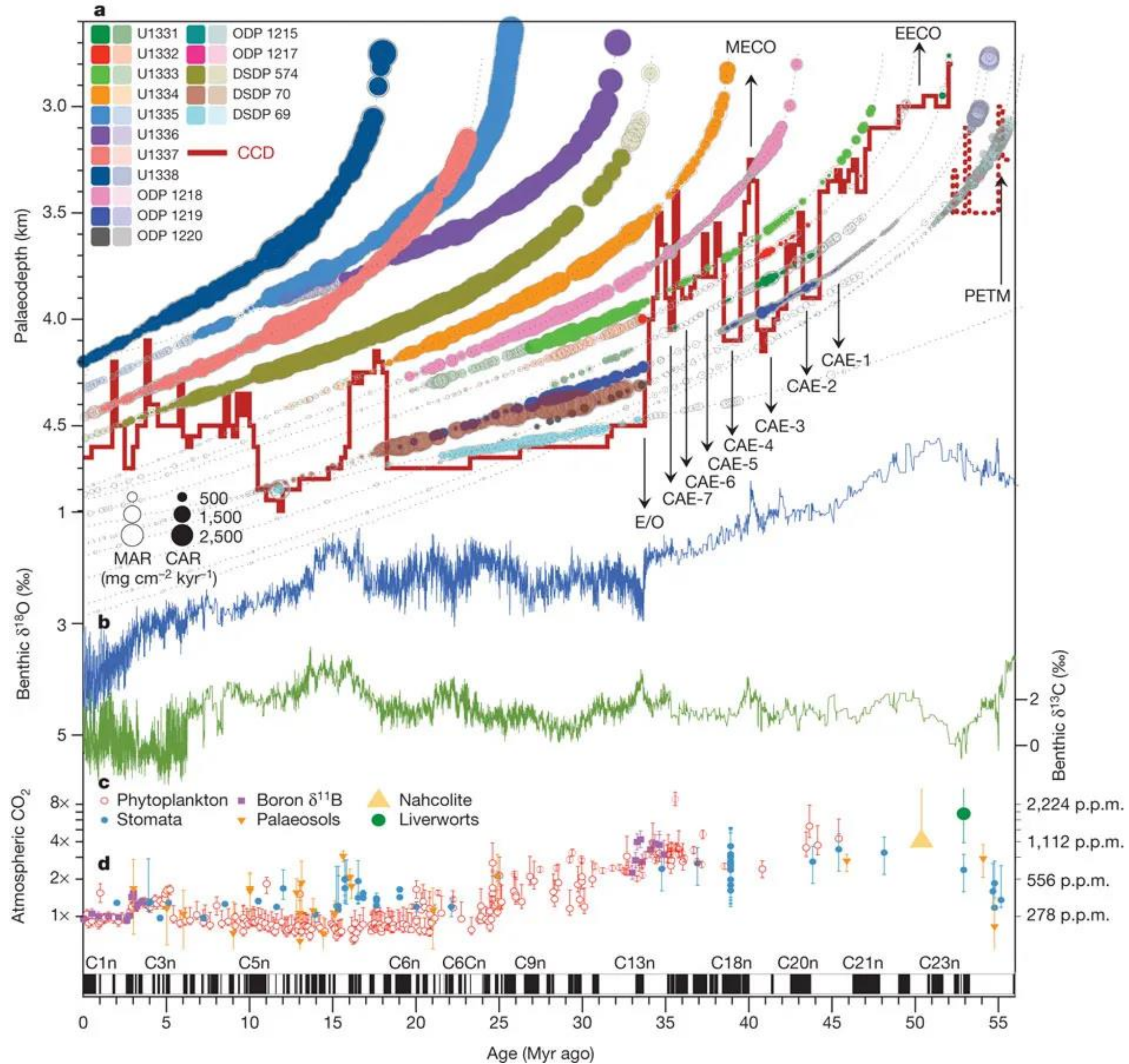


Gun deaths in Florida

