Philosophy

Having a philosophy for analyzing data makes it
faster, **cheaper** and **more fun**
to achieve sound and meaningful results.

Fundamental theorem of data analysis

Frameworks/theories/models

Descriptive/mechanistic/normative levels

*Aside: Marr’s levels of analysis

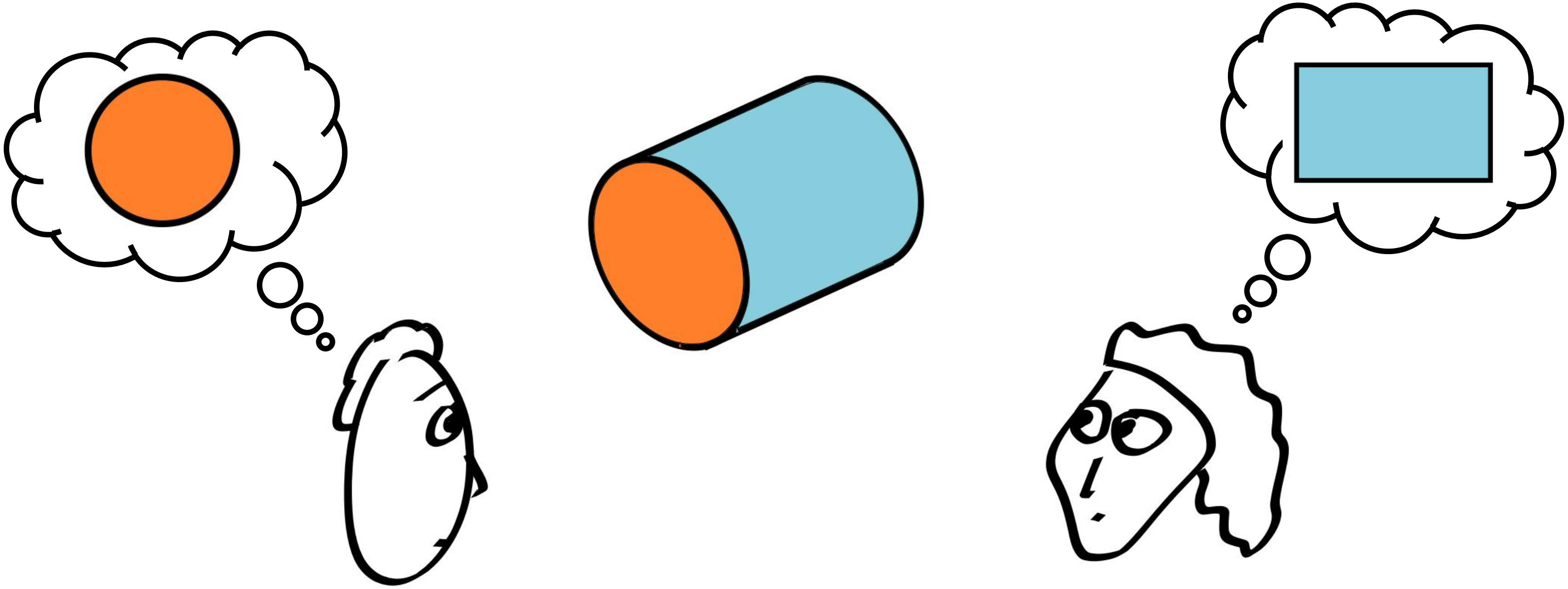
Mental model of a dataset

Transforming data into science

Predictions make results meaningful

McNamara fallacy

What you see depends on how you look



What you see depends on how you look

What you see depends on how you look

Rarely a “best” way of looking (especially in neuroscience)

You just see different things when you look in different ways

BUT it’s useful to think about which ways of looking will reveal something interesting

Frameworks, theories and models

A neuroscientifically useful way of thinking

Framework: language/terminology, way of thinking/looking

- *NOT falsifiable: no framework is wrong (unless logically inconsistent)*
- *But certain frameworks are more USEFUL, because they seem to produce much insight*
- *Example A: Tuning curves*
- *Example B: Dynamical systems*

Theory: “story”, in language of framework, explaining system/process

- *Falsifiable (can envision experiment that would prove theory wrong)*
- *Often mathematical*
- *Example A: Neurons in hippocampus are tuned to spatial position*
- *Example B: Hippocampal networks implement fixed point attractors*

(sometimes a bit fuzzy)

Model: Specific instantiation of theory that allows comparison to data

- *Usually requires specification of parameters*
- *Example A: Firing rate is squared exponential function of distance to “preferred” location*
- *Example B: Synaptic weights + dynamics update rule implementing fixed point attractor*
- *Models do not need equations, but we will consider ones that do*

On the role of theory and modeling in neuroscience (Levenstein et al 2023)
https://www.jneurosci.org/content/43/7/1074

