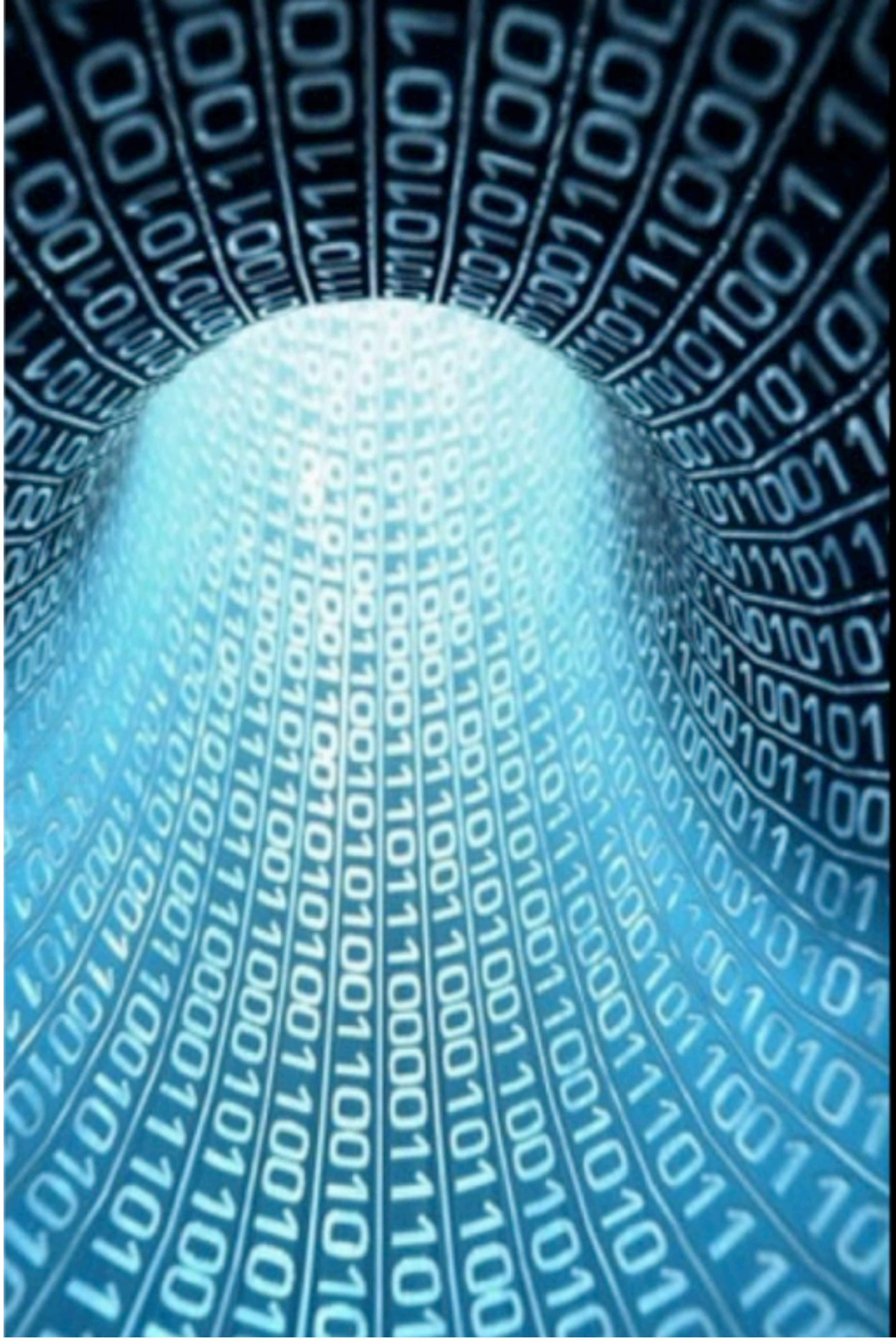
Reasoning under uncertainty

Ronojoy Adhikari
The Institute of Mathematical Sciences
Chennai, India

PyCon 2015
Bangalore, India



Logic
Probability

Logic
Probability

Probabilistic
programming

Logic
Probability

Probabilistic
programming

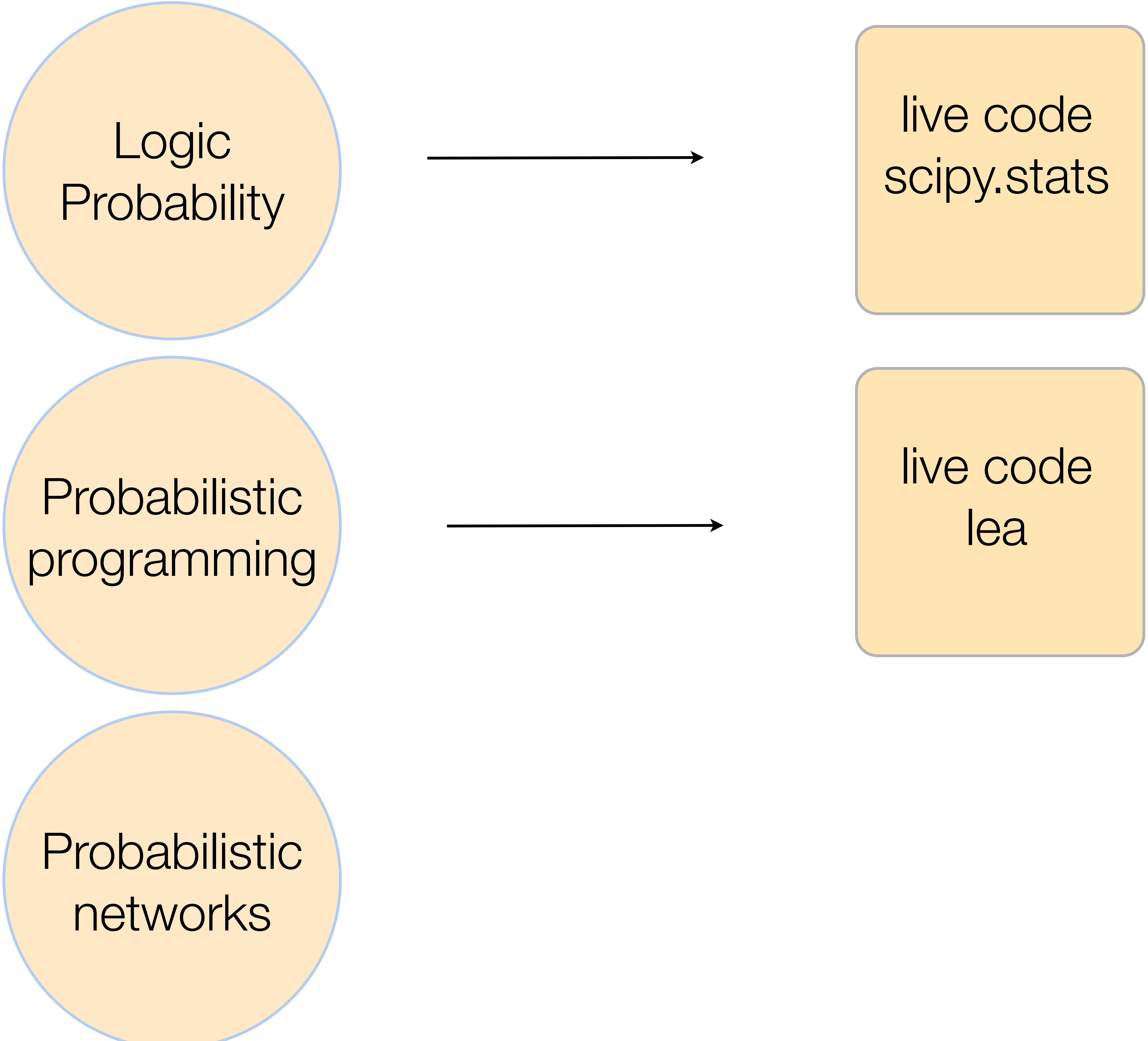
Probabilistic
networks

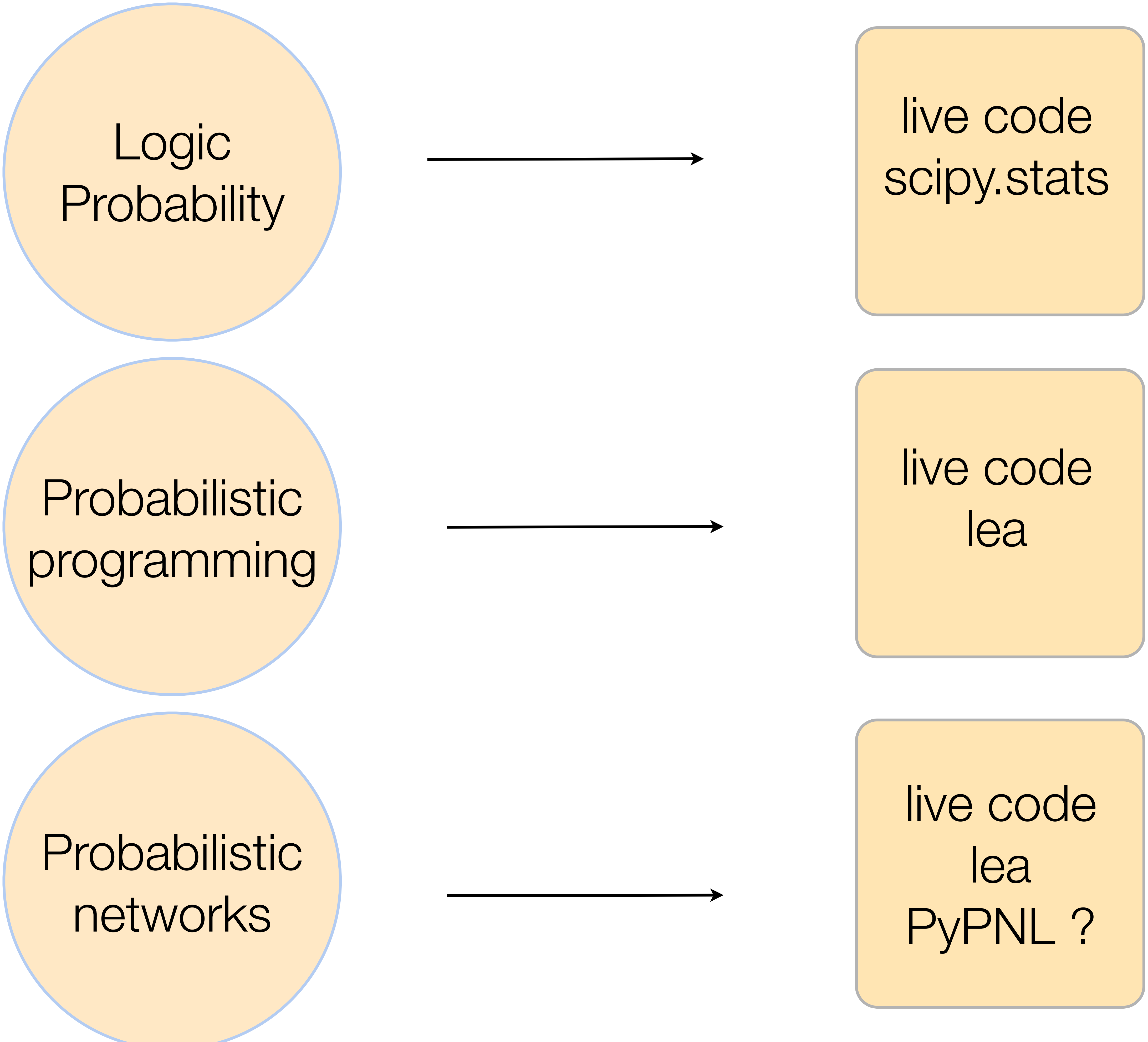
Logic
Probability

live code
scipy.stats

Probabilistic
programming

Probabilistic
networks

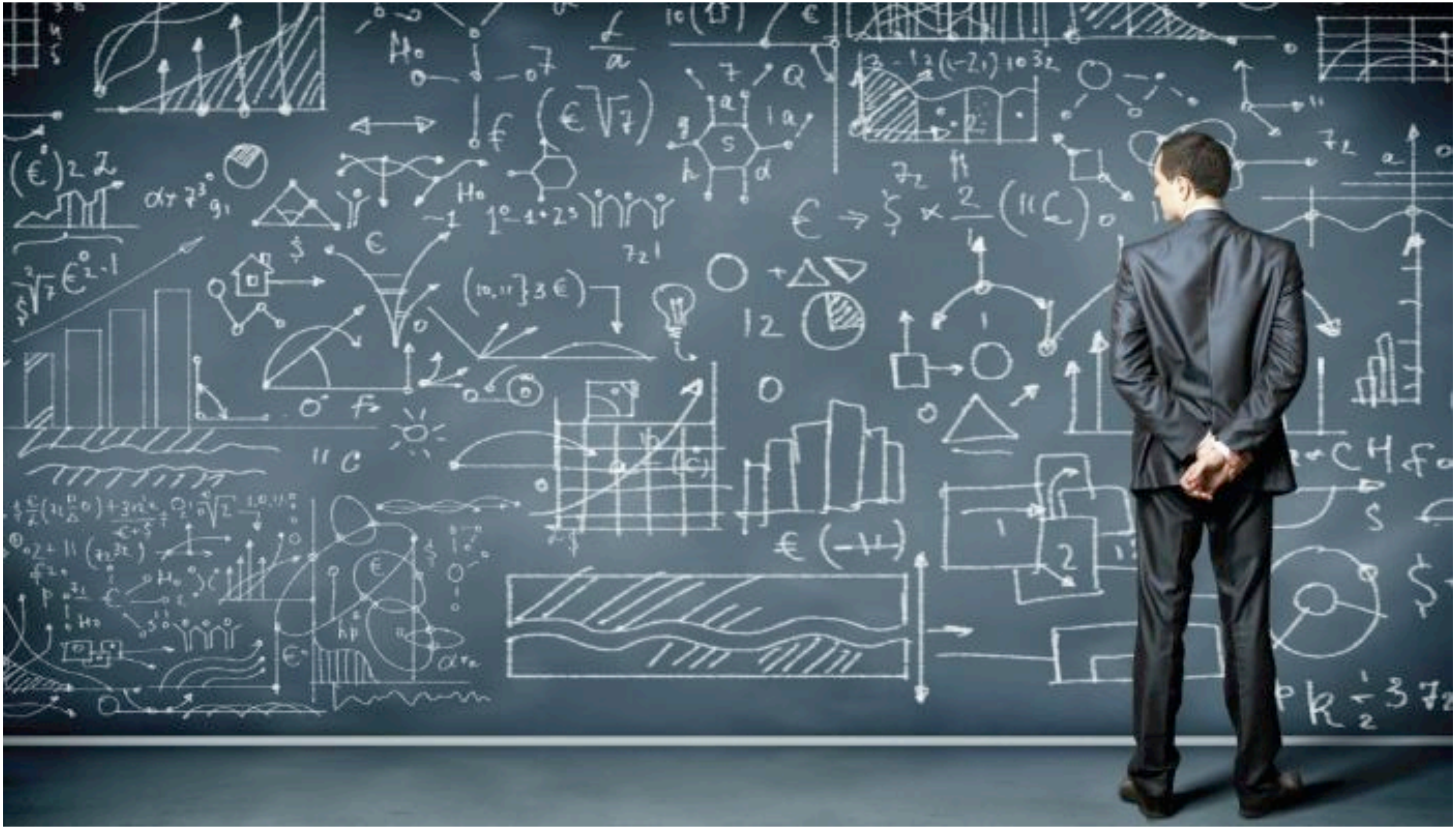

Logic Probability
live code scipy.stats
Probabilistic programming
live code lea
Probabilistic networks


Logic
Probability
live code
scipy.stats
Probabilistic
programming
live code
lea
Probabilistic
networks
live code
lea
PyPNL ?

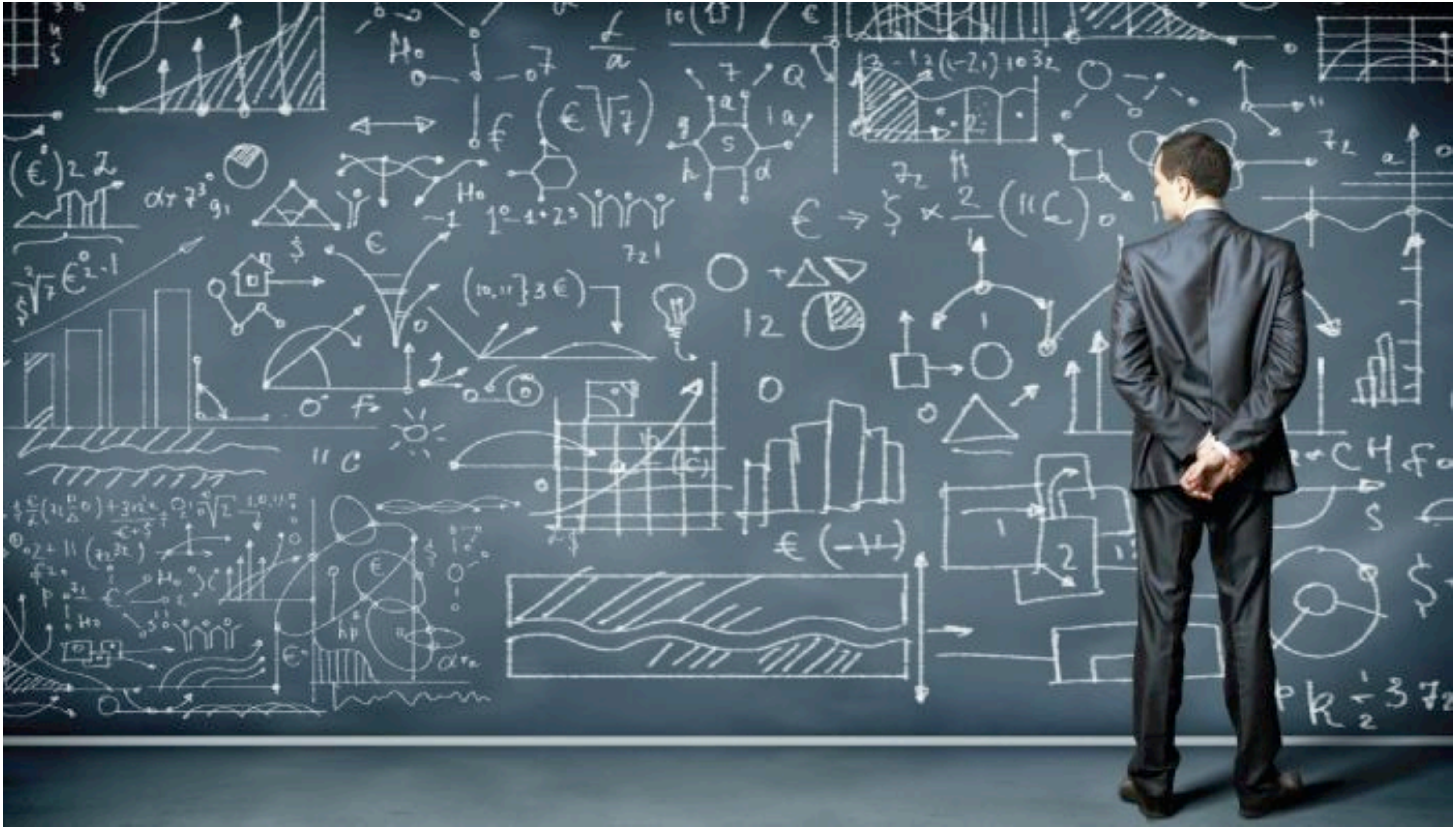
The purpose of computing is insight, not numbers.

The purpose of computing is insight, not numbers.

The purpose of computing is insight, not numbers.



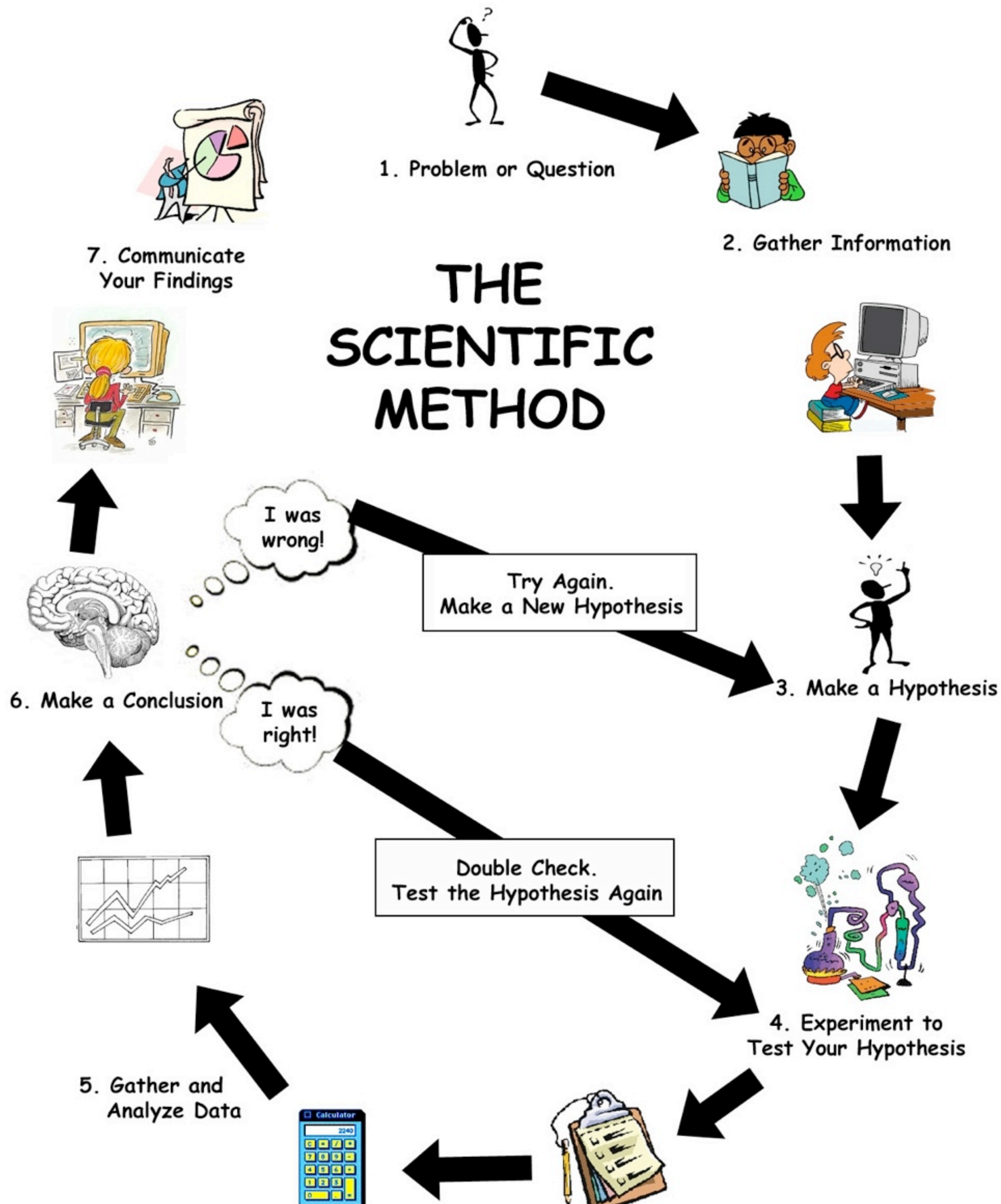
What is the purpose of data science ?

