

An Introduction to

Data insight

## Overview

- What is data insight?
- What is the method &  
+ obstacles to, data insight?
- What roles & topics are  
related to data insight?

- Data  
→ points of information (of any kind)?
- “Data Science”  
Computable: some Numerical Representation.
- Insight  
↓ Conclusions “derived”  
from Data  
insight implies quality  
of decisions
- Methodology  
Most datasets  
don't contain  
info. you  
need.

# Data Insight

def. high quality (or optimal)  
decision-making & inference  
from data

Data insight :

Some observations  
on method

Method: The greatest challenge is getting useful data..

- How do you get it?

→ Derive from existing dataset

How?

→ Algorithm or Model.

E.g. use weight / height

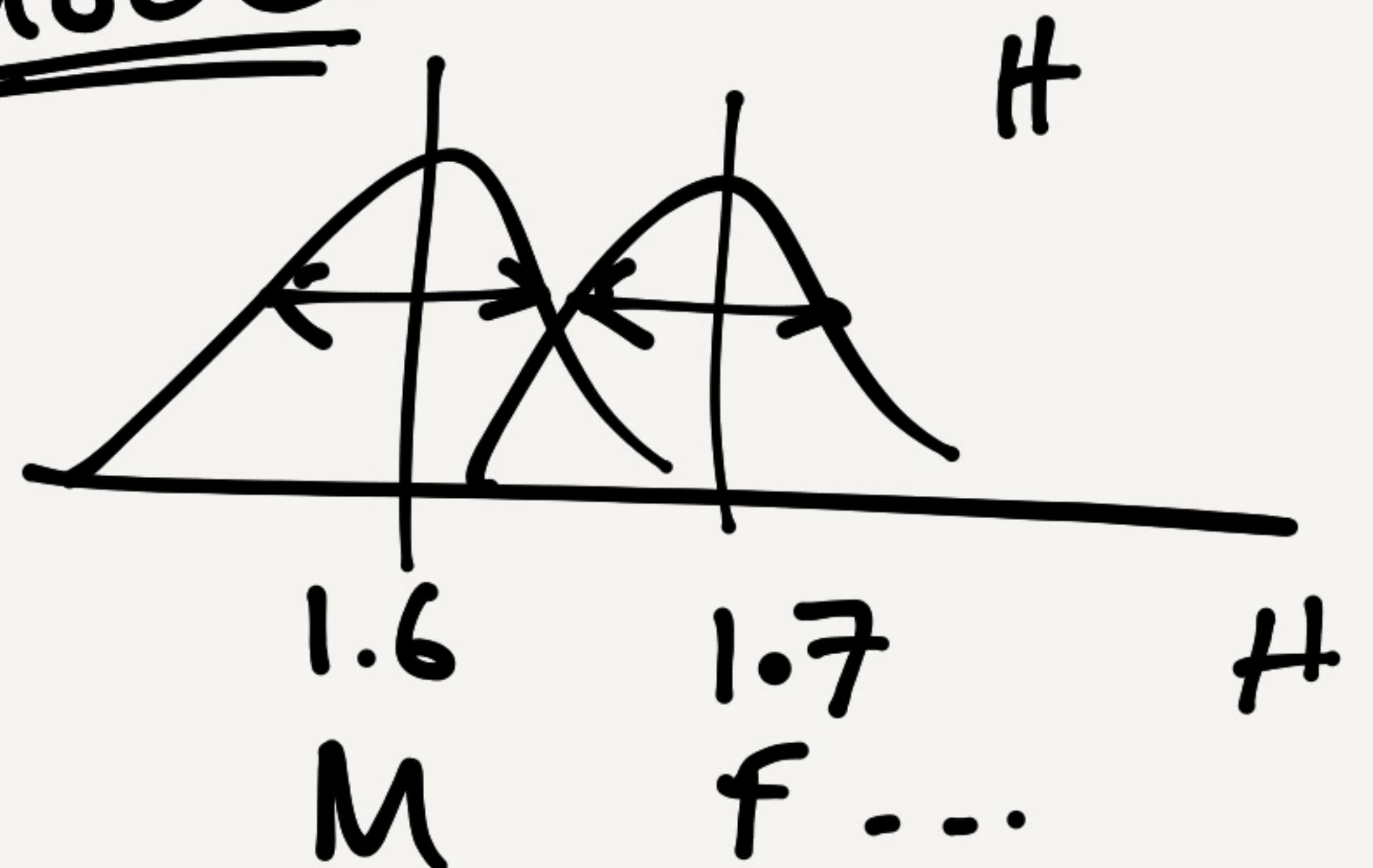
to infer sex:

W	H	...	sex
80	180		Male

- Approximate Answer, / "Chape Gließen"

