From Statistical Inference to Probabilistic Reasoning

Using Bayesian Networks for Reasoning with Small Samples, Missing Values, and No Data.

BAYESIALAB

Hello

my name is

Stefan Conrady

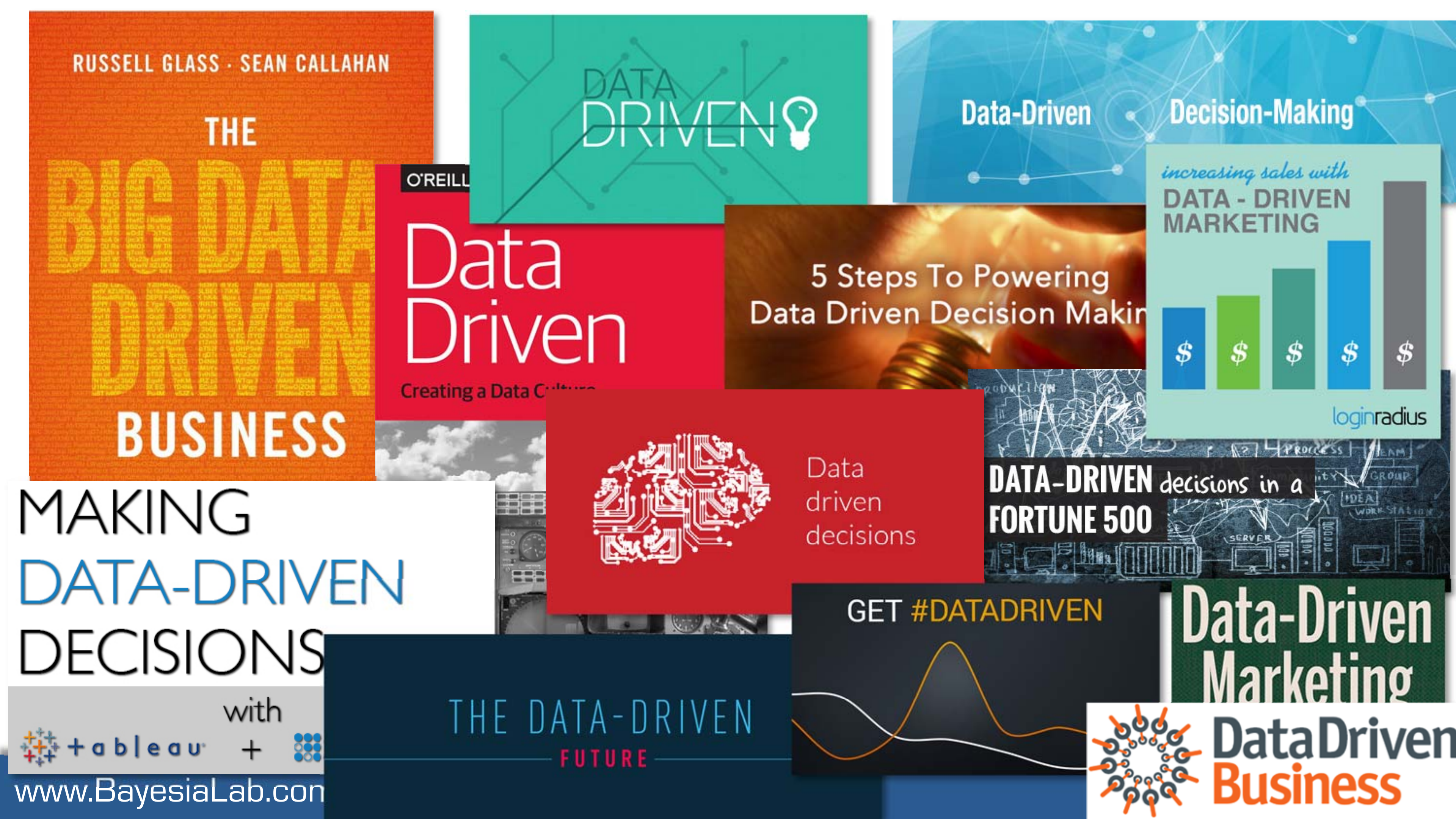
The BayesiaLab Software Platform

U.S.ARMY
NASA
Booz | Allen | Hamilton
GEORGETOWN UNIVERSITY
DASSAULT AVIATION
LOCKHEED MARTIN
bp
LA POSTE
ArcelorMittal
groupm
WPP
gsk GlaxoSmithKline
University of Virginia
RENAULT
sodexo
DELL
Lilly
Mu Sigma
DO THE MATH
EADS
Holiday Inn
SONY
IHG
Unilever
synovate
SOCIETE GENERALE
STANFORD UNIVERSITY
BBDO
HYATT
GfK
Nestlé
Fractal
NTT
MAi
Firmenich
National Analysts WORLDWIDE
SAMSUNG
france telecom
Cargill
eDF
Ipsos
McGill
SNCF
P&G
CDC
eni
Hilton HOTELS & RESORTS
LANCASTER UNIVERSITY
VirginiaTech
orange
NISSAN
DGA
SAINT-GOBAIN
Polk
IMRB international
RUSH UNIVERSITY
AIRBUS
Microsoft
Team Detroit
sanofi pasteur
tns
SMUCKER'S
MillwardBrown
American Diabetes Association.
PENNSTATE
Deloitte.
ACXIOM
cancer care ontario | action cancer ontario
UCLA
L'ORÉAL
Rotman UNIVERSITY OF TORONTO
intel
AIRFRANCE
SAFRAN Turbomeca
Neiman Marcus
TOTAL
PSA PEUGEOT CITROËN
MEDSOLUTION
LSU LOUISIANA STATE UNIVERSITY
nielsen
OmnicomGroup
Capgemini

Bayesian Networks & BayesiaLab

A Practical Introduction for Researchers

- Free download:
  www.bayesia.com/book
- Hardcopy available on Amazon:
  http://amzn.com/0996533303


RUSSELL GLASS · SEAN CALLAHAN
THE BIG DATA DRIVEN BUSINESS
DATA DRIVEN
Data-Driven Decision-Making
Data Driven
5 Steps To Powering
increasing sales with DATA - DRIVEN MARKETING
loginradius
Data driven decisions
DATA-DRIVEN decisions in a FORTUNE 500
MAKING DATA-DRIVEN DECISIONS with
tableau +
GET #DATADRIVEN
Data-Driven Marketing
THE DATA-DRIVEN FUTURE
Data Driven Business

The End of Theory?

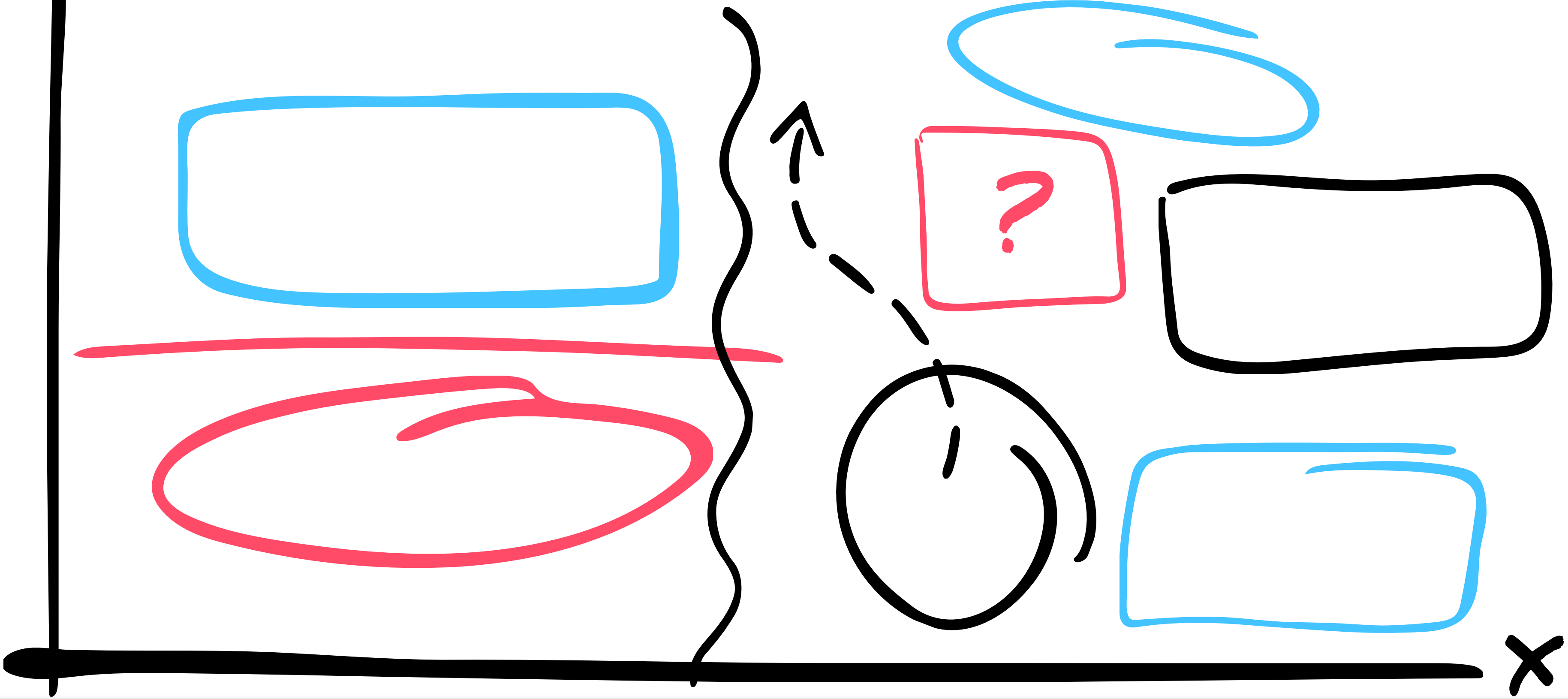
What if I have little or no data?

Small Data in a Big-Data World

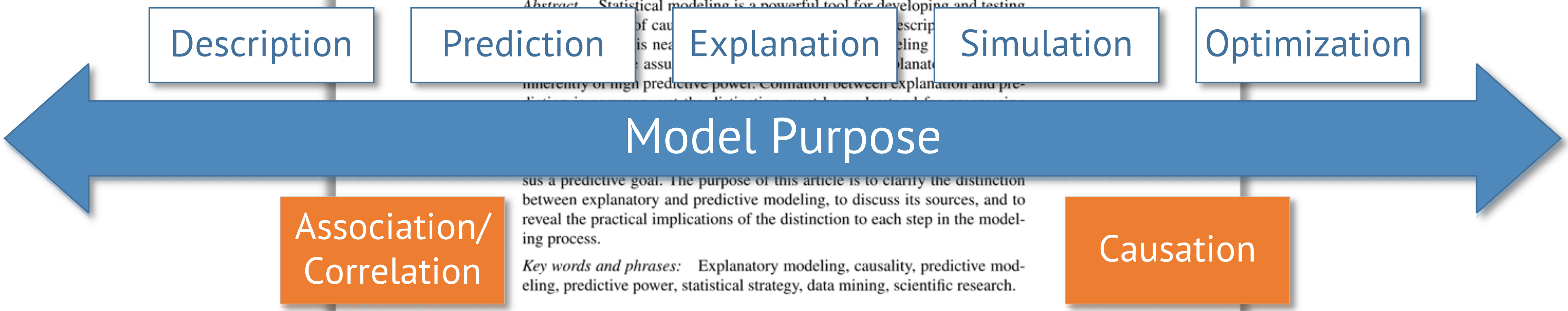
Small-Data Challenges

- Generating knowledge from “small data”
    - Overparameterization
    - Variable selection
- Applying knowledge to “small data”
    - Incomplete observations
    - Uncertain observations
    - Hypothetical scenarios
    - Cost of observations

A Map of Analytic Modeling



The Purpose of Models



Statistical Science
2010, Vol. 25, No. 3, 289–310
DOI: 10.1214/10-STS330
© Institute of Mathematical Statistics, 2010

To Explain or to Predict?

Galit Shmueli

sus a predictive goal. The purpose of this article is to clarify the distinction between explanatory and predictive modeling, to discuss its sources, and to reveal the practical implications of the distinction to each step in the modeling process.

