

DESIGN FOR DATA VIZ

KIRELL BENZI, PH.D.



www.kirellbenzi.com

What is design?

Design is a innovative and creative process to solve a problem

It takes into account the aesthetic, functional, economic, and sociopolitical dimensions

It is used in many fields from architecture to user interfaces



What is design?

THIS

Design is different from art

something must have a function or purpose to users

it should be based on justification or evidence



**IS A SPECIAL SHAPED
KNIFE FOR CLEANING OUT
NUTELLA JARS ENTIRELY...**

Why do we care in data viz?



THE COLLECTION

The Uncomfortable is a collection of deliberately inconvenient everyday objects by Athens-based architect Katerina Kamprani

Huge design space

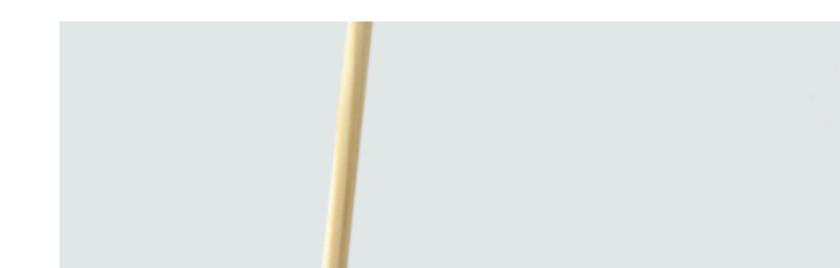
[ALL](#)

3D VISUALIZATION

PROTOTYPES

Many ineffective combinations

Many different datasets with unique user requirements

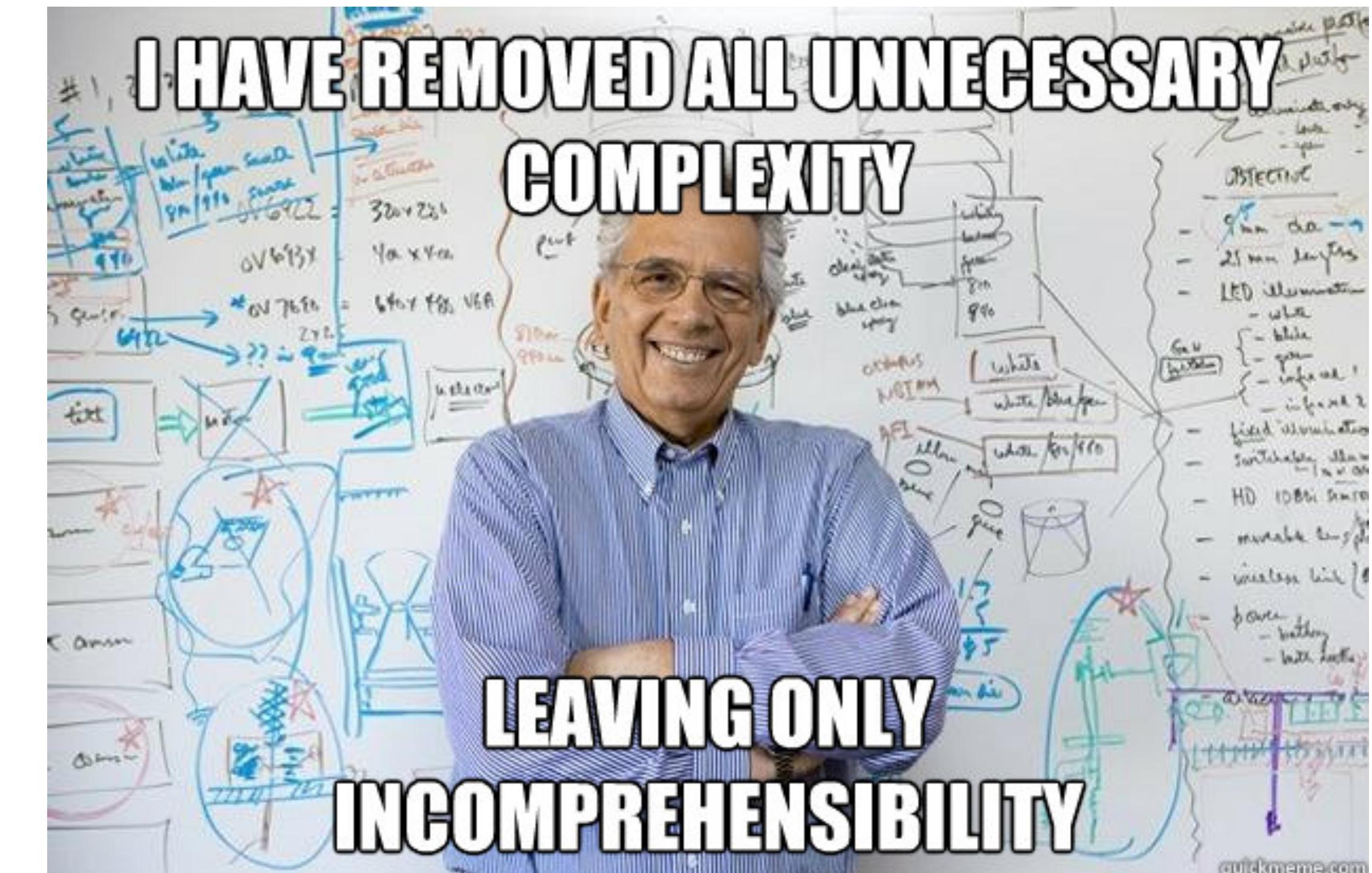


When do we design?