Model

via a model,



- Experiment
- collect data

## Method

'useful' data is ambiguous

Almost any hypothesis can be  
evidenced

Eg. H: The Sun is Blue (FALSE)

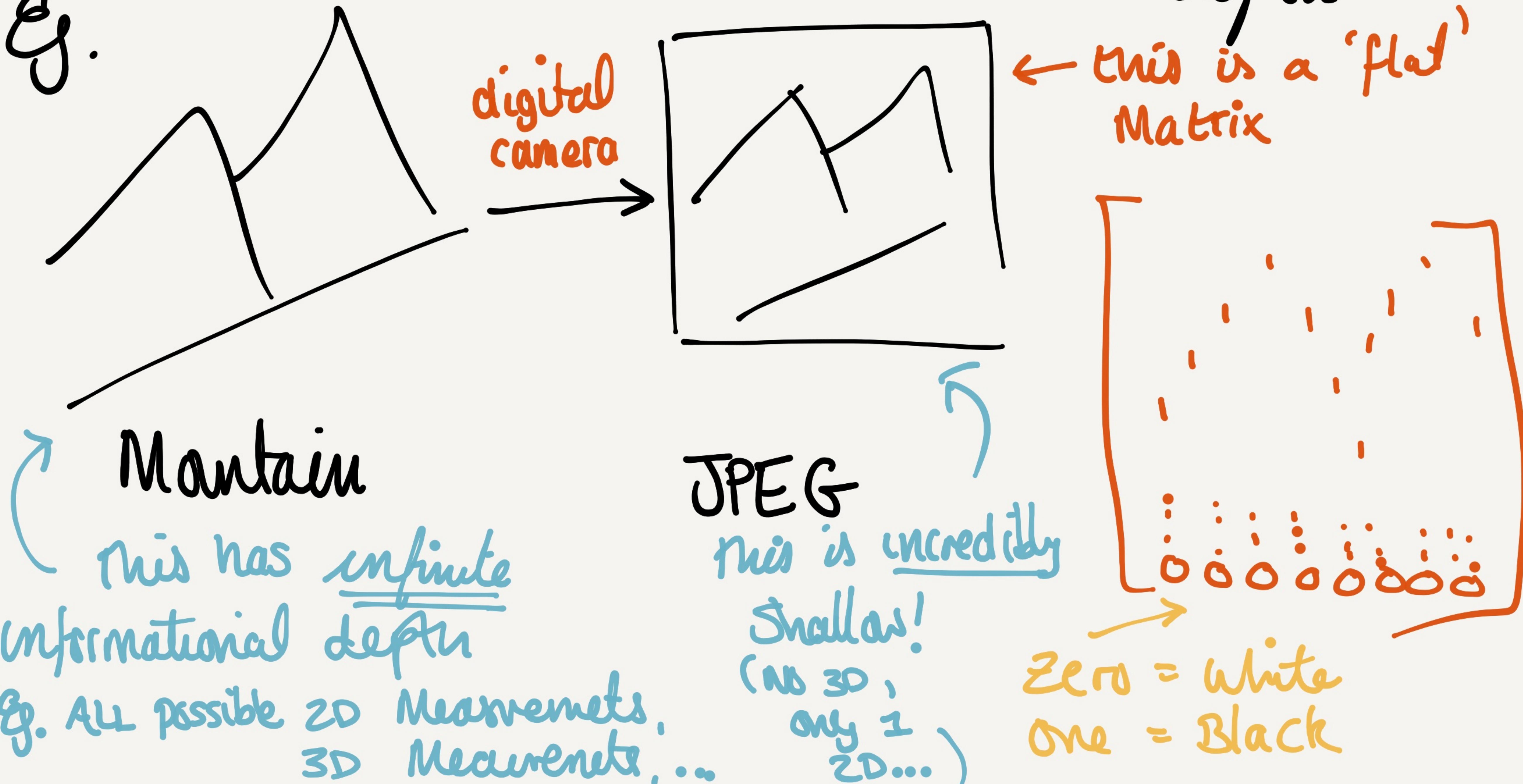
↳ Evidence: My blue light detector  
reads blue light (TRUE)

↳ Evidence-based decision making is hard!  
(hence: Science)

# Method

Physical objects are  
Encoded as data  
& therefore loose "informational depth"

Eg.