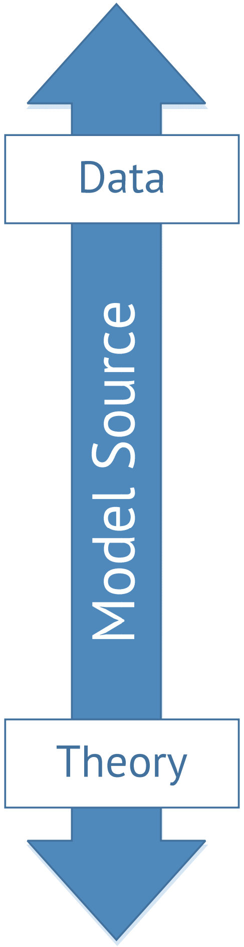
Key words and phrases: Explanatory modeling, causality, predictive modeling, predictive power, statistical strategy, data mining, scientific research.

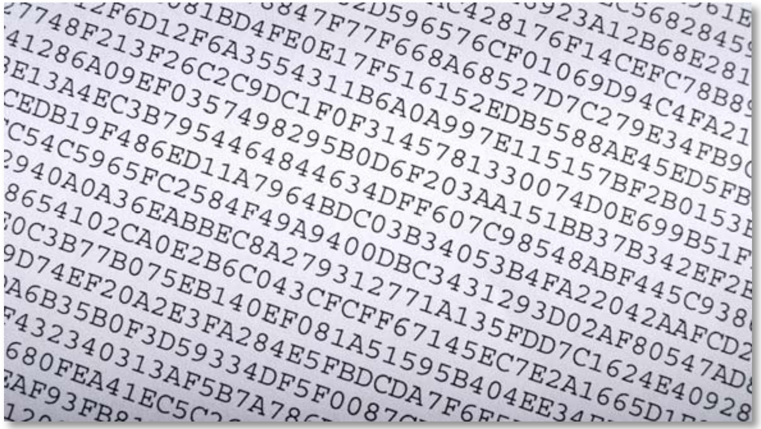
1. INTRODUCTION

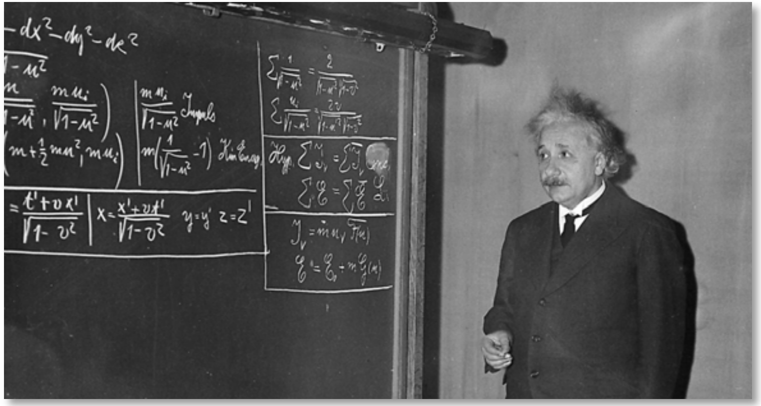
Looking at how statistical models are used in different scientific disciplines for the purpose of theory building and testing, one finds a range of perceptions regarding the relationship between causal explanation and empirical prediction. In many scientific fields such as economics, psychology, education, and environmen-

focus on the use of statistical modeling for causal explanation and for prediction. My main premise is that the two are often conflated, yet the causal versus predictive distinction has a large impact on each step of the statistical modeling process and on its consequences. Although not explicitly stated in the statistics methodology literature, applied statisticians instinctively sense that predicting and explaining are different. This article