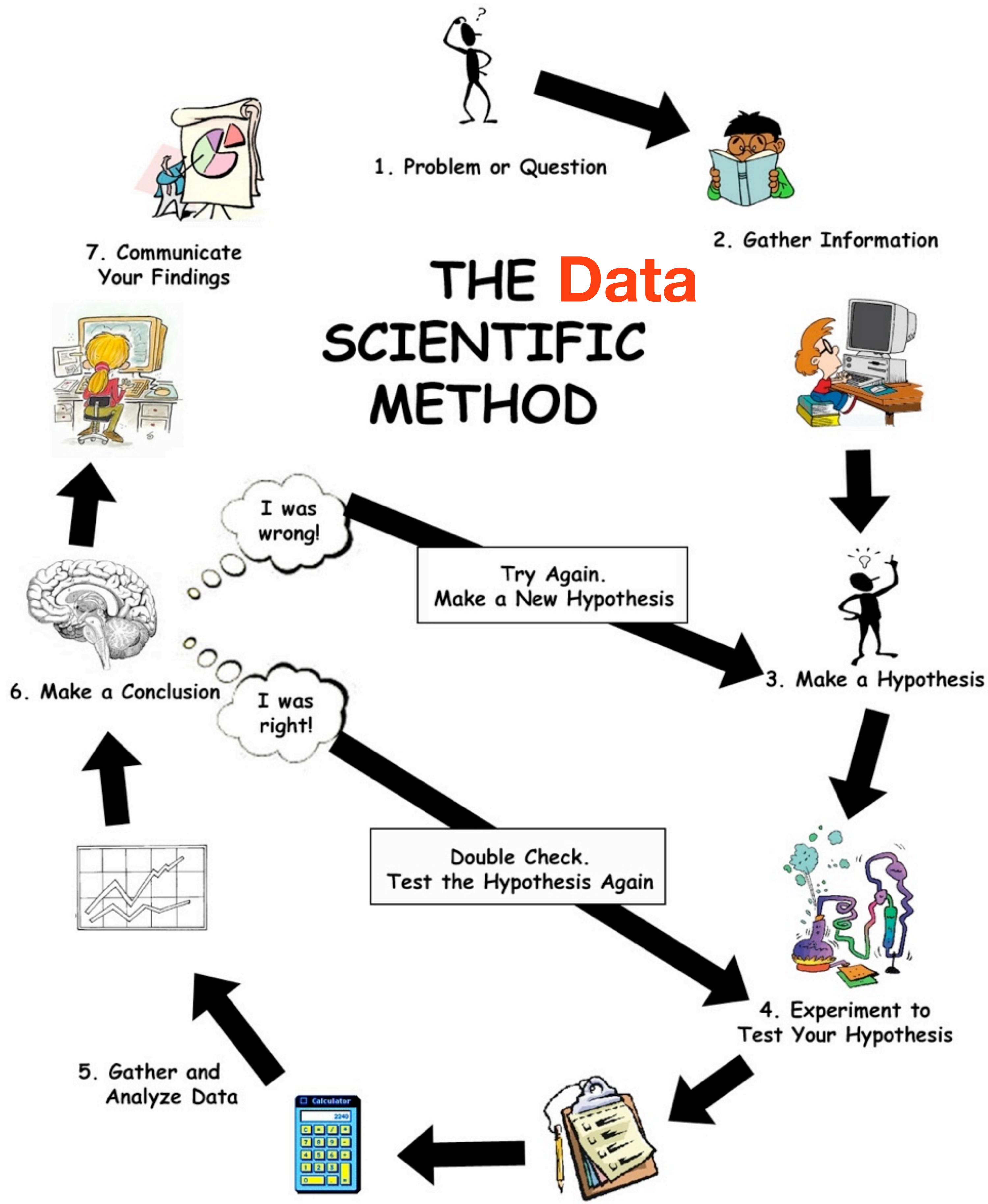


What is the purpose of data science ?

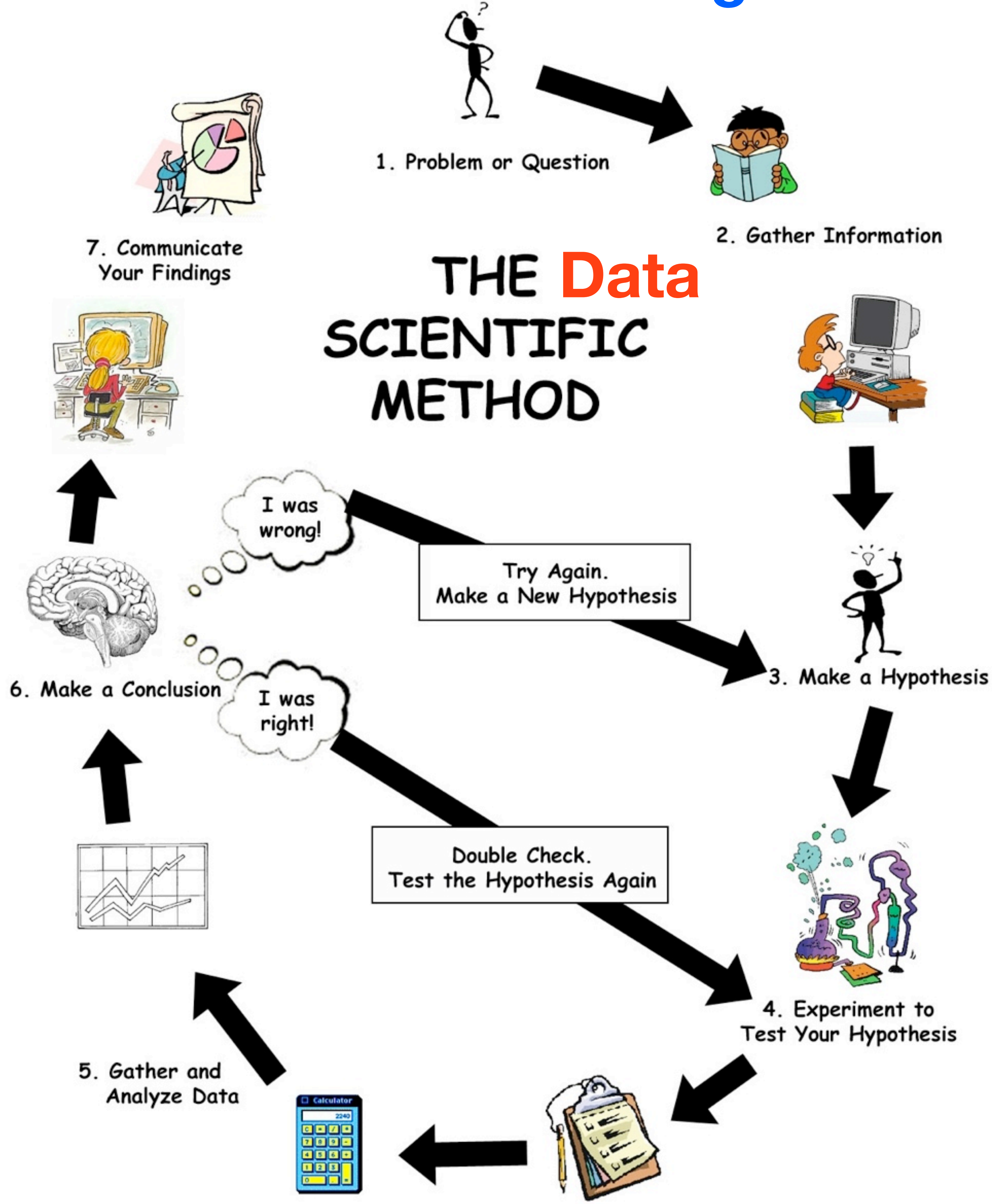
Insight, not numbers!

Data science


THE SCIENTIFIC METHOD
1. Problem or Question
2. Gather Information
3. Make a Hypothesis
4. Experiment to Test Your Hypothesis
5. Gather and Analyze Data
6. Make a Conclusion
7. Communicate Your Findings
I was wrong!
Try Again. Make a New Hypothesis
I was right!
Double Check. Test the Hypothesis Again
Calculator

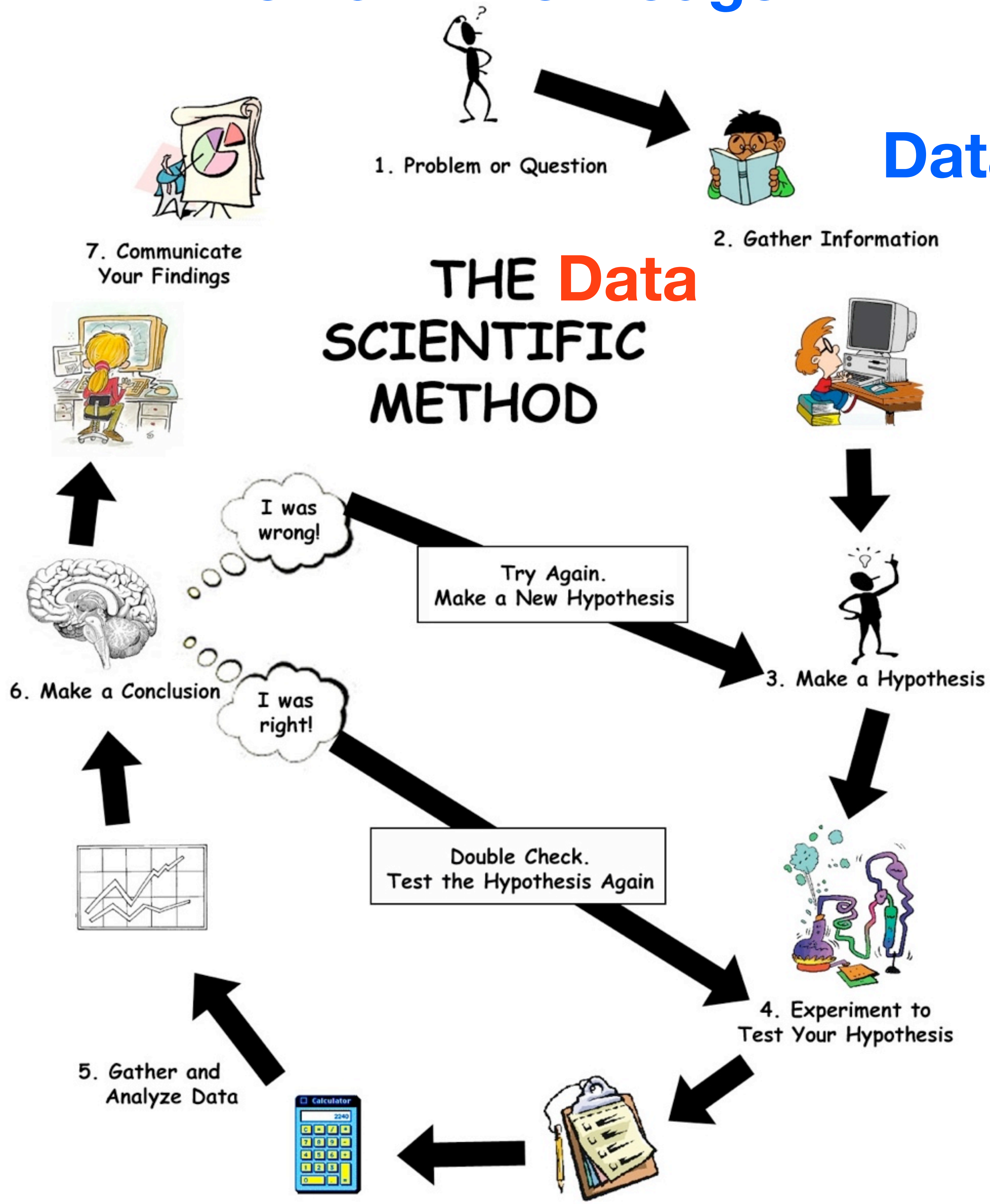

THE Data
SCIENTIFIC
METHOD
1. Problem or Question
2. Gather Information
3. Make a Hypothesis
4. Experiment to
Test Your Hypothesis
5. Gather and
Analyze Data
6. Make a Conclusion
7. Communicate
Your Findings
I was
wrong!
Try Again.
Make a New Hypothesis
I was
right!
Double Check.
Test the Hypothesis Again