# Method

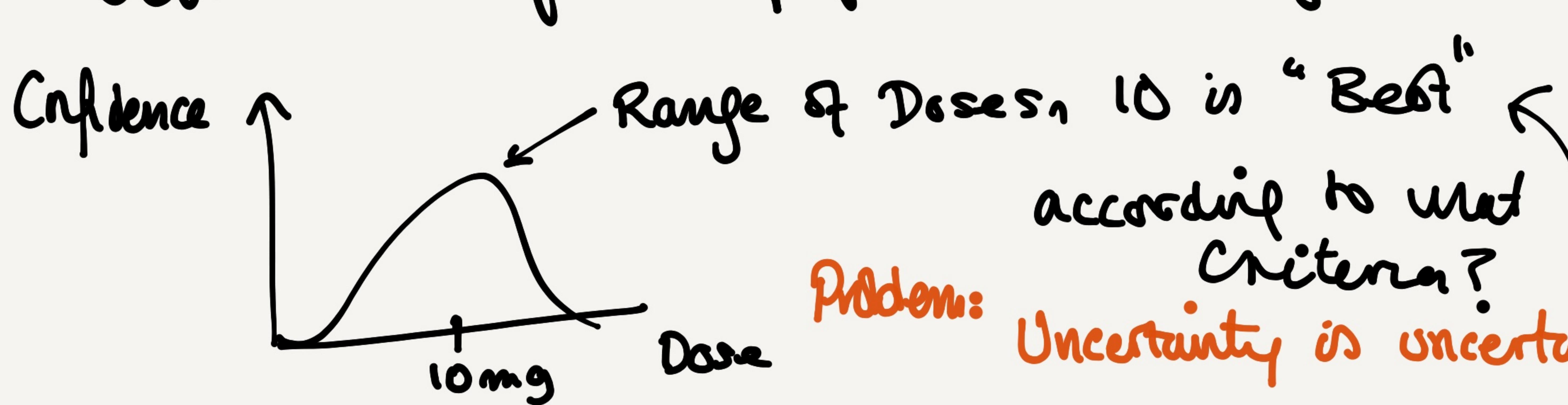
Results are hard to  
Explain & justify

- Can we explain our conclusion?