No clear problem definition

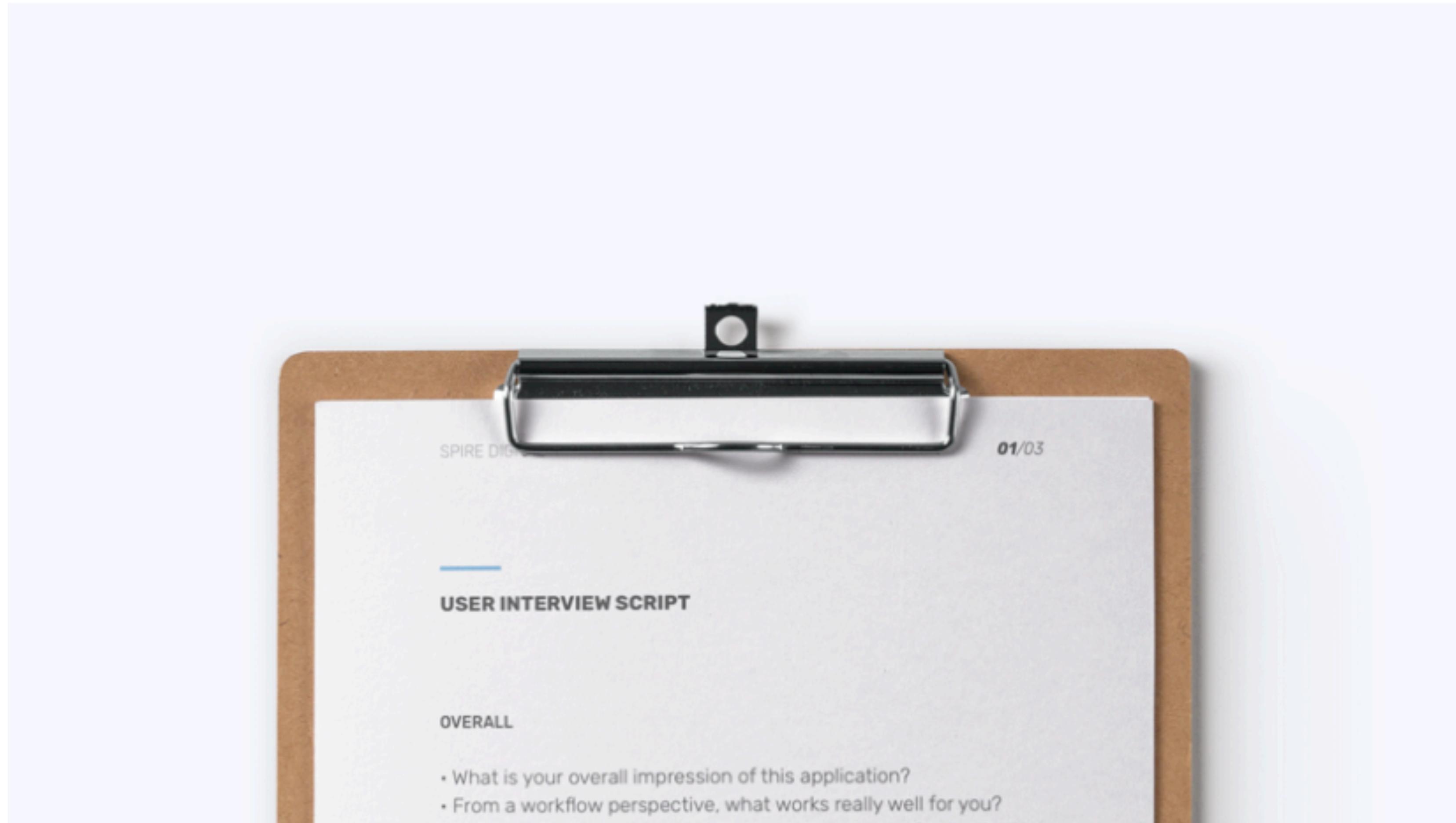
No obvious best solution, trade-off with good and bad points

No clear point to stop designing (except time and money)



Questions UX designers should be asking

Designing is easier when you start with the answers.



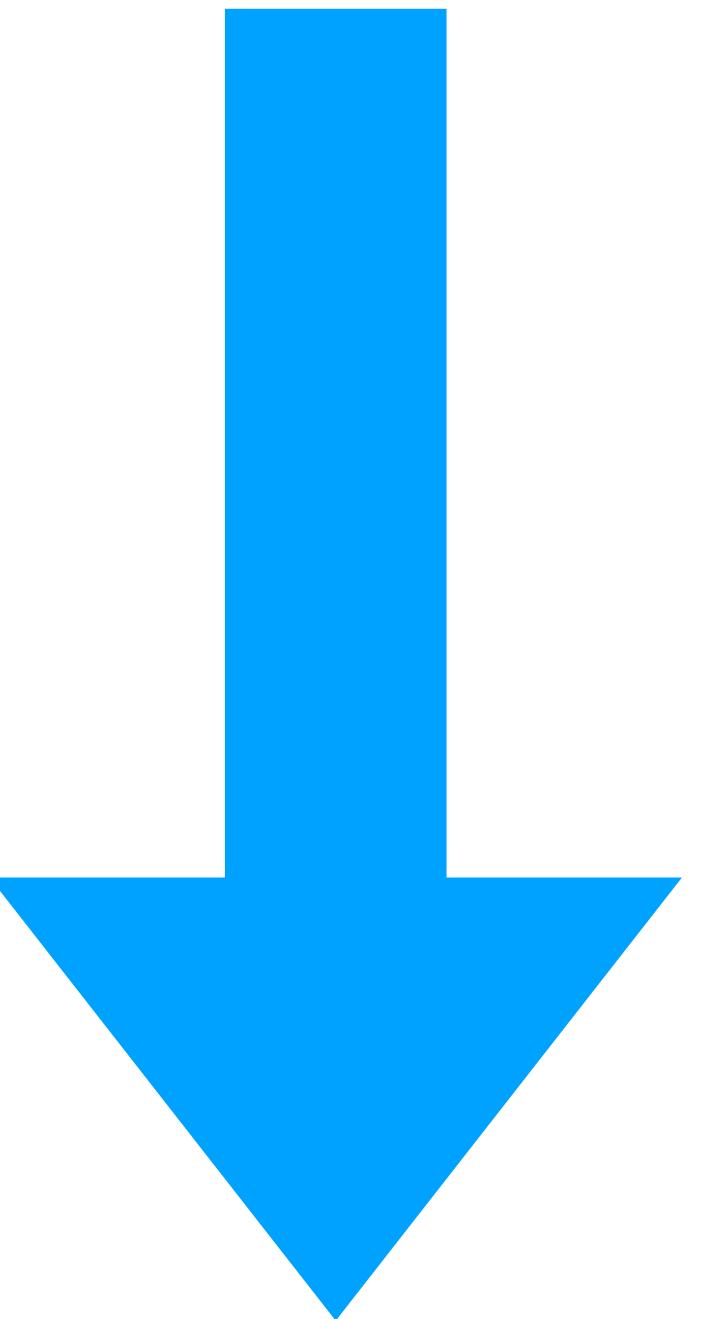
Problem-driven design

Top-down approach: design study

Identify problems encountered by users

Design a solution to ease their pain and make them more efficient

Use a set of well-defined visual representations or interactions techniques called idioms