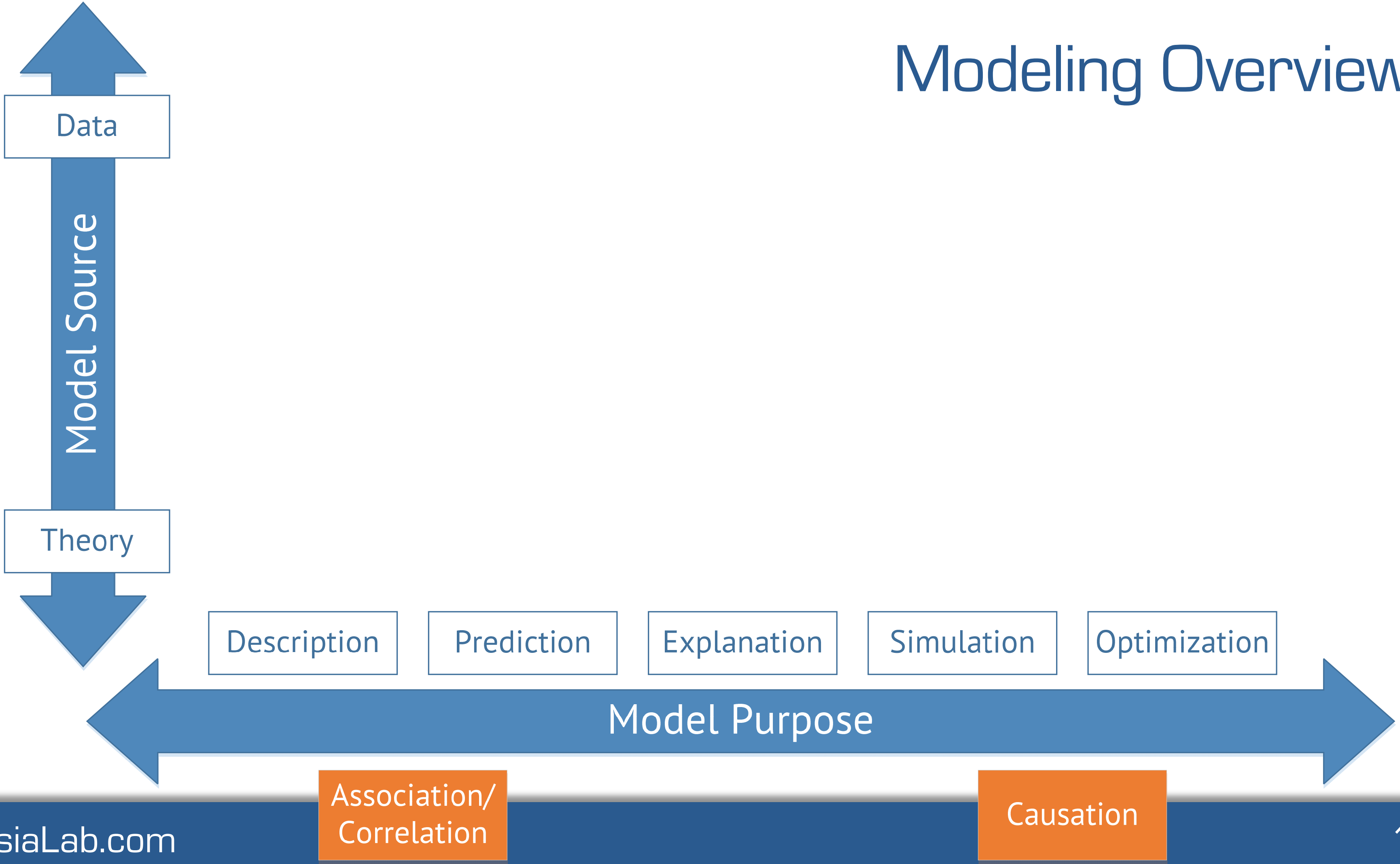
Source of Models



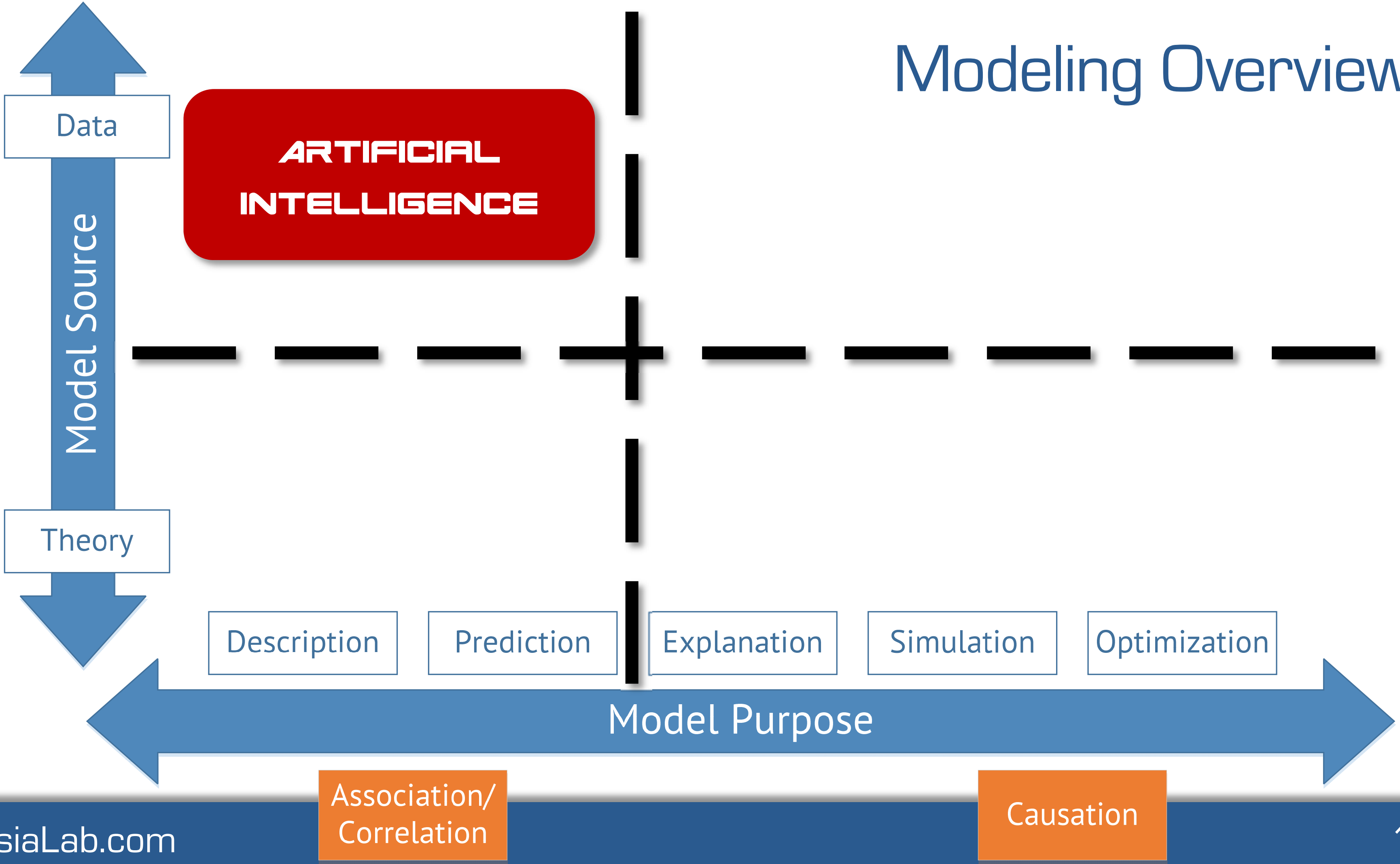




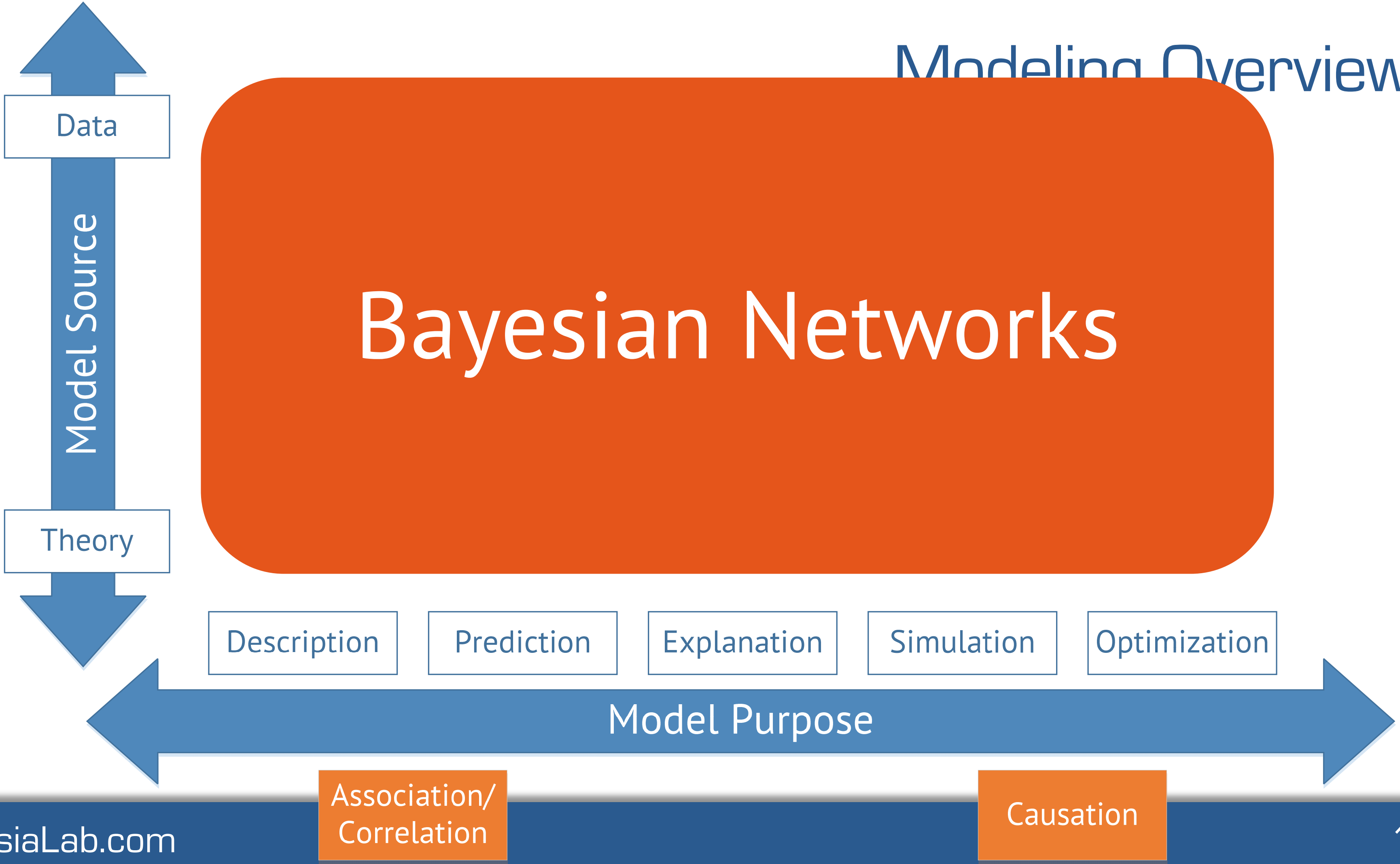
Modeling Overview



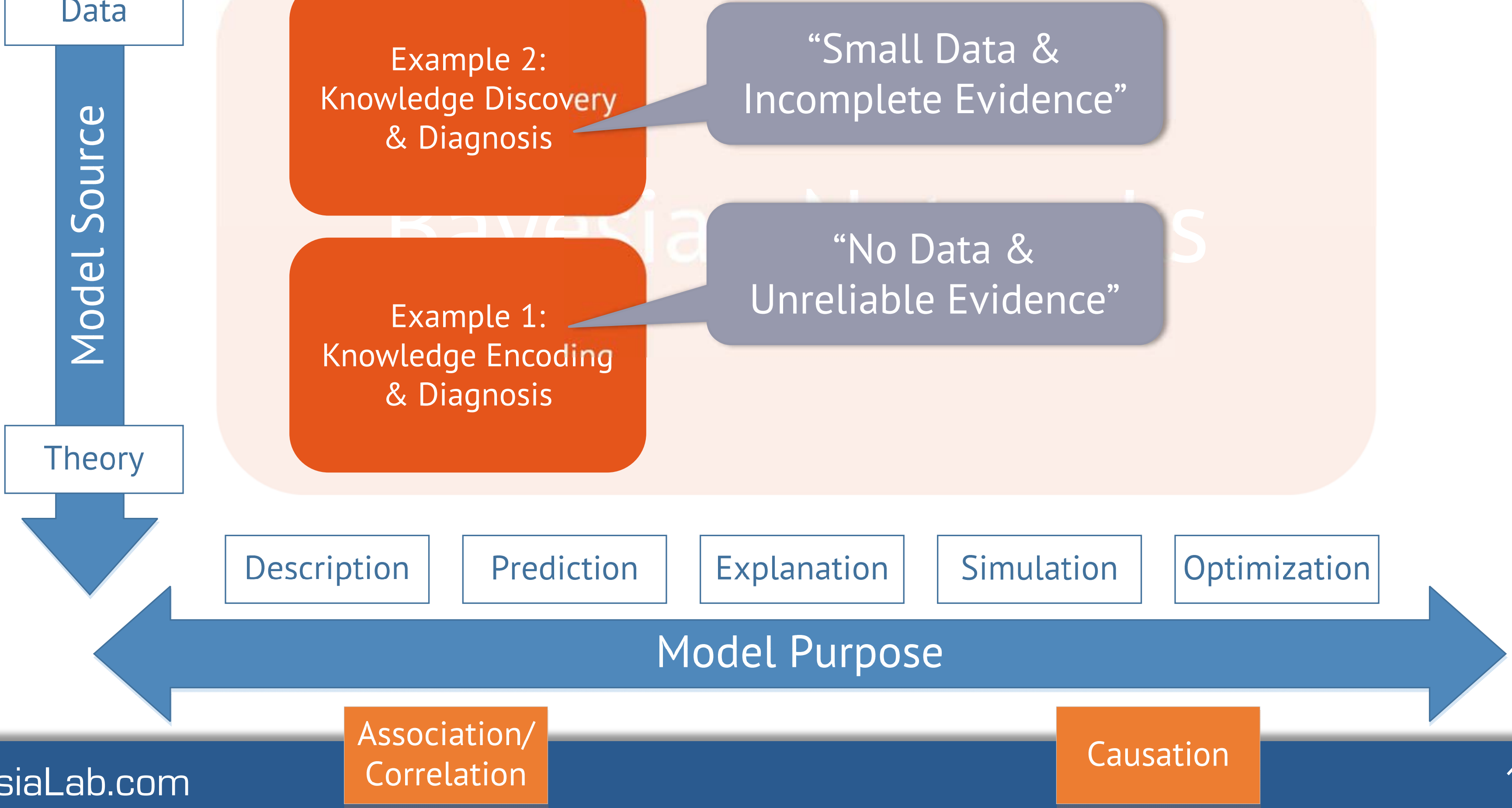
Modeling Overview



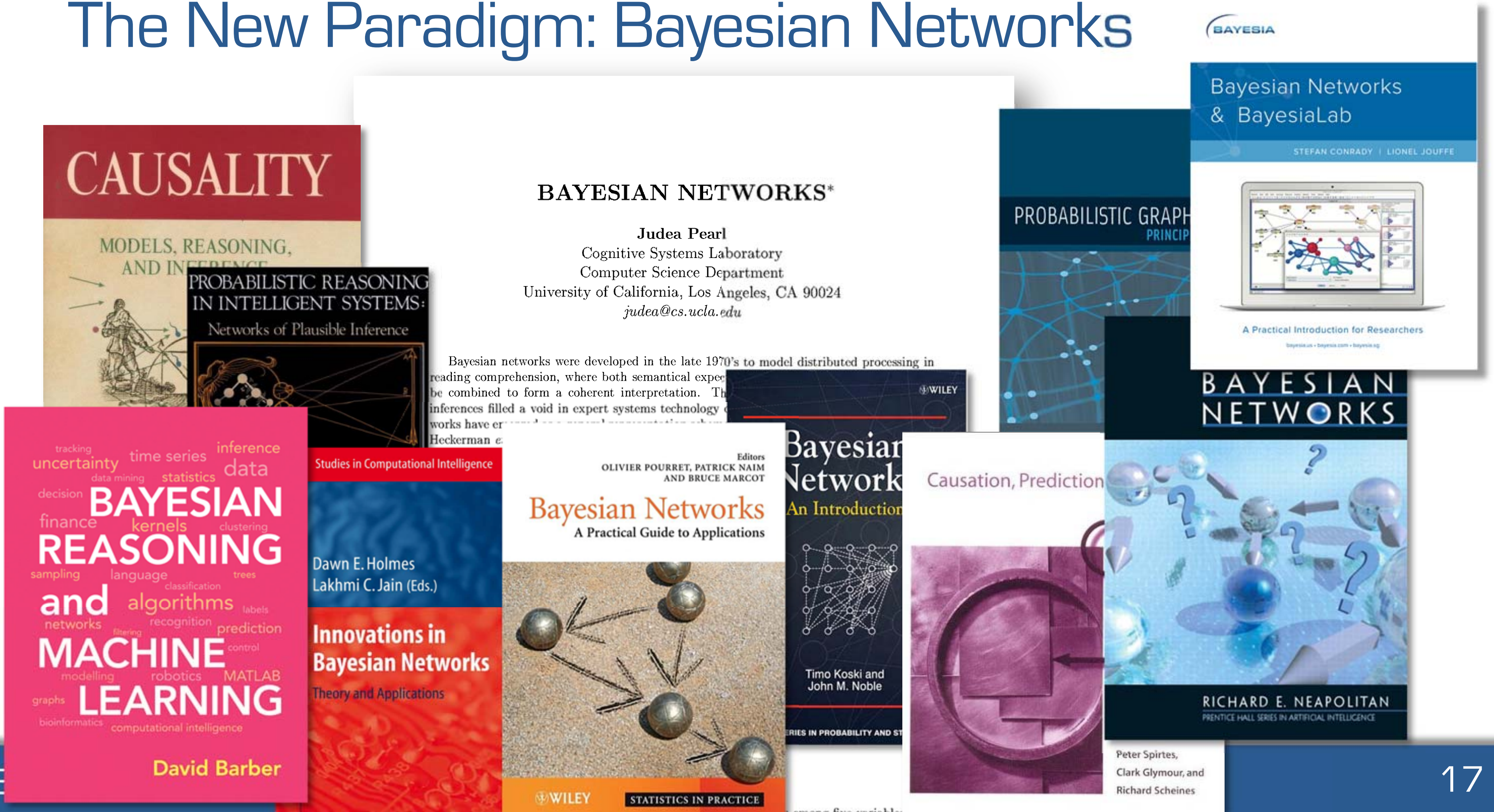
Modeling Overview



Modeling Overview



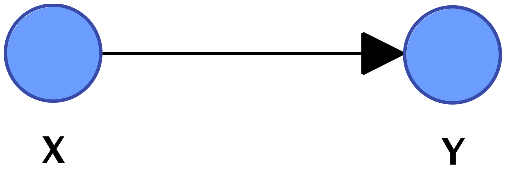
The New Paradigm: Bayesian Networks

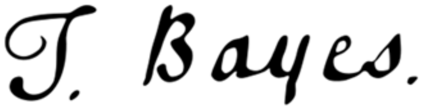


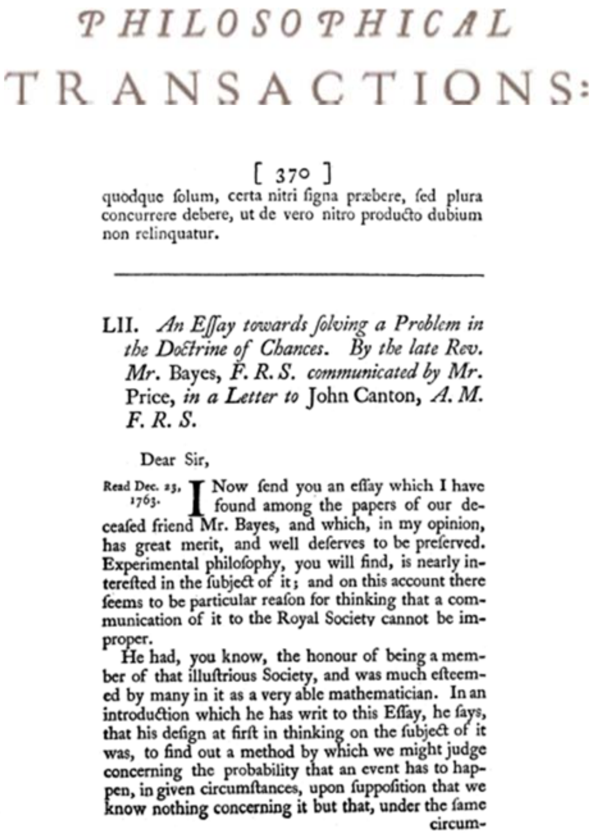
Rev. Thomas Bayes

Bayes's Theorem for Conditional Probabilities

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)}$$







PHILOSOPHICAL

TRANSACTIONS:

[370]

quodque ſolum, certa nitri ſigna præbere, ſed plura concurrere debere, ut de vero nitro producto dubium non relinquatur.