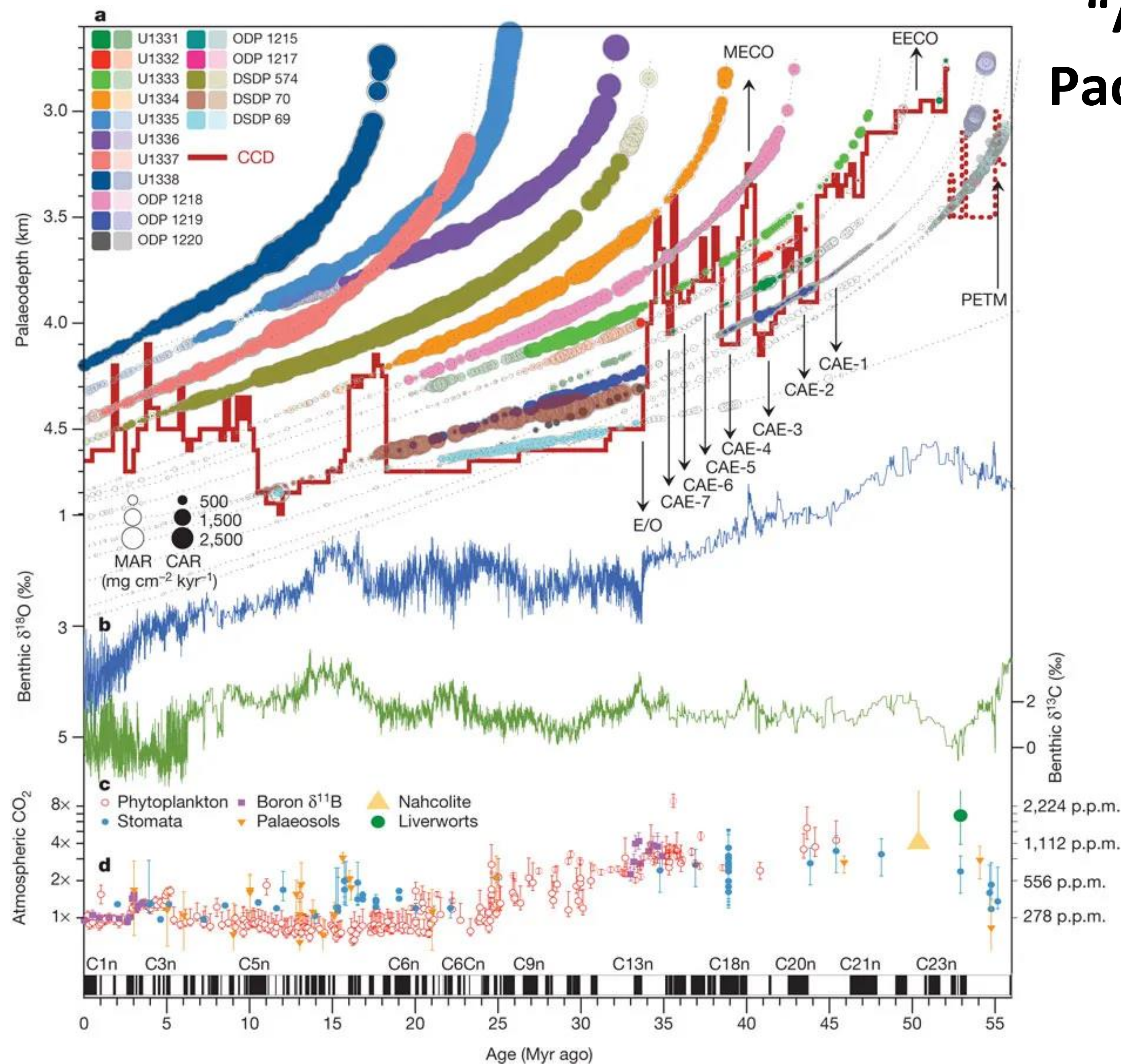
Number of murders committed using firearms



