Metodologie

(Module 2)

Science IFT3700 data /

IFT6700 Fall 2018 ©

Alain Tapp

Contents

- scientfque method
- Dataset
- Model (parameters)
- Bayesian probability
- Hypothesis testing (P-values)
- statstque recall and Corrélaton
- causal Relaton

scientifc

method

Psychological Assessment
2000, Vol. 12, No. 1, 19–30

1040-3590/00/$5.00 DOI: 10.1037//1040-3590.12.1.19

Clinical Versus Mechanical Prediction: A Meta-Analysis

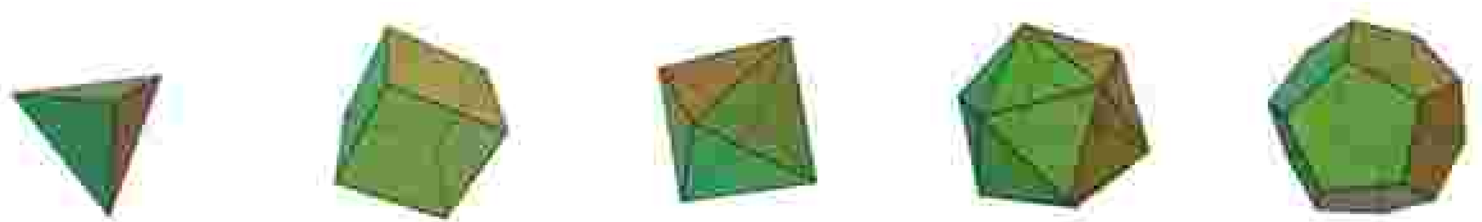
William M. Grove, David H. Zald, Boyd S. Lebow, Beth E. Snitz, and Chad Nelson
University of Minnesota, Twin Cities Campus

The process of making judgments and decisions requires a method for combining data. To compare the accuracy of clinical and mechanical (formal, statistical) data-combination techniques, we performed a meta-analysis on studies of human health and behavior. On average, mechanical-prediction techniques were about 10% more accurate than clinical predictions. Depending on the specific analysis, mechanical prediction substantially outperformed clinical prediction in 33%–47% of studies examined. Although clinical predictions were often as accurate as mechanical predictions, in only a few studies (6%–16%) were they substantially more accurate. Superiority for mechanical-prediction techniques was consistent, regardless of the judgment task, type of judges, judges' amounts of experience, or the types of data being combined. Clinical predictions performed relatively less well when predictors included clinical interview data. These data indicate that mechanical predictions of human behaviors are equal or superior to clinical prediction methods for a wide range of circumstances.

The axiomatic method

Euclid's Elements

- around 300 BC. AD
- Masterpiece of axiomatque method
- Pythagore's theorem
- Only five Platonic solids



Organum of Aristotle (384 BC. 322 BC. AD)

If I state the following: " **All Men Are Mortal, gold Leon is a man, so Leo is mortal "** no one will find nothing wrong. But if I say "All cats fours, or my dog fours, so my dog is a cat," we retort that I lack of logic! Similarly, if I say, "All men are mortal, gold Leon is deadly, so Leo is a man", my conclusion is not correct because Leo could be an animal (there are not that men die). What diference there-he has diferent between these sentences? Why only the first can it be considered a valid argument? These statements are syllogisms. Aristotle formalized them precisely in Book III of the Organum, which was named The First Analytques. It defnes syllogism thus: "The syllogism is a reasoning where some things are proven,

HTPS: //lagazetedeventamicena.blogspot.com/2015/10/les-syllogismes-dans-la-logique-d.html