LII. *An Eſſay towards ſolving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *F. R. S. communicated by Mr.* Price, *in a Letter to* John Canton, *A. M. F. R. S.*

Dear Sir,

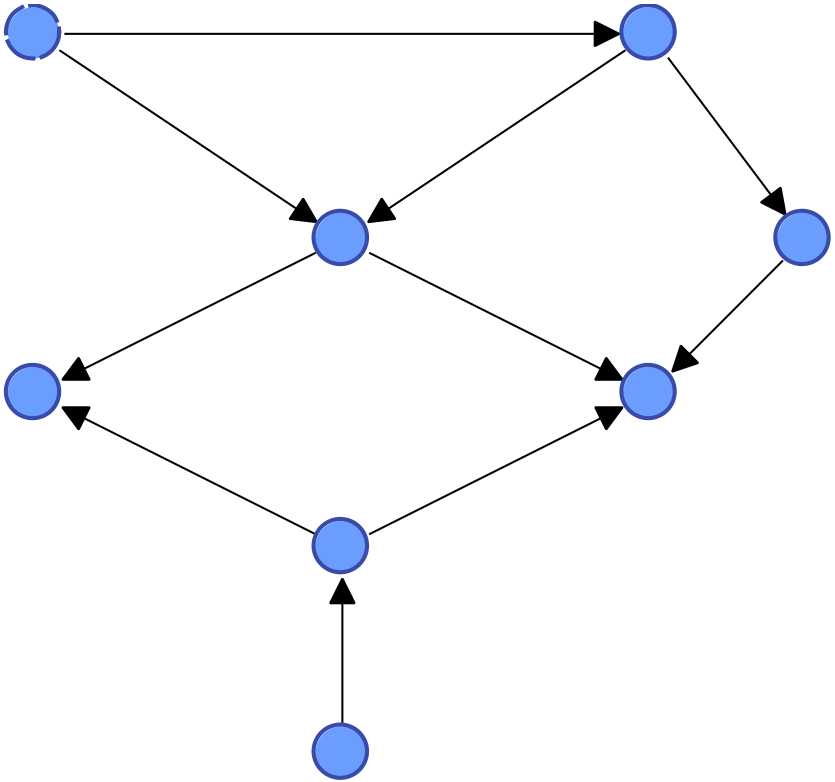
Read Dec. 23, 1763. I Now ſend you an eſſay which I have found among the papers of our deceaſed friend Mr. Bayes, and which, in my opinion, has great merit, and well deſerves to be preſerved. Experimental philoſophy, you will find, is nearly intereſted in the ſubject of it; and on this account there ſeems to be particular reaſon for thinking that a communication of it to the Royal Society cannot be improper.

He had, you know, the honour of being a member of that illuſtrious Society, and was much eſteemed by many in it as a very able mathematician. In an introduction which he has writ to this Eſſay, he ſays, that his deſign at firſt in thinking on the ſubject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumſtances, upon ſuppoſition that we know nothing concerning it but that, under the ſame circum-

The New Paradigm: Bayesian Networks

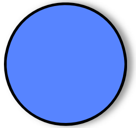
Probabilistic Graphical Model

- The graph ***is*** the model
- No formulas, no equations!



The New Paradigm: Bayesian Networks

Two Components:

- Node 
- Arc 

The New Paradigm: Bayesian Networks

Example

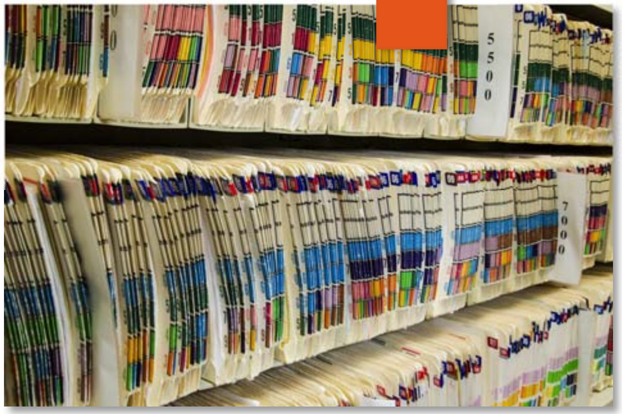
- A specialist in respiratory medicine summarizes his knowledge about his patients.

The New Paradigm: Bayesian Networks

Node:

Interest

Node Name	State No.	State Value	Probability
Age		<30	25%
		30-65	40%
	3	>65	35%



The New Paradigm: Bayesian Networks

Node: Variable of Interest

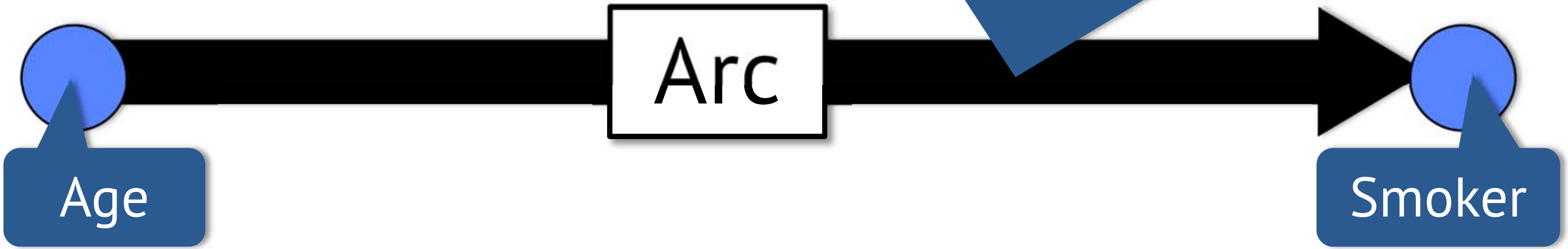
Node Name	State No.	State Value	Probability
Smoker	1	TRUE	53.75%
	2	FALSE	46.25%

Age

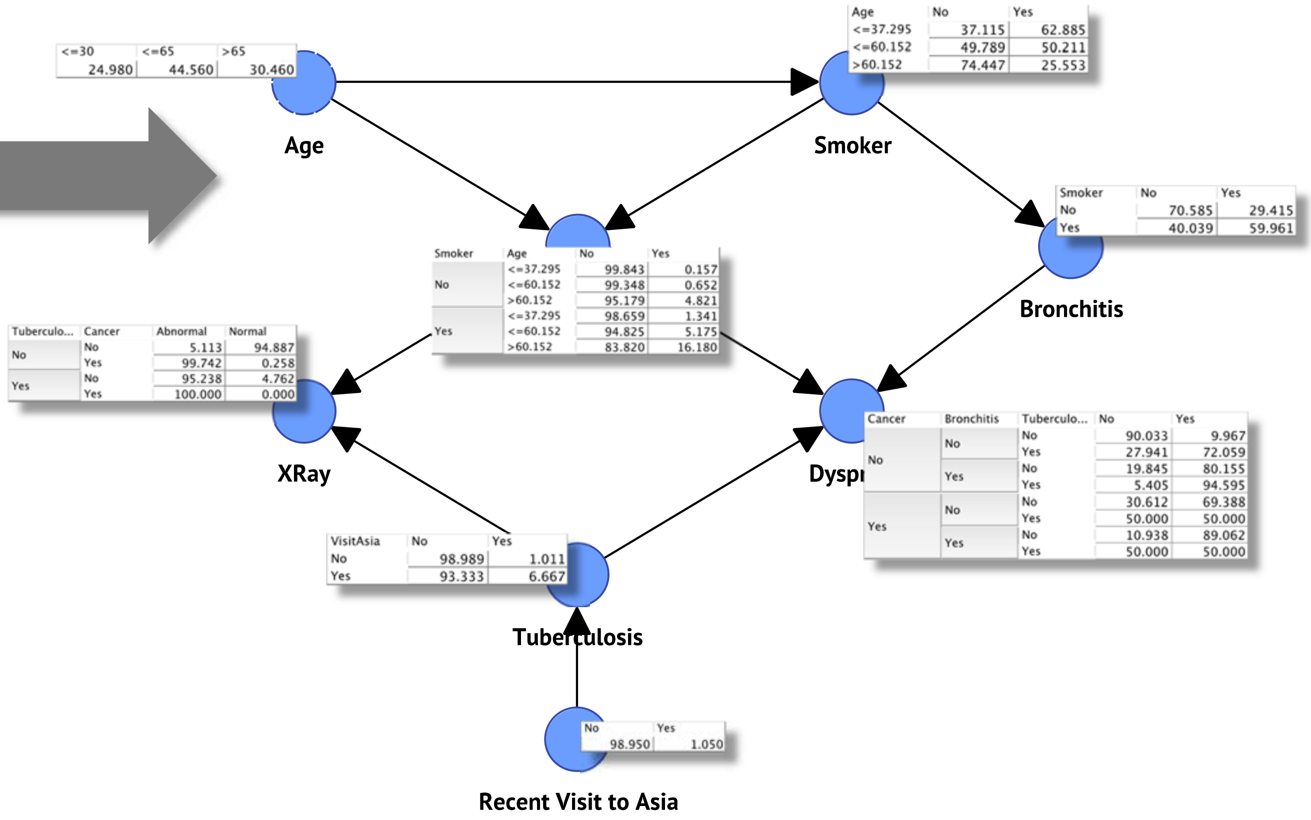
The New Paradigm: Bayesian Networks

Discrete & Nonparametric Probabilistic Relationship
P(Y|X)

	Smoker	
Age	**FALSE**	**TRUE**
<30	31%	69%
30-65	53%	47%
>65	74%	26%



The New Paradigm: Bayesian Networks



The New Paradigm: Bayesian Networks

Key Properties of Bayesian Networks

- Representation (or approximation) of the joint probability distribution of all variables.
- Numerical and categorical variables are treated identically.
- No distinction between dependent and independent variables.
- Nonparametric.

Compare to algebraic formula:

Representation of **one** variable of the joint probability distribution, i.e. y=f(x)

Dependent

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$$

Independent

Independent

The New Paradigm: Bayesian Networks

Key Properties of Bayesian Networks

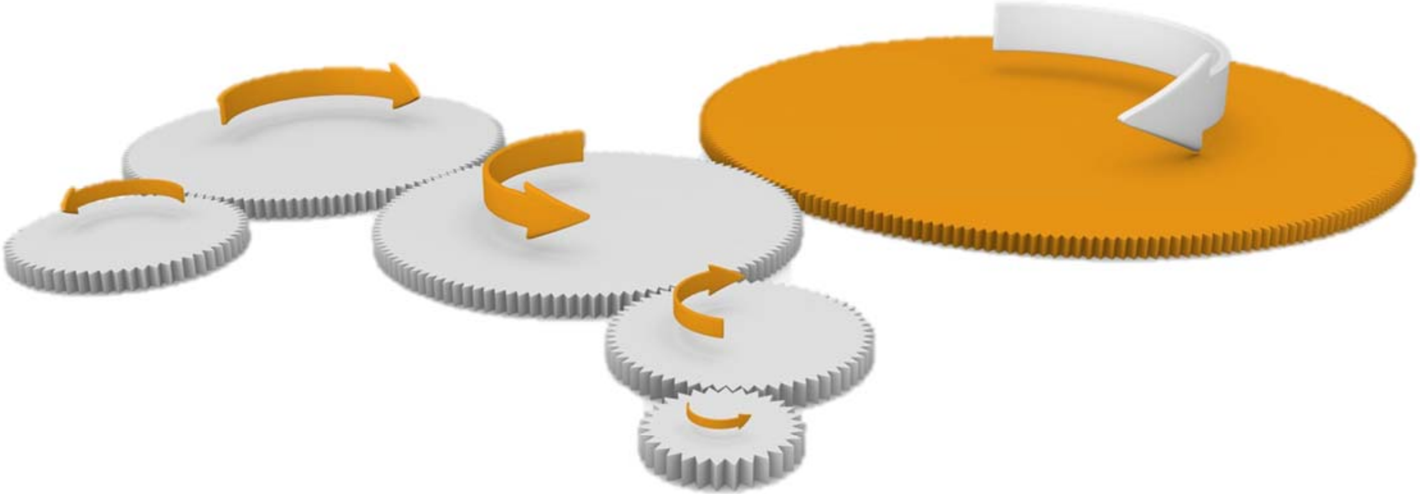
- Omni-directional Inference, i.e. evaluation is always performed in all directions.