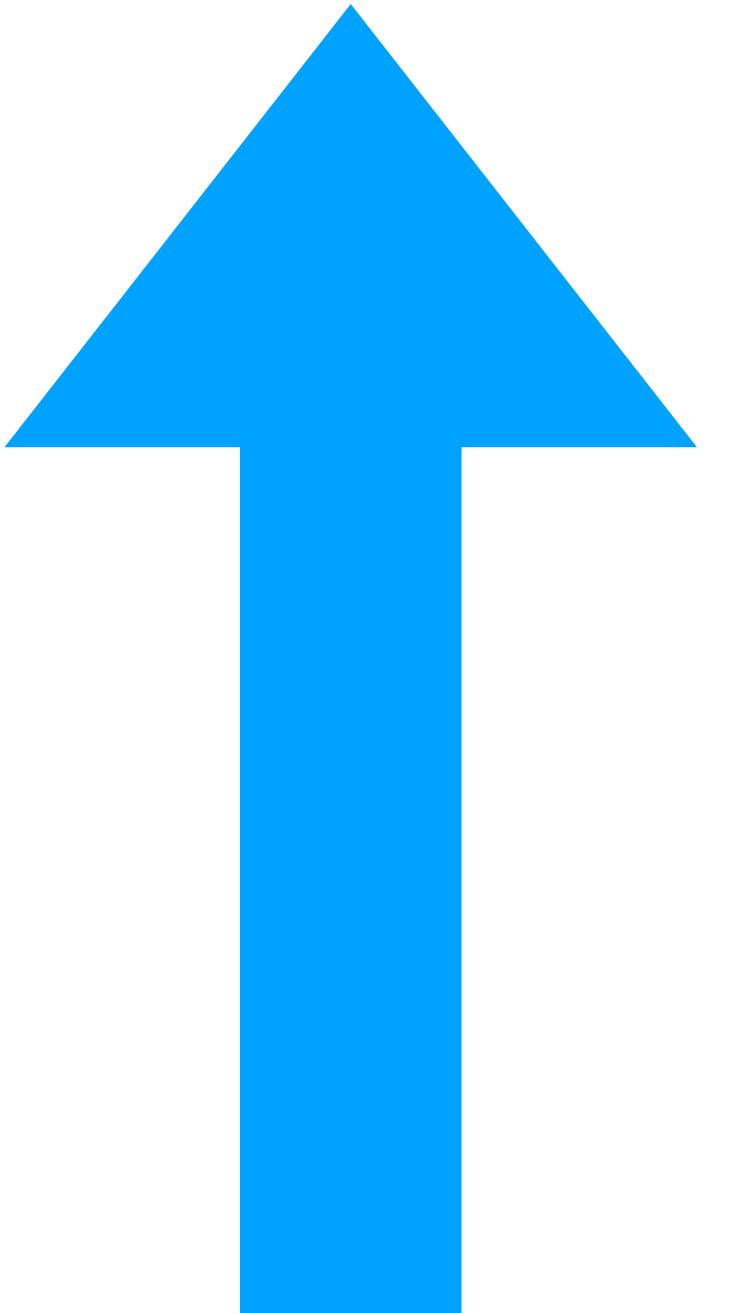


Technique-driven design

Bottom-up approach

Start with the idiom or algorithm design, to invent a better one

Classify or compare against other idioms and algorithms



Key factors impacting the design

People: audience, stakeholders

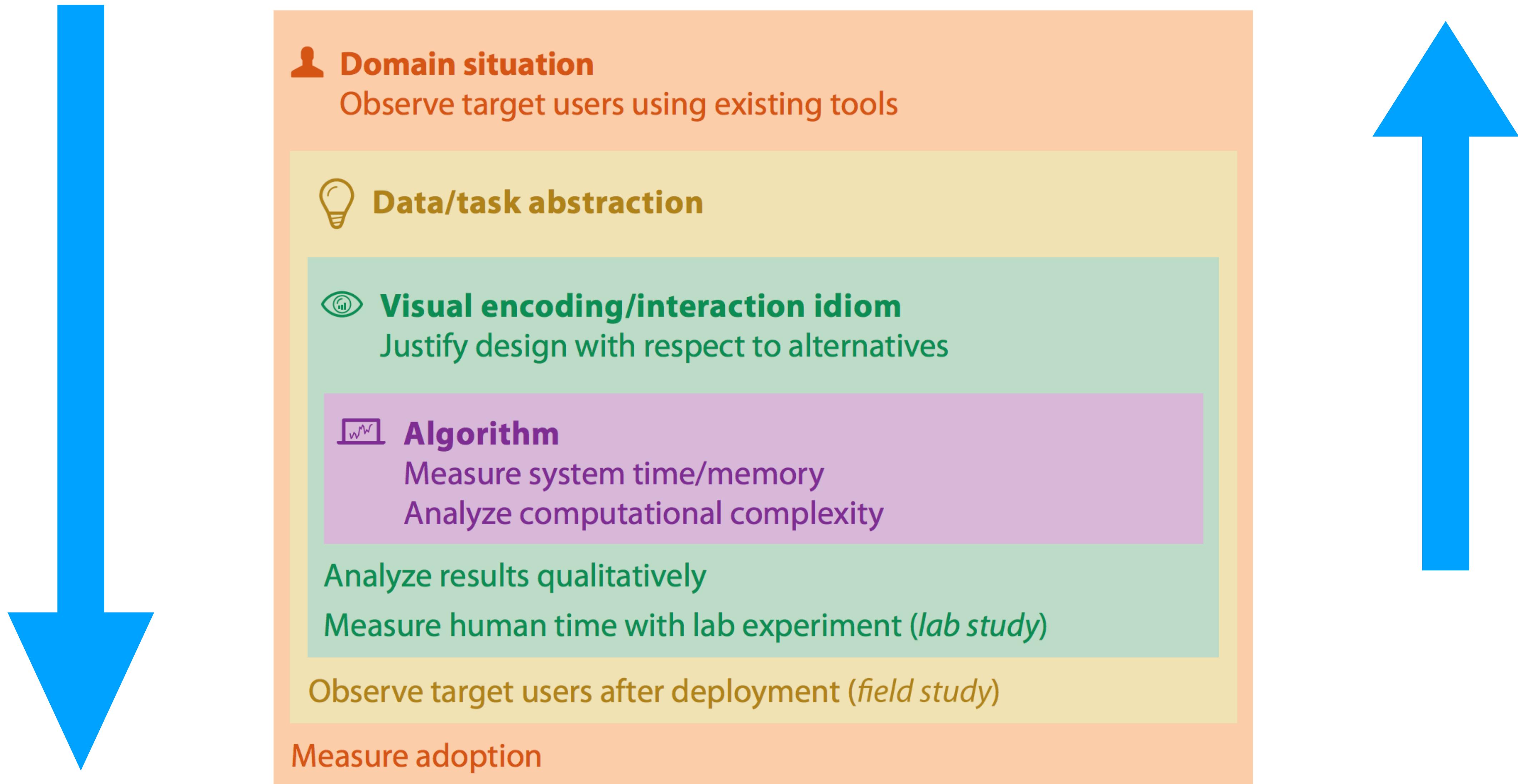
Constraints: pressure, rules, team, money

Deliverables: format, quantity

Skills, technology

Question: What are we accomplishing with this data viz?

The 4 nested levels of vis design



Purpose of the nested model

pragmatic approach to the design process

analyse each aspect of the pipeline

justify a design choice

quickly iterate on the results

helps avoiding ineffective solutions

👤 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

Domain situation

What type of curiosity has motivated the decision to do this project?