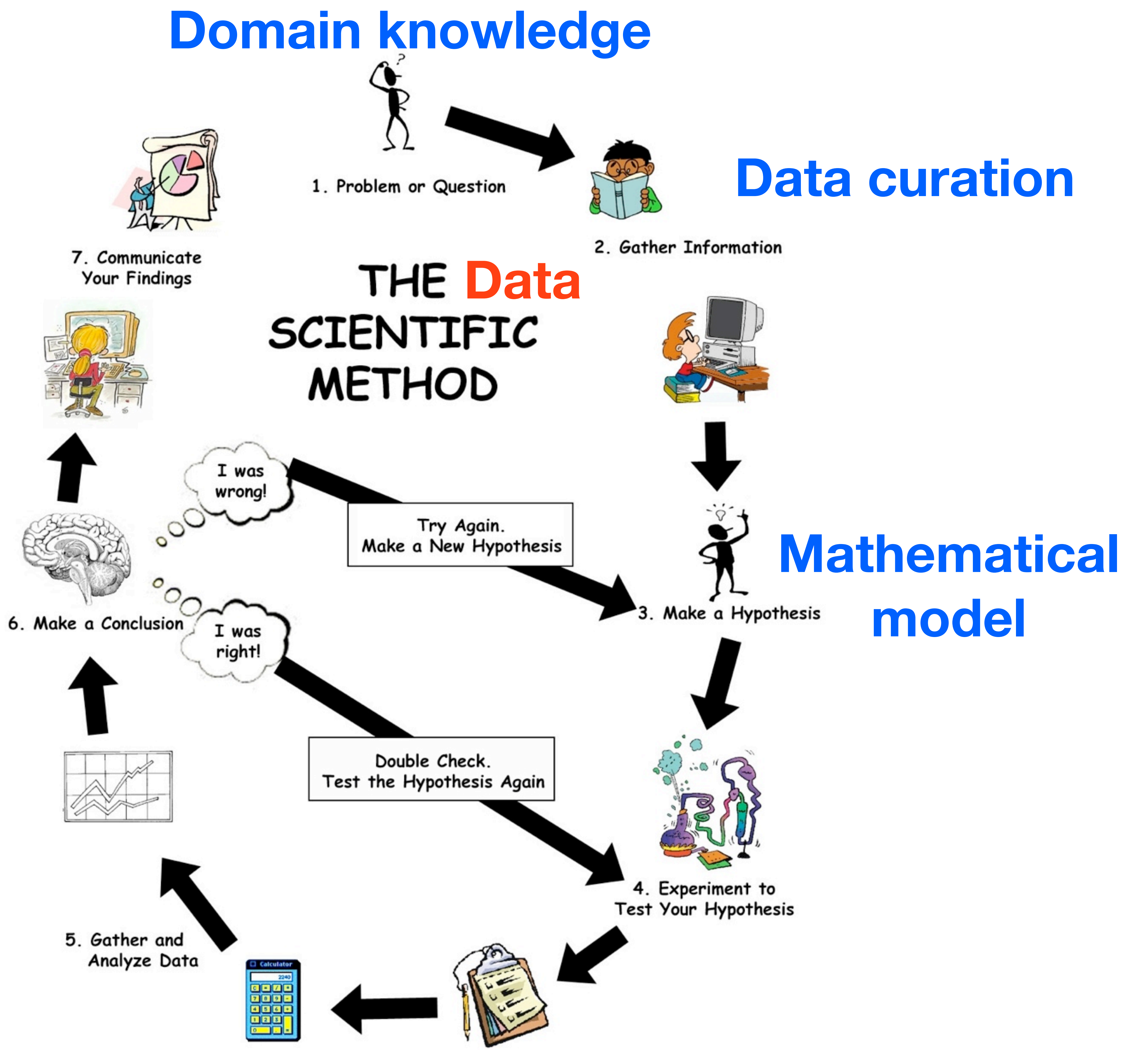
Domain knowledge



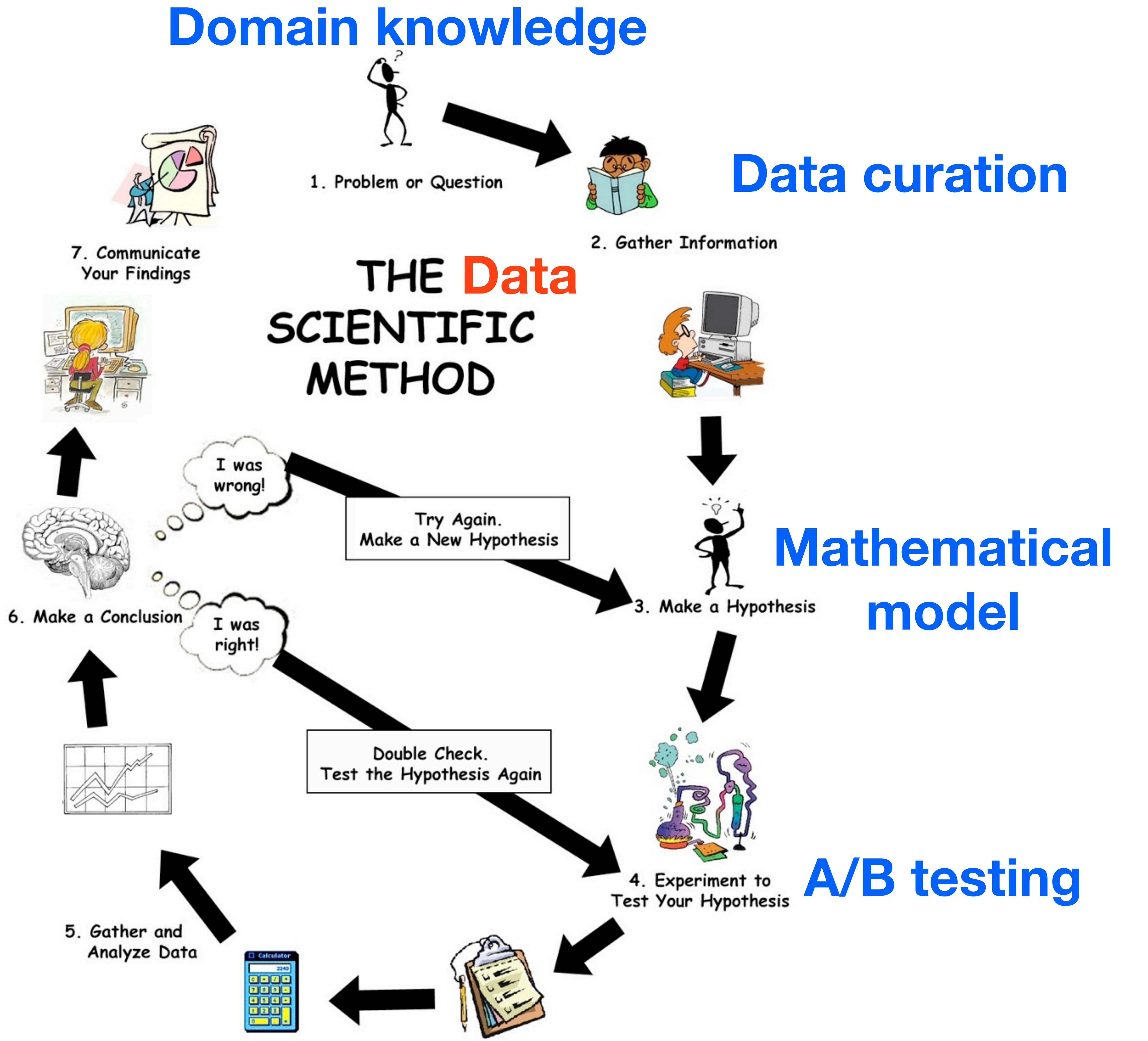
Domain knowledge

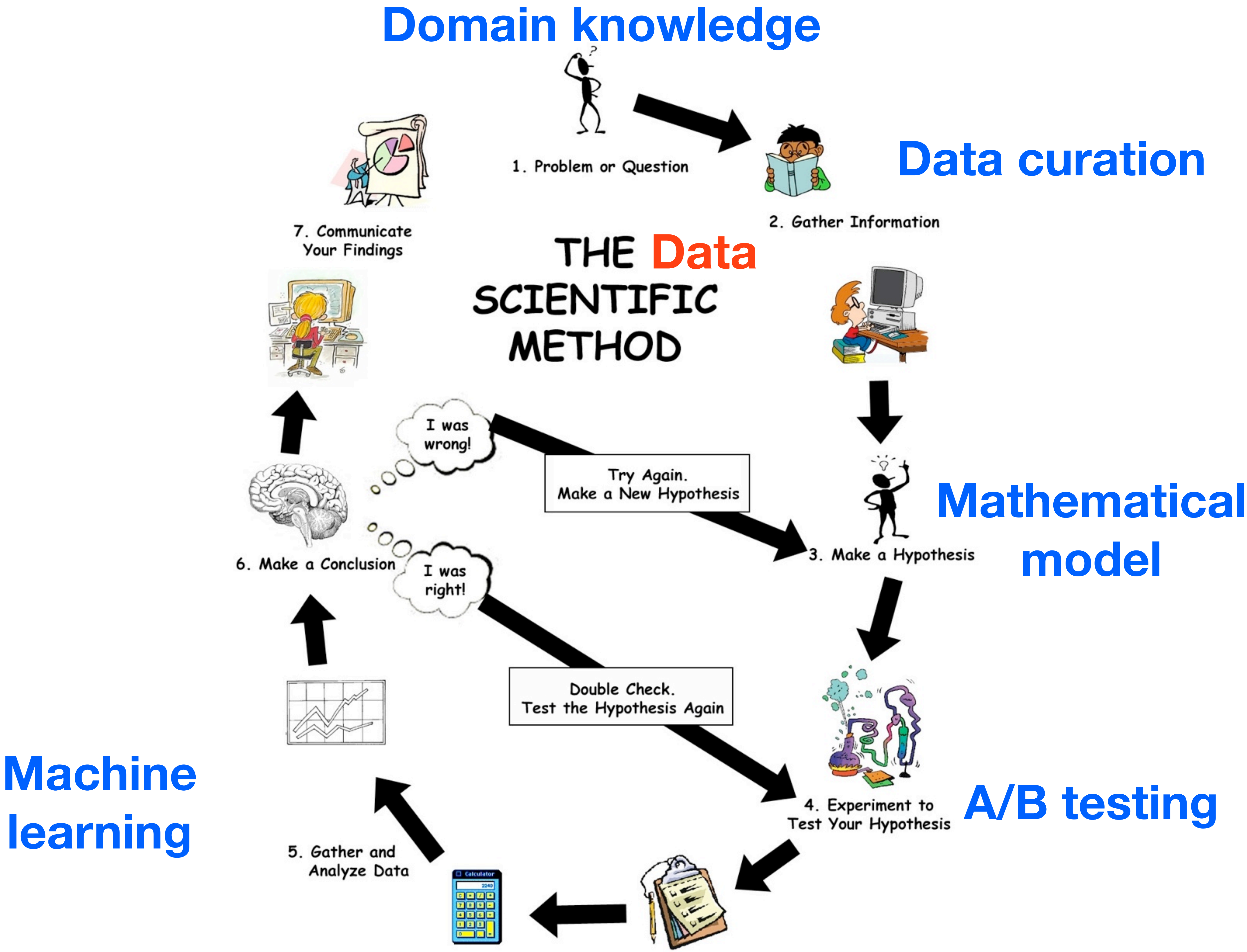
Data curation

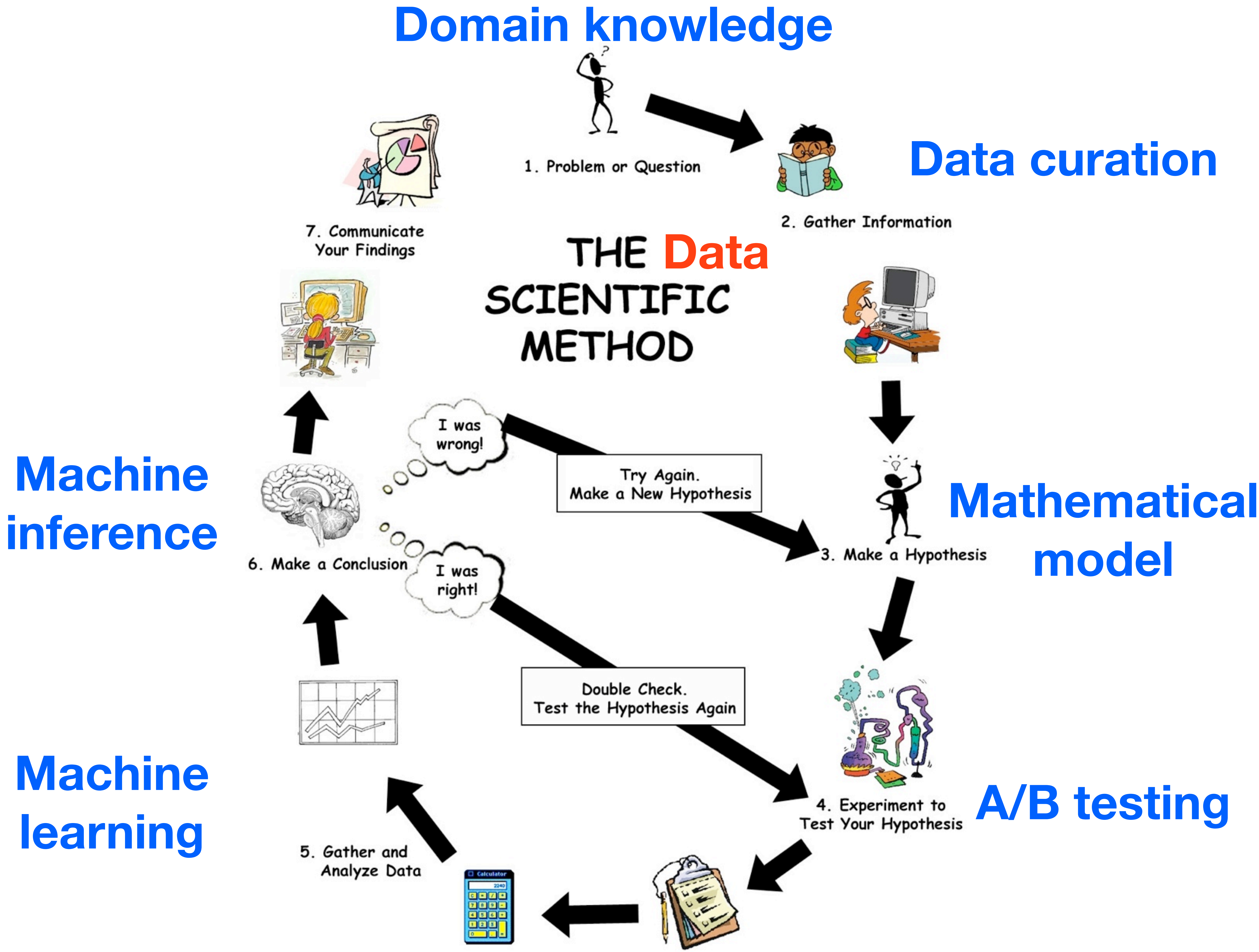


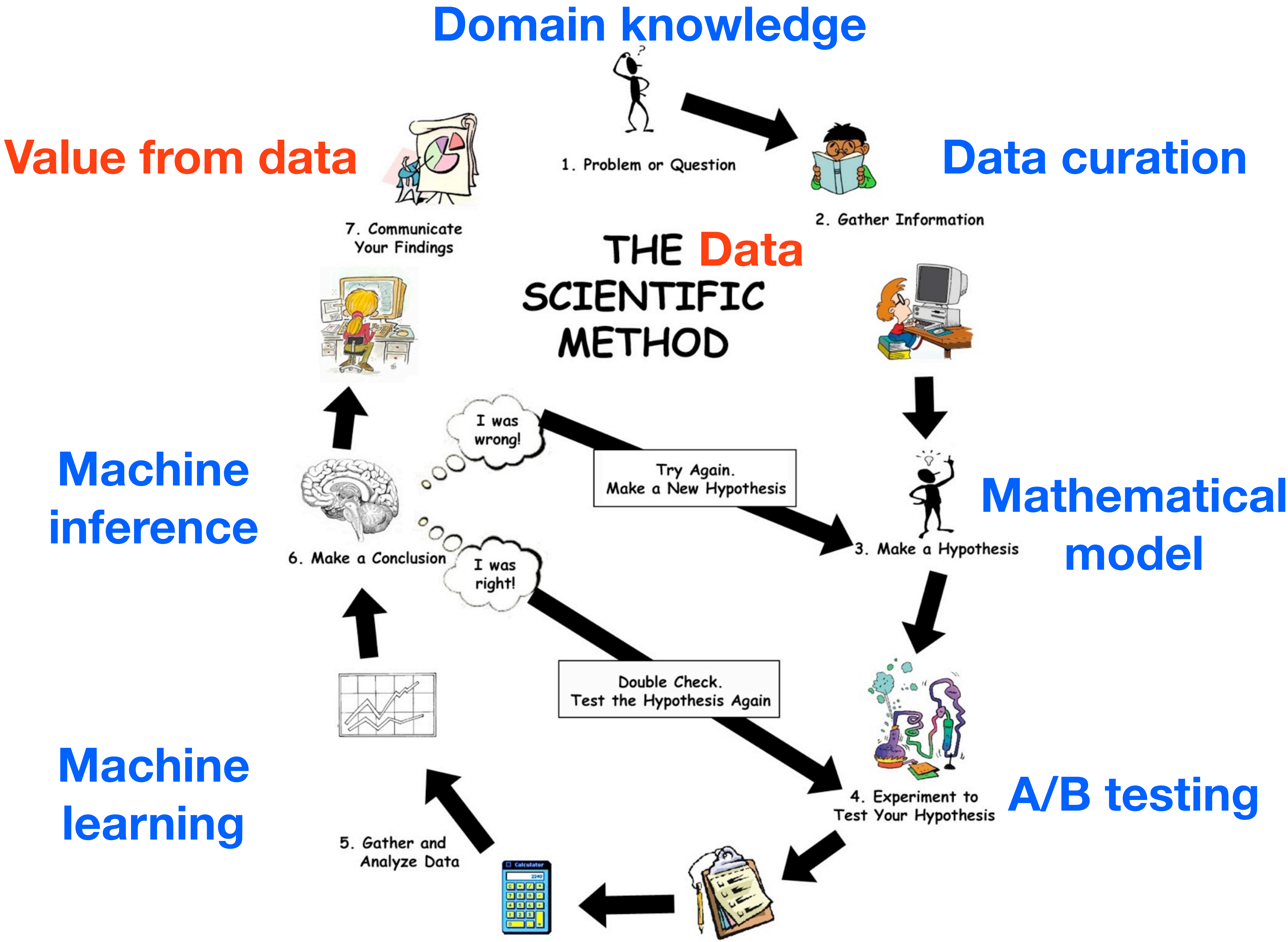


Domain knowledge
1. Problem or Question
Data curation
2. Gather Information
7. Communicate Your Findings
THE Data SCIENTIFIC METHOD
I was wrong!
Try Again. Make a New Hypothesis
Mathematical model
3. Make a Hypothesis
6. Make a Conclusion
I was right!
Double Check. Test the Hypothesis Again
4. Experiment to Test Your Hypothesis
5. Gather and Analyze Data


Domain knowledge
1. Problem or Question
2. Gather Information
Data curation
7. Communicate Your Findings
THE Data SCIENTIFIC METHOD
I was wrong!
Try Again. Make a New Hypothesis
3. Make a Hypothesis
Mathematical model
6. Make a Conclusion
I was right!
Double Check. Test the Hypothesis Again
4. Experiment to Test Your Hypothesis
A/B testing
5. Gather and Analyze Data

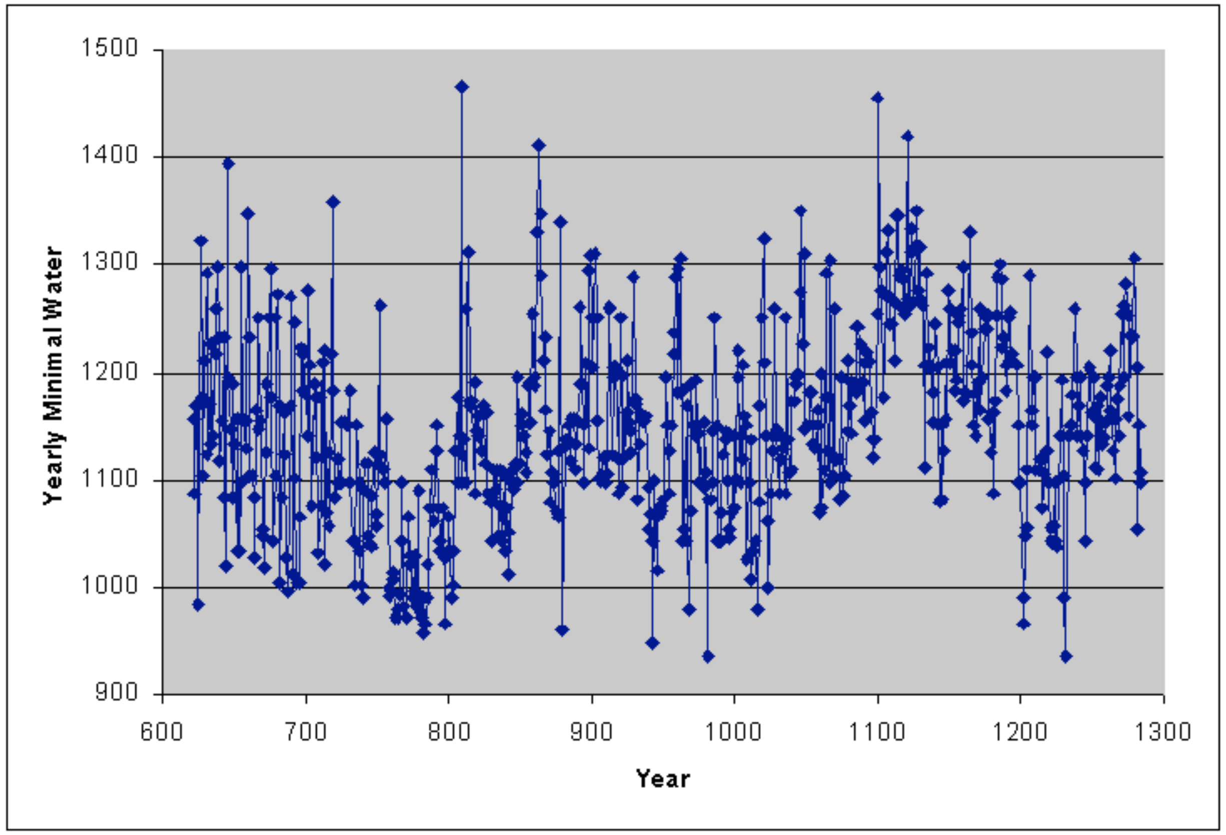

Domain knowledge
1. Problem or Question
2. Gather Information
Data curation
THE Data
SCIENTIFIC
METHOD
3. Make a Hypothesis
Mathematical
model
4. Experiment to
Test Your Hypothesis
A/B testing
5. Gather and
Analyze Data
Machine
learning
6. Make a Conclusion
I was
wrong!
Try Again.
Make a New Hypothesis
I was
right!
Double Check.
Test the Hypothesis Again
7. Communicate
Your Findings
Calculator
2240


Domain knowledge
Data curation
Mathematical model
A/B testing
Machine learning
Machine inference
THE Data SCIENTIFIC METHOD
1. Problem or Question
2. Gather Information
3. Make a Hypothesis
4. Experiment to Test Your Hypothesis
5. Gather and Analyze Data
6. Make a Conclusion
7. Communicate Your Findings
I was wrong!
I was right!
Try Again. Make a New Hypothesis
Double Check. Test the Hypothesis Again

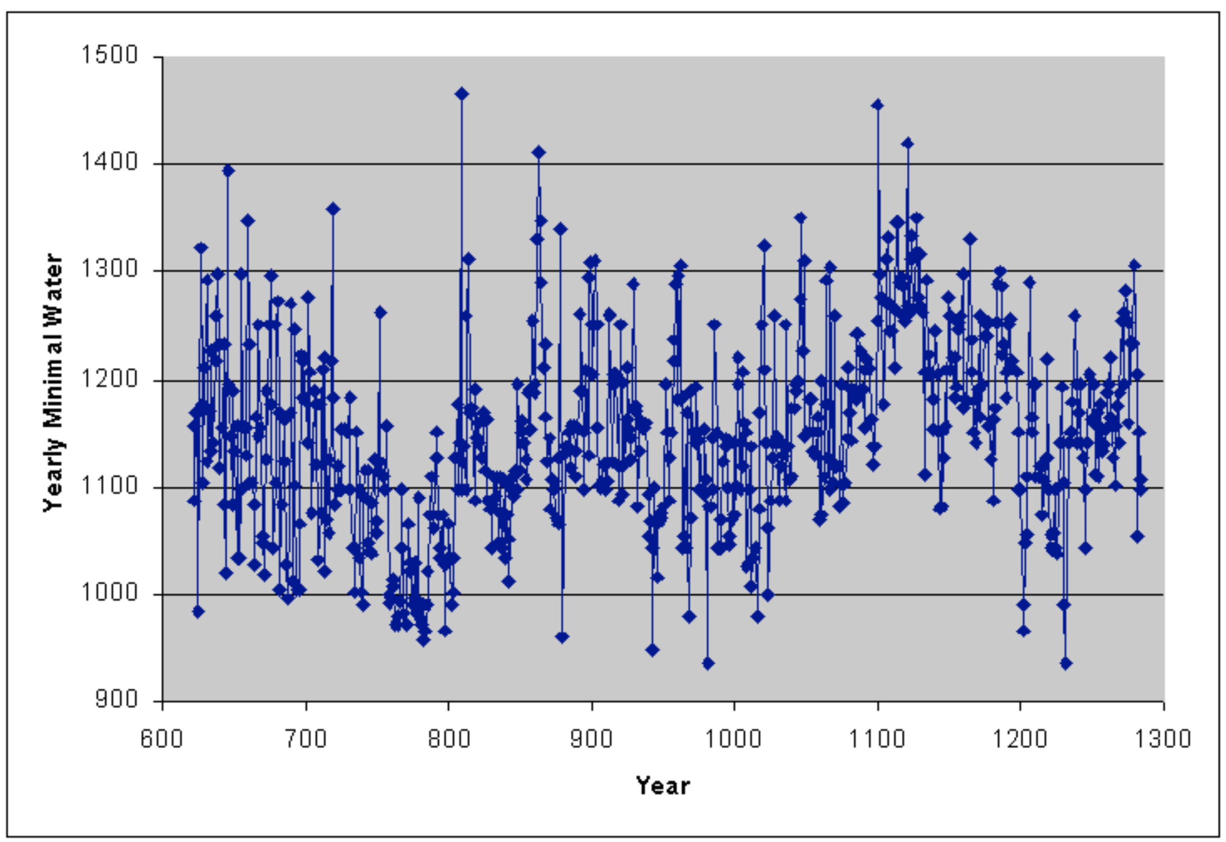

Domain knowledge
Value from data
Data curation
1. Problem or Question
2. Gather Information
7. Communicate Your Findings
THE Data SCIENTIFIC METHOD
I was wrong!
Try Again. Make a New Hypothesis
Machine inference
Mathematical model
3. Make a Hypothesis
6. Make a Conclusion
I was right!
Double Check. Test the Hypothesis Again
Machine learning
4. Experiment to Test Your Hypothesis
A/B testing
5. Gather and Analyze Data


© Patricia Felix

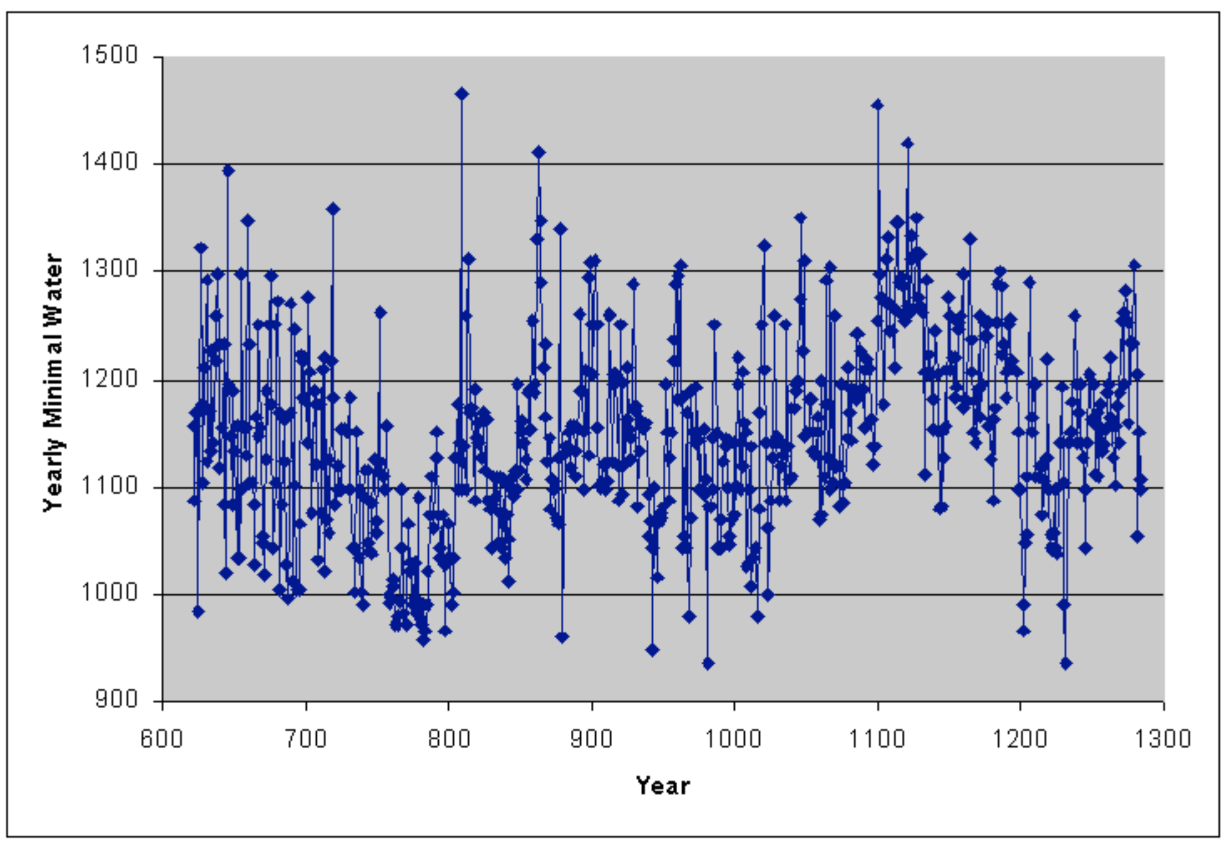

© Patricia Felix


1500
1400
1300
1200
1100
1000
900
Yearly Minimal Water
600
700
800
900
1000
1100
1200
1300
Year

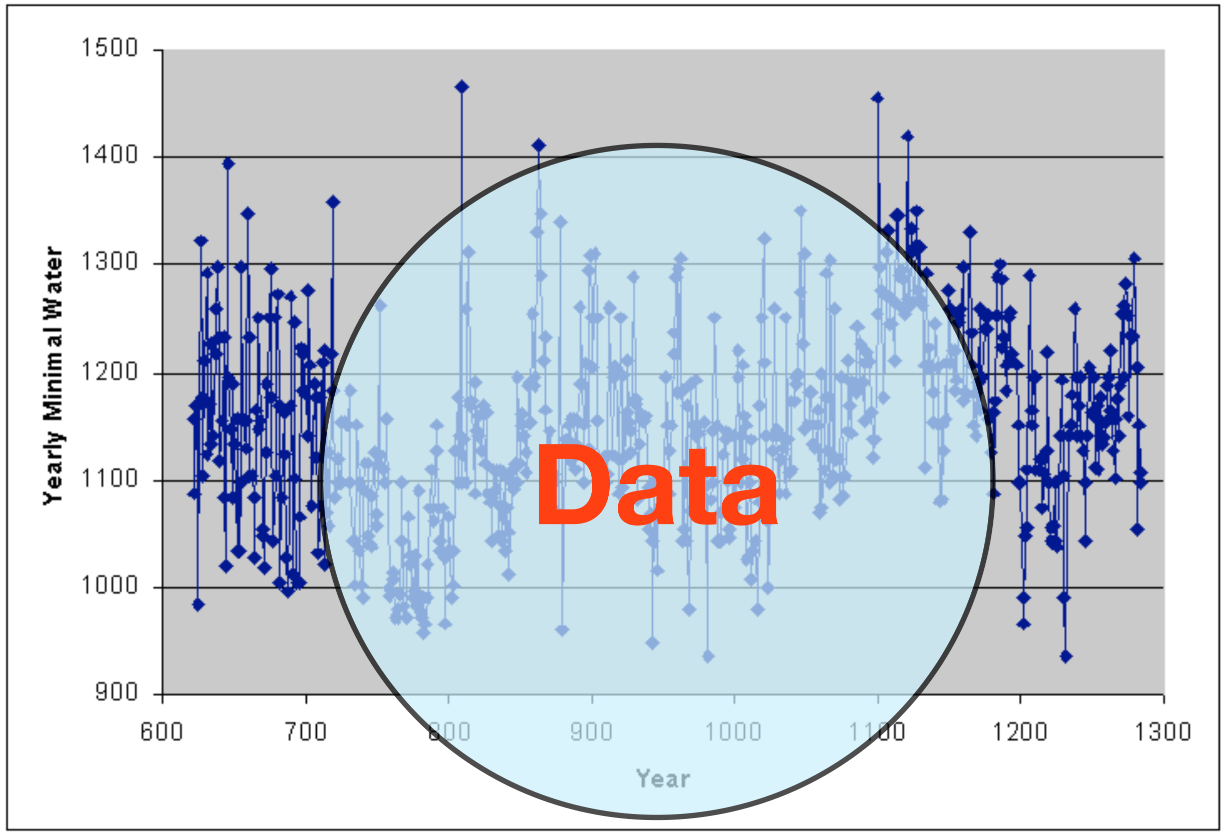

© Patricia Felix


1500
1400
1300
1200
1100
1000
900
Yearly Minimal Water
600
700
800
900
1000
1100
1200
1300
Year

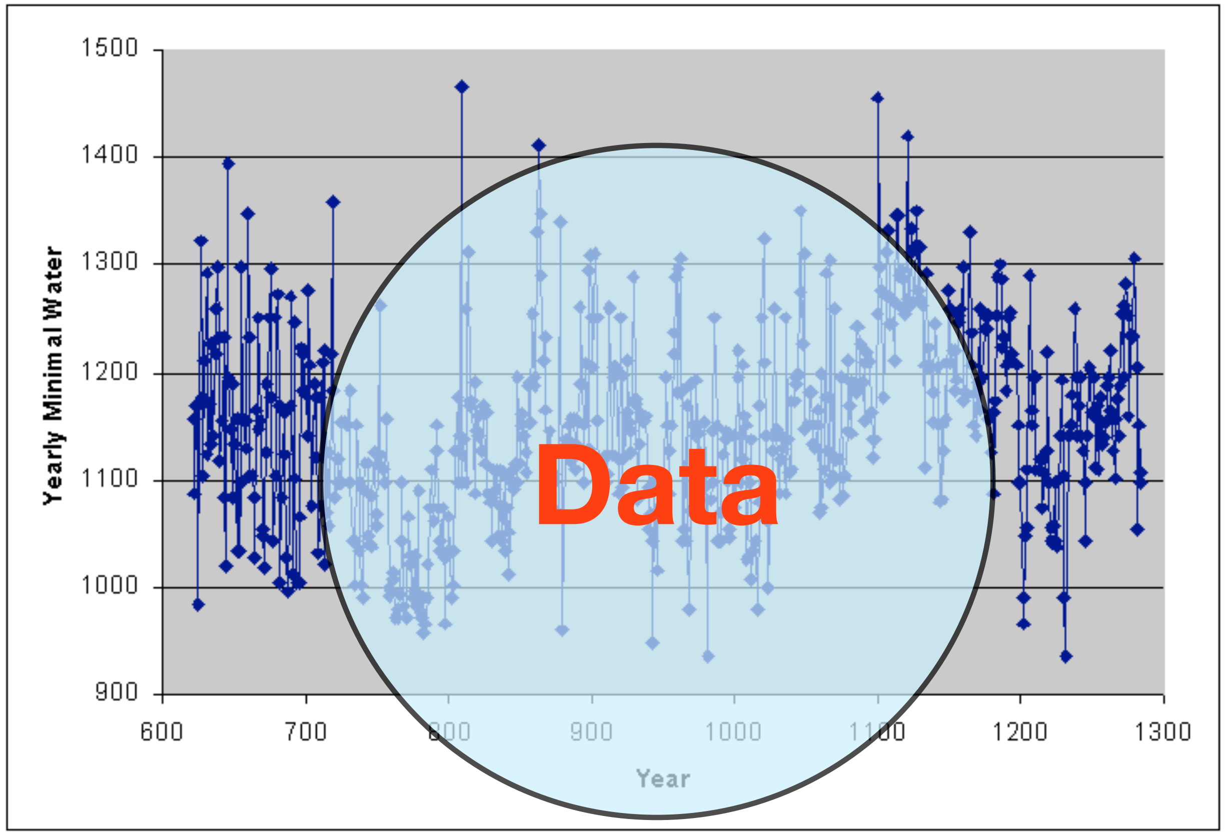

Question
© Patricia Felix

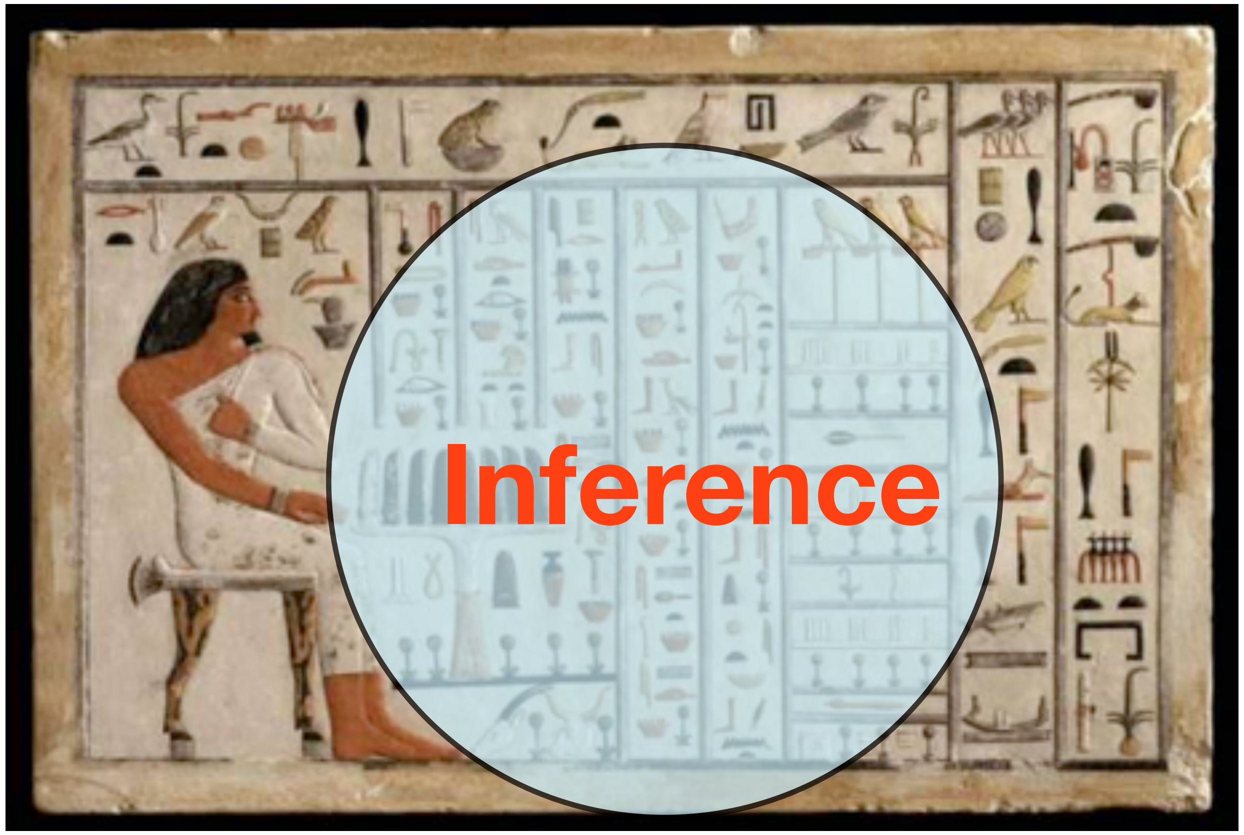

1500
1400
1300
1200
1100
1000
900
600
700
800
900
1000
1100
1200
1300
Yearly Minimal Water
Year