Compare to "uni-directional" algebraic formula and human intuition

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$$

The New Paradigm: Bayesian Networks

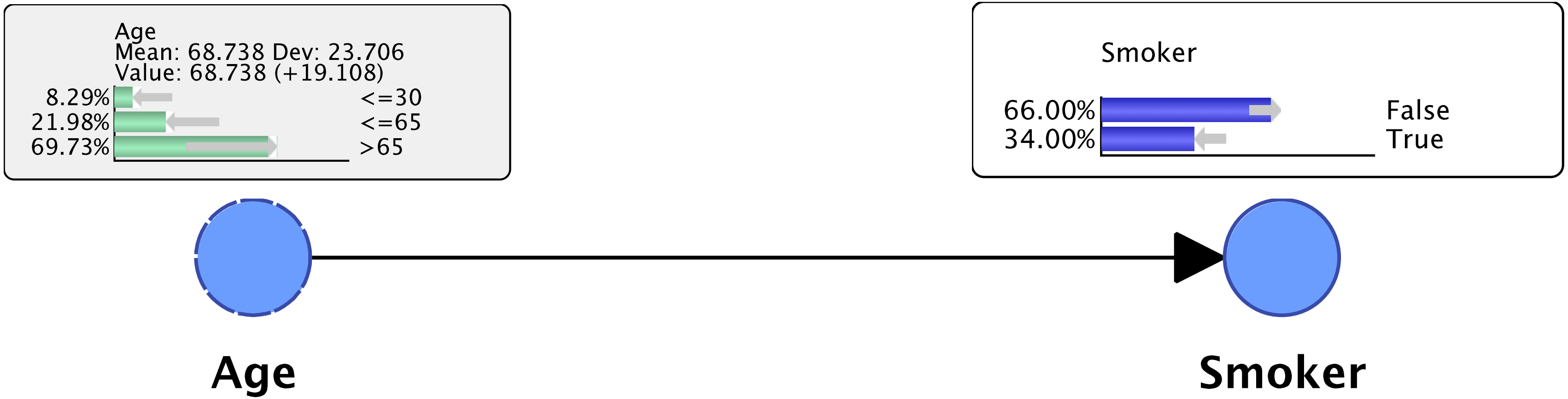
Omni-Directional Inference



The New Paradigm: Bayesian Networks

Key Properties of Bayesian Networks

- Bayesian networks are inherently probabilistic.
- Evidence and inference are represented as distributions.
- Inference can be performed with partial evidence.



The New Paradigm: Bayesian Networks

Key Properties of Bayesian Networks

- Bayesian networks are inherently probabilistic.
- Evidence and inference are represented by distributions.
- Inference can be performed with partial evidence.

Compare to algebra

Deterministic Point Estimate

Single Value Input

Single Value Input

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$$

The New Paradigm: Bayesian Networks

Key Properties of Bayesian Networks

- Bayesian networks can encode causal direction, algebra cannot.

Limitations of Algebra

$$a = \frac{F}{m}$$

Causal Interpretation Possible

$$m = \frac{F}{a}$$

Causal Interpretation **Not** Possible

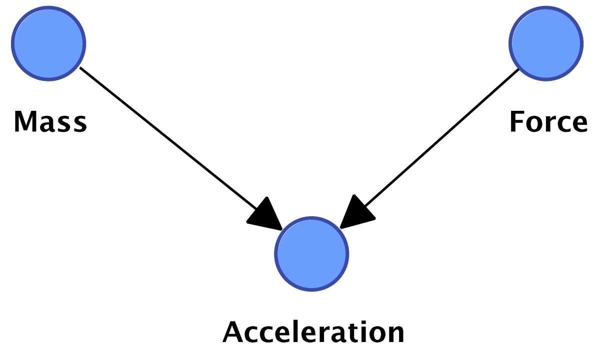
The New Paradigm: Bayesian Networks

Key Properties of Bayesian Networks

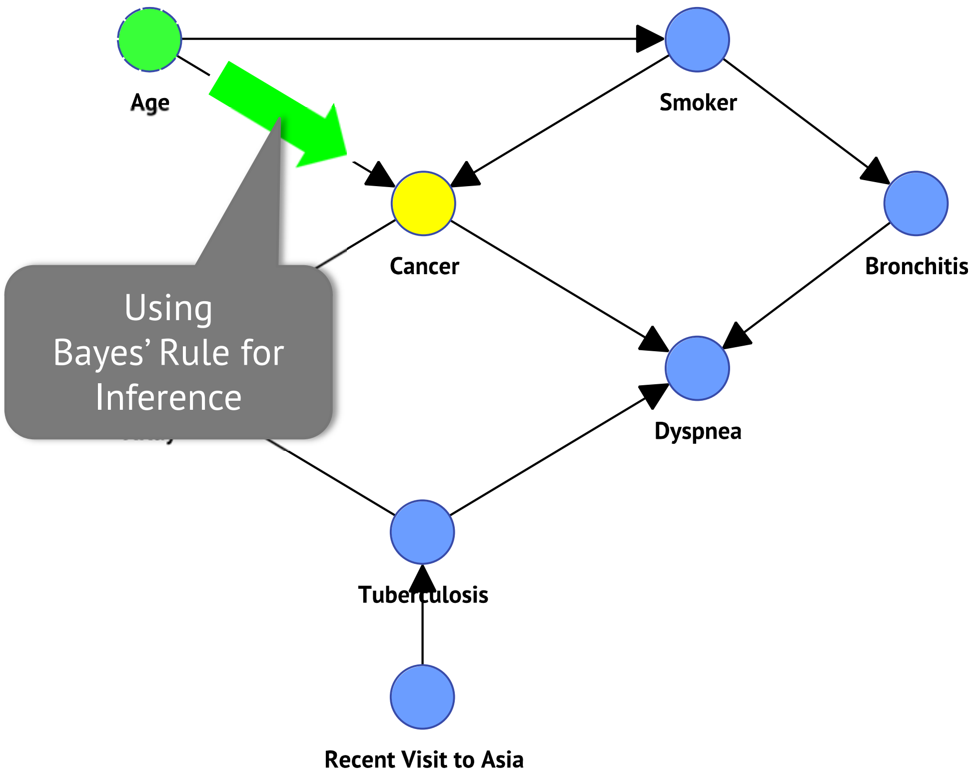
- Bayesian networks can encode causal direction, algebra cannot.