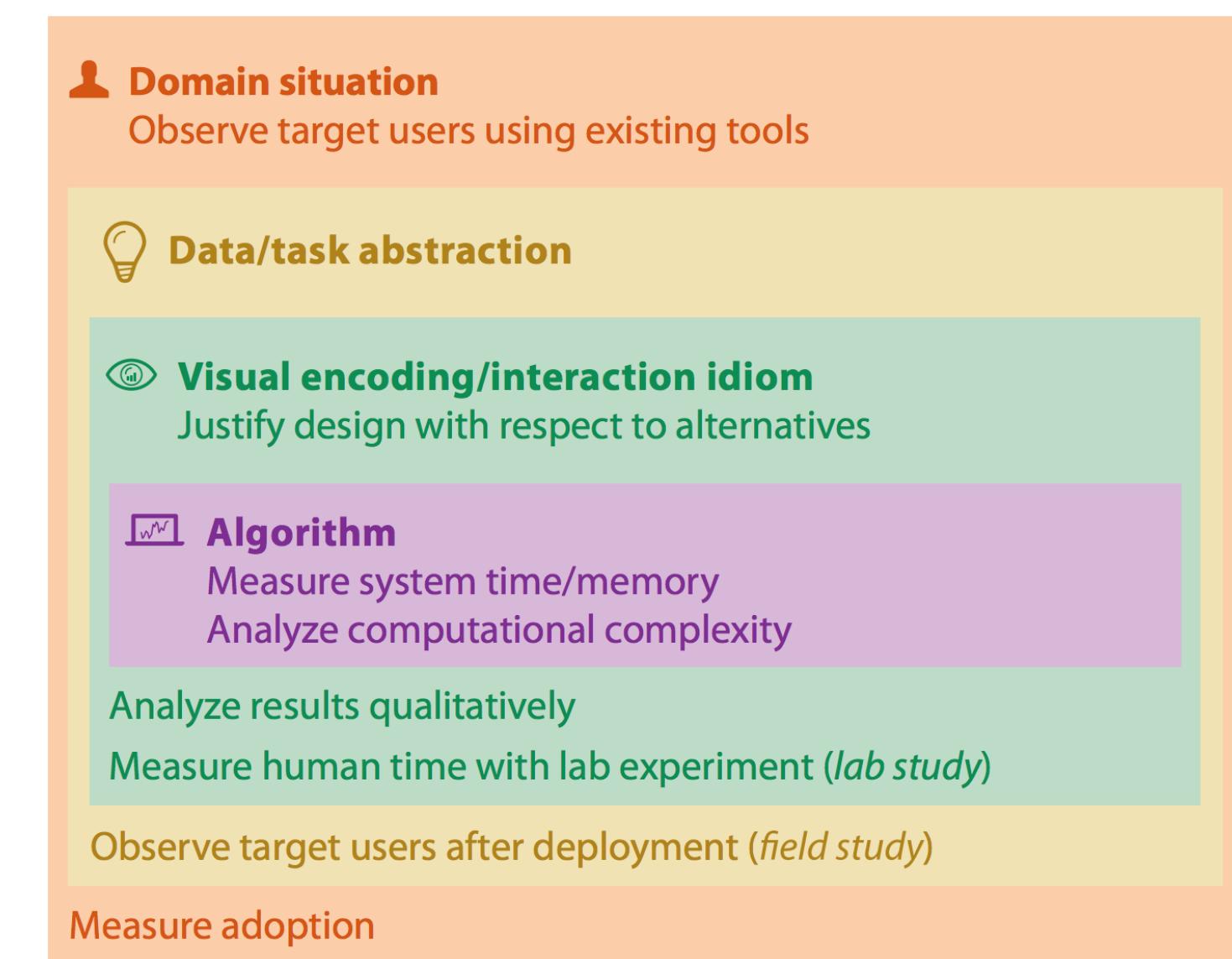
What is the target usergroup of the vis?

What essential questions need to be answered?

What kind of data can we expect?

Purely depends on domain and should be specific

yet, you cannot just ask people what they do because introspection is hard!



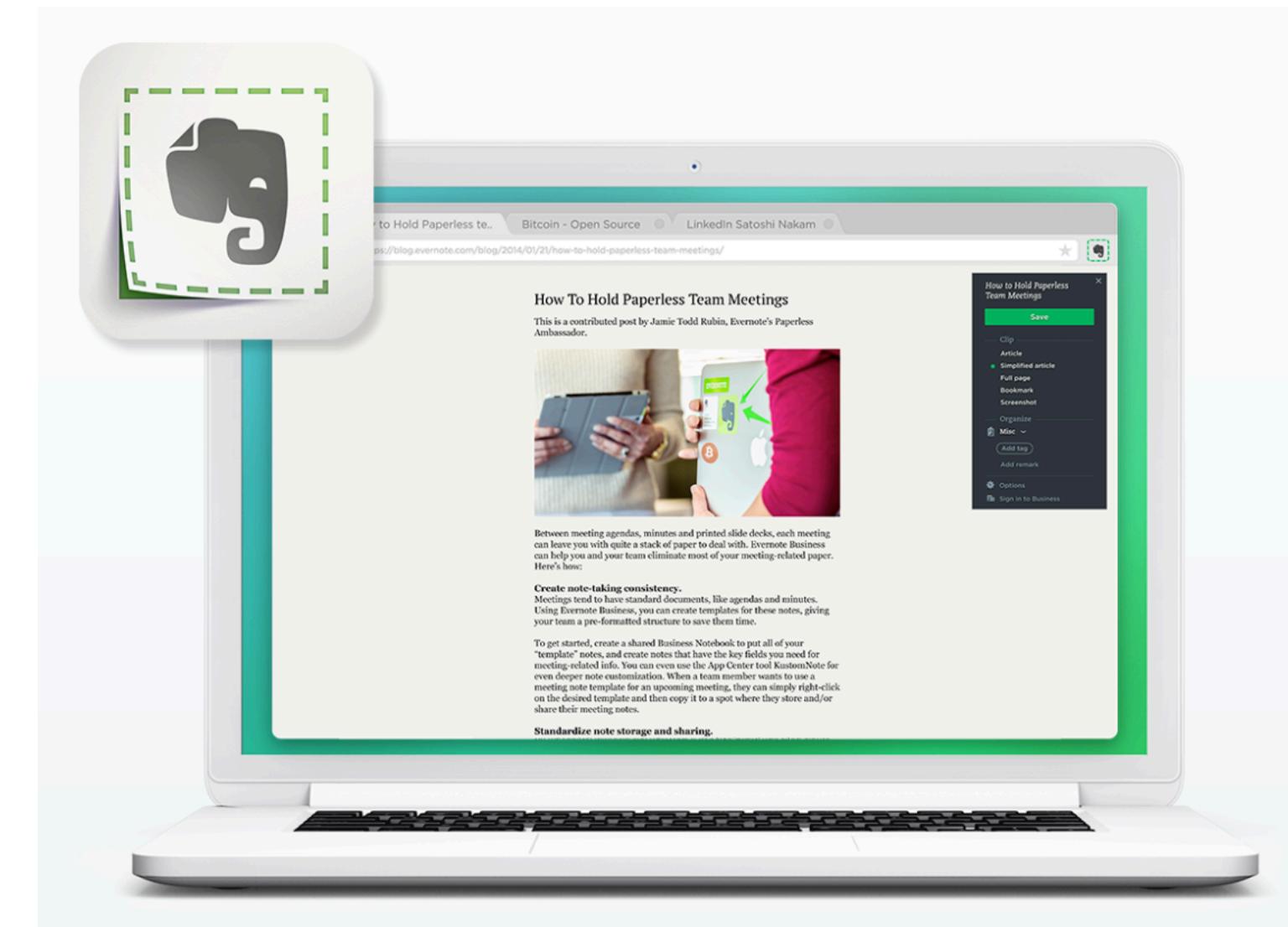
Gain knowledge on the problem

Read a lot on the field: use Google Scholar for scientific articles

Search for other designer's solution to related problems

Keep a digital scrapbook of everything you can find

Create personas: create reliable and realistic representations of your users segments for reference.