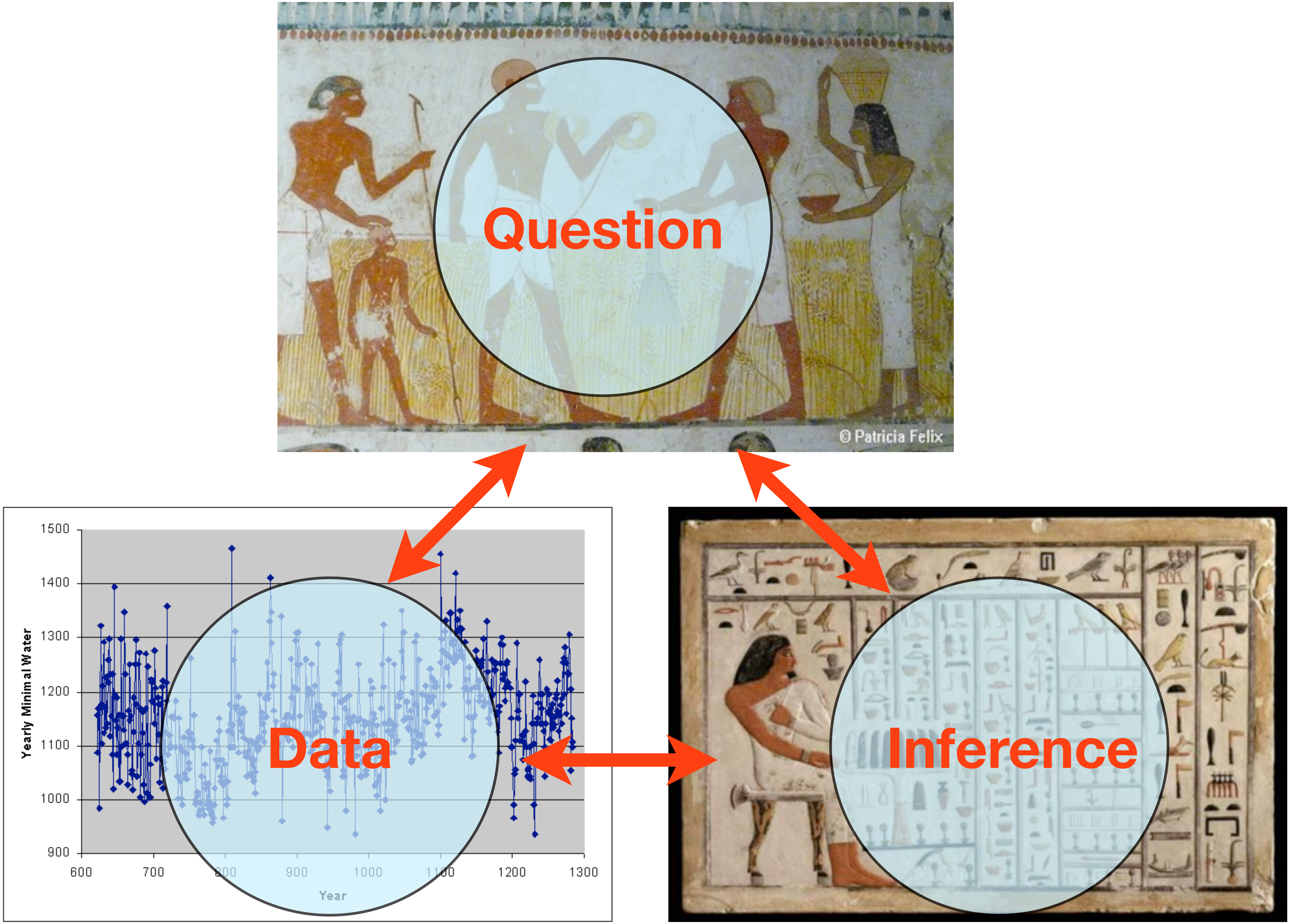

Question
© Patricia Felix

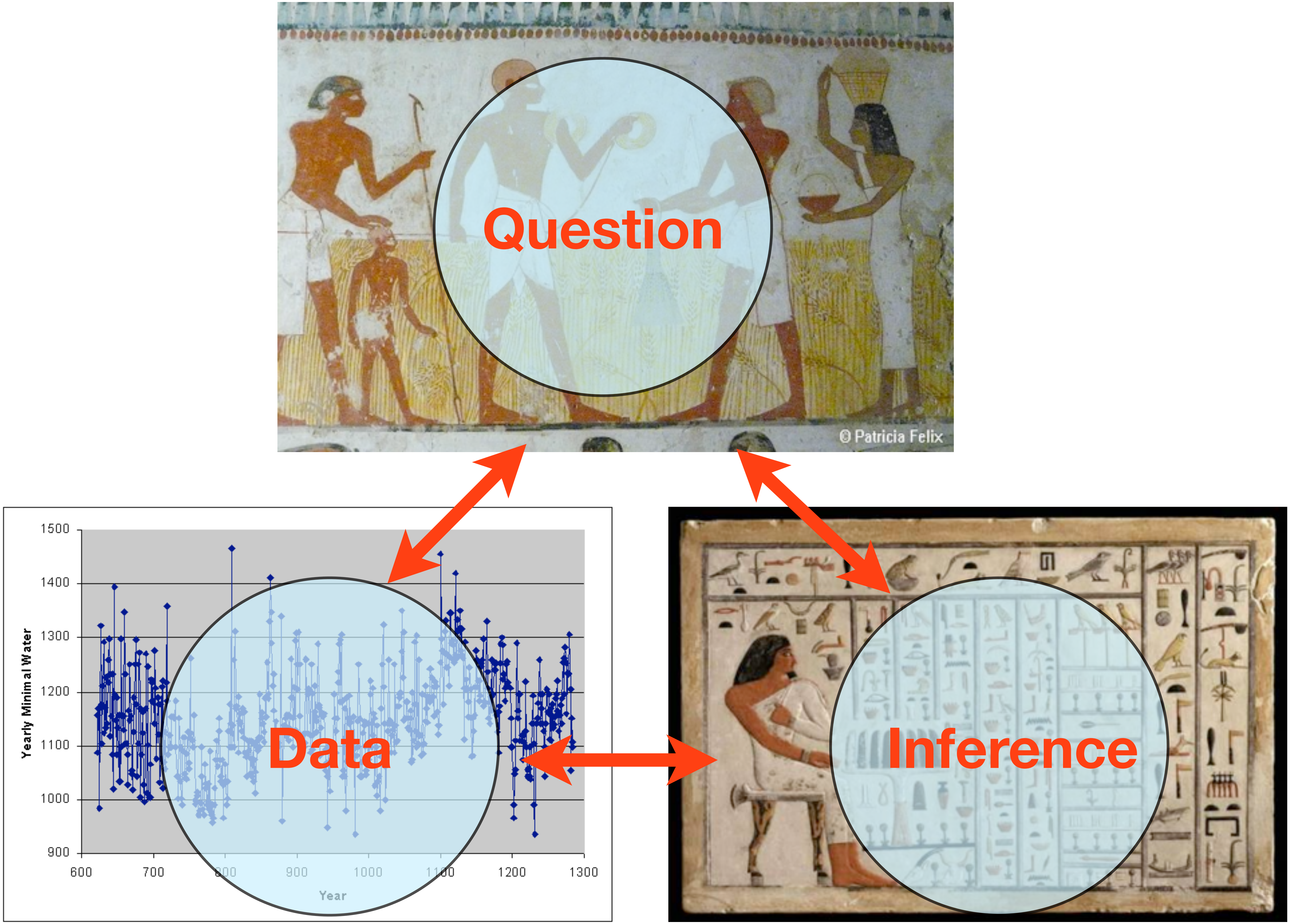

Yearly Minimal Water
Year
1500
1400
1300
1200
1100
1000
900
600
700
800
900
1000
1100
1200
1300
Data


Question
© Patricia Felix


Data
Yearly Minimal Water
Year
1500
1400
1300
1200
1100
1000
900
600
700
800
900
1000
1100
1200
1300


Inference


Question
© Patricia Felix
Data
Inference
Yearly Minimal Water
1500
1400
1300
1200
1100
1000
900
600
700
800
900
1000
1100
1200
1300
Year



Automated Inference ~ Machine Learning

Uncertainty

- how much will the Nile flood ?
- when will this equipment fail ?
- is this email spam ?
- is this applicant a good hire ?

Decisions

- should we invest in dams ?
- should we build redundancy ?
- should i delete without reading?
- should we look more ?

We need to make reasoned decisions in the face of uncertainty

Reasoning : Logic - Boolean algebra
Uncertainty : Chance, probability
Combine : Bayesian probability

Logic and probability

Logic and probability

- Reasoning as the basis for a science of data
- Reasoning under certainty and under uncertainty
- Boolean logic and probability
- Rules of probability theory
- Assigning probabilities - indifference and maximum entropy
- Inference and learning
- Is this a fair coin ? Elementary example of reasoning under uncertainty

The scientific method

The scientific method

The scientific method

"Now this is the peculiarity of scientific method, that when once it has become a habit of mind, that mind converts all facts whatsoever into science. The field of science is unlimited; its solid contents are endless, every group of natural phenomena, every phase of social life, every stage of past or present development is material for science. The unity of all science consists alone in its method, not in its material.

The man who classifies facts of any kind whatever, who sees their mutual relation and describes their sequence, is applying the scientific method and is a man of science. The facts may belong to the past history of mankind, to the social statistics of our great cities, to the atmosphere of the most distant stars, to the digestive organs of a worm, or to the life of a scarcely visible bacillus. It is not the facts themselves which form science, but the method in which they are dealt with."

The scientific method

The

Grammar of Science

KARL PEARSON, M.A., F.R.S.

PART I.—PHYSICAL

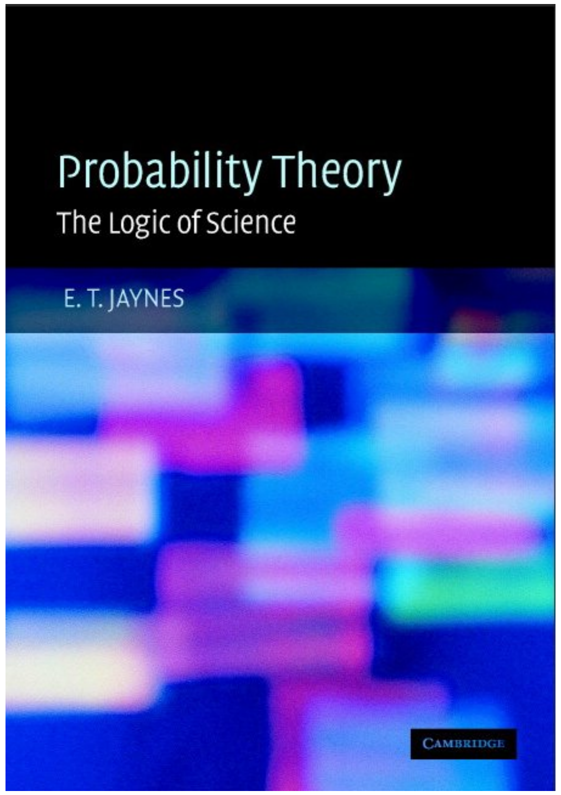
LONDON

ADAM AND CHARLES BLACK

“Now this is the peculiarity of scientific method, that when once it has become a habit of mind, that mind converts all facts whatsoever into science. The field of science is unlimited; its solid contents are endless, every group of natural phenomena, every phase of social life, every stage of past or present development is material for science. The unity of all science consists alone in its method, not in its material.

The man who classifies facts of any kind whatever, who sees their mutual relation and describes their sequence, is applying the scientific method and is a man of science. The facts may belong to the past history of mankind, to the social statistics of our great cities, to the atmosphere of the most distant stars, to the digestive organs of a worm, or to the life of a scarcely visible bacillus. It is not the facts themselves which form science, but the method in which they are dealt with.”

The scientific method

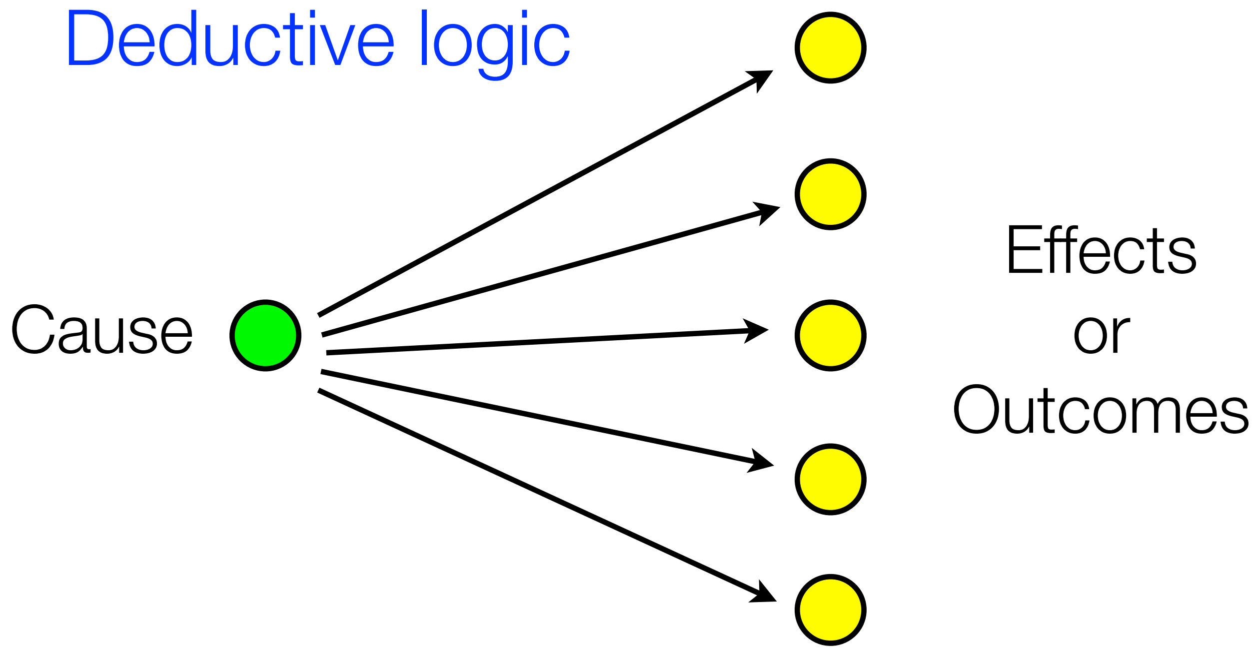


"Now this is the peculiarity of scientific method, that when once it has become a habit of mind, that mind converts all facts whatsoever into science. The field of science is unlimited; its solid contents are endless, every group of natural phenomena, every phase of social life, every stage of past or present development is material for science. The unity of all science consists alone in its method, not in its material.

The man who classifies facts of any kind whatever, who sees their mutual relation and describes their sequence, is applying the scientific method and is a man of science. The facts may belong to the past history of mankind, to the social statistics of our great cities, to the atmosphere of the most distant stars, to the digestive organs of a worm, or to the life of a scarcely visible bacillus. It is not the facts themselves which form science, but the method in which they are dealt with."

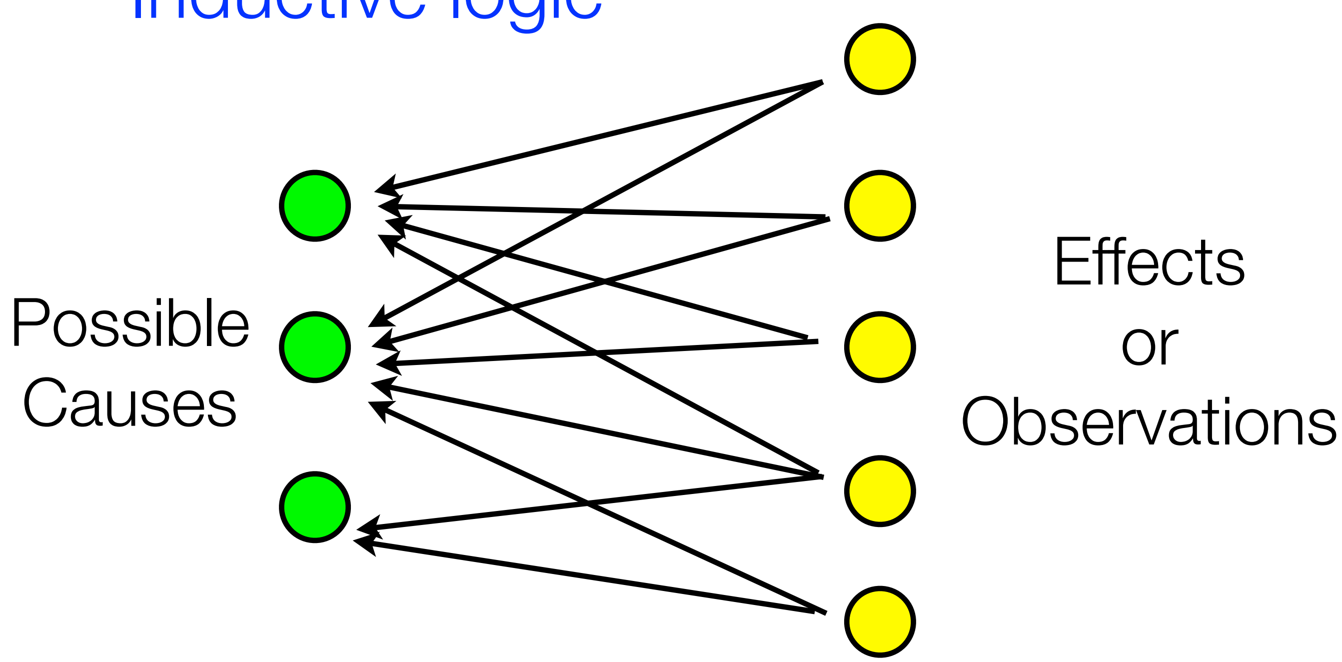
Reasoning in the scientific method

Deductive logic



Boolean algebra

Inductive logic



Bayesian probability

Boolean algebra : reasoning under certainty

- Formalization of Aristotelian logic
- Propositions : are either TRUE or FALSE
- Operations : conjunction (AND), disjunction (OR), negation (NOT)
- Laws : algebraic identities between compound propositions
- Ex.1 : NOT(A AND B) = (NOT A) OR (NOT B)
- Ex. 2 : NOT(A OR B) = (NOT A) AND (NOT B)
- Rules for reasoning consistently with certain propositions.

Probability : reasoning under uncertainty

- Generalization of Boolean logic
- Propositions have a truth value p, with p = 0 (FALSE) and p = 1 (TRUE)
- Operations : conjunction (AND), disjunction (OR), negation (NOT)
- sum rule : P(A) + P(NOT A) = 1
- product rule : P(A AND B) = P(A|B)P(B) = P(B|A) P(A)
- => P(A OR B) = P(A) + P(B) - P(A AND B)
- independent => P(A|B) = P(A) ; mutually exclusive => P(A OR B) = P(A) + P(B)

Assigning probabilities

- Probabilities are ALWAYS conditioned on information P(A) = P(A | I)

- Consider a set of propositions A1, A2, ... An, that are exhaustive and mutually exclusive. In the absence of any other information, the principle of indifference says that P(Ai) = 1/N (Laplace)

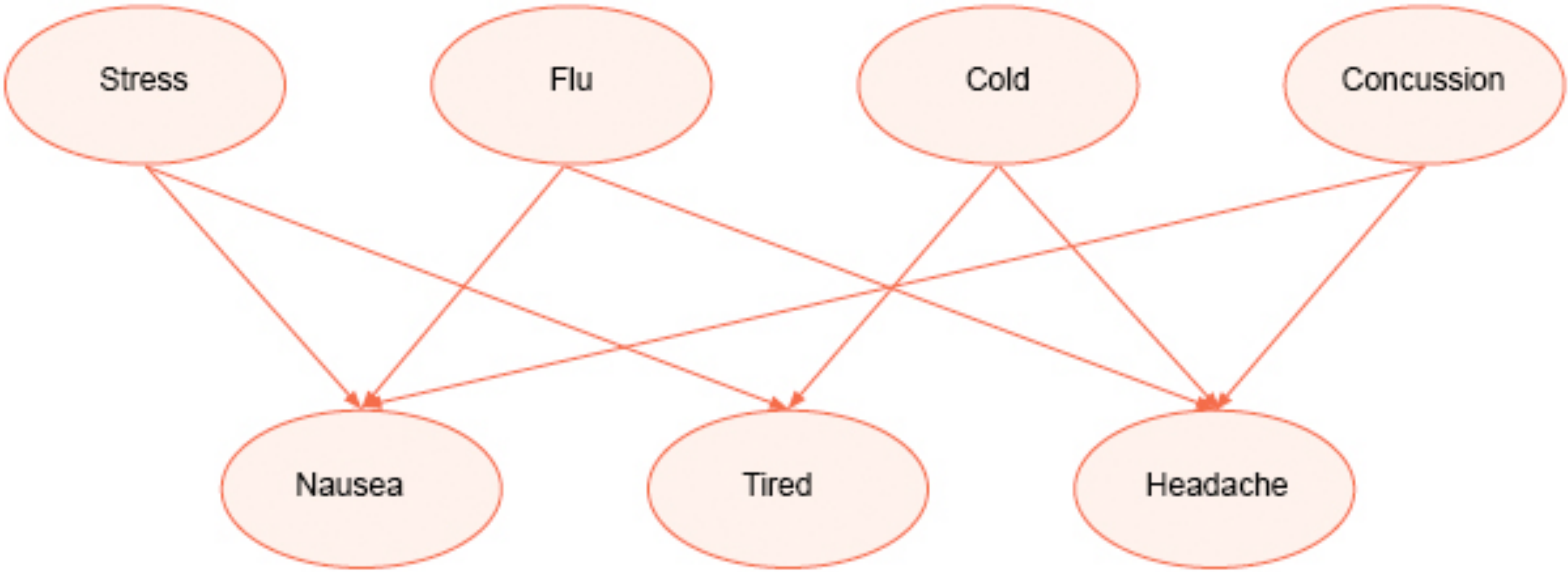
- When additional information is available, probabilities are assigned taking the additional information into account. The principle of maximum entropy says that P should be assigned by maximizing \sum P_i log P_i, subject to the constraints that derive from the additional information.

- Maximum entropy reduces to indifference when there are no constraints.