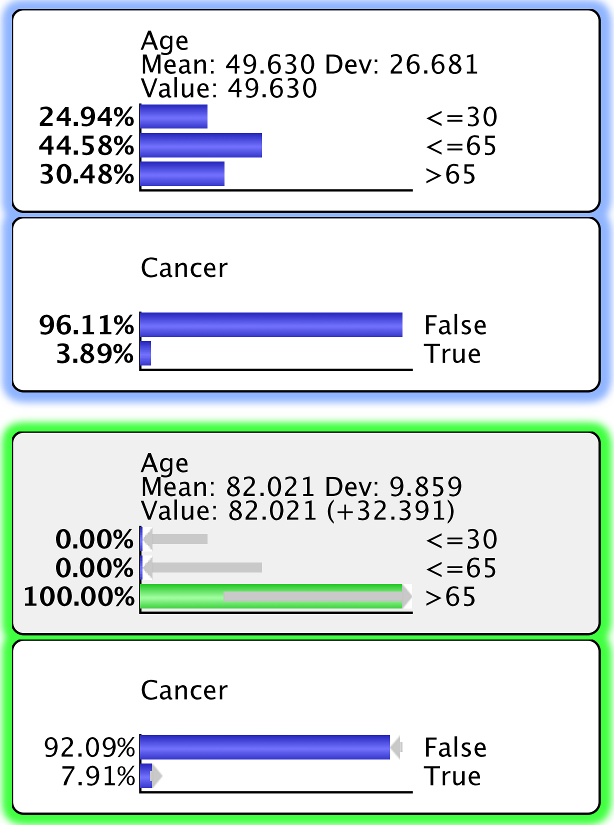
Algebra vs. Bayesian Network

$$a = \frac{F}{m}$$

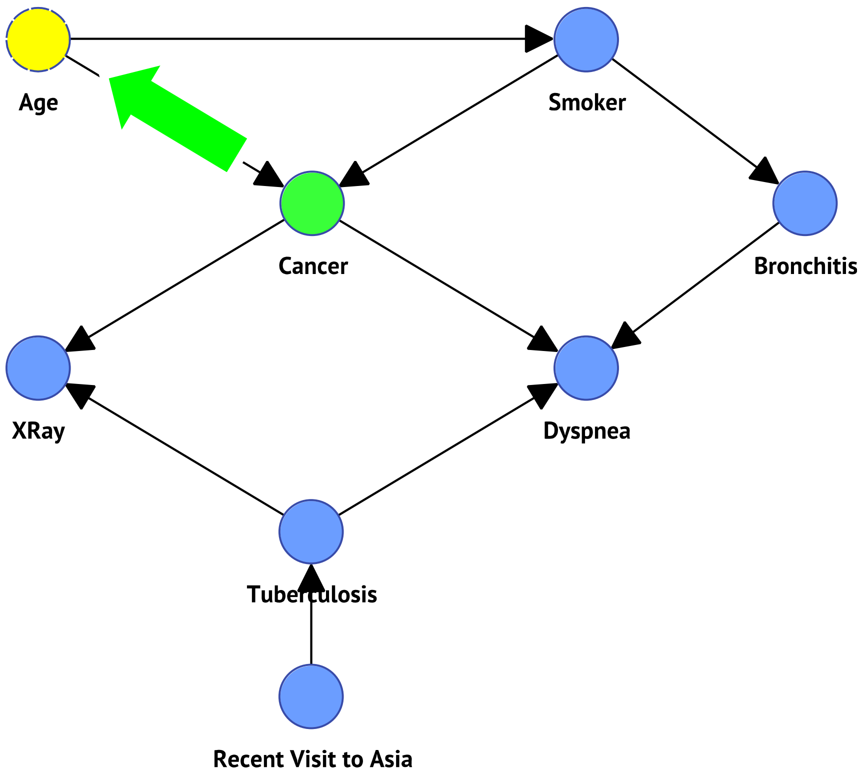


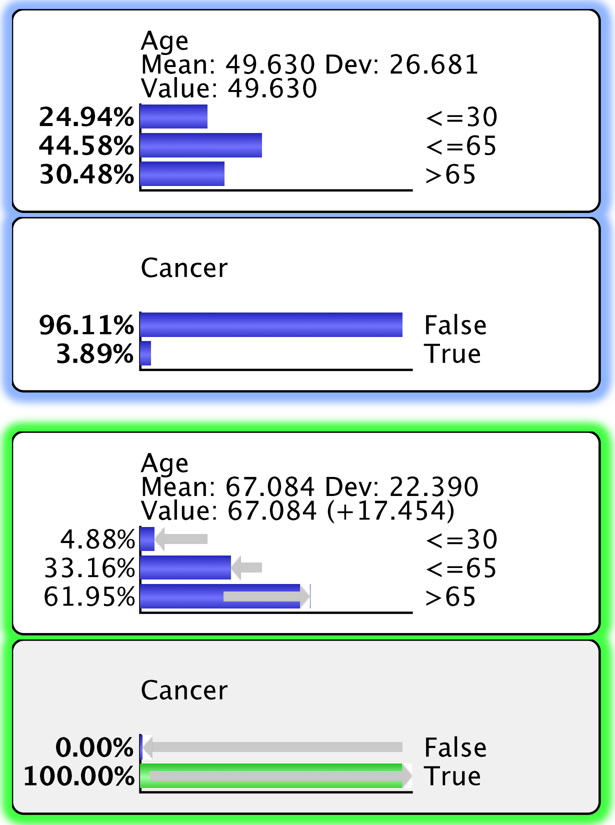
The New Paradigm: Bayesian Networks



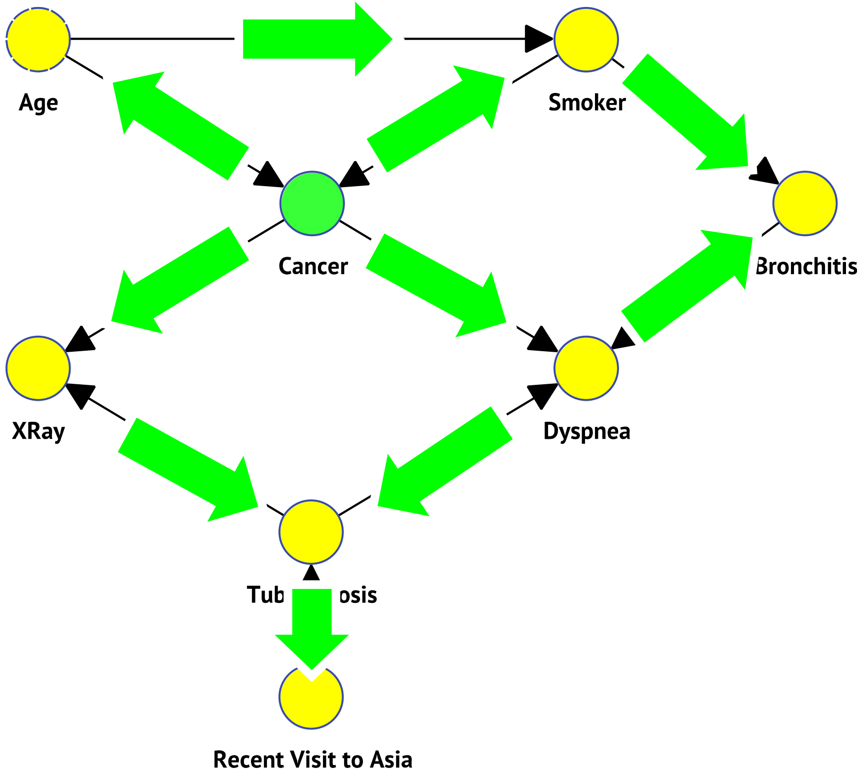


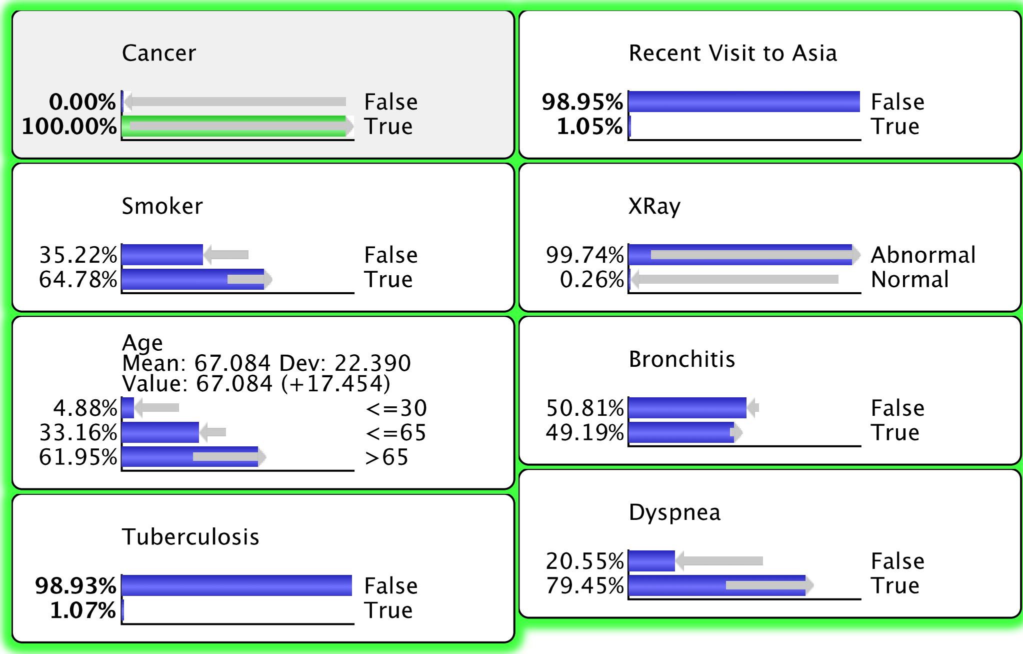
The New Paradigm: Bayesian Networks





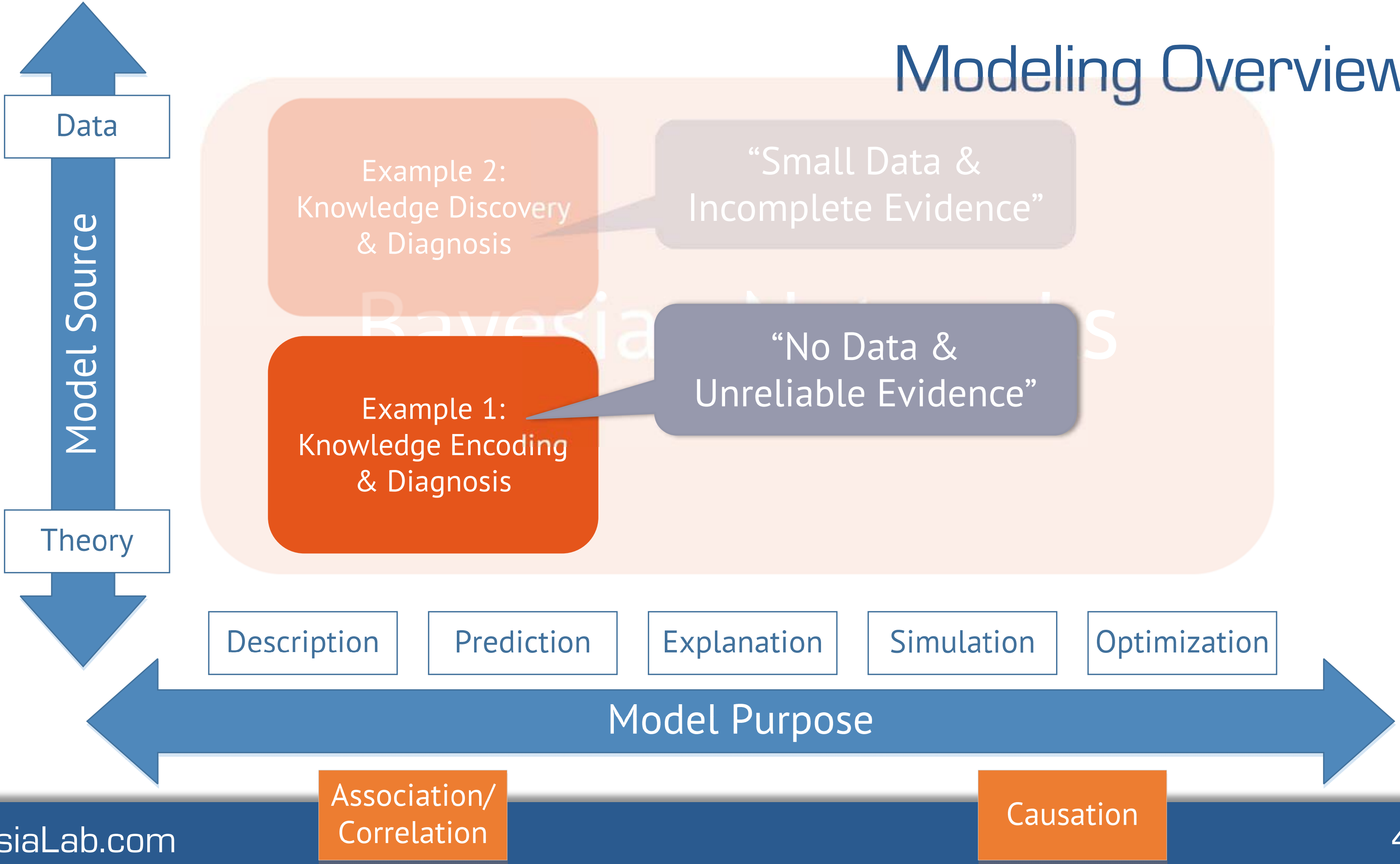
The New Paradigm: Bayesian Networks





Inference Engine

Modeling Overview



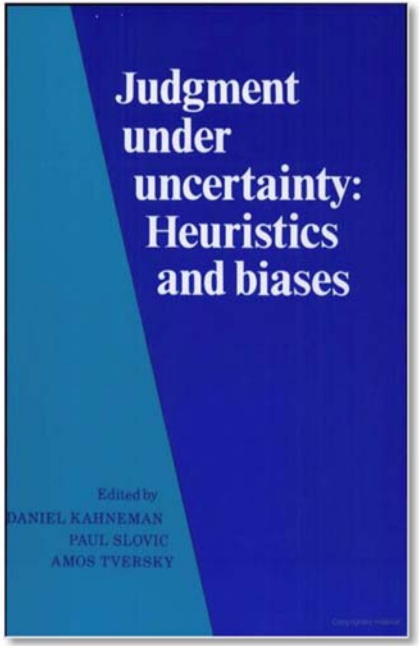
BAYESIALAB

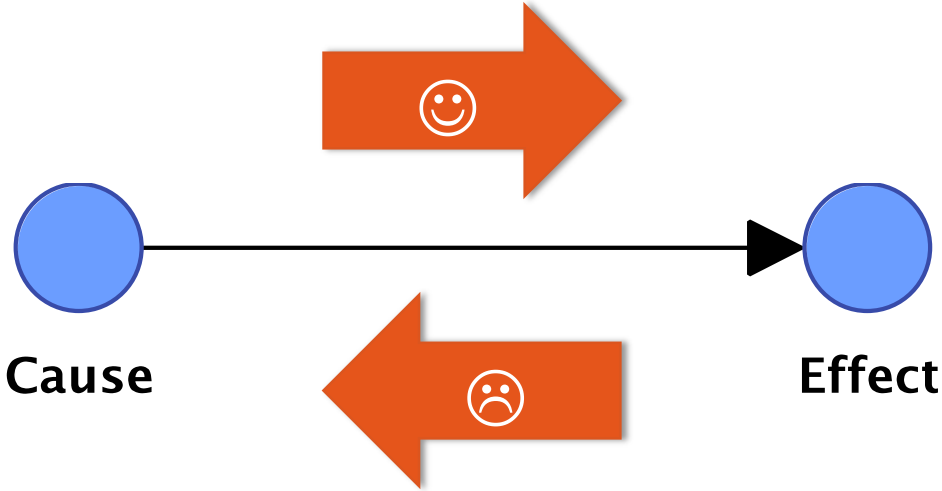
Example 1:
Probabilistic Inference
Taxi Cab Example

Example 1

See Chapter 4

Motivation: Human Biases in Diagnostic Reasoning

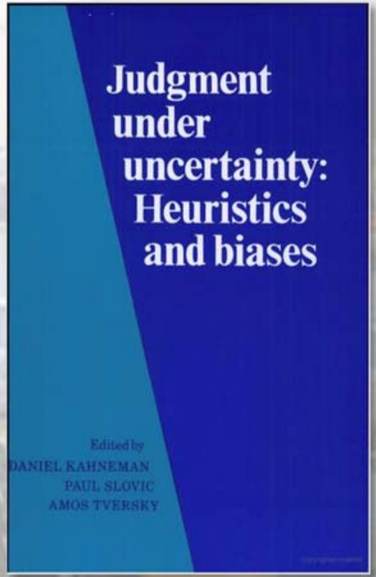




Probabilistic Inference

Human Reasoning Experiment (adapted from Kahneman & Tversky, 1980)

- A cab was involved in a hit-and-run accident at night.
- Two taxicab companies are operating in the city, one with yellow and one with white taxis:
    - 85% are yellow and 15% are white



Probabilistic Inference

- A witness identified the taxi involved in the accident as white…

Probabilistic Inference

At the Trial

- An expert witness explains that human vision has an 80% sensitivity in terms of distinguishing between white and yellow given light conditions at the time of the accident.
- What is the probability that the taxi was actually white?