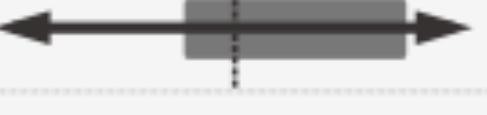


CEO (decision-making)

| | | | |
|---------------|---|-------|---|
| Goals | Coordinate personnel and operations | | |
| Knowledge | Operations | ●●●●● | Cyber ●○○○○ |
| Cyber SA | Attention | ●○○○○ | Temporal Window  |
| Key Questions | <ul style="list-style-type: none"> • How can we maintain ongoing operations? • What could happen if a critical system is impacted? • What are the most critical systems at risk of attack? • What cyber resources will be needed in the future? | | |

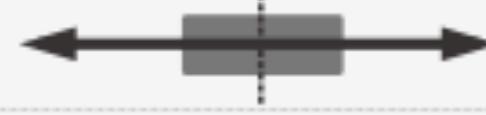
Decisions

Director of IT (decision-making)

| | | | |
|---------------|--|-------|---|
| Goals | Maintain cyber situational awareness | | |
| Knowledge | Operations | ●●●○○ | Cyber ●●●○○ |
| Cyber SA | Attention | ●●○○○ | Temporal Window  |
| Key Questions | <ul style="list-style-type: none"> • Does this attack matter? • How serious is the attack? • What do I do about the attack? • Are there any negative effects? • What did the bad guys do/take? • Is it a good day on the network? • How is my network different from last week? | | |

Information

NOC Manager (information-synthesis)

| | | | |
|---------------|---|-------|---|
| Goals | Communicate impact on operations | | |
| Knowledge | Operations | ●○○○○ | Cyber ●○○○○ |
| Cyber SA | Attention | ●●●○○ | Temporal Window  |
| Key Questions | <ul style="list-style-type: none"> • Does this attack matter? • How serious is the attack? • What do I do about the attack? • Are there any negative effects? • How successful was the attack? • What did the bad guys do? • What did the bad guys take? | | |

Information

Cyber Analyst (information-gathering)

| | | | |
|---------------|---|-------|---|
| Goals | Identify anomalous network behavior | | |
| Knowledge | Operations | ●○○○○ | Cyber ●●●●● |
| Cyber SA | Attention | ●●●●○ | Temporal Window  |
| Key Questions | <ul style="list-style-type: none"> • What does my network look like? • What happened on the network last night? What's different? • Is something bad happening? • How was my network attacked? • Who is attacking my network? • Does this attack matter? • What did the bad guys do? | | |

Data abstraction

How to model the data, what data types to use?

What good transformation of the data could be useful?

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*)

What?

Datasets

→ Data Types

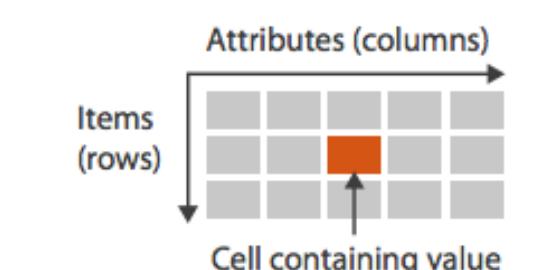
→ Items → Attributes → Links → Positions → Grids

→ Data and Dataset Types

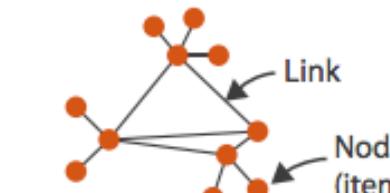
| Tables | Networks & Trees | Fields | Geometry | Clusters, Sets, Lists |
|------------|------------------|-----------|-----------|-----------------------|
| Items | Items (nodes) | Grids | Items | Clusters, Sets, Lists |
| Attributes | Links | Positions | Positions | Items |

→ Dataset Types

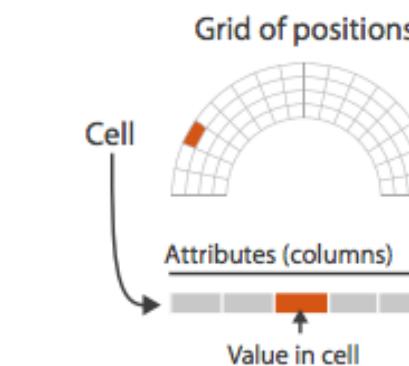
→ Tables



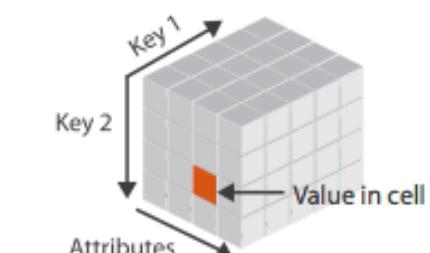
→ Networks



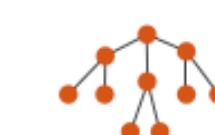
→ Fields (Continuous)



→ Multidimensional Table



→ Trees



→ Geometry (Spatial)



→ Dataset Availability

→ Static



→ Dynamic



Attributes

→ Attribute Types

→ Categorical



→ Ordered

→ Ordinal



→ Quantitative



→ Ordering Direction

→ Sequential



→ Diverging



→ Cyclic



What?

Why?

How?

Task abstraction: actions

What tasks user already do?

What tasks users wish to perform?