Bayes theorem

- P(A and B) = P(A|B) P(B) = P(B|A)P(A)
- P(A|B) = P(B|A) P(A) / P(B)
- Looks trivial but is extremely deep!
- P(disease | symptom) = want to know this.
- P(symptom | disease) = can estimate this (even empirically!)
- P(disease | symptom) = P(symptom | disease) P (disease) / P(symptom)

Probabilistic networks



P(disease | symptoms) = diagnosis

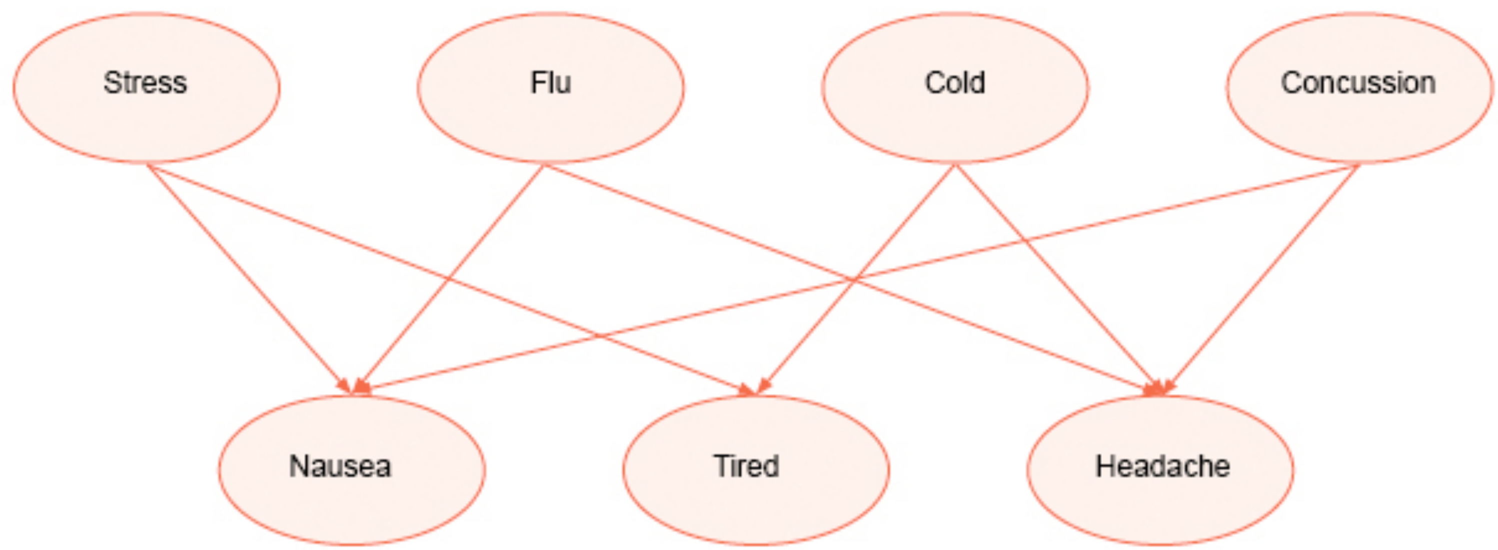
RILACS

Representation

Inference

Learning

Actions



Bayesian inference : is this a fair coin ?

Bayesian inference : is this a fair coin ?

A coin is thrown 6 times with the outcome $D = \{hhhhtt\}$

Bayesian inference : is this a fair coin ?

A coin is thrown 6 times with the outcome $D = \{hhhhtt\}$

What is the probability of the proposition $H = \text{The coin is fair}$

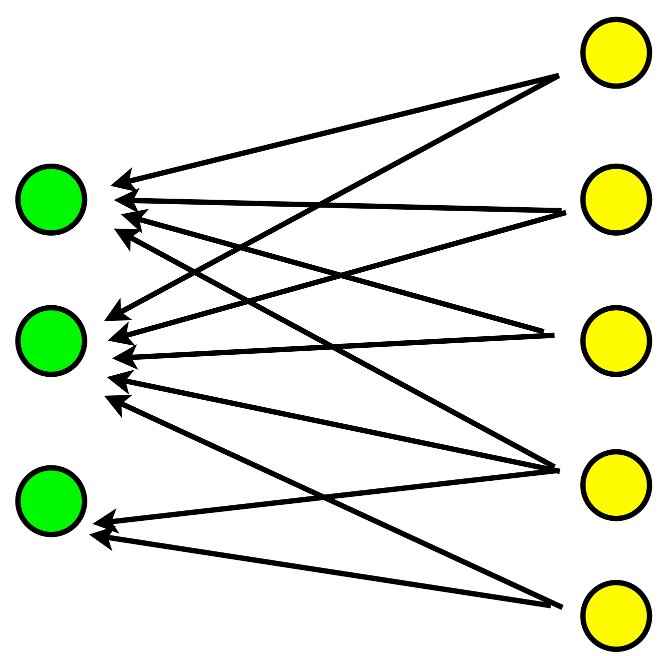
Bayesian inference : is this a fair coin ?

A coin is thrown 6 times with the outcome $D = \{hhhhtt\}$

What is the probability of the proposition $H =$ The coin is fair

$H =$ The coin is fair

$\bar{H} =$ The coin is not fair



$D = \{hhhhtt\}$

Possible "causes"

Effects or Observations

Bernoulli trials

$$P(D|H) = \text{Probability of D } \textit{given} \text{ H}$$
$$= \text{Probability of outcome } \textit{given} \text{ a fair coin}$$

$$P(h|H) = P(t|H) = \frac{1}{2}$$

$$P(D|H) = P(hhhhtt|H) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$$

$$P(h|\bar{H}) = \theta, P(t|\bar{H}) = 1 - \theta$$

$$P(D|\bar{H}) = P(hhhhtt|\bar{H}) = \theta \cdot \theta \cdot \theta \cdot \theta \cdot (1-\theta) \cdot (1-\theta)$$

Bernoulli trials

$$P(D|\theta) = P(hhhhtt|\theta) = \theta \cdot \theta \cdot \theta \cdot \theta \cdot (1-\theta) \cdot (1-\theta)$$

$$P(\theta|D) = P(\theta|hhhhtt)$$

$$= \text{Probability of the hypothesis } \theta \textit{ given } \text{the data}$$

$$H \longrightarrow \theta = \frac{1}{2}$$

$$\bar{H} \longrightarrow \theta \neq \frac{1}{2}$$

$P(D|\theta)$ know this

$P(\theta|D)$ want this

Bernoulli trials

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

Bernoulli trials

$$P(n_1|\theta, N) = \frac{N!}{n_1!(N-n_1)!}\theta^{n_1}(1-\theta)^{N-n_1}$$

P(D|H) - likelihood

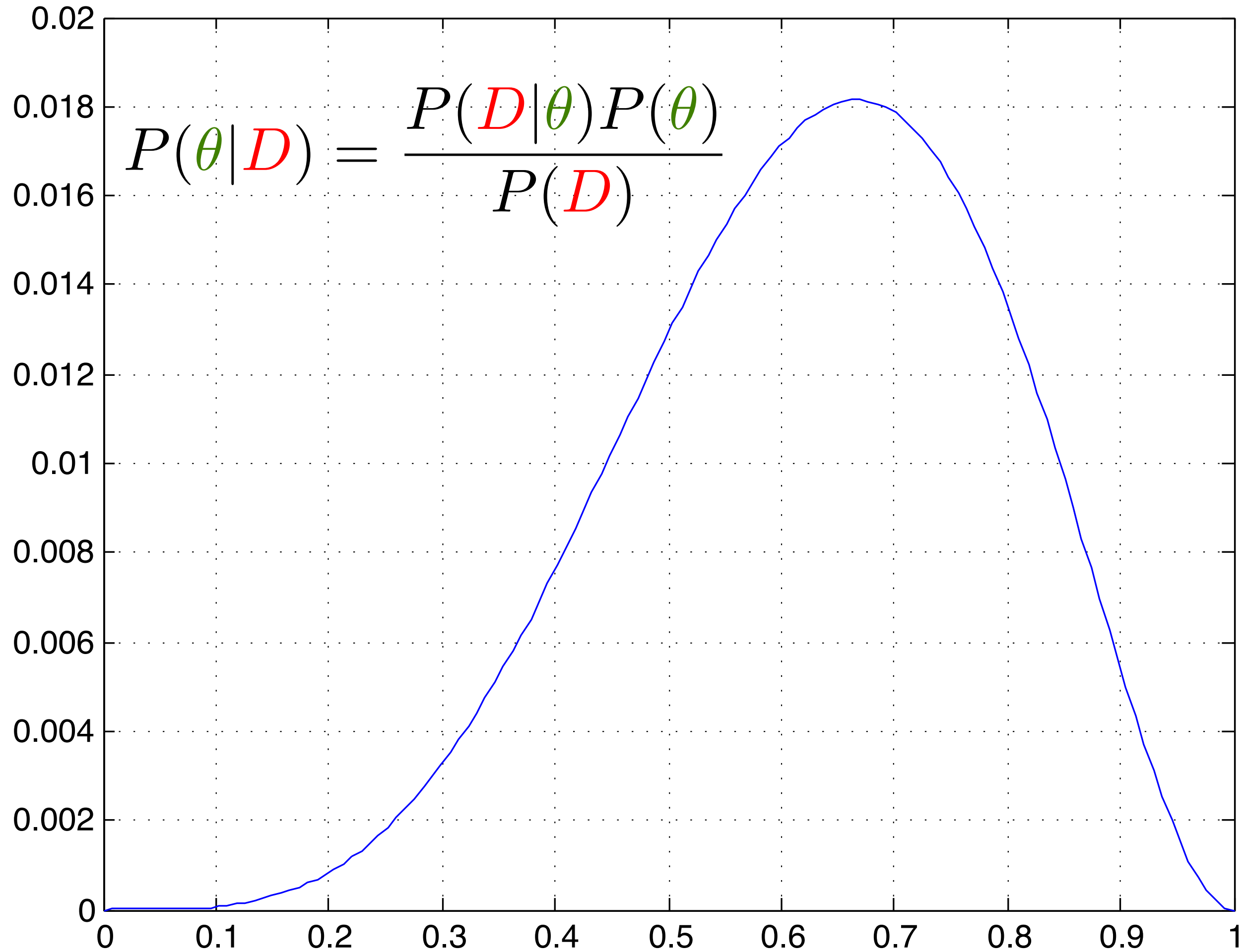
$$P(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$$

P(H) = prior $\langle\theta\rangle = \frac{a}{a+b}$

$$P(\theta|n_1, N) \sim \theta^{n_1+a-1}(1-\theta)^{n-n_1+b-1}$$

P(H|D) = posterior

Bernoulli trials



Posterior distribution with uniform prior

Live coding 1

Probabilistic programming

Let the data speak for themselves!

Let the data speak for themselves!

The data *cannot* speak for themselves; and they never have, in any real problem of inference.

Classification

Clustering

Regression

Dimensionality reduction

Classification

predict class, given attributes

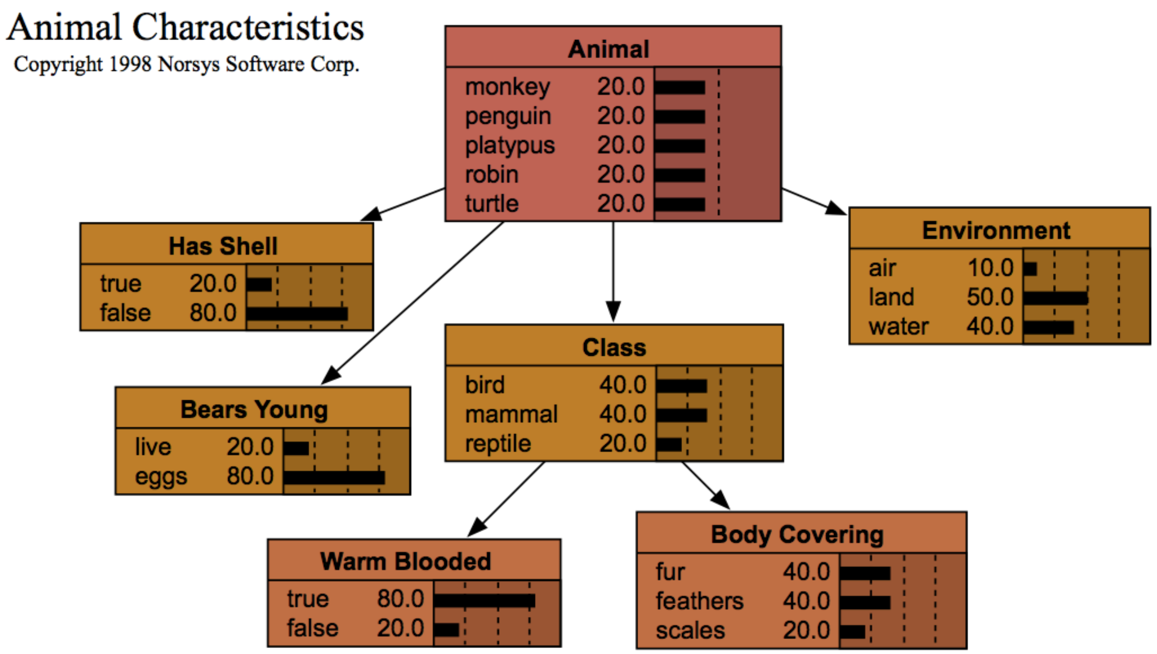
Clustering

Regression

Dimensionality reduction

Classification

predict class, given attributes



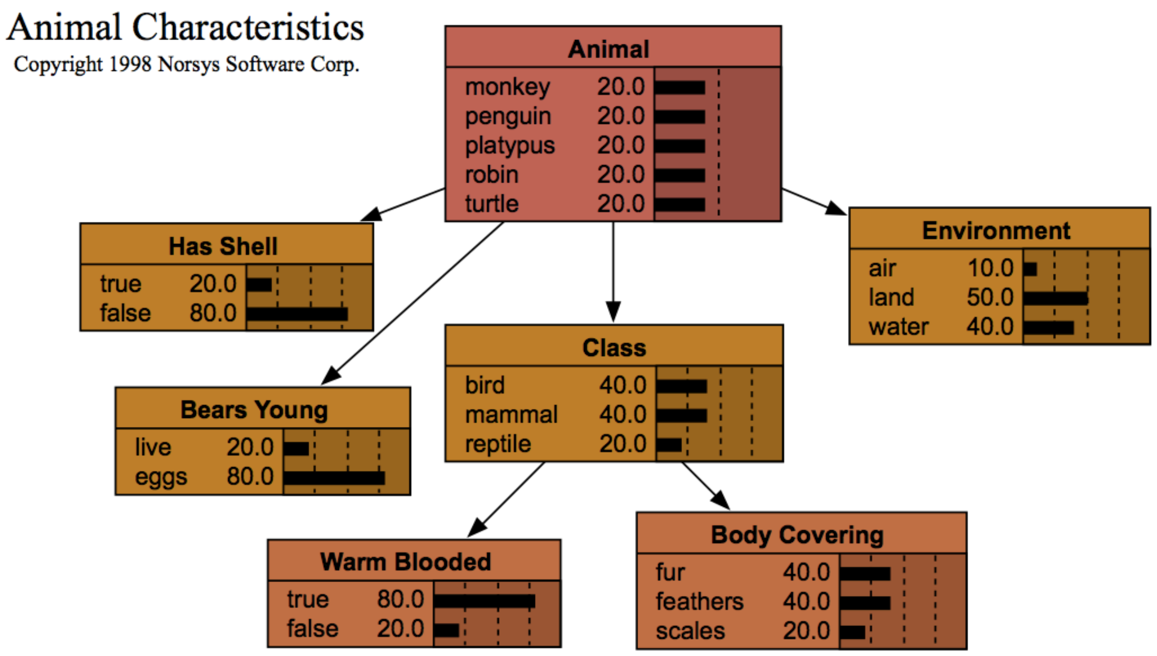
Clustering

Regression

Dimensionality reduction

Classification

predict class, given attributes



Clustering

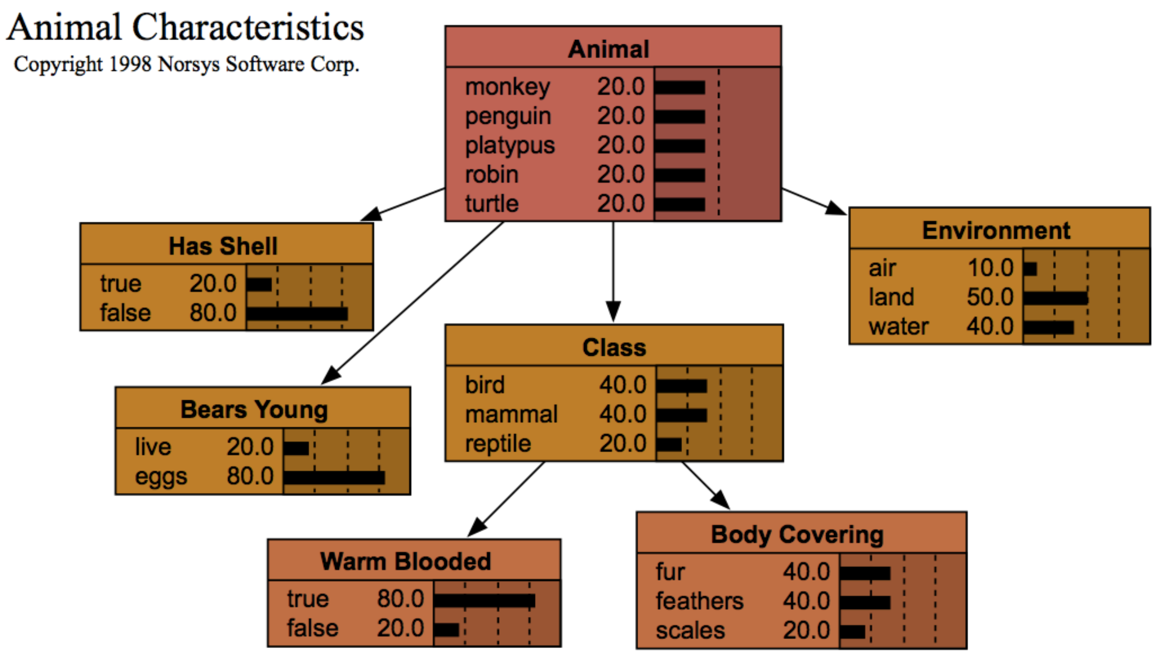
Regression

predict values, given other values

Dimensionality reduction

Classification

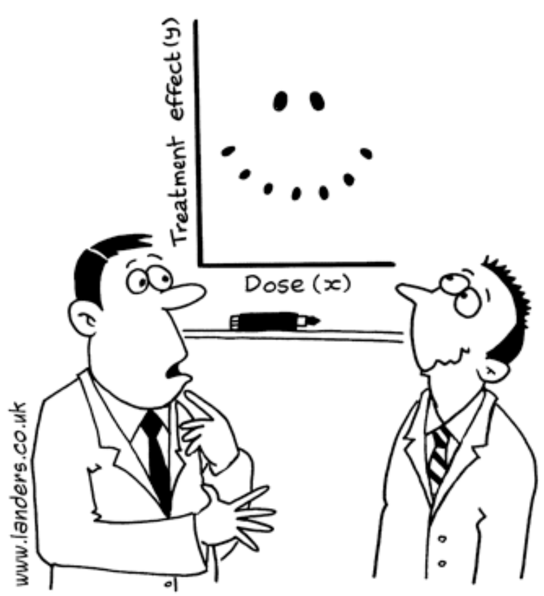
predict class, given attributes



Clustering

Regression

predict values, given other values

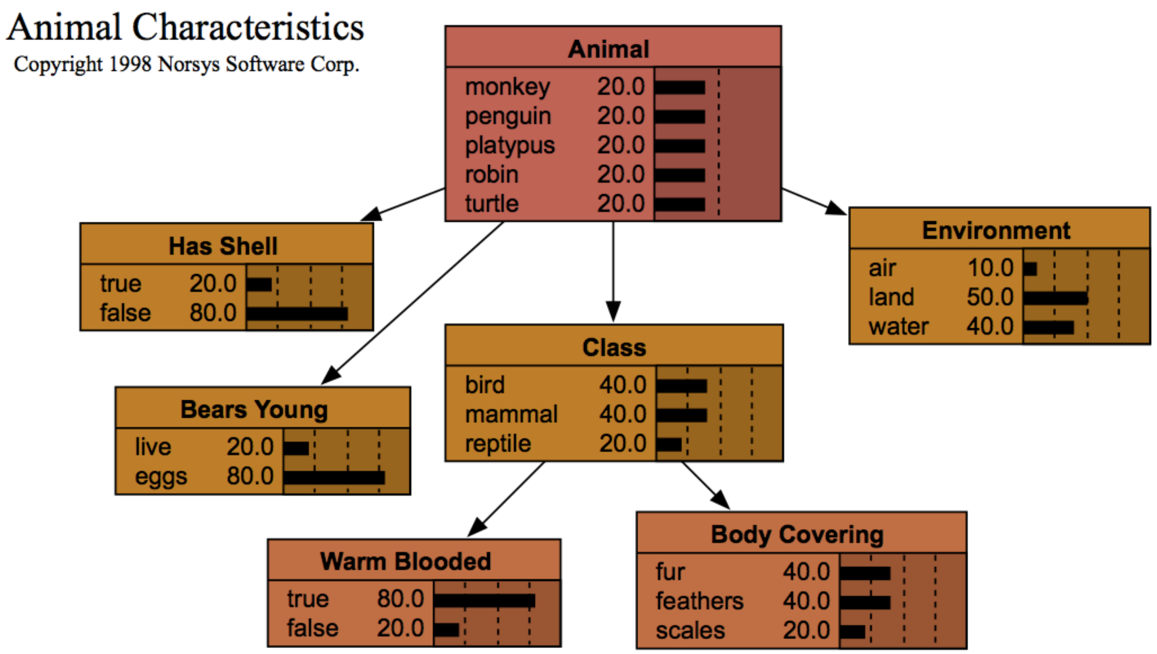


"It's a non-linear pattern with outliers.....but for some reason I'm very happy with the data."

Dimensionality reduction

Classification

predict class, given attributes

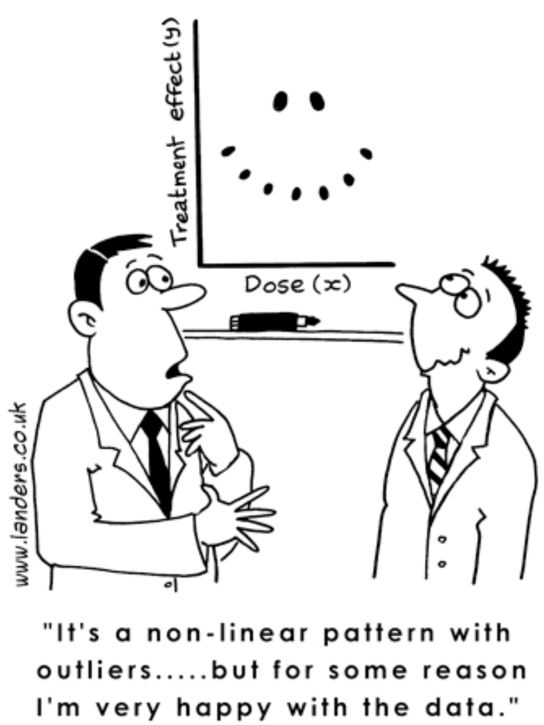


Clustering

group similar things together

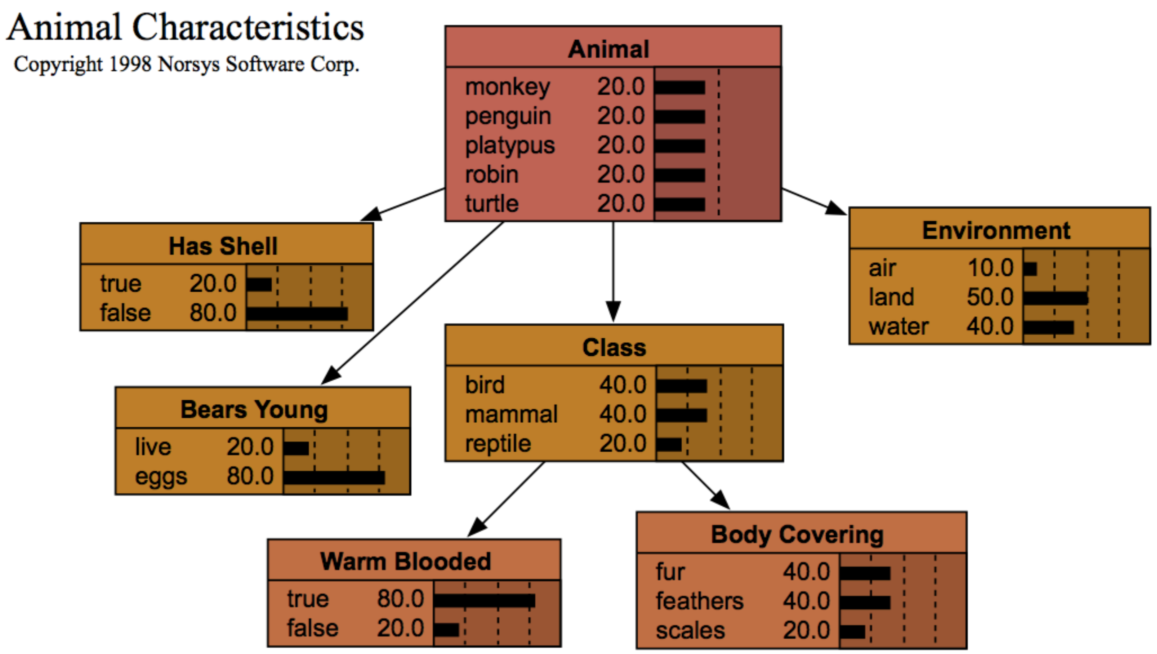
Regression

predict values, given other values

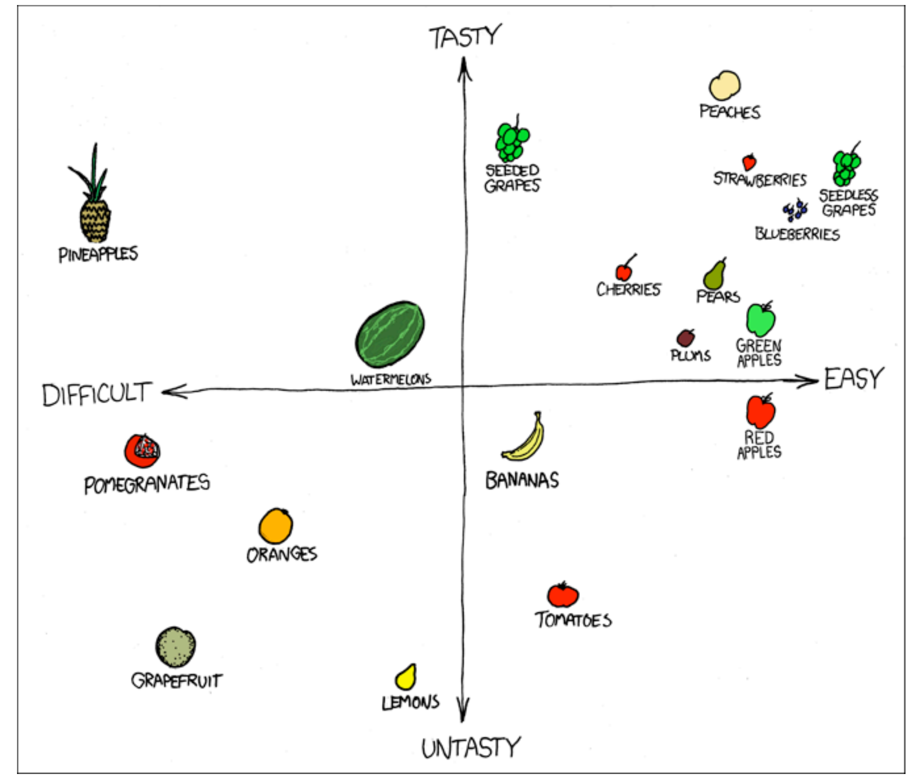


Dimensionality reduction

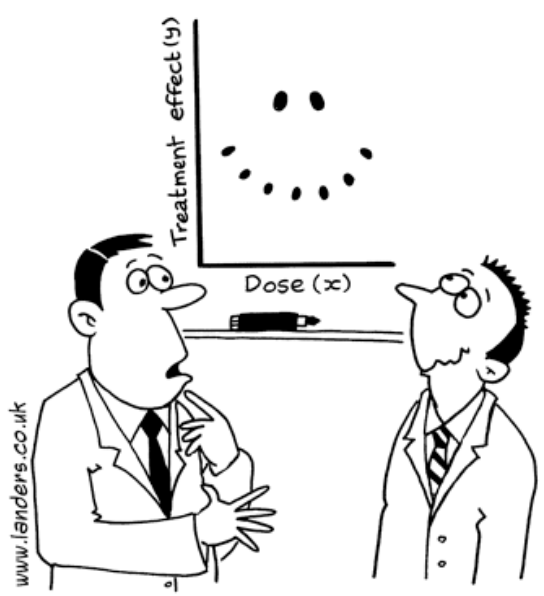
Classification
predict class, given attributes