Data  $\rightarrow$  Model  $\rightarrow$  Prediction

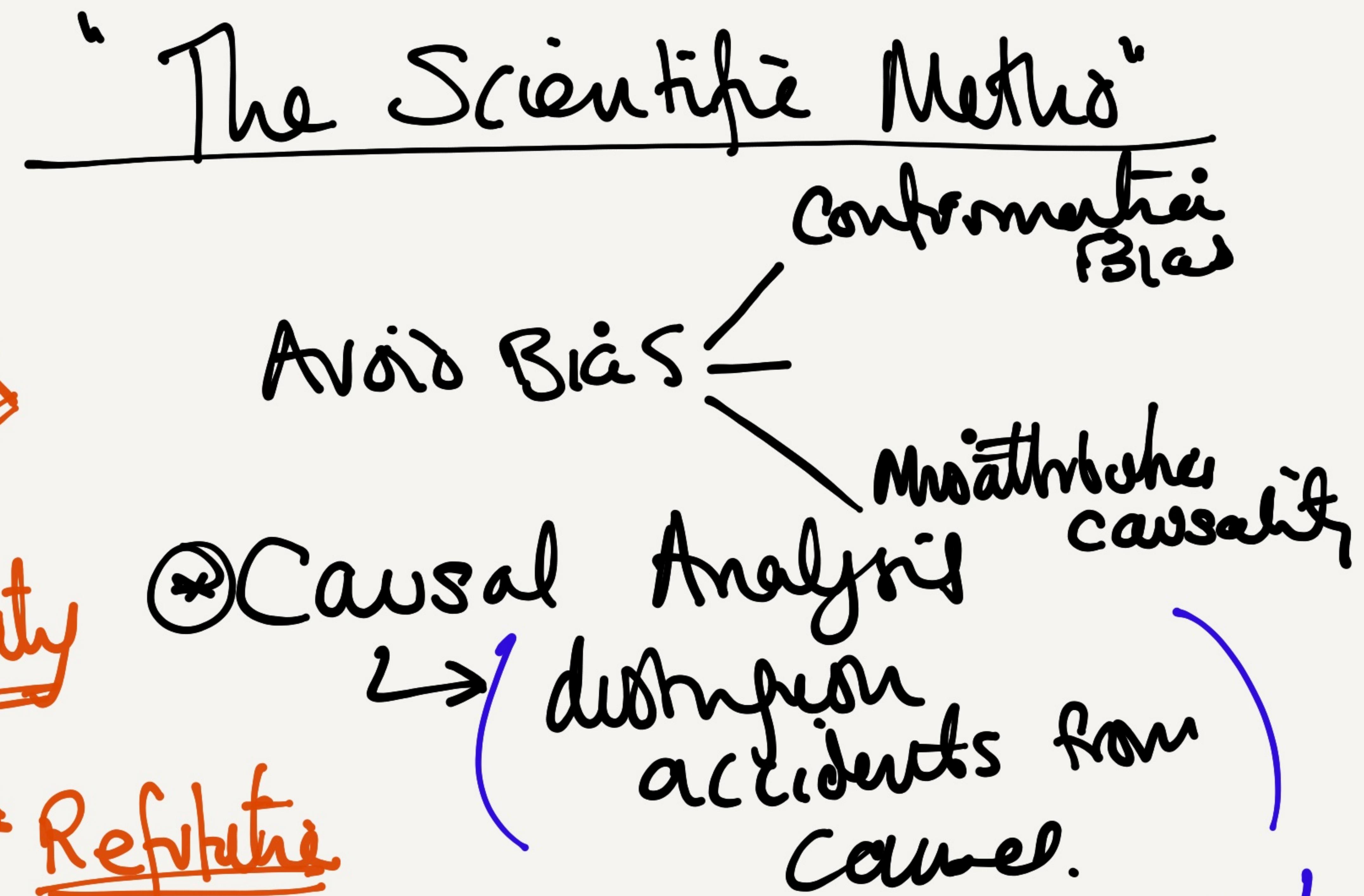
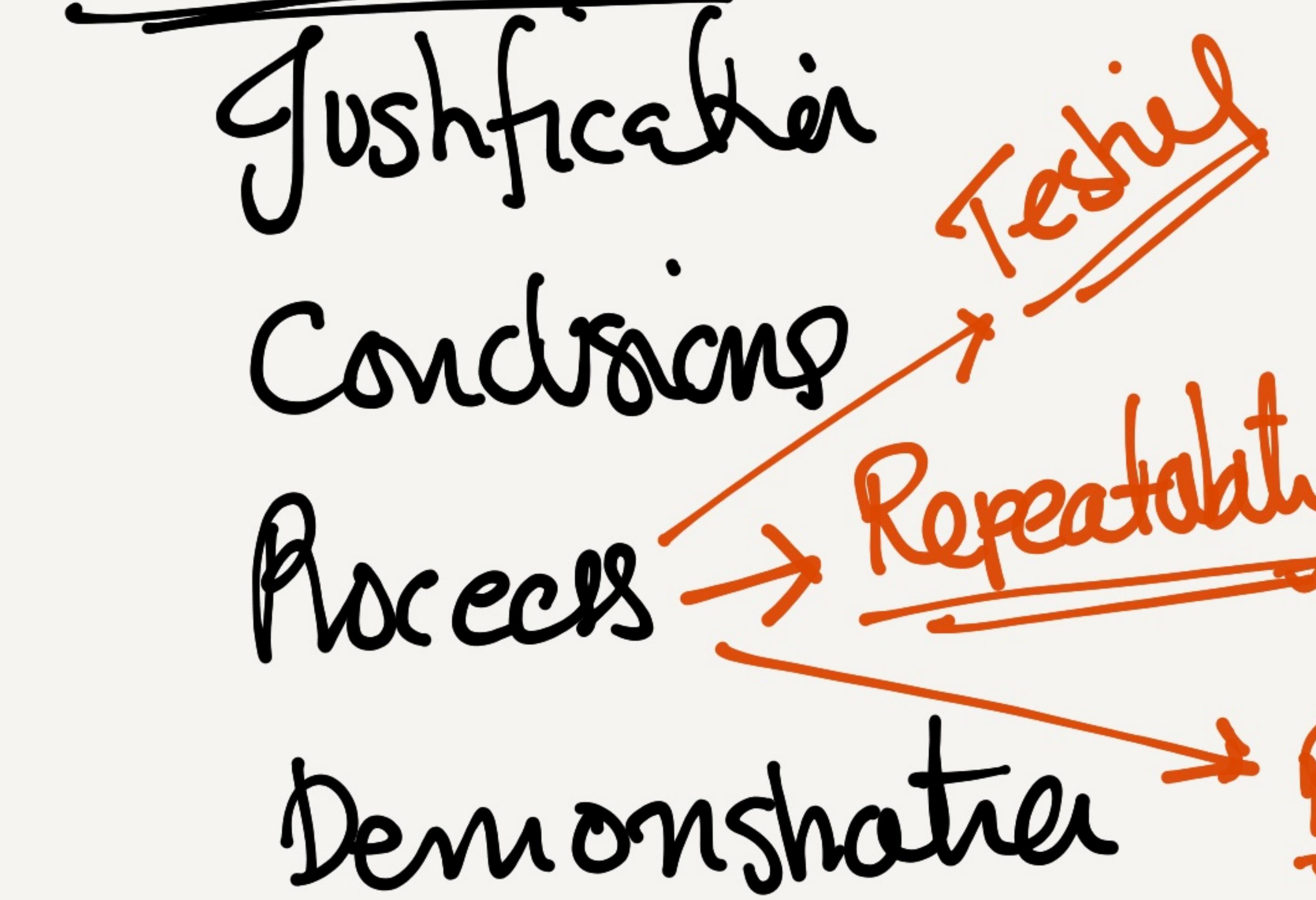
(HR, BP, ..)  $\rightarrow \hat{f}(x) \rightarrow \underline{\text{Drug Dose}}$   
 $\hookrightarrow$  why 10mg?