→ 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



Analyze

Consume data

Discover: find new knowledge that was not previously known

Present: telling a story with data

Enjoy: casual encounters with vis, no pressing goals

Produce data

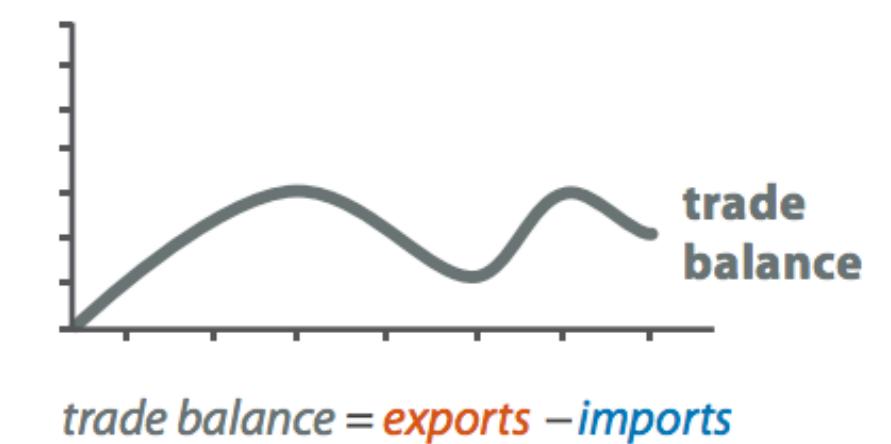
Annotate: addition of elements via manual action of a user

Record: captures visualization elements as persistent artifacts

Derive: produce new data elements based on existing data elements



Original Data



Derived Data

Search

Target known

Lookup: I know what I am looking for and where

Locate: find out where the specific object is

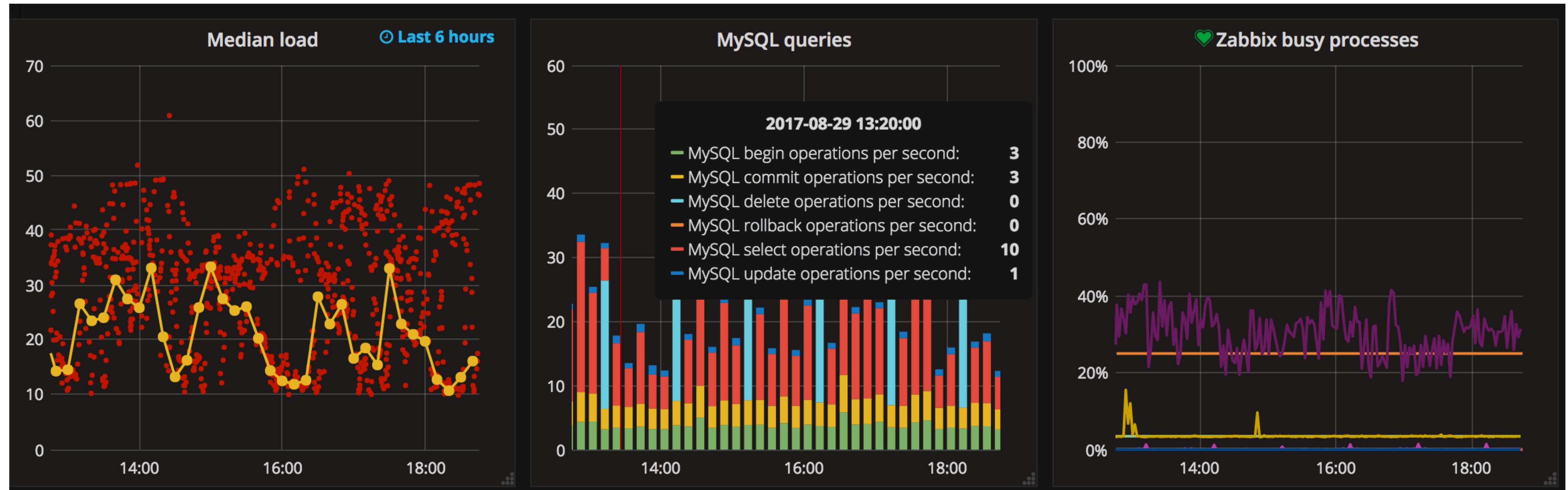
Target unknown

Browse: I know approximately where I could search

Explore: I am not sure where nor what to expect

Query

- **Identify:** a single target
- **Compare:** multiple targets
- **Overview:** summarize all possible targets



Task abstraction: targets

- Target: some aspect of the data that is of interest to the user.
- Targets are nouns, whereas actions are verbs.