Clustering
group similar things together



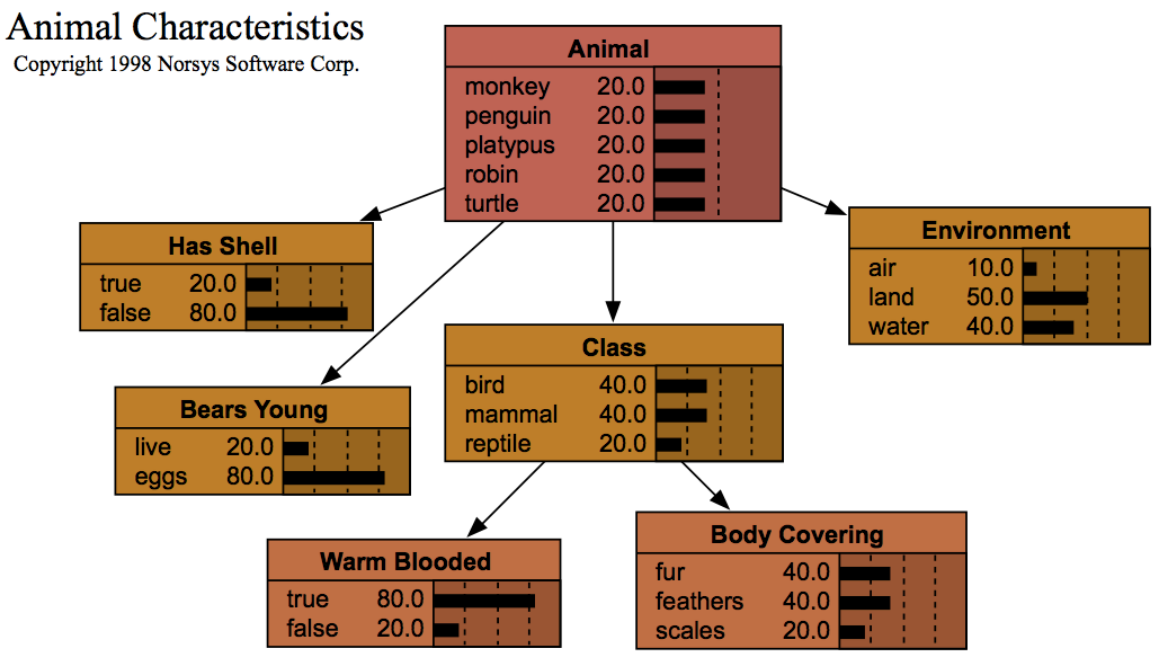
Regression
predict values, given other values



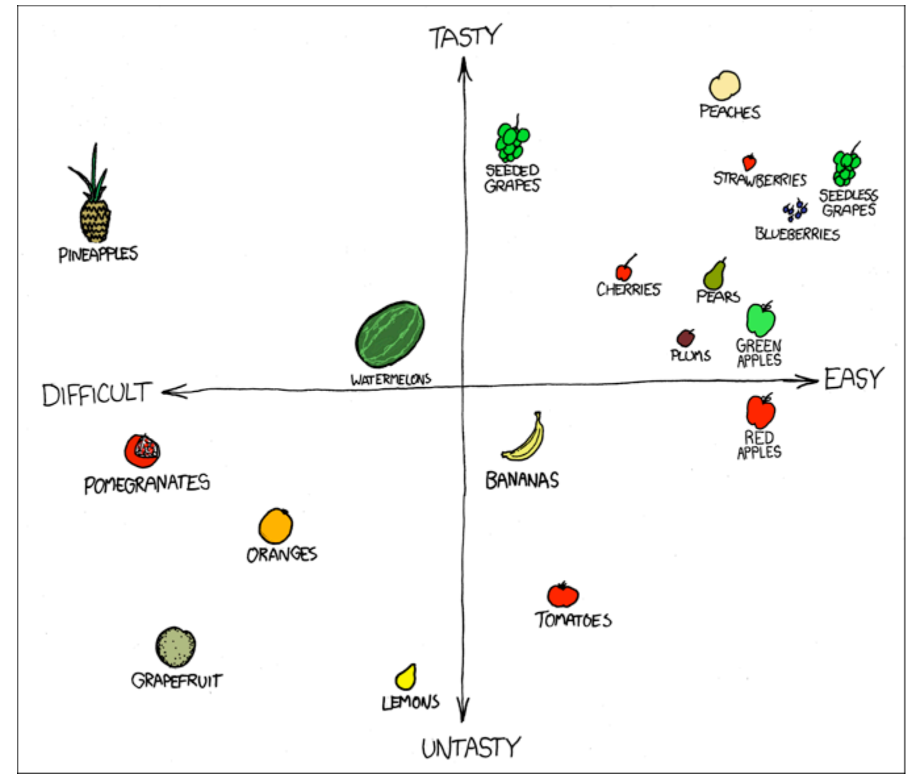
"It's a non-linear pattern with outliers.....but for some reason I'm very happy with the data."

Dimensionality reduction

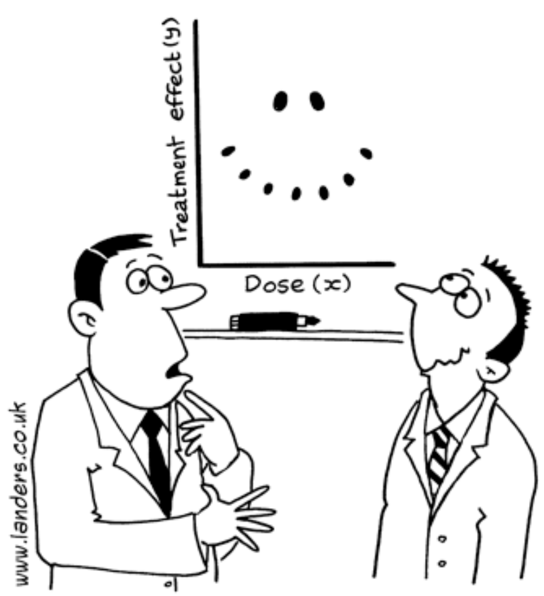
Classification
predict class, given attributes



Clustering
group similar things together



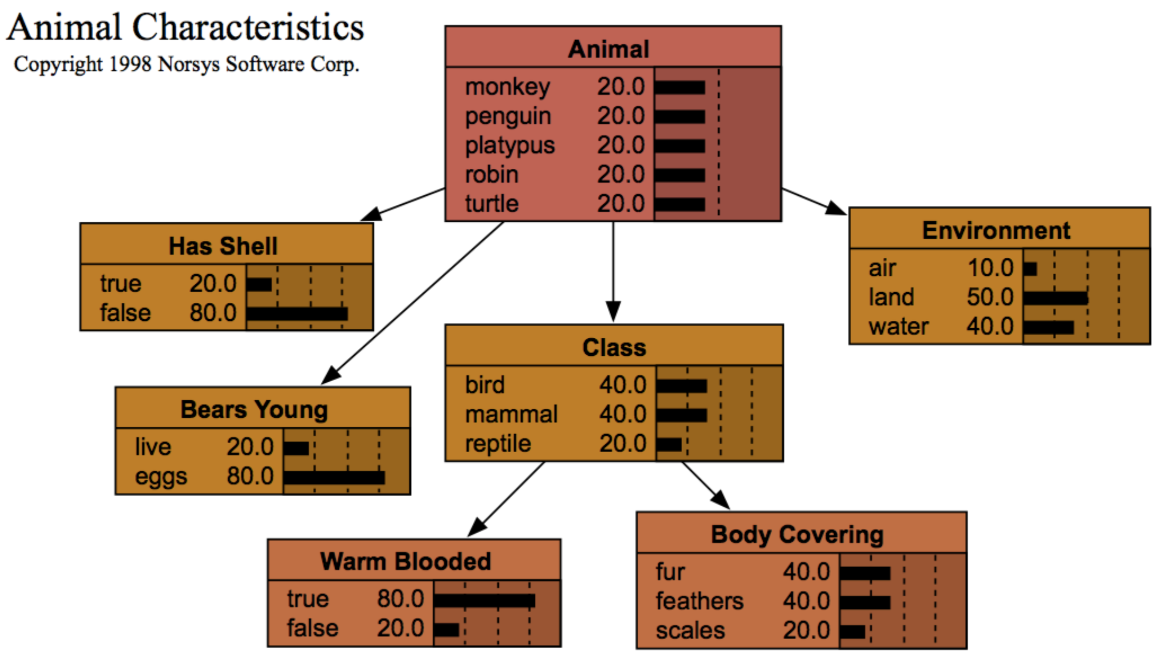
Regression
predict values, given other values



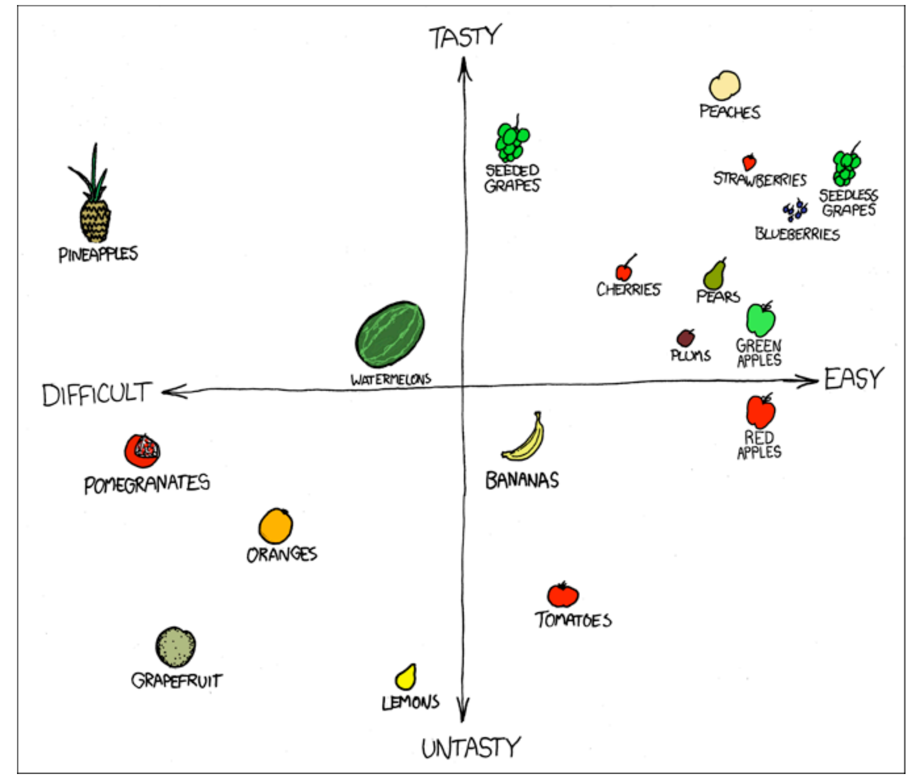
"It's a non-linear pattern with outliers.....but for some reason I'm very happy with the data."

Dimensionality reduction
keeping only the relevant variables

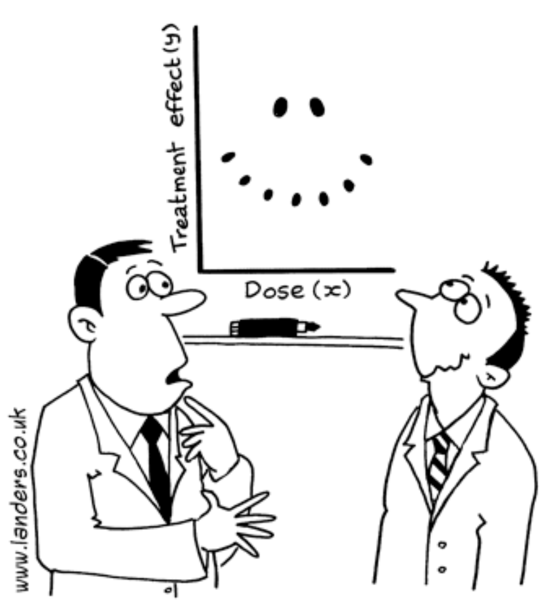
Classification
predict class, given attributes



Clustering
group similar things together

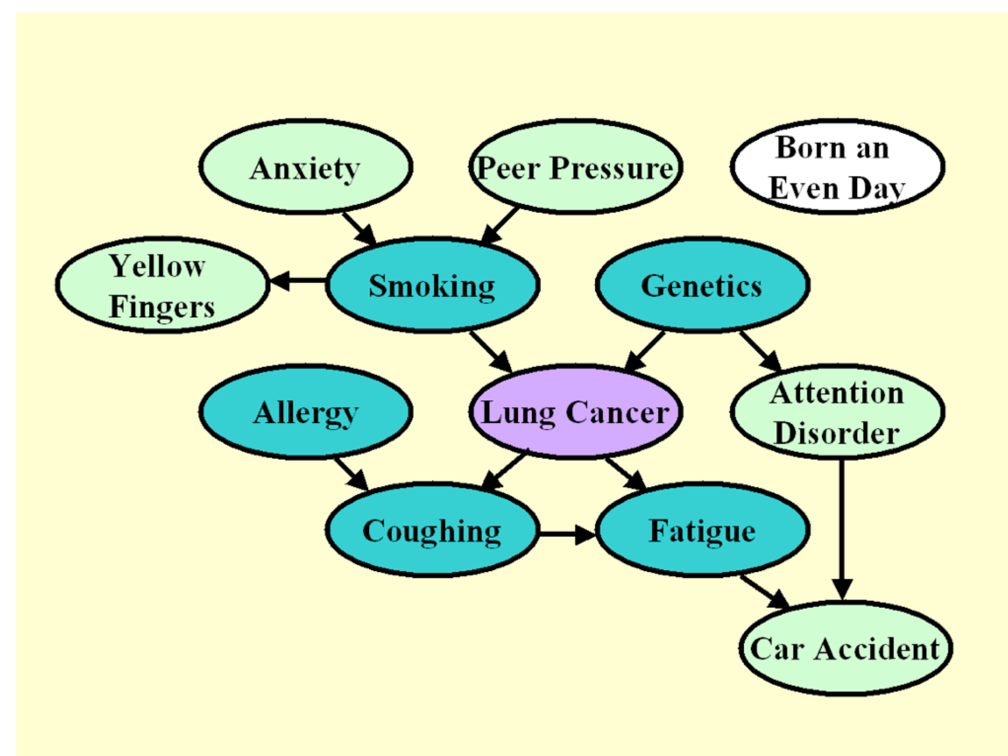


Regression
predict values, given other values



"It's a non-linear pattern with outliers.....but for some reason I'm very happy with the data."

Dimensionality reduction
keeping only the relevant variables



Hypotheses

require

mathematical models

Bayesian **Frequentist**

Blackbox **Causal**

Bayesian

probability is a state of knowledge

Frequentist

Blackbox

Causal

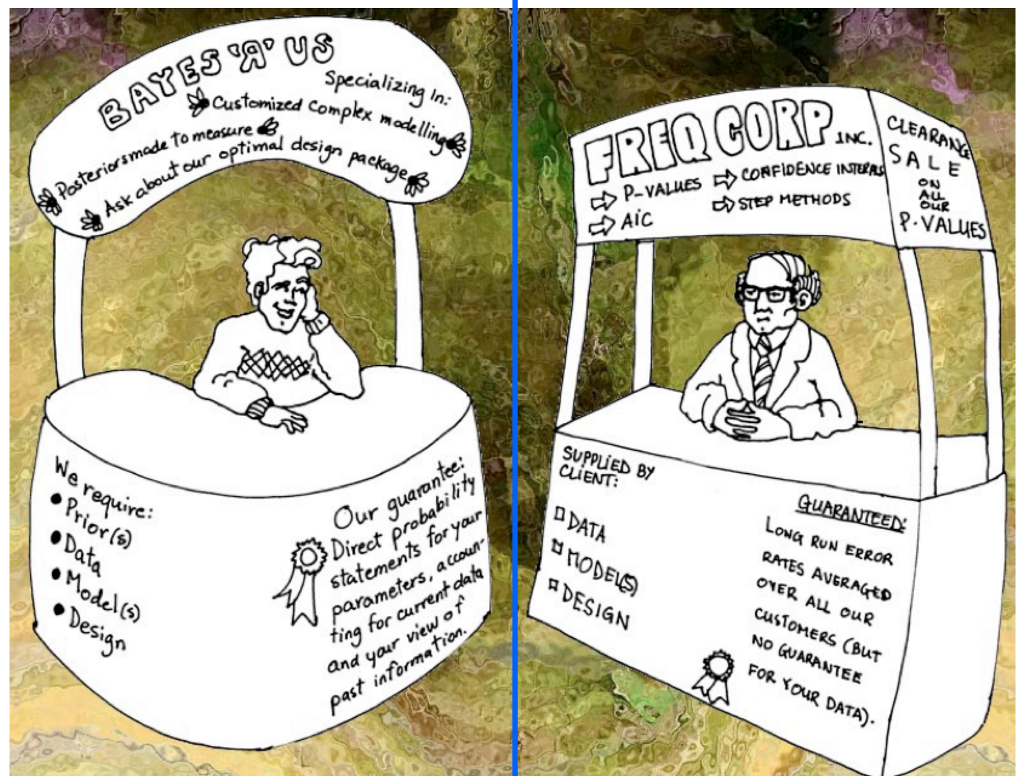
Bayesian	Frequentist
probability is a state of knowledge	probability is a frequency
Blackbox	**Causal**