- Can we quantify uncertainty?



Problem: Uncertainty is uncertain!

Science  
"Normative"



Conformatiue Bias  
→ false consensus

Replication Crisis

Physics ABC/DEF...  
Society  
free association

```
graph TD; A[Physics ABC/DEF...] --> B[Society  
free association]; C[X] --> D[A0...AM]; E[X] --> F[X]
```

Data Science ← disagreeable.

↳ often testif. strategy / Practices

Executives may feel undermined / threatened

Science → "Settles debates"

=====

↳ Precise questions

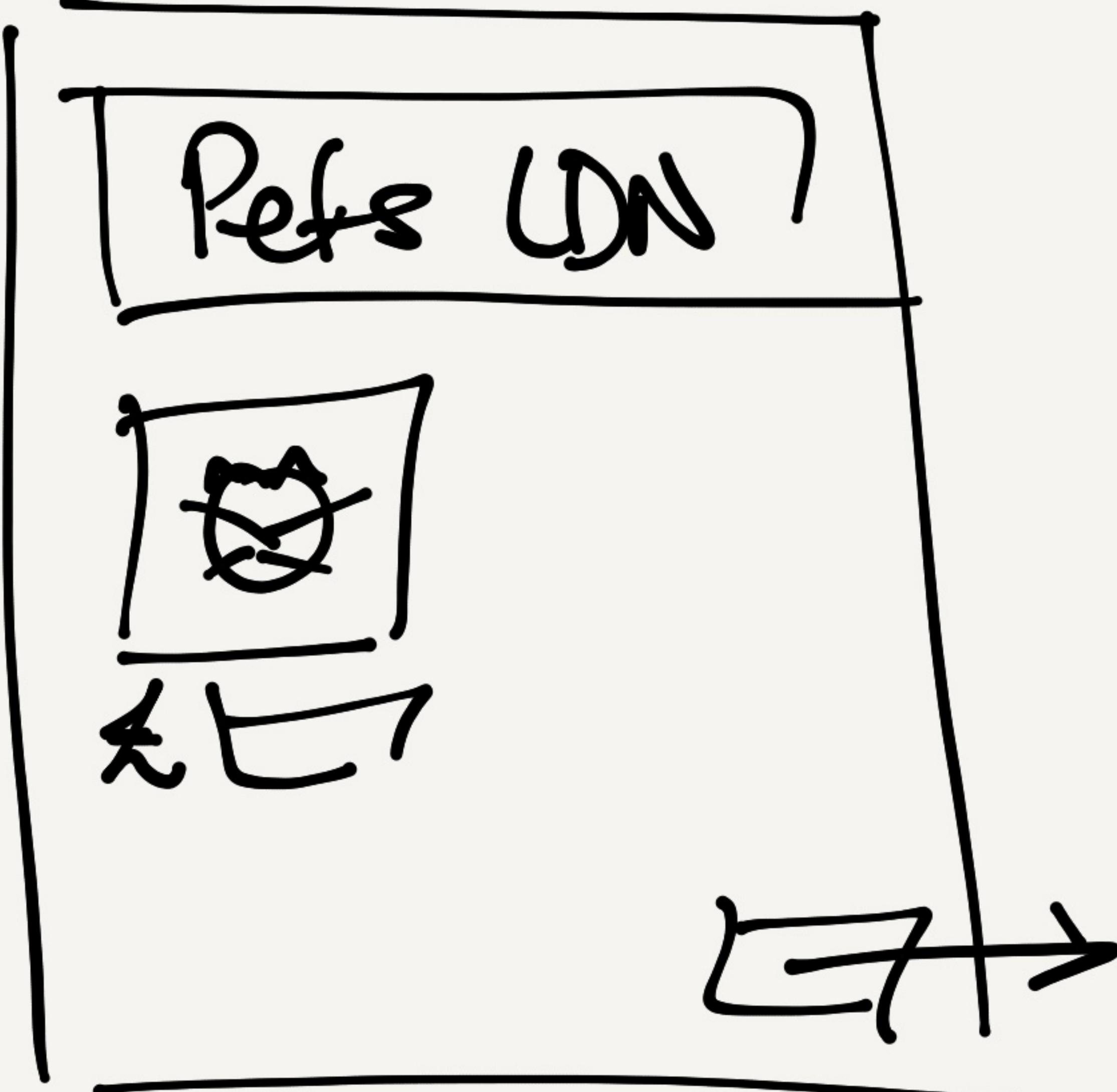
Challenging within org.