→ All Data

→ Trends



→ Outliers



→ Features



→ Attributes

→ One



→ Many

→ Distribution



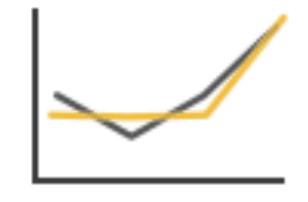
→ Dependency



→ Correlation



→ Similarity



→ Network Data

→ Topology



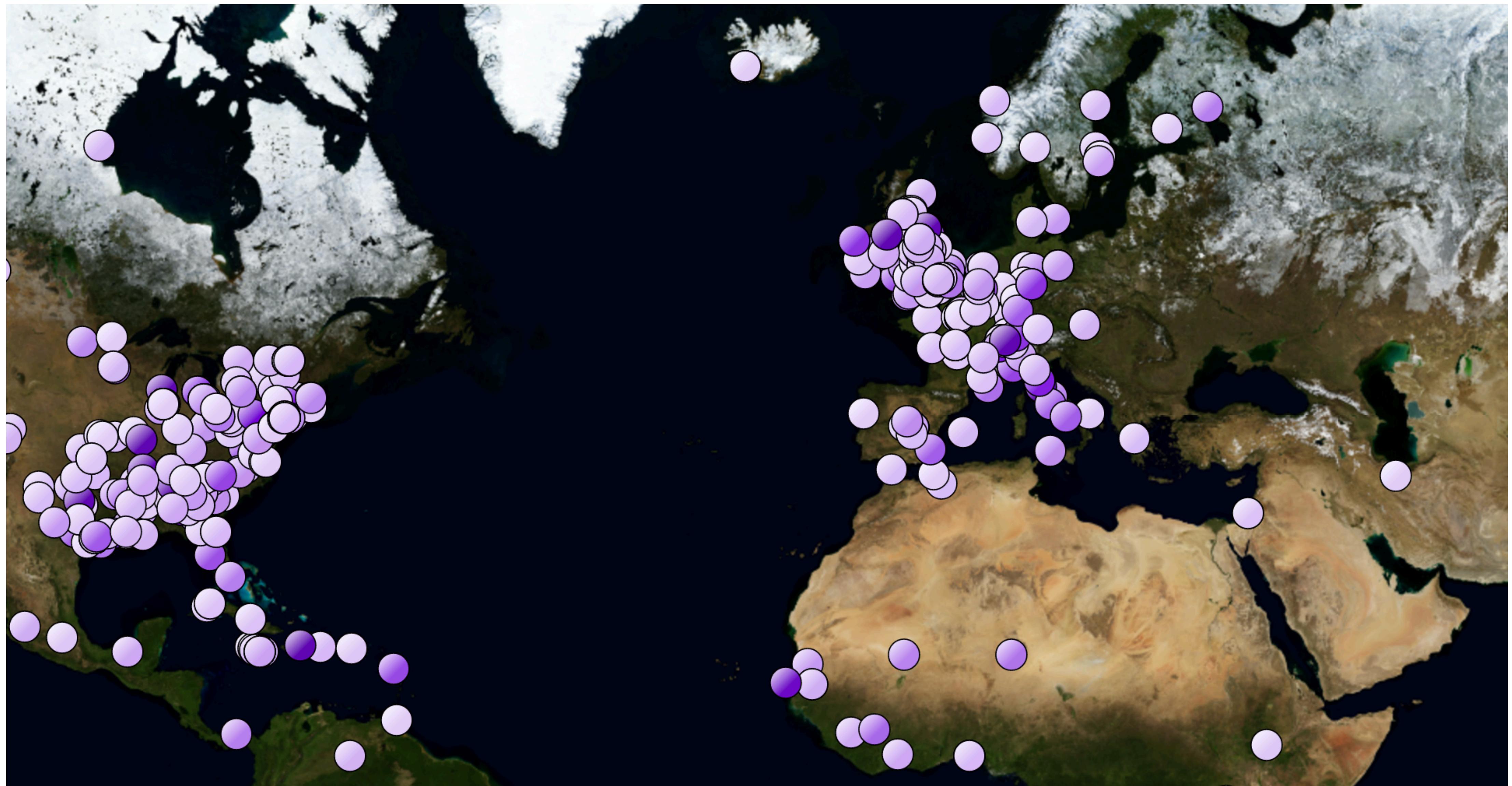
→ Paths



→ Spatial Data

→ Shape





Choosing a visual encoding / interaction idom

- What the ways to create and manipulate the data?
 - Rely on visual perception principles

Visual encoding/interaction idiom

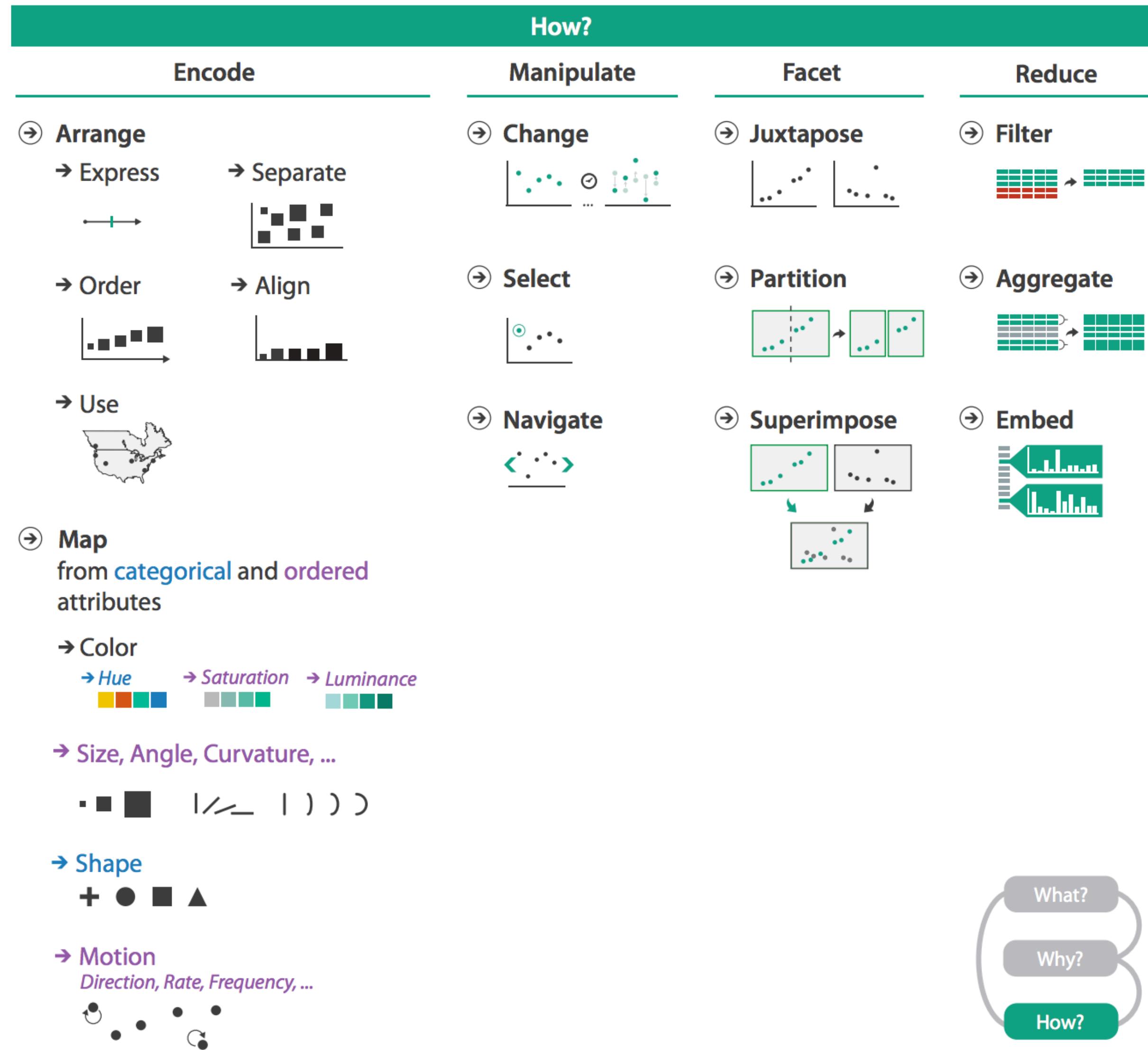
Justify design with respect to alternatives

Algorithm

- Measure system time/memory
- Analyze computational complexity

Analyze results qualitatively

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



Visual Vocabulary

There are so many ways to visualise data - how do we know which one to pick? Click on a category below to decide which data relationship is most important in your story, then look at the different types of charts within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

Click any section below to view the charts



Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative).

Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e., one causes the other).

Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

Distribution

Show values in a dataset and how often they occur. The shape (or 'skew') of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

Change over Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries: Choosing the correct time period is important to provide suitable context for the reader.

Part-to-Whole

Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead.

Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see fine differences). Usually these show a 'counted' number (for example, barrels, dollars or people) rather than a calculated rate or per cent.