Source: Florida Department of Law Enforcement

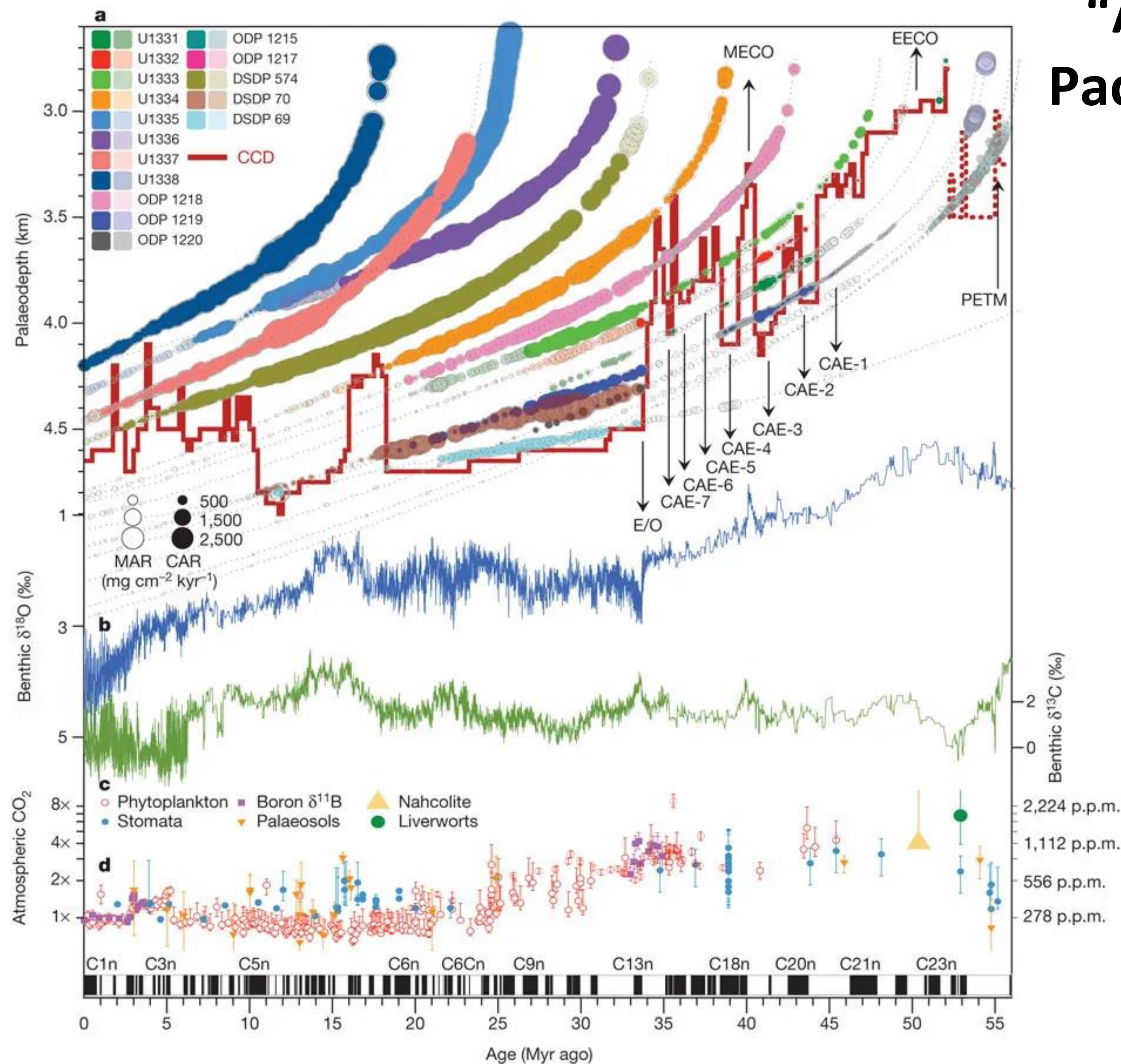


“A Cenozoic record of the equatorial Pacific carbonate compensation depth”