Probabilistic Inference

- We need to perform diagnostic probabilistic inference, i.e. from effect to cause, to answer this question.
- The Bayes Rule allows us to compute the probability:

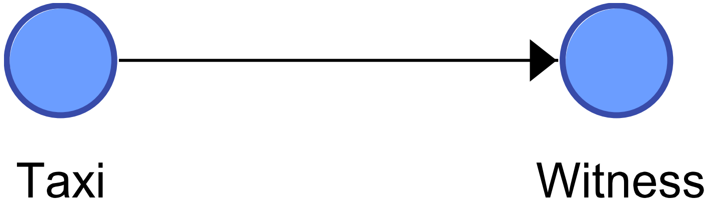
$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)}$$

$$P(Taxi = white \mid Witness = white) = \frac{P(Witness = white \mid Taxi = white)P(Taxi = white)}{P(Witness = white)} =$$

$$\frac{P(Witness = white \mid Taxi = white)P(Taxi = white)}{P(Witness = white \mid Taxi = white)P(Taxi = white) + P(Witness = white \mid Taxi = yellow)P(Taxi = yellow)}$$

Probabilistic Inference

Representing our domain knowledge in the form of a simple Bayesian network



Yellow	White
85.000	15.000

Marginal Distribution

Taxi	Yellow	White
Yellow	80.000	20.000
White	20.000	80.000

Conditional Probability Table

Probabilistic Inference

Carrying out inference based on observing evidence

Taxi

Yellow	White
85.000	15.000

Witness

Taxi	Yellow	White
Yellow	80.000	20.000
White	20.000	80.000

Probabilistic Inference

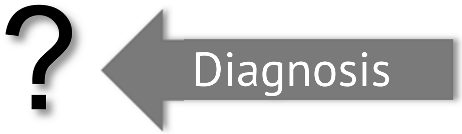
Carrying out inference based on observing evidence

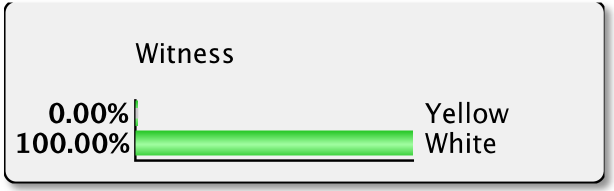
Taxi

Yellow	White
85.000	15.000

Witness

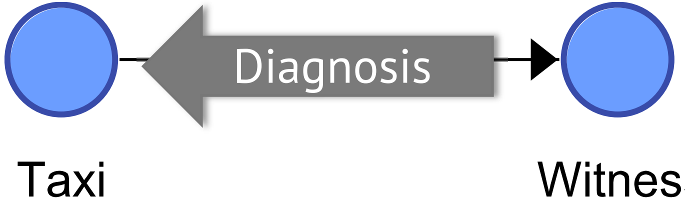
Taxi	Yellow	White
Yellow	80.000	20.000
White	20.000	80.000





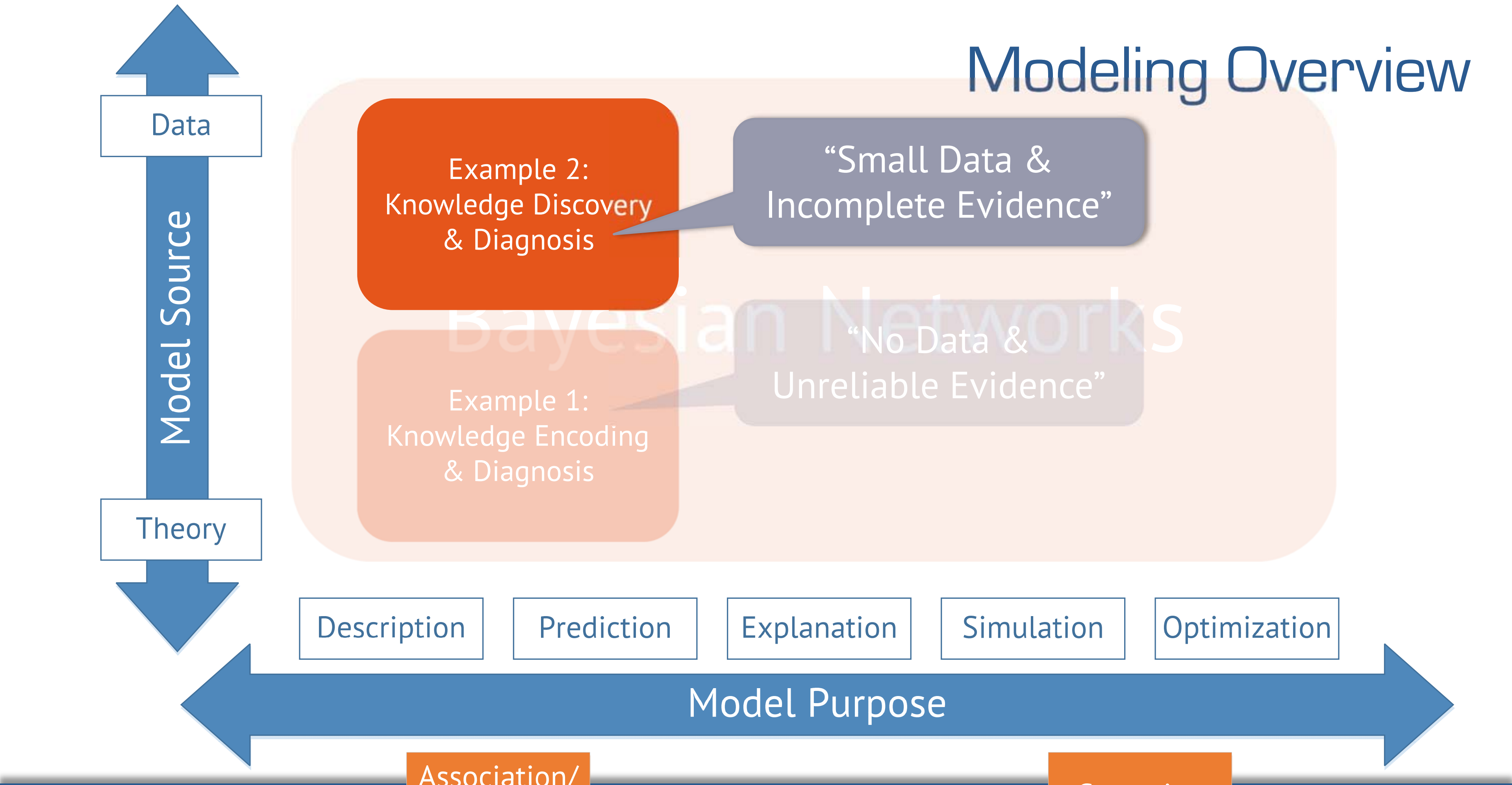
Probabilistic Inference

Carrying out inference based on observing evidence



Yellow	White
85.000	15.000

Taxi	Yellow	White
Yellow	80.000	20.000
White	20.000	80.000



Modeling Overview
Data
Model Source
Theory
Example 2:
Knowledge Discovery
& Diagnosis
“Small Data &
Incomplete Evidence”
Bayesian Networks
Example 1:
Knowledge Encoding
& Diagnosis
“No Data &
Unreliable Evidence”
Description
Prediction
Explanation
Simulation
Optimization
Model Purpose
Association/
Correlation
Causation

Example 2: Breast Cancer Diagnostics

Supervised Learning

See Chapter 6

Breast Cancer Diagnostics

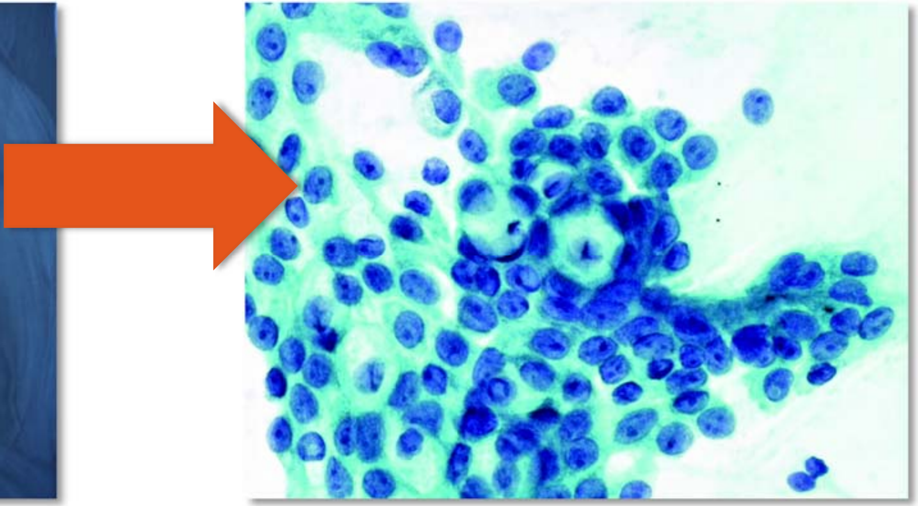
Background for Original Study (Wolberg et al.)

- Challenge in Breast Cancer Diagnostics:
    - Mammography lacks sensitivity (i.e. true positive rate): 68% to 79%;
    - Surgical biopsy has high sensitivity (>98%), but invasive, time-consuming and costly;

Breast Cancer Diagnostics

Image Analysis of Fine Needle Aspirates

- Sensitivity of Fine Needle Aspiration with visual interpretation varies widely (65% to 98%)



Breast Cancer Diagnostics

Image Analysis of Fine Needle Aspirates

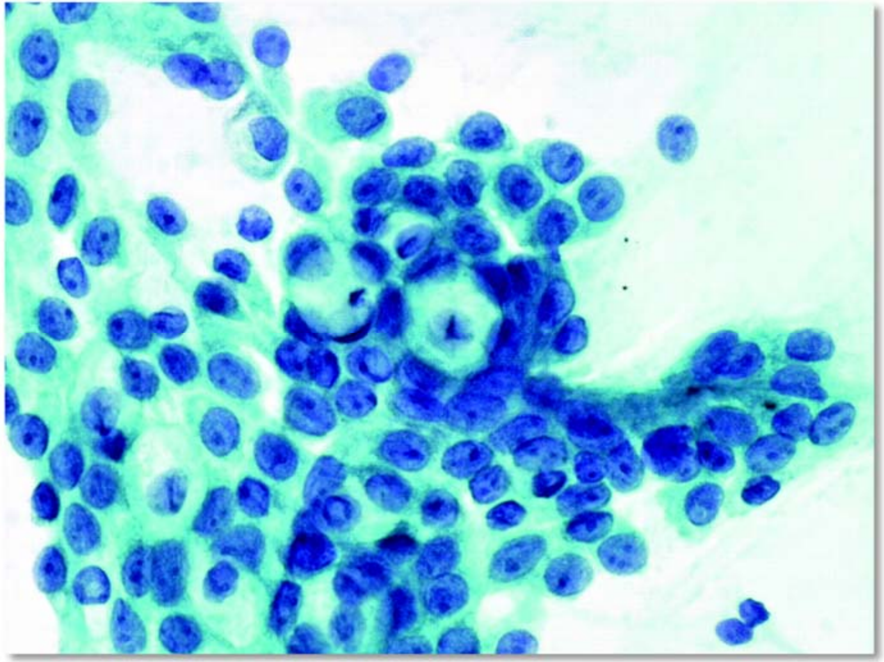
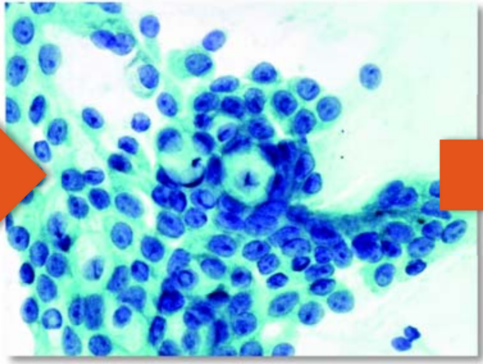
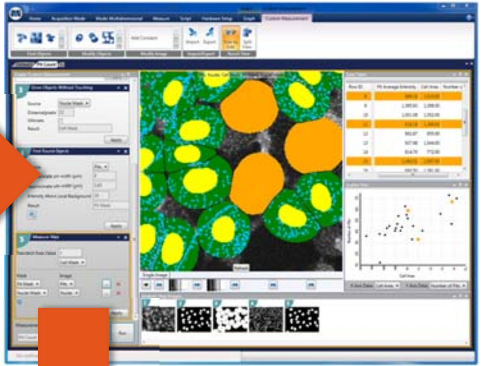


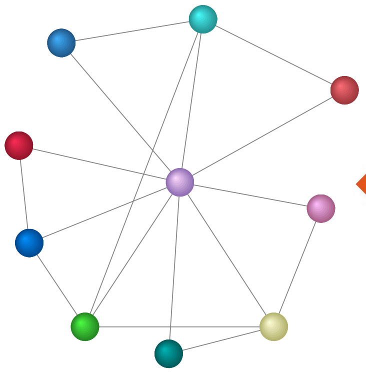
Image Attributes

- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape
- Marginal Adhesion
- Single Epithelial Cell Size
- Bare Nuclei
- Bland Chromatin
- Normal Nucleoli
- Mitoses