Descriptive, mechanistic, and normative levels

A neuroscientifically useful way of thinking

Descriptive: “what”

- E.g. hippocampal neurons are tuned to 2-D position

Mechanistic: “how”

- E.g. tuning emerges from layered network transformations

Normative: “why”

- E.g. spatial tuning creates useful code for navigation
- Specific to research where makes sense to talk about “function” of systems, e.g. biology

Can exist at various levels of detail, e.g.:

High-level description: neurons are tuned to position

Low-level theory: Tuning curve equations with parameters drawn from specific class of distribution

On the role of theory and modeling in neuroscience (Levenstein et al 2023)
https://www.jneurosci.org/content/43/7/1074

***Not absolute**
(one person’s mechanism is another’s description)

Examples

	Framework	Theory	Model
Descriptive	Neural coding via tuning curves	Hippocampal neurons are tuned to position	Tuning curve equation (e.g. firing rates vs position)
Mechanistic	Network transformations produce tuning	Recurrent neural network transforms sensory input into spatial codes	Instantiation of network (weights, update rule, etc.)
Normative	Neural computations for navigation	Spatial codes advantage route-finding	Equations showing how route is found using spatial codes

Very useful for keeping track of different ideas when forming scientific argument

Aside: Marr’s levels of analysis

Computational: What computation is being done?

- E.g. V1 is computing edges from images
- Also called “content”

Algorithmic

- E.g., successive network layers extract features from image that produce edge-detector filters
- Also called “form”

Implementational

- E.g. synaptic weights from retina to LGN, then LGN to V1, + single-neuron nonlinearities
- Also called “medium”

*Somewhat controversial over how useful vs limiting they are

Exist within a computational framework (neurons perform computation by transforming information)

- So don’t apply to every problem about e.g. emergent dynamics in complex systems where computational language isn’t used

https://www.albany.edu/~ron/papers/marrlevl.html

Building a mental model of a dataset

Simplifies envisioning analysis routines and outcomes

Can be much more selective about which analyses to run (they take time and are complex, even in a dry lab)

Otherwise surprisingly easy to apply sophisticated analysis only to realize months later that it was going to yield inconclusive results due to data artifact

Prerequisite for having a brilliant shower thought for how to analyze data

Makes advanced analyses more meaningful

A rough guide

Spend serious time *just looking* at raw/ lightly processed data

- Write down anything weird/interesting

Internalize dataset structure and scope (e.g. file sizes and organization)

Compute and internalize all the basic stats (obvious correlations, key timescales, etc)

Note all artifacts and obvious confounds for analysis (e.g. nonstationarities)

Make a report/PDF summarizing all of the above that you can come back to or share with others

DON'T do anything fancy

Transforming data into science

The light side

Clearly something interesting going on, just need to quantify

The twilight zone

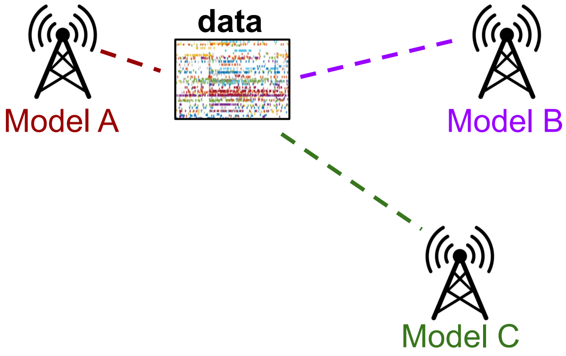
Apply “reduction” to data (e.g. dimensionality reduction) to see if something interesting pops out*

Then quantify

**Be CAREFUL: Easy to mistake artifact of analysis method for property of data*

The dark side

Good data, but complex and hard to tell what’s going on.



Employ model comparison to gain insight.

Predictions

Why is making predictions important?

- Ensures self-consistency of analysis
- Sanity checks
- Tests whether analysis can produce meaningful results (can one imagine two different realistic outcomes of analysis?)
- Forms prior to ensure analysis result will contain information
- Useful for avoiding likely inconclusive/useless analyses
- **If one can predict multiple reasonable yet different outcomes, guarantees analysis results will be meaningful**

Literary analogy

Irony := meaningful surprise

AND-BUT-THEREFORE

Transmitter of meaning

- Bessy predicts winning the lottery will make her happy but it doesn’t.

Surprise ~ prediction error

Meaningful prediction error ⇒ *learning*

Prediction error requires *predictions*

E.g. prediction 1: chaos, prediction 2: oscillations.

McNamara Fallacy

- Named after U.S. secretary of defence Robert McNamara during the Vietnam War
- McNamara Fallacy = over-reliance on quantitative observations
- In context of data & science: Easiest to quantify ≠ most important
- When analyzing data in biological science, always keep in mind what else is going on in the system, even if not measured/quantified or not easy to measure