a, Equatorial Pacific accumulation rate history as a function of geological age at the backtracked and unloaded palaeo-water depth, and using a palaeomagnetic polarity age scheme² plotted as black and white bars below the figure. Circle area is scaled by accumulation rate: carbonate accumulation rate (CAR; filled circles), total mass accumulation rate (MAR; open circles). Data are plotted with a lighter colour outside a $\pm 3.5^\circ$ band around the palaeo-equator. The position of the equatorial Pacific CCD is indicated by a solid red line (dashed red line marks reconstruction from off-equatorial sites). See text for abbreviations. **b**, **c**, Benthic oxygen (**b**; blue curve, left-hand vertical axis) and carbon (**c**; green curve, right-hand vertical axis) isotope values from a global compilation⁹, reported relative to the VPDB (Vienna Pee Dee Belemnite) standard. **d**, Atmospheric CO₂ compilation and error bars from refs 16, 40; left-hand vertical axis, log CO₂ scale relative to pre-industrial CO₂ (1x = 278 p.p.m.v.); right-hand vertical axis, log CO₂ scale in absolute values. Error bars are as in ref. 16; for example, for boron $\delta^{11}\text{B}$ error bars reflect long-term analytical reproducibility or internal precision, whichever is larger (at 95% confidence). PETM, Palaeocene-Eocene Thermal Maximum.

“A Cenozoic record of the equatorial Pacific carbonate compensation depth”



a, Equatorial Pacific accumulation rate history as a function of geological age at the backtracked and unloaded palaeo-water depth, and using a palaeomagnetic polarity age scheme² plotted as black and white bars below the figure. Circle area is scaled by accumulation rate: carbonate accumulation rate (CAR; filled circles), total mass accumulation rate (MAR; open circles). Data are plotted with a lighter colour outside a $\pm 3.5^\circ$ band around the palaeo-equator. The position of the equatorial Pacific CCD is indicated by a solid red line (dashed red line marks reconstruction from off-equatorial sites). See text for abbreviations. **b**, **c**, Benthic oxygen (**b**; blue curve, left-hand vertical axis) and carbon (**c**; green curve, right-hand vertical axis) isotope values from a global compilation⁹, reported relative to the VPDB (Vienna Pee Dee Belemnite) standard. **d**, Atmospheric CO₂ compilation and error bars from refs 16, 40; left-hand vertical axis, log CO₂ scale relative to pre-industrial CO₂ (1x = 278 p.p.m.v.); right-hand vertical axis, log CO₂ scale in absolute values. Error bars are as in ref. 16; for example, for boron $\delta^{11}\text{B}$ error bars reflect long-term analytical reproducibility or internal precision, whichever is larger (at 95% confidence). PETM, Palaeocene-Eocene Thermal Maximum.

Yellow = explicit encodings

Olive = implicit encodings

Upcoming Assignments & Communication

A look at the upcoming assignments and deadlines

- Textbook, Readings & Reading Quizzes
- *! PAST !*
 - [Assignment 4a — D3 Basic Charts](#)
 - [Assignment 4b — Altair & JupyterLab Setup](#)
 - [Assignment 4c — Register for IEEE VIS 2020](#)
 - [Assignment 5 — Altair Basic Plots](#)
 - [Project 2 — Proposal, Related Work, & Group Charter](#)
- 2020-10-20
 - [Assignment 6 — D3 Event Handling](#)
 - [Project 3 — Interview & Task Analysis](#)
- 2020-10-26 **No Class — Attend IEEE VIS**
- 2020-10-27
 - [Assignment 7a — Critique "Energy Portfolio Analysis"](#)
 - [Assignment 7b — Critique "Color Theory"](#)
 - [Project 4 — Data Collection & Exploration, Sketches](#)
- 2020-10-28 **No Class — Attend IEEE VIS**

<https://c.dunne.dev/ds4200f20>

Everyday Required Supplies:

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

Use **Canvas Discussions** for general questions, email the **instructor & TAs** for questions **specific to you**.