Should be resolved

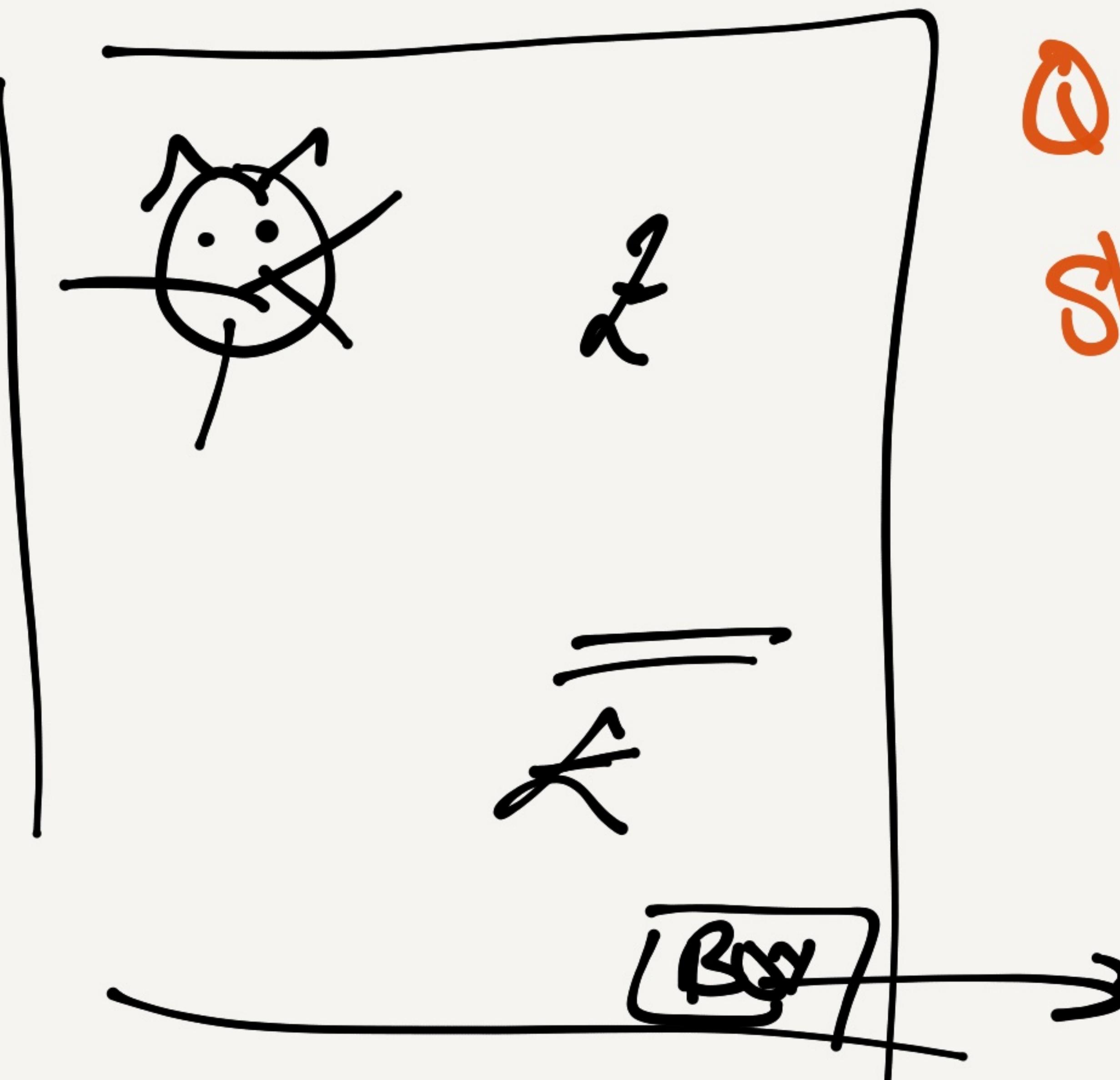
Note: Whether to act on recommendation is

NOT "settled"

Case Study → History-driven, 'merely analytical'  
("Traditional") Website done. (c. 2000s)