Bayesian

probability is a state of knowledge

Frequentist

probability is a frequency



Blackbox

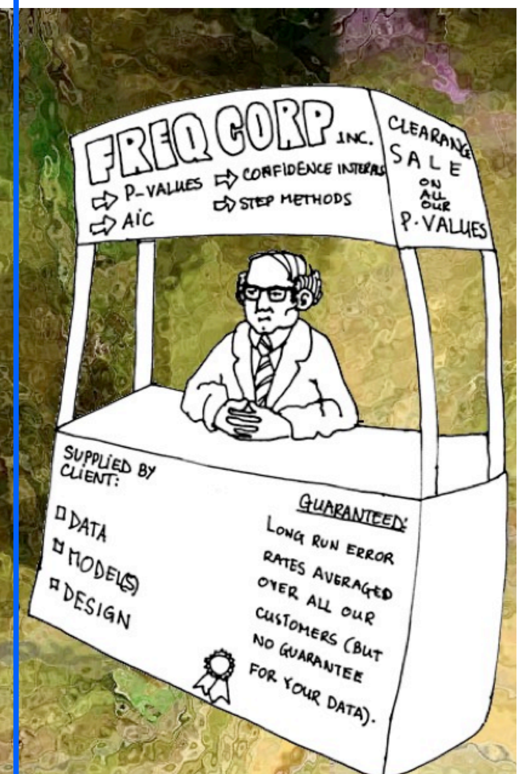
Causal

Bayesian

probability is a state of knowledge

Frequentist

probability is a frequency



Blackbox

ML : toolbox for processing data

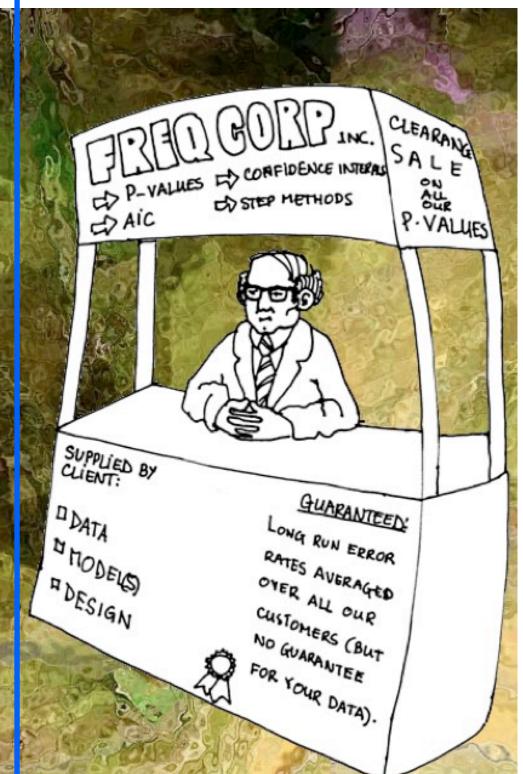
Causal

Bayesian

probability is a state of knowledge

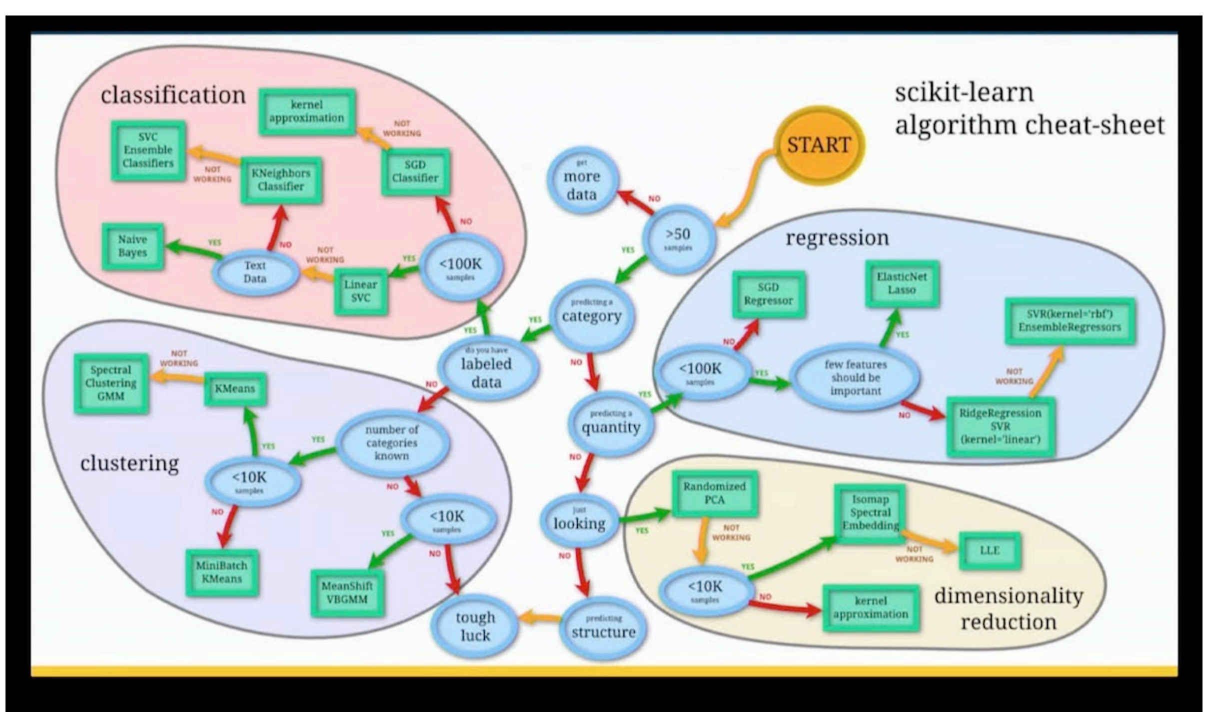
Frequentist

probability is a frequency



Blackbox

ML : toolbox for processing data



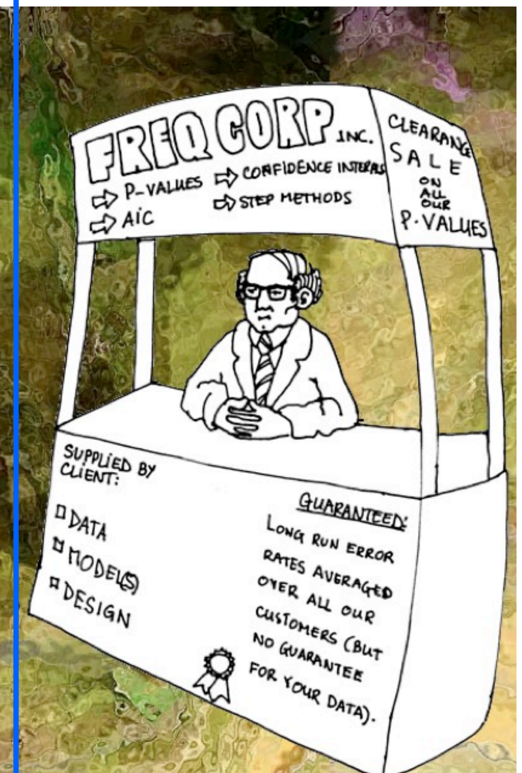
Causal

Bayesian

probability is a state of knowledge

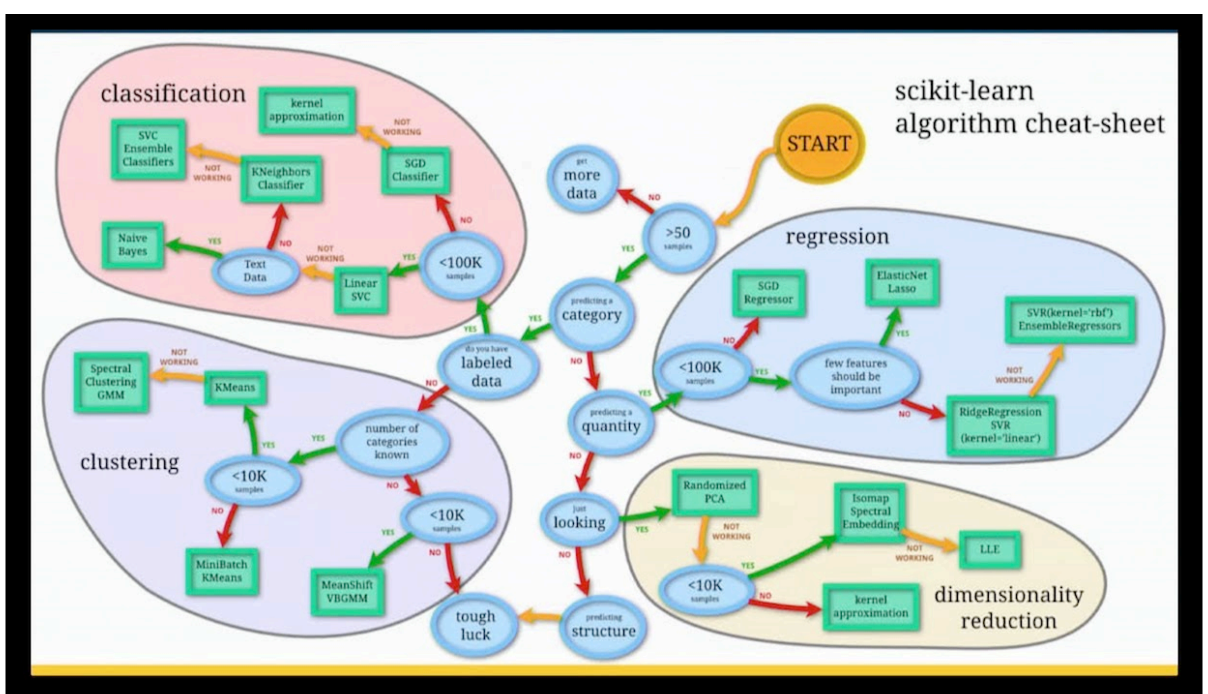
Frequentist

probability is a frequency



Blackbox

ML : toolbox for processing data



Causal

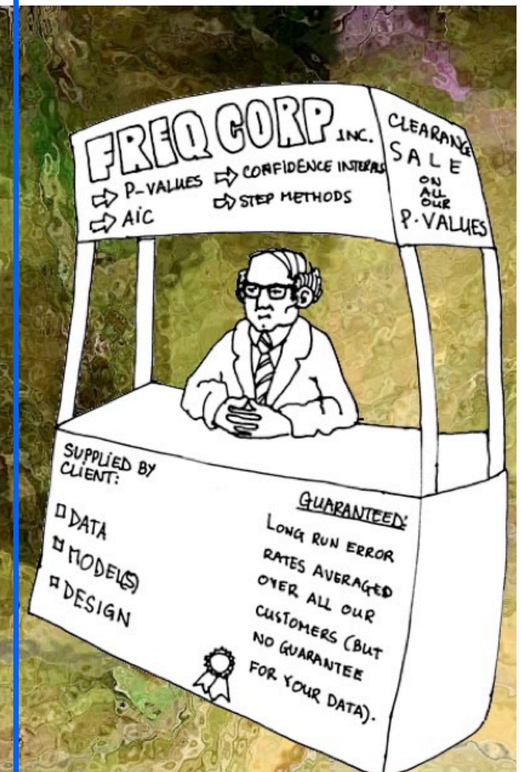
ML : learning generative models of data

Bayesian

probability is a state of knowledge

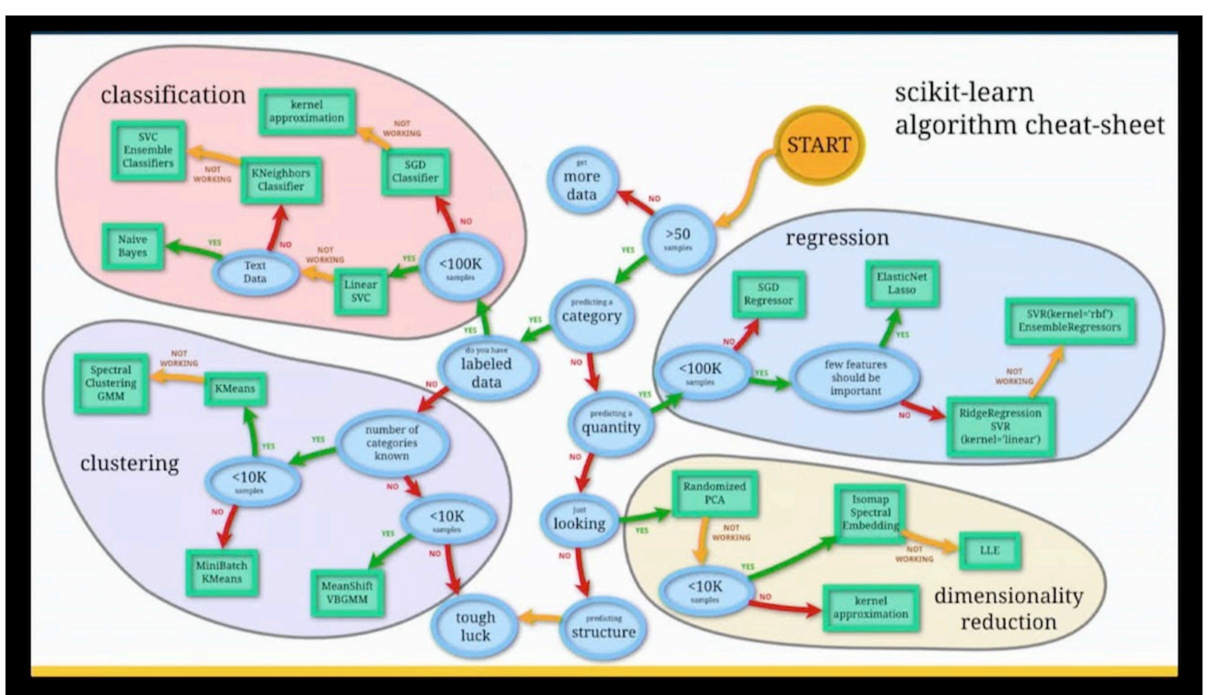
Frequentist

probability is a frequency



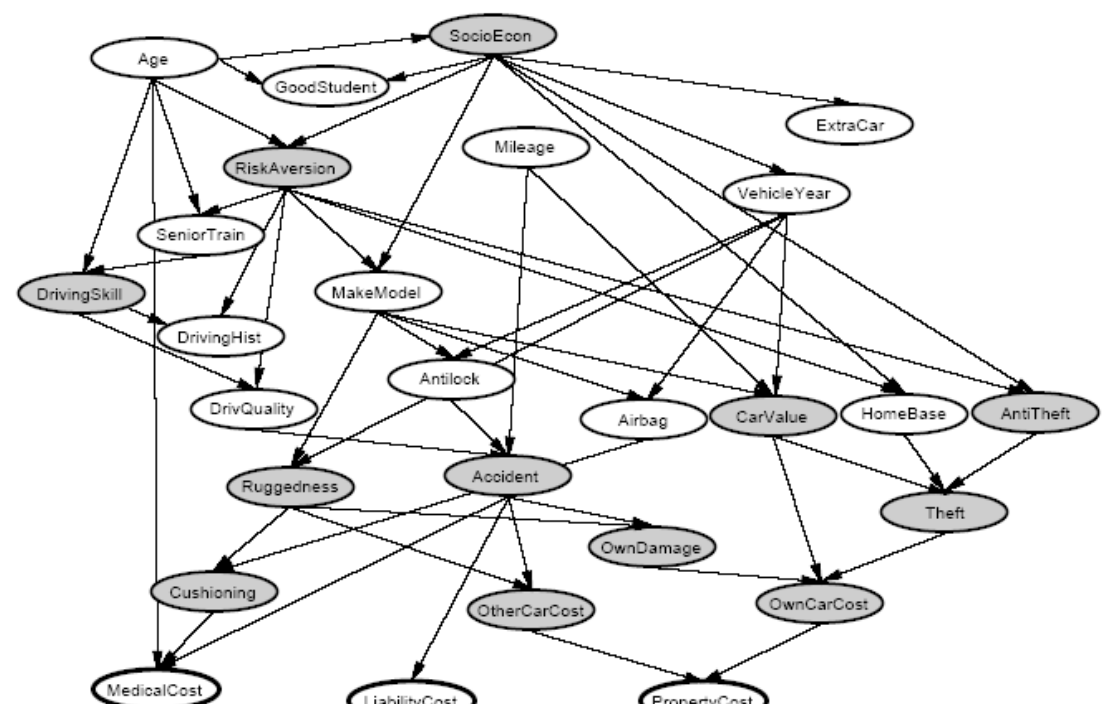
Blackbox

ML : toolbox for processing data



Causal

ML : learning generative models of data

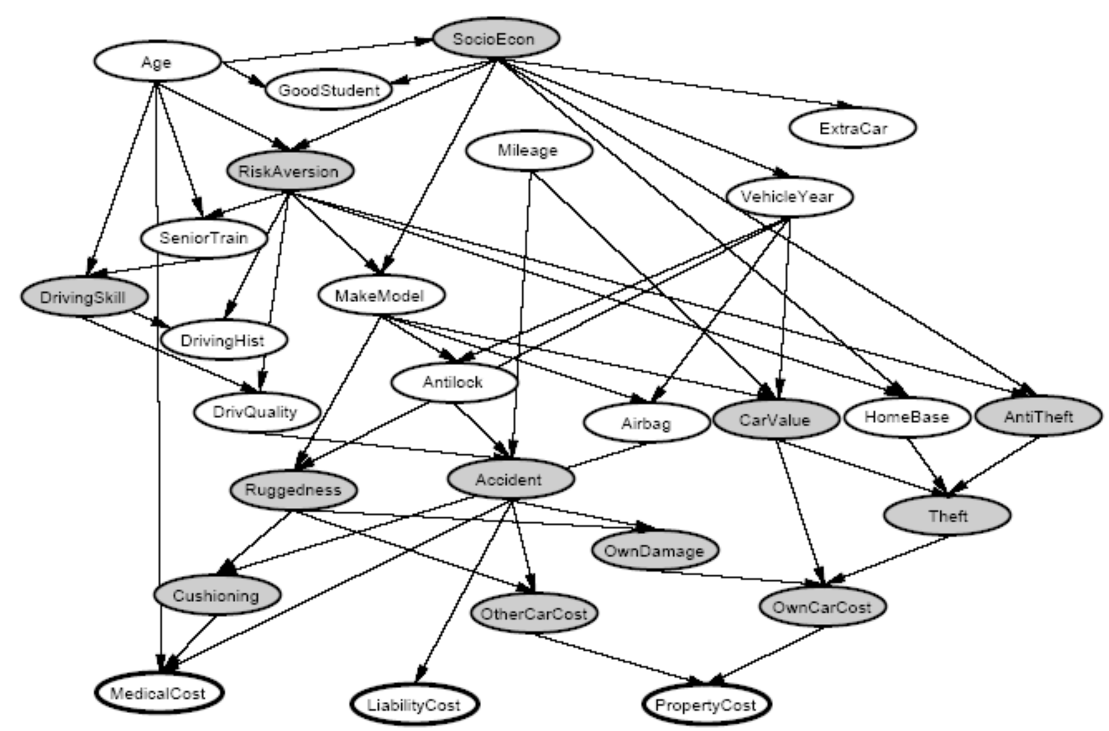


Bayesian

probability is a state of knowledge

Causal

ML : learning generative models of data



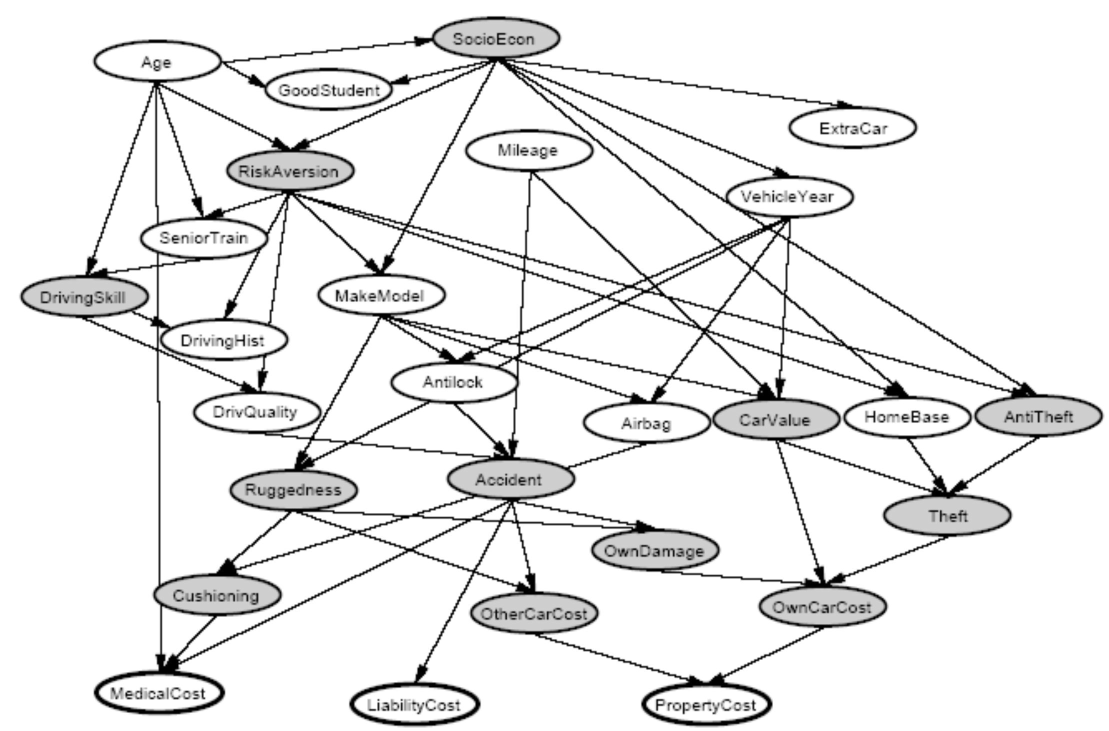
Bayesian

probability is a state of knowledge

Nearly all approaches to probabilistic programming are Bayesian since it is hard to create other coherent frameworks for automated reasoning about uncertainty - Ghahramani, Nature, 2015

Causal

ML : learning generative models of data



Machine Learning is Ubiquitous and Very Useful

ISR

DARPA Grand Challenge
Fully autonomous ground vehicles competition

Nuclear Test Ban Treaty Compliance: Deduce set of seismic events given detections and misdetections on a network of stations

Image Search/Activity Detection: Find and identify objects and actions in video

Object Tracking: Follow vehicles as they move through a city and are recorded in multiple video streams (DARPA CZTS)

Patterns of Life: Process wide area aerial surveillance data and associated tracks to infer location types and object dynamics

Bird Migration Patterns: Model spatio-temporal distribution of birds (by species); involves large-scale sensor integration

DARPA LAGR: Vision-based robot navigation

Google Glasses: Perform searches based on images taken by user cell phone cameras

Natural Language Processing

© Apple

Siri
Voice recognition and Natural Language Processing (NLP)

Watson: Computer system capable of answering questions posed in natural language

Topic Models: Statistical model for discovering the abstract "topics" that occur in a collection of documents

Distributed Topic Models: Asynchronous distributed topic discovery

Citation Analysis: Given citations, extract author, title, and venue strings and identify co-reference citations

Entity Resolution: Discovering entities that are mentioned in a large corpus of news articles

NLP Sequence Tagging: Tagging parts of speech and recognizing named entities in newspaper articles

Predictive Analytics →

.....

© Netflix

Netflix Challenge
Predict user ratings for films based on previous ratings

Microsoft Matchbox: Match players based on their gaming skill set

Predictive Database: Understand information based on causal relationships in data

Bing Image Search: Search for images on the web by selecting text in word document

Amazon Recommendation Engine: Recommend items based on consumer data

Cyber and Other

ORNL's Attack Variant Detector: Discover compromised systems

Yahoo's Bayesian Spam Filtering: Self-adapting system based on word probabilities

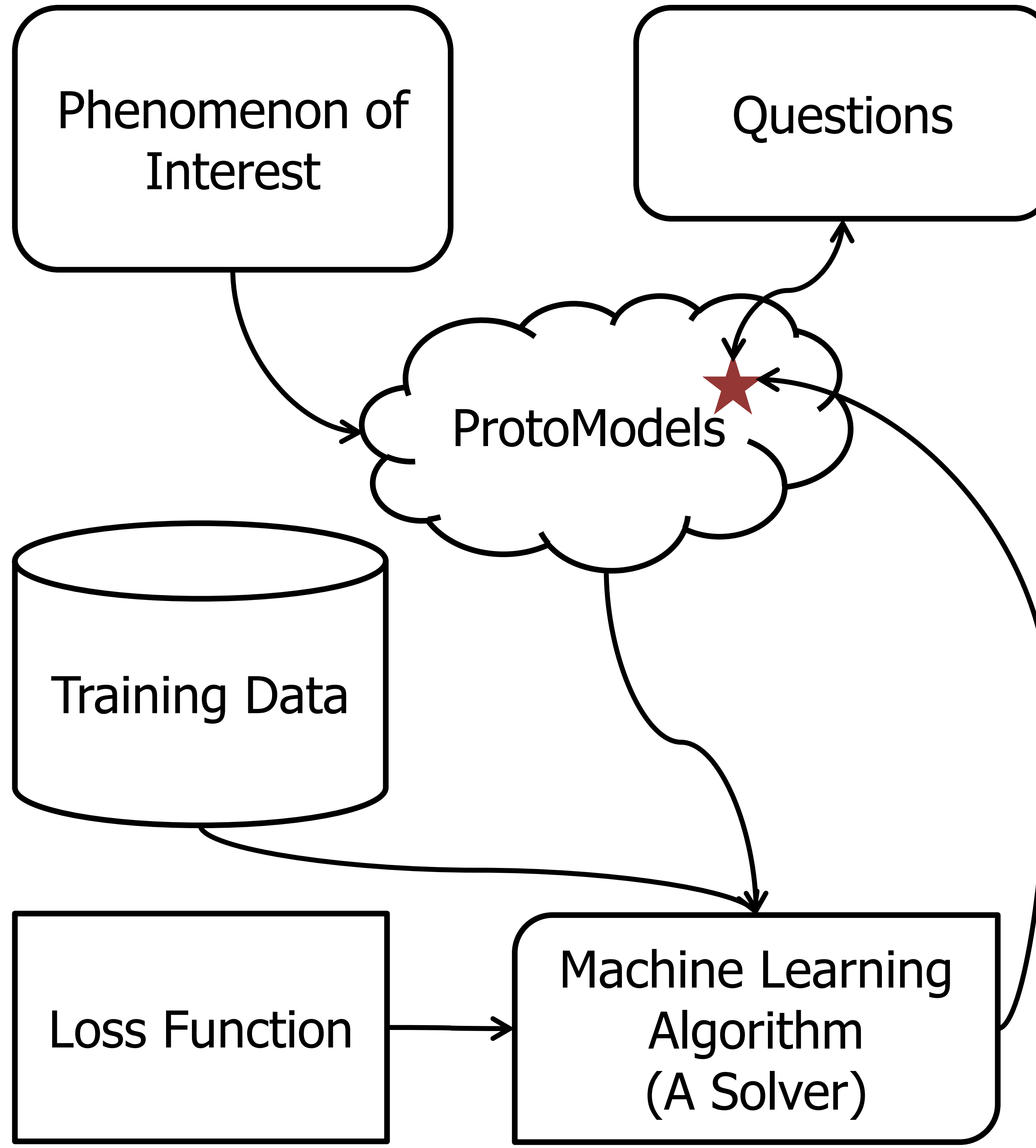
Cyber Genome Lineage: Reverse engineer malware samples to find shared "genetic" features between different malware samples

Gene Sequencing: Determine order of nucleotides in a DNA molecule

Disclaimer: Images of specific products are used for illustration only. Use of these images does not imply endorsement of inherent technical vulnerabilities.

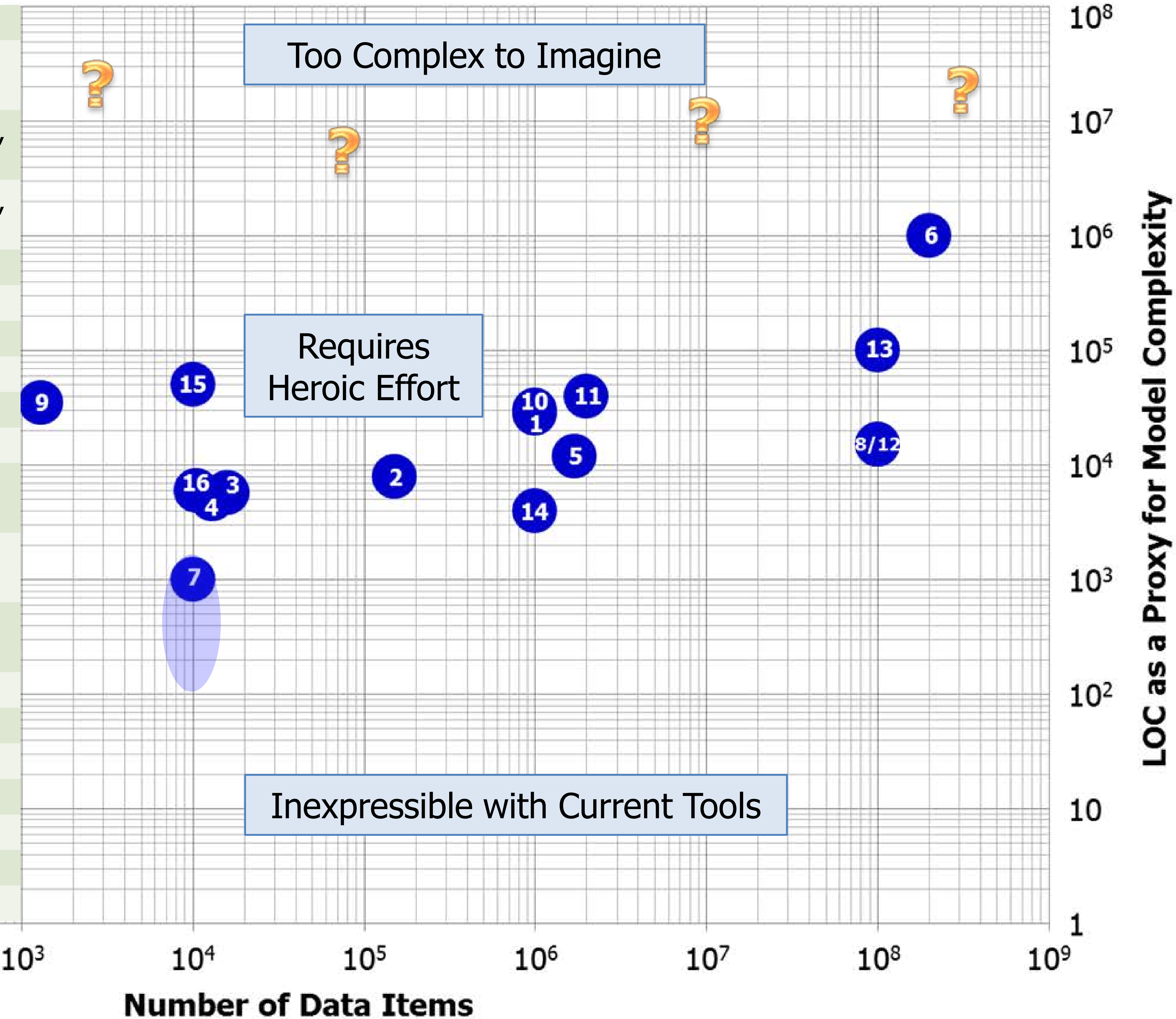
Why Is It Hard?