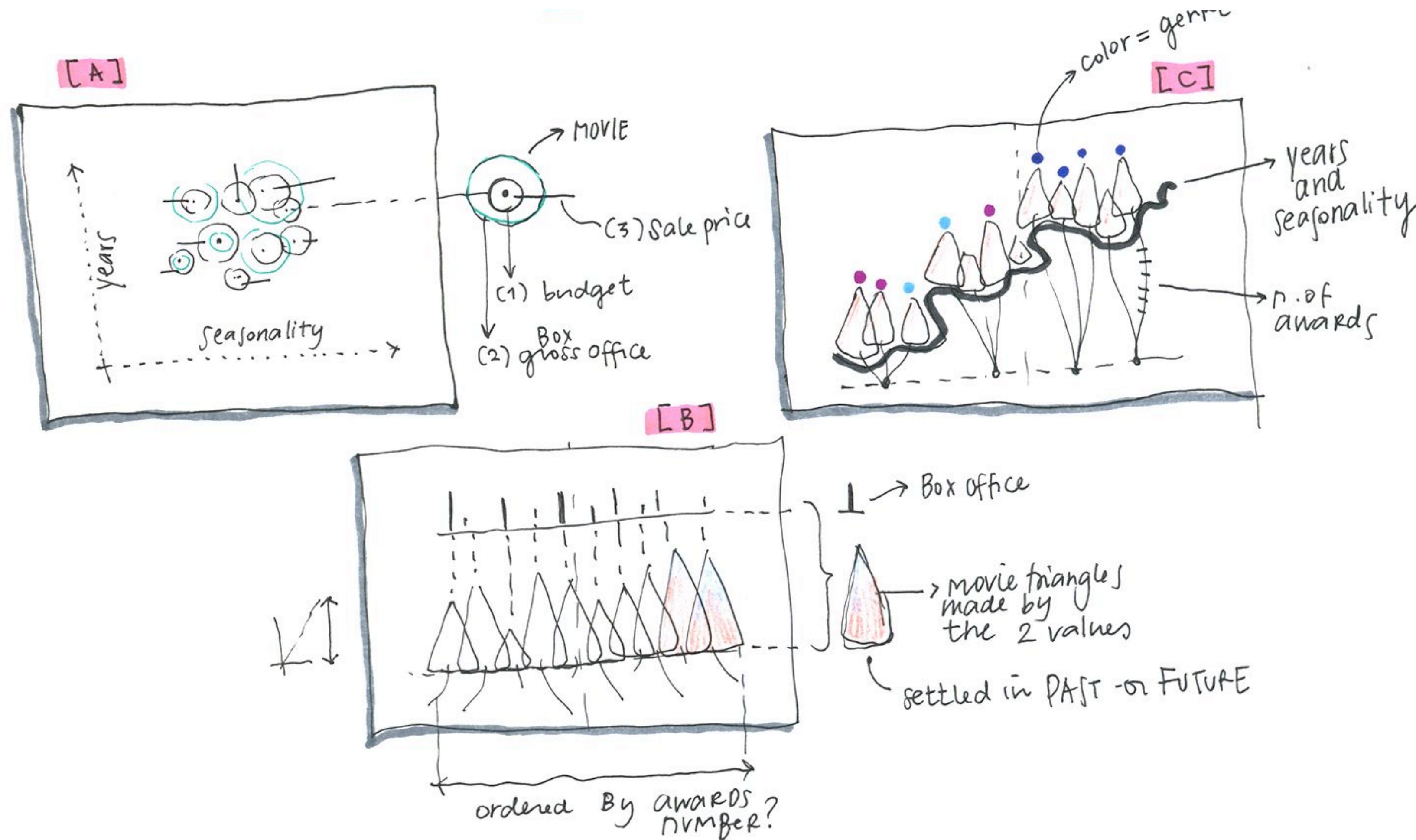
Spatial

Used only when precise locations or geographical patterns in data are more important to the reader than anything else.

Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

Sketching



Wireframing

Ask Question **Browse Questions**

Ask Question > Patient Rate / Count Question

Selection Criteria

- Demographics**
 - Gender: Female
 - Age: Any
 - Race / Ethnicity: Hispanic
- Diagnosis**
 - Patient Diagnosis: Multiple Selected
 - Any Selected (radio button)
 - All Selected (radio button)
- Treatment**
 - Stimulants (any)
- Date Range**: 2009 to 2011
- Data Type**
 - Count of Patients
 - Percentage of Patients within total enrolled population
- Location**
 - Mental Health Research Network (MHRN)
 - Group Health Cooperative (GHC)

Question Summary

Your question is:

How many and what percentage of hispanic female patients between 2009 and 2011 were diagnosed with Possible Self-Inflicted Injury or Self-Inflicted Injury and were taking Stimulants (any)?

Results

Show Breakdown by Age
Show Breakdown by Diagnosis
 Show Annual Breakdown

Table View

| Location | Year | Count of Patients | Percentage |
|----------|------|-------------------|------------|
| MHRN | 2009 | z,zzz | z.z% |
| MHRN | 2010 | z,zzz | z.z% |
| MHRN | 2011 | z,zzz | z.z% |
| GHC | 2009 | y,y | y.y% |
| GHC | 2010 | y,y | y.y% |
| GHC | 2011 | y,y | y.y% |

Statistics

| Location | Year | Mean Annual Count | Std Dev of Count | Mean Percentage |
|----------|---------|-------------------|------------------|-----------------|
| MHRN | Overall | z,zzz | zz | z.z% |
| GHC | Overall | y,y | y | y.y% |

Share Link to Results **Print View** **Download Table**

Mental Health Research **My Settings**

My Dashboard Ask Question Browse Questions

Ask Another Question

Which site(s) do you want to view?

Mental Health Research Network Group Health Both

Over what time frame? 2009 to 2012

What types of results are you interested in?

Count of Patients Rate within a population

Who is in your population?

age: 0 - 18
dx: Attention Deficit Disorder

What do you want the rate of?
ex: Stimulants (any)

xxx,xxx patients at MHRN
y,yyy patients at Group Health
x,x% at MHRN
y,y% at Group Health

Show Full Results

What types of results are you interested in?

Count of Patients Rate within a population

Rate (Numerator)
diagnosis or treatment

Nothing Selected

Population (Denominator)
gender or race/ethnicity

Not Filtering
All patients selected

At: Mental Health Research Network (MHRN)
Group Health Cooperative (GHC)

Over: 2011 to 2011

2011 Quick Results

| Numerator | Rate | Denominator |
|-----------|------------|-----------------|
| MHRN: 0 | MHRN: 0.0% | MHRN: x,xxx,xxx |
| GHC: 0 | GHC: 0.0% | GHC: zzz,zzz |

Algorithms

Must efficiently support all idioms

Managing algorithm complexity is important to reach interactive framerates:

60 fps: 16ms loop



Algorithm

Measure system time/memory

Analyze computational complexity

Strive for “fast-enough” response

Have a look on the CPU / memory utilization

Learn about 2D graphics if you can

Evaluation / validation

The only way to identify and justify your choices is to test them in-vivo

In-vivo: actual physical demonstration of the viz in front of real users

Measuring success:

User feedback

Key Performance Indicators (KPI)

Algorithmic runtime/complexity

Domain situation

You misunderstood their needs

Data/task abstraction

You're showing them the wrong thing

Visual encoding/interaction idiom

The way you show it doesn't work

Algorithm

Your code is too slow

Deploying your viz

Maintenability

Strive for reusable code parts

What happens if new data comes in?

Hosting

Amazon s3, Github pages, static-page generator

Communication

Write a blog post around the viz

Leverage social networks: LinkedIn, Twitter, Medium

Be friends with celebrities...