Overview







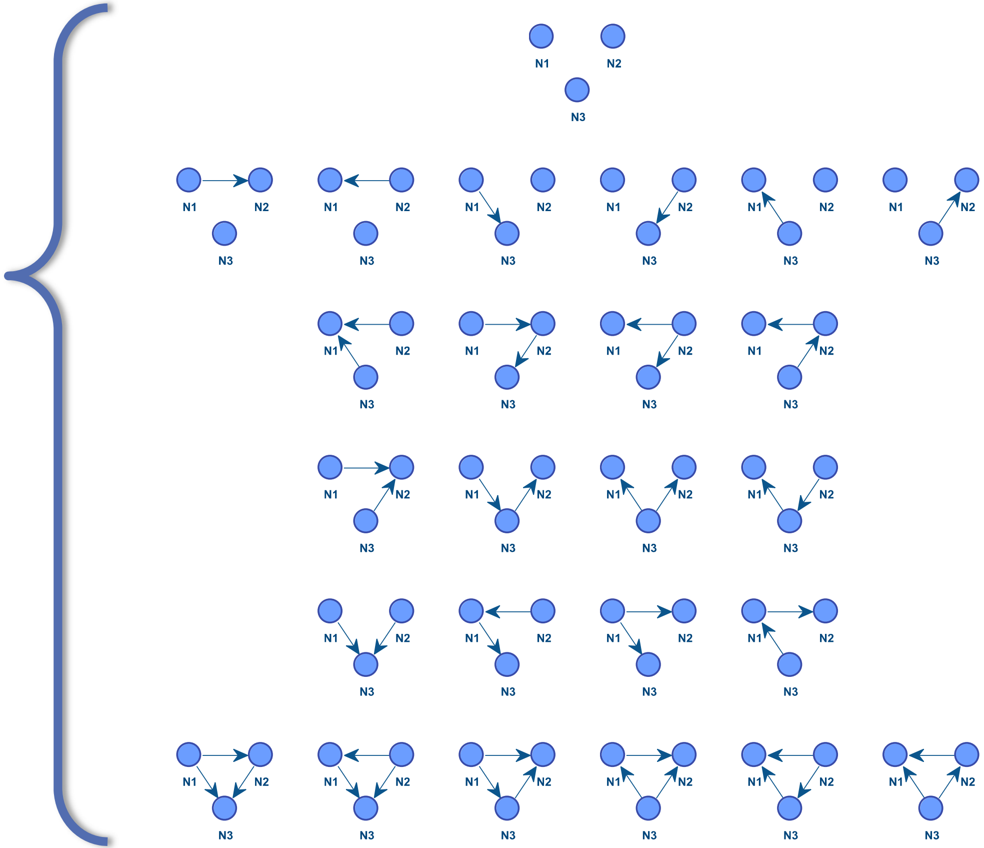
Sample Code number	Clump Thickness	Uniformity of Cell Size	Uniformity of Cell Shape	Marginal Adhesion	Single Epithelial Cell Size	Bare Nuclei	Bland Chromatin	Normal Nucleoli	Mitoses	Class
1000025	5	1	1	1	2	1	3	1	1	2
1002945	5	4	4	5	7	10	3	2	1	2
1015425	3	1	1	1	2	2	3	1	1	2
1016277	6	8	8	1	3	4	3	7	1	2
1017023	4	1	1	3	2	1	3	1	1	2
1017122	8	10	10	8	7	10	9	7	1	4
1018099	1	1	1	1	2	10	3	1	1	2
1018561	2	1	2	1	2	1	3	1	1	2
1033078	2	1	1	1	2	1	1	1	5	2

Wisconsin Breast Cancer Database

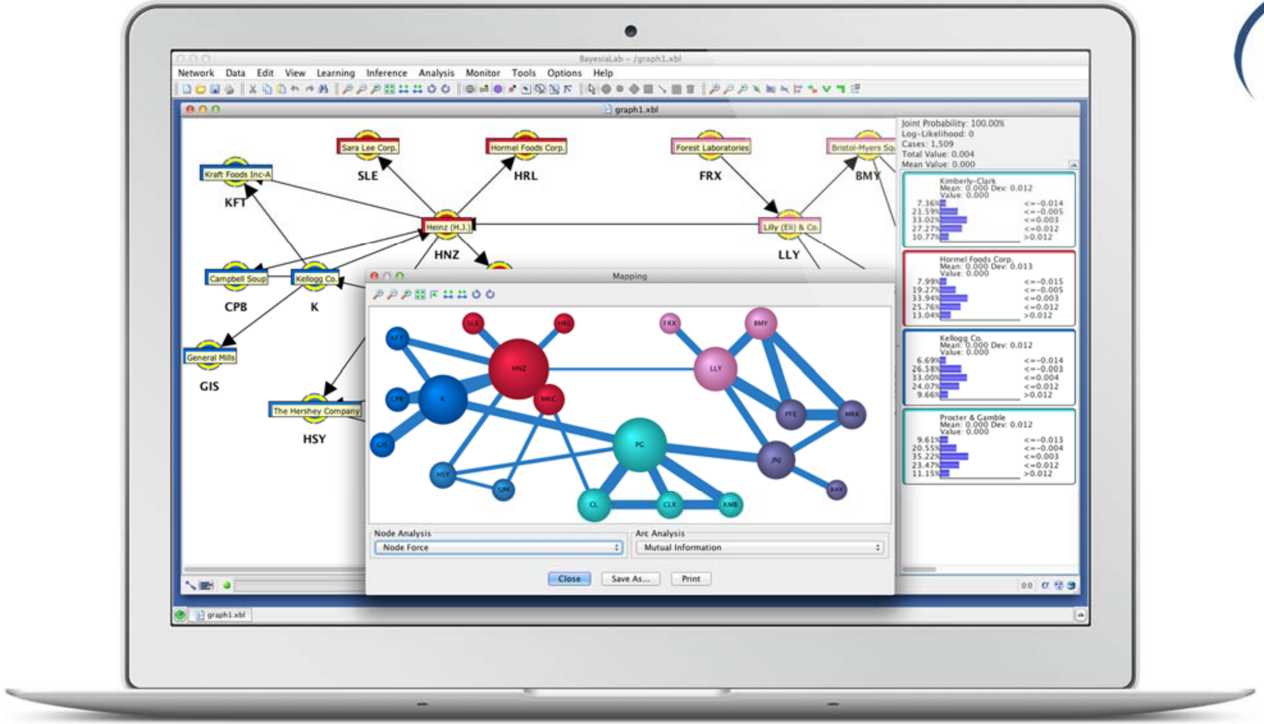
Challenges

Number of Possible Networks

- 2 Nodes: 3
- **3 Nodes: 25**
- 4 Nodes: 543
- 5 Nodes: 29,281
- 6 Nodes: 3.8×10^{6}
- 7 Nodes: 1.1×10^{9}
- 8 Nodes: 7.8×10^{11}
- 9 Nodes: 1.2×10^{15}
- 10 Nodes: 4.2×10^{18}
- 11 Nodes: 3.2×10^{22}
- 12 Nodes: 5.2×10^{26}



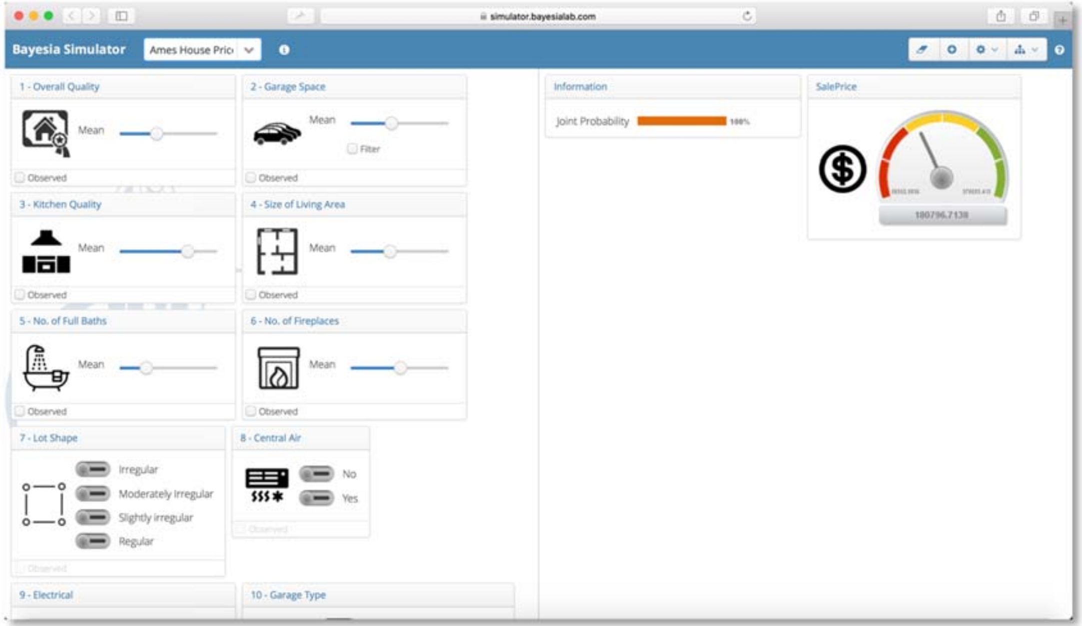
The BayesiaLab Software Platform

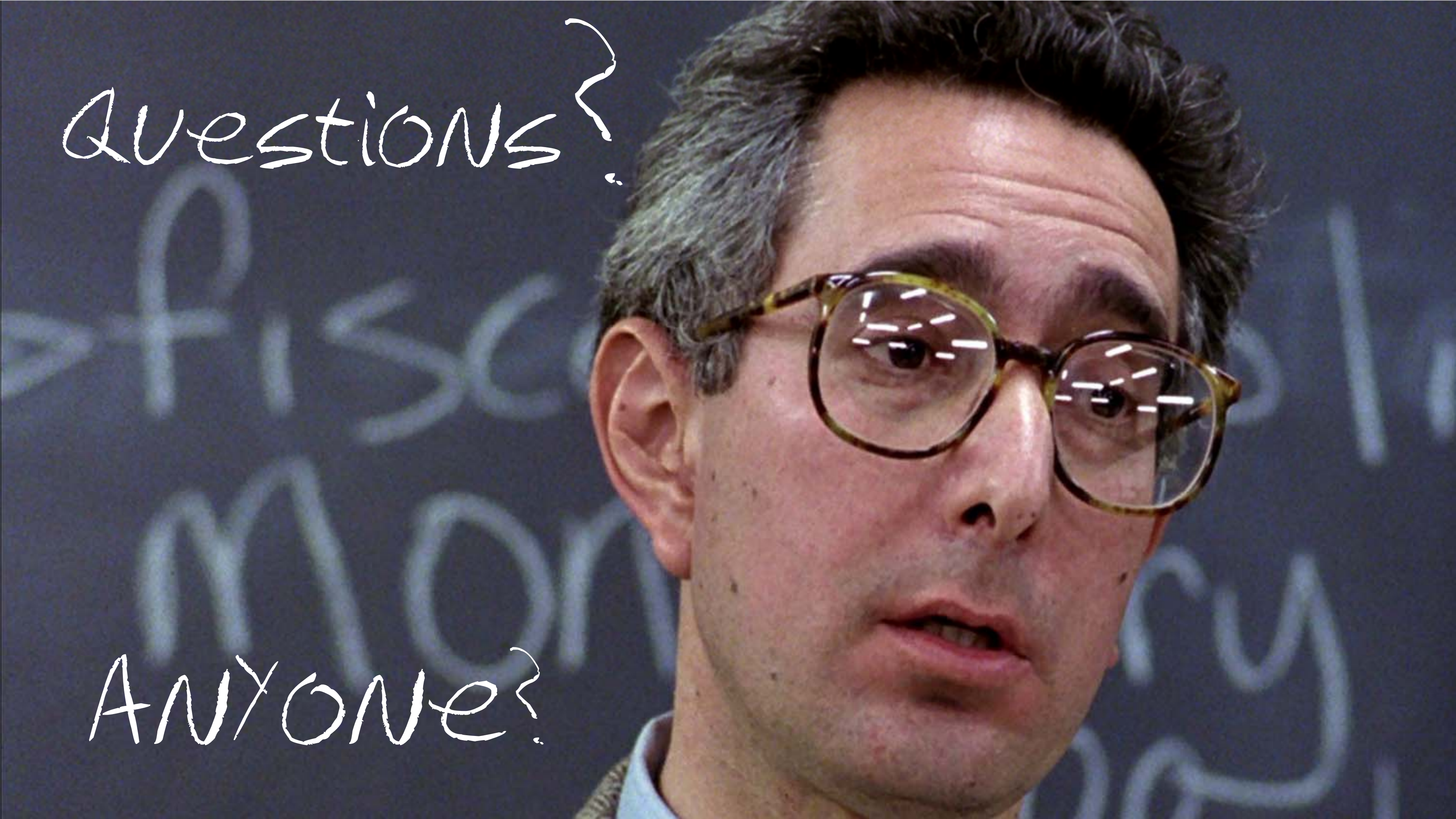


- learning
- editing
- inference
- analysis
- simulation
- optimization
- publication

Breast Cancer Diagnostics

BayesiaLab WebSimulator




Questions?
Anyone?

Bayesian Networks & BayesiaLab

A Practical Introduction for Researchers

- Free download:
  www.bayesia.com/book
- Hardcopy available on Amazon:
  http://amzn.com/0996533303

amazon

Thank You!

stefan.conrady@bayesia.us

BayesianNetwork

linkedin.com/in/stefanconrady

facebook.com/bayesia