Consider a local pet store website, selling e.g. CAT FOOD



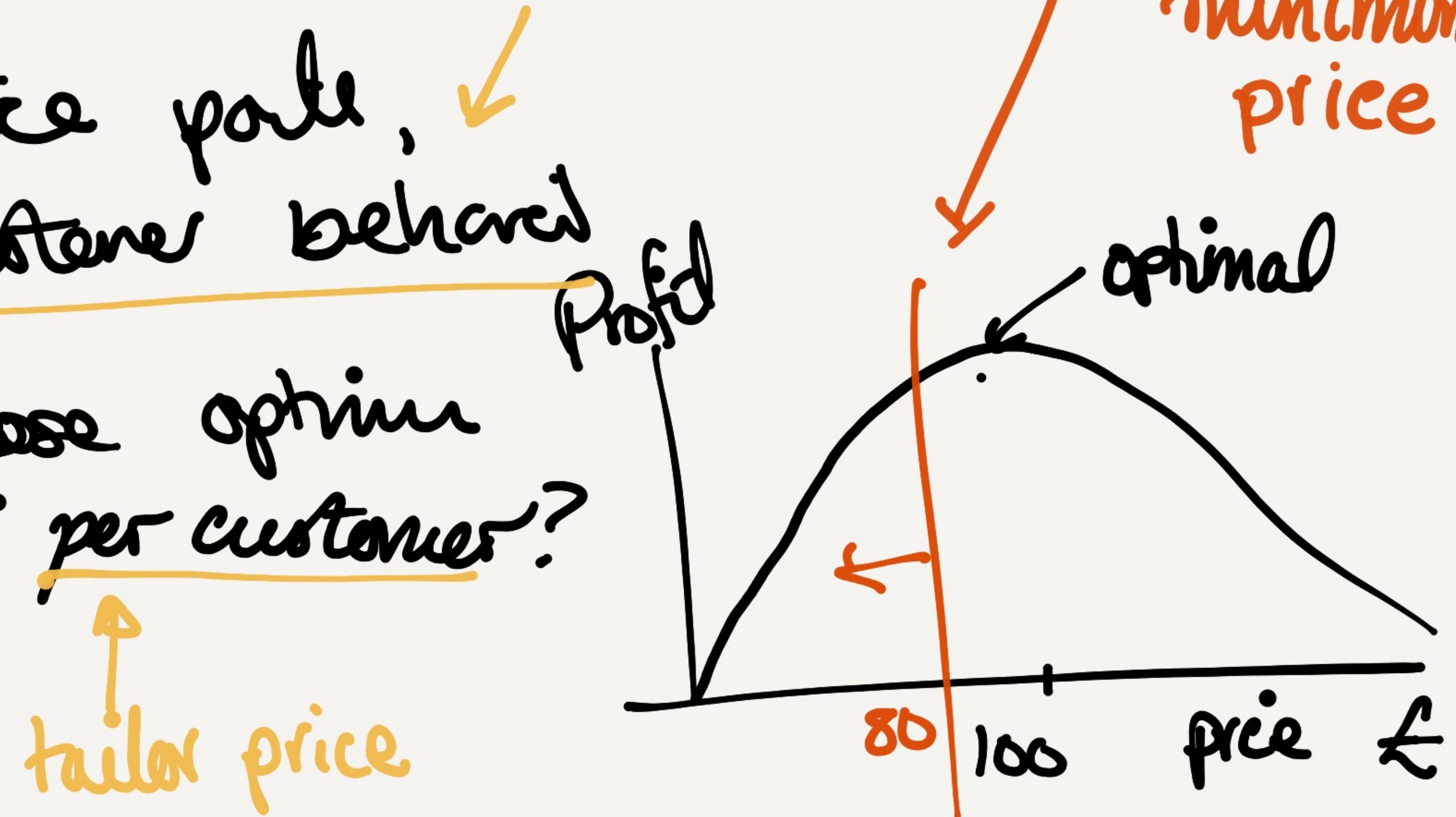
Q. What should their products be price?

### TRADITIONAL PRICING:

- Competitor Analysis
- Intuition
- Historical Analysis

You are hired as a data scientist to answer the Q. "How should we price a \_\_\_?"

- \* How much it costs us to keep one  
*Dynamic, Exponential* → sets a minimum price
- \* Try various price points, observe customer behavior & choose optimal profit per customer?
  - tailor price
  - Experiment!



Related

Rules &

Topics

# Data Ethics

"Casuistry"

def.

Evaluating case X by moral  
cavity of related cases, W & R

- What is the dataset?
- What "biases" are "baked in"?

"Statistical"  
Accurate  
 $\neq$  Useful!

X: SHOULD WE PREDICT GRADES?

R:

RIGHT CASES

Predict tumor  
without Dr, but  
Dr Review &  
contamination

W:

WRONG CASES

Without a fudge,  
Predict a sentence  
& actin/implement.

Eg. US Example Reproduces  
Bias w/o control.

which cases  
are most  
similar?

CONSIDER  
MANY!

# Data Analysis

Historical

Fact - Based

Reports

└ Aggregation

└ Intuition

Q. What were  
the most liked products?

A. "In 2000, X,  
2001, Y, ..."

Report:

"TOP previous  
best sellers"

Insight: "It seems  
X will sell well"

# Data Science includes DA.

⊕ Future - Orientated → Achön  
→ Predictions.