- Brittleness of implementations & lack of reusable tools.
  PHY decoder: $100M/standard;
  Infer.NET: max 20% on model
- High level of required expertise
  10K solvers,
  100s of grad student hours per model
- Painfully slow & unpredictable solvers
  Massive data sets, complex algorithms, tricky coding for graph traversal and numeric stability
- Challenges constructing models
  Limited modeling vocabulary;
  models entwined with solvers



We're Missing a Tool to Write These Applications

ISR	
1. Seismic Monitoring for Nuclear Test Ban Treaty	
2. Image Search and Activity Detection	PerSEAS, VIRAT, VMR
3. Object tracking (video)	PerSEAS, VIRAT, CZTS
4. Patterns of Life	ONR
5. Bird Migration Patterns	
Natural Language Processing	
6. Watson	
7. Topic Models	CALO
8. Distributed Topic Models	CALO
9. Citation Analysis	BOLT, DEFT, MADCAT, RATS
10. Entity Resolution	BOLT, DEFT, MADCAT
11. NLP Sequence Tagging	BOLT, DEFT, MADCAT
Predictive Analytics	
12. Microsoft Matchbox	
13. Netflix Challenge	
14. Predictive Database	XDATA
Cyber and Other	
15. Cyber Genome Lineage	Cyber Defense
16. Gene Sequencing	



The Missing Tool (Explained by Example)

Source: Noah Goodman, POPL (2013)

Model:

```
mem strength person = gaussian 100 10
lazy person = flip 0.1
pulling person = if lazy person then (strength person) / 2
                                 else strength person
total-pulling team = sum (map pulling team)
winner team1 team2 = greater (total-pulling team1)
                             (total-pulling team2)
```

Query:

```
strength Bob
```

Facts:

```
[Bob, Mark] = winner [Bob, Mark] [Tom, Sam]
[Bob, Fred] = winner [Bob, Fred] [Jon, Jim]
```

System will calculate probability distribution for Bob's strength given known facts

The user describes the model at a high level. An inference engine analyzes the program, query, data, and available hardware resources to produce best solution

What is happening in the Python space ?

PyMC3
Stan
Emcee

stochastic inference

Infer.py
BayesPy

deterministic inference

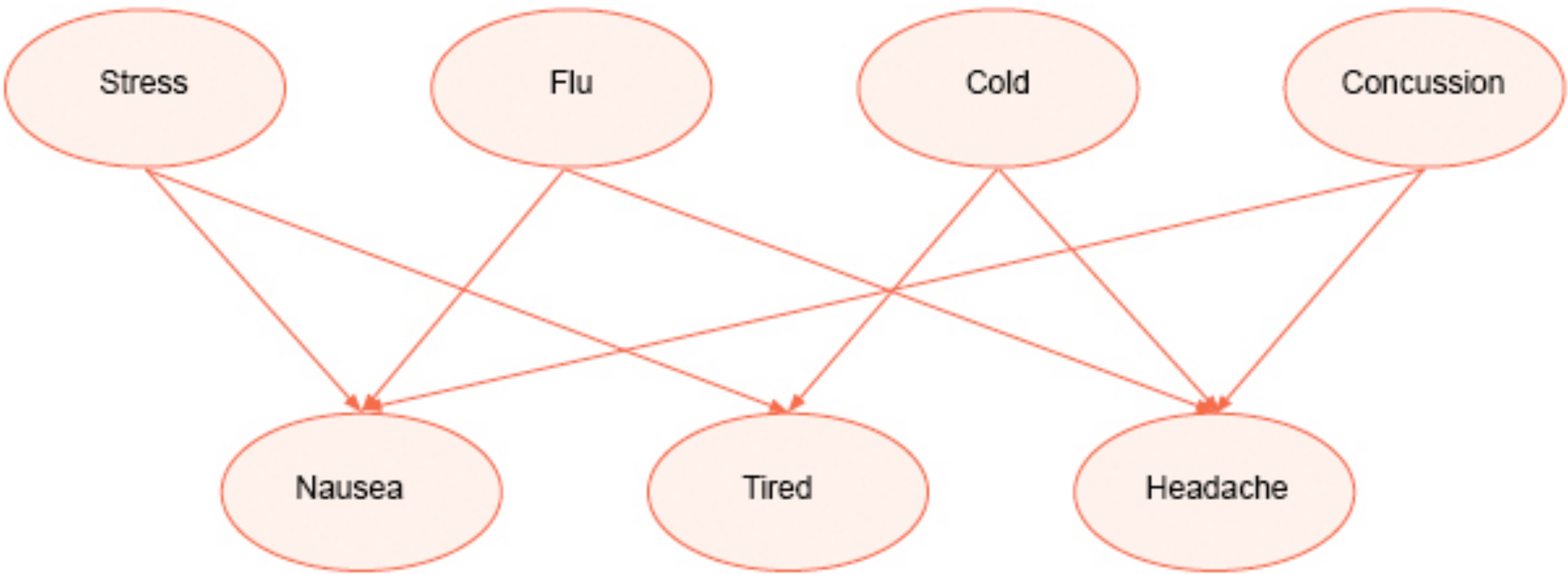
Church
Lea

black box inference

Live coding 2

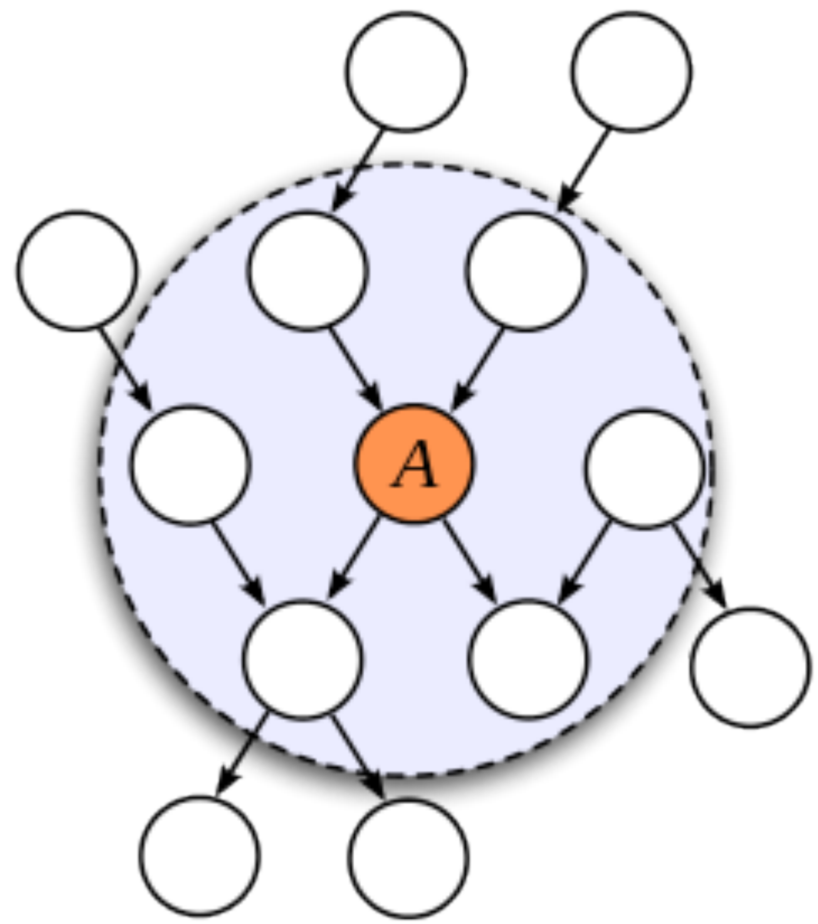
Probabilistic networks

Probabilistic network : conditional independences



causal reasoning : P(tired | flu)
diagnostic reasoning : P(flu | tired)

Markov blanket



Live coding 3

This repository Search Pull requests Issues Gist

ValueFromData / **PyPNL**

Unwatch 4

Python Probabilistic Network Library — Edit

4 commits **1** branch **0** releases **1** contributor

Branch: **master** PyPNL / +

add shorter copy

ronojoy authored 2 minutes ago latest commit 689586f250

README.md	add shorter copy	2 minutes ago

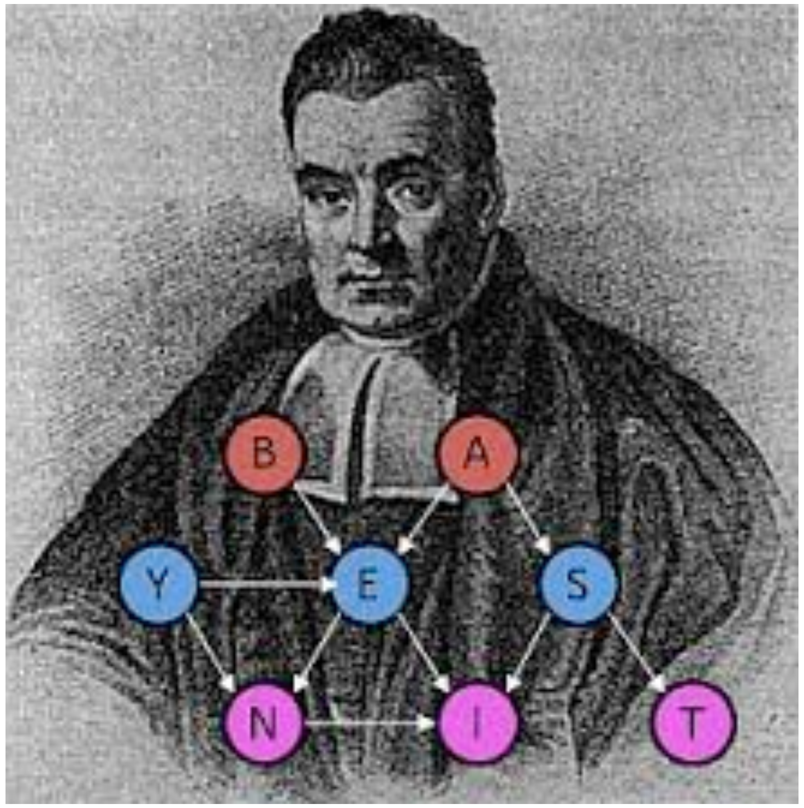
README.md

PyPNL

PyPNL is a probabilistic network library in Python. The library consists of a Python API which interfaces to state of the art C/C++ inference and learning engines. The API design is modular and allows for plug-ins which augment functionality. PyPNL comes with documentation, examples, and a test suite.

Features

The PyPNL library has a rich set of features for **representation**, **inference**, **learning** and **actions** using probabilistic networks.



This repository Search

Pull requests Issues Gist

ValueFromData / **PyPNL**

Unwatch 4

Python Probabilistic Network Library — Edit

4 commits | **1** branch | **0** releases | **1** contributor

Branch: **master** PyPNL / +

add shorter copy

ronojoy authored 2 minutes ago latest commit 689586f250

README.md add shorter copy 2 minutes ago

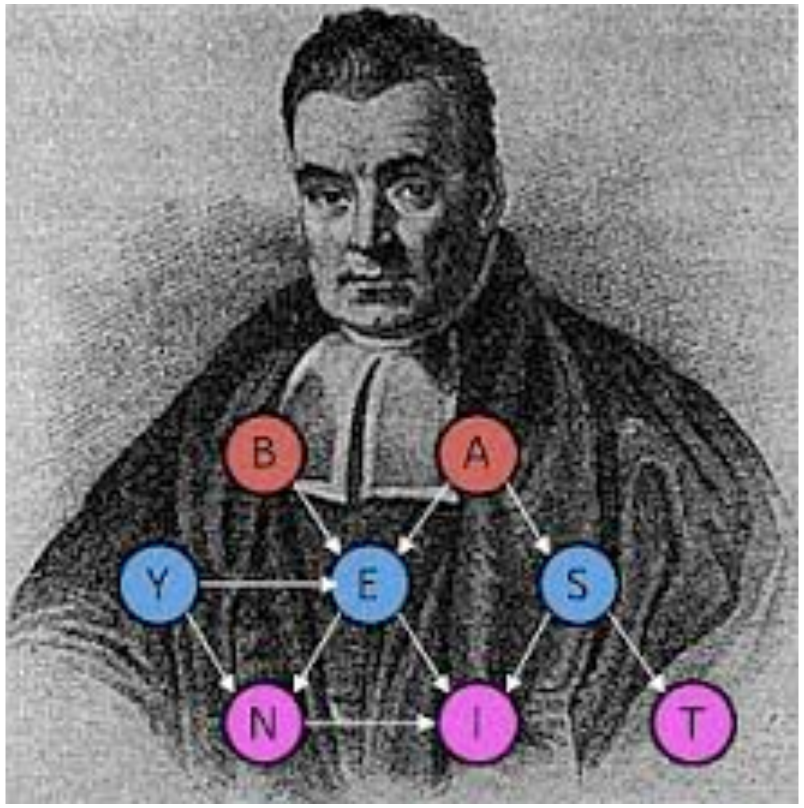
README.md

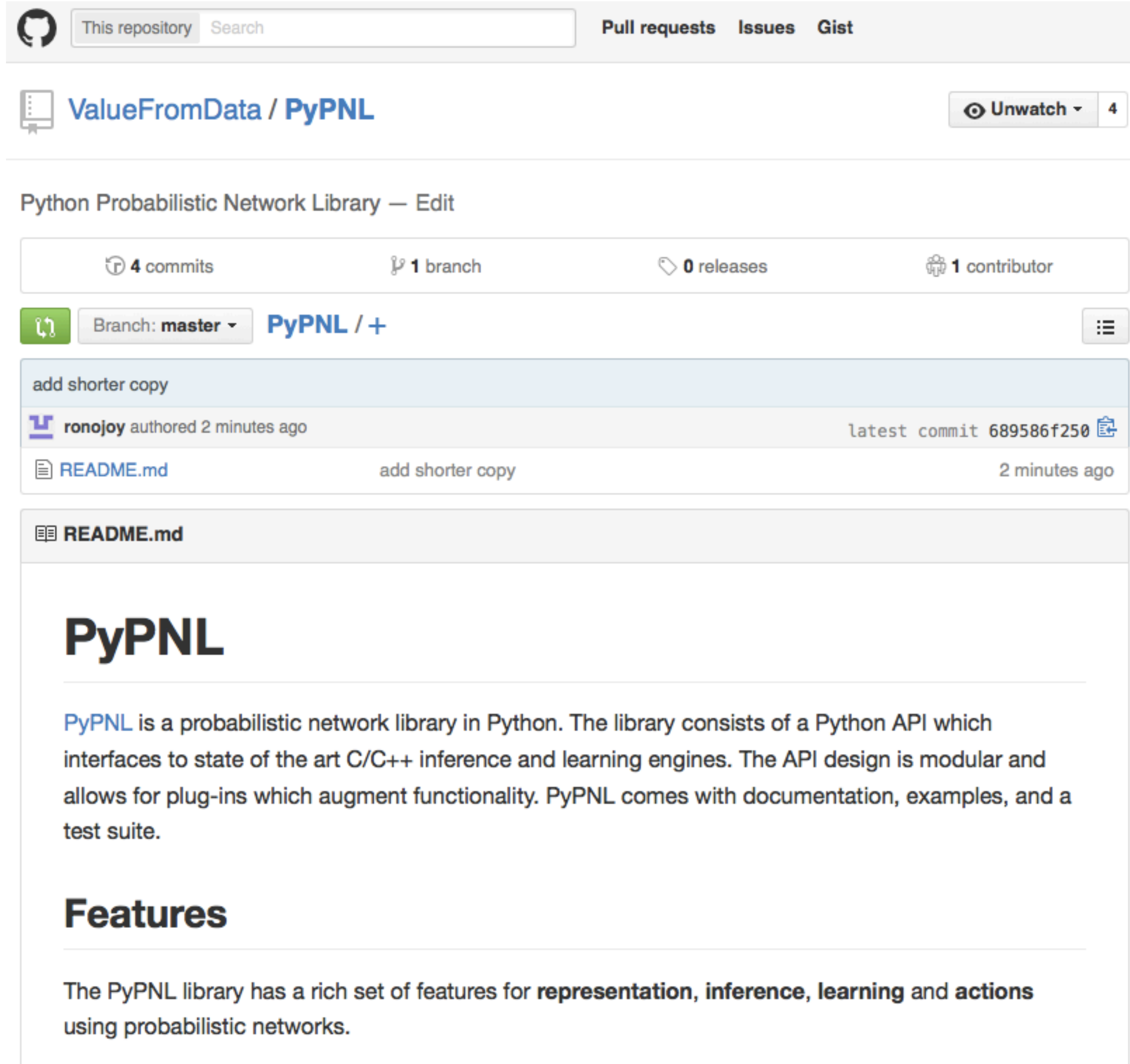
PyPNL

PyPNL is a probabilistic network library in Python. The library consists of a Python API which interfaces to state of the art C/C++ inference and learning engines. The API design is modular and allows for plug-ins which augment functionality. PyPNL comes with documentation, examples, and a test suite.

Features

The PyPNL library has a rich set of features for **representation**, **inference**, **learning** and **actions** using probabilistic networks.





This repository Search | Pull requests | Issues | Gist

ValueFromData / **PyPNL**

Unwatch 4

Python Probabilistic Network Library — Edit

4 commits | **1** branch | **0** releases | **1** contributor

Branch: **master** PyPNL / +

add shorter copy

ronojoy authored 2 minutes ago latest commit 689586f250

README.md add shorter copy 2 minutes ago

README.md

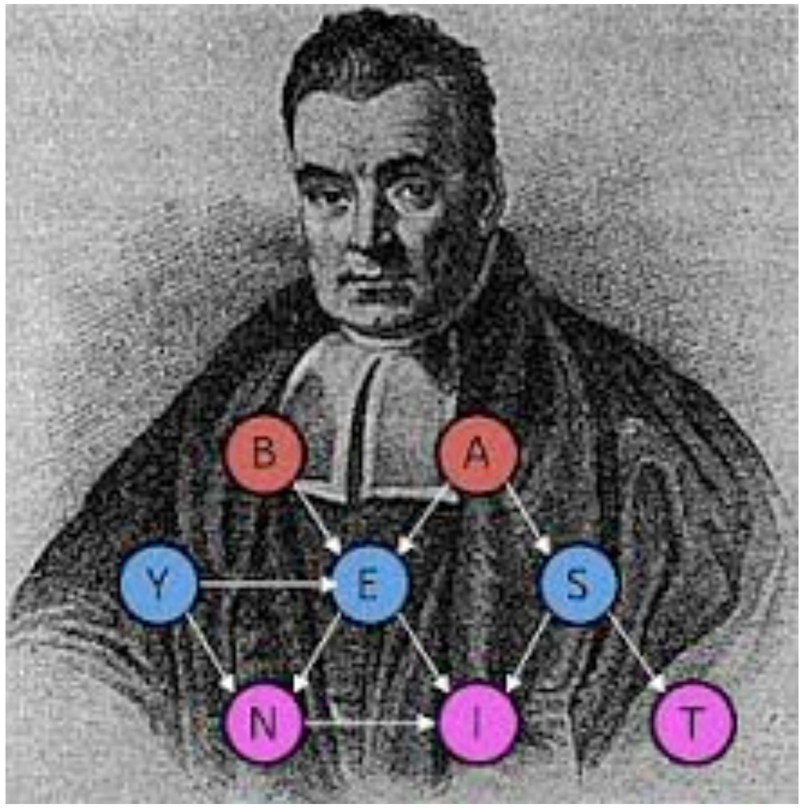
PyPNL

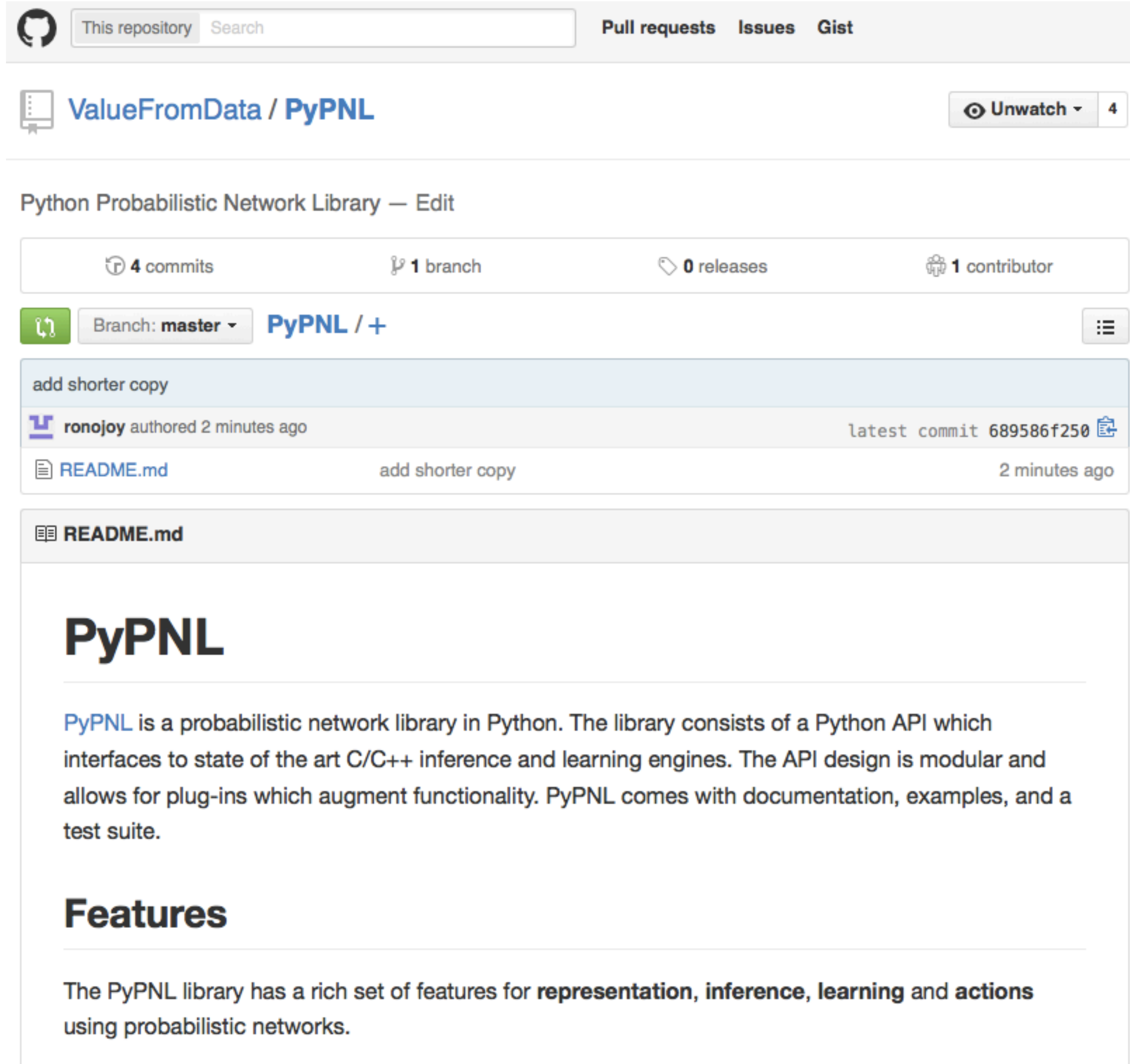
PyPNL is a probabilistic network library in Python. The library consists of a Python API which interfaces to state of the art C/C++ inference and learning engines. The API design is modular and allows for plug-ins which augment functionality. PyPNL comes with documentation, examples, and a test suite.

Features

The PyPNL library has a rich set of features for **representation**, **inference**, **learning** and **actions** using probabilistic networks.

We are building a causal learning and inference engine that will beat the current state-of-art!





We are building a causal learning and inference engine that will beat the current state-of-art!

Thank you for your attention!