Guess / Estimation  
→ Uncertainty

Proposie

→ Achins part of &  
finished by Project

Explaining & Modelling what will happen

# Big Data

Common definition:

Volume → N° Rows / Data size

Veracity → Accuracy.

Velocity → Rate

Variety → N° Cols. / Complexity

Actual definition:

You have a non-traditional Problem

def. Relational ('Tabular') DB + Data Analytics  
+ Non-Inferential Softw'

# Data Engineering

↳ operational support for "Data Insight"

↓  
tools, software, APIs, "technical infrastructure"  
↳ Query, Optimization

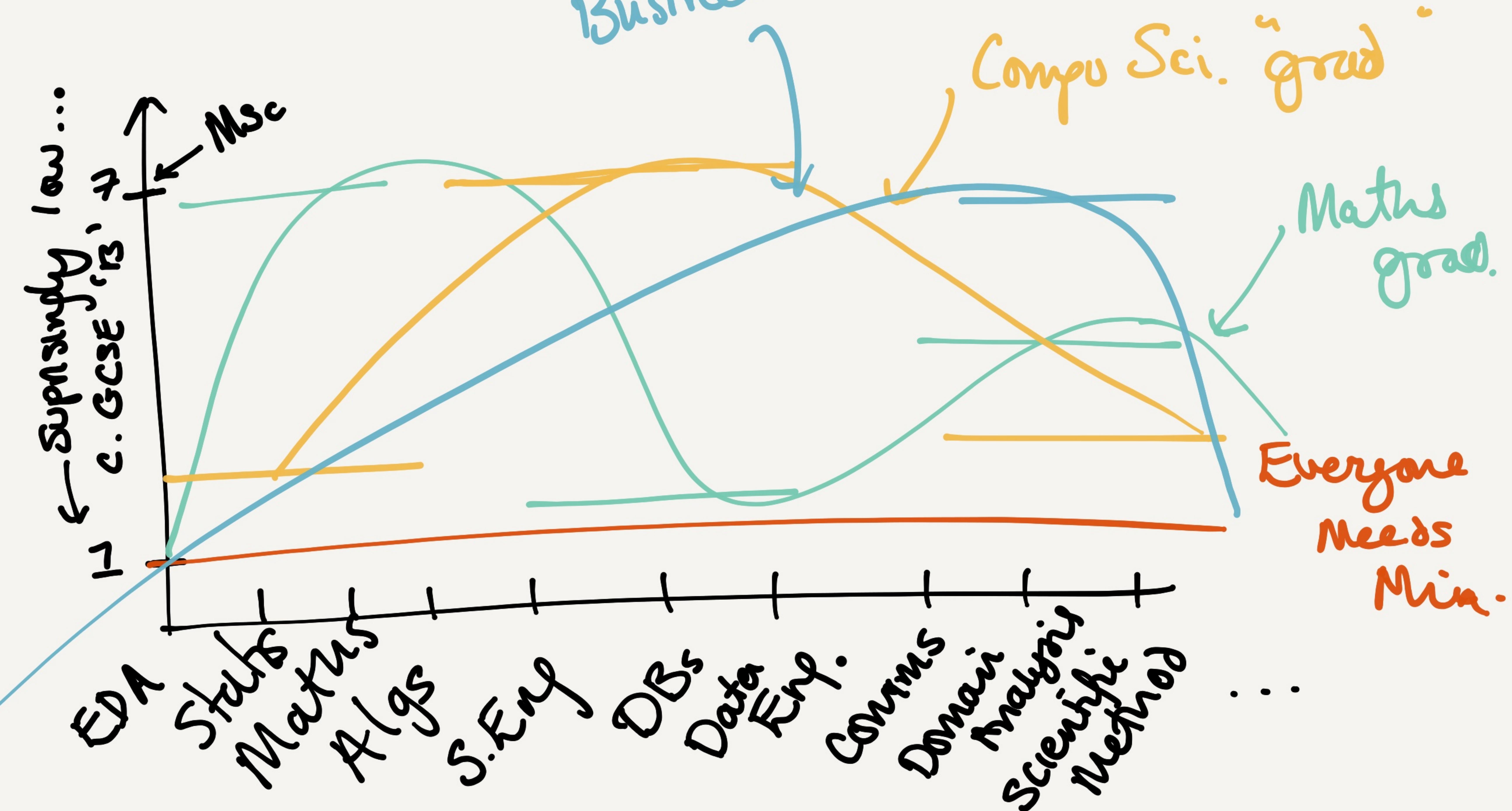
\* DE should precede DI

"Democratizing Data" ① ← cultural.

Tools Should  
Enforce ② → Excel?

# Data Scientist

Business Analysis "Background"



Every D/S team need a mixed membership.