Argument

$$2 + 3 = 5$$

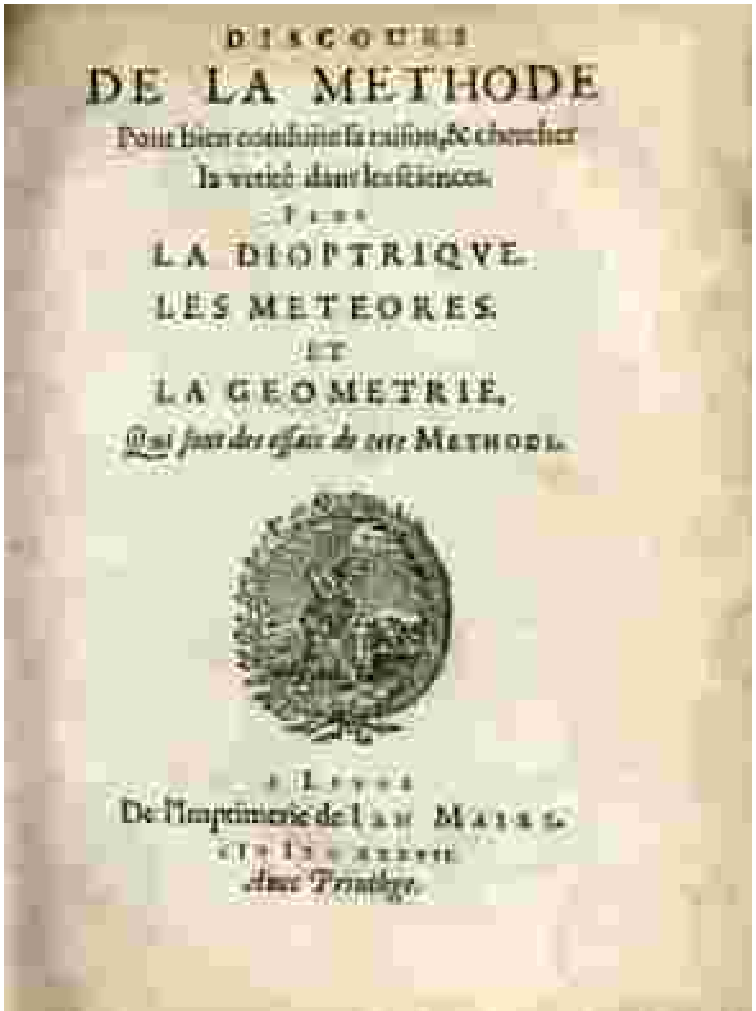
for 2, 3 and 5 are the first three primes consécutfs

invalid argument

	TRUE	FALSE
VALID	YES	NO
INVALID	YES	YES

Propaganda: show a fool with incorrect proof of non-A to convince A.

Discourse on Method
René Descartes
1637



DISCOURS
DE LA METHODE
Pour bien conduire sa raison, & chercher
la verité dans les sciences.
PLUS
LA DIOPTRIQUE.
LES METEORES.
ET
LA GEOMETRIE.
Qui sont des essais de cete METHODE.

A LEYDE
De l'Imprimerie de IAN MAIRE.
Avec Privilege.

Discourse on Method
René Descartes
1637

1. Do not accept anything as true as his spirit aura clearly and distinctly treated previously.

2. Divide each difficulty to better address and resolve them.

3. Establish an order of thoughts, starting with the simplest objects to the most complex and diverse, and so retain all and in order.

4. Place all the things reviewed to omit nothing.

Discourse on Method
René Descartes
1637

"Furthermore, I would like them ft consider that the great artery and the arterial vein are of compositon much harder and firmer than are venous artery and the vena cava; and that these two expand before entering the heart, and there are like two scholarships named Heart earrings, which are composed of flesh like his; and **there is always more heat in the heart than in any other part of the body,** and fnally that cete heat is able to do that, if he enters a drop of blood in its cavities, it promptly ENFE and expands, and generally do all the liquors, when allowed to fall drop by tasted in a vessel which is very hot.

William Harvey, English physician, is famous for the discovery of the blood circulaton exposed in his book *De motu cordi (* 1628).

"The science is straightorward contre box: much of the literature scientfc, Perhaps half, May simply be untrue. Aficted by studies with small sample sizes, tny efects, invalid exploratory analyzes, and fagrant conficts of interest, together with an obsession for Pursuing fashionable trends of dubious importance, science HAS taken a turn Towards Darkness. "

Richard Horton editor of *The Lancet.*
www.thelancet.com Vol 385 April 11, 2015

http: //rsos.royalsocietypublishing.org/content/1/3/100216

"If you use p = 0.05 to suggest That You-have made a discovery, you will be wrong at least 30% of the tme. If, as is ofen the box, experiments are underpowered, you will be wrong MOST of the tme. "

David Colquhoun

An investgaton of the false discovery rate and the p-values of misinterpretaton Published November 19 2010.DOI: 10.1098 / rsos.100216

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the nature international weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

News & Comment > News > 2015 > September > Article

E-alert | RSS | Facebook | Twitter

NATURE | NEWS

Over half of psychology studies fail reproducibility test

Largest replication study to date casts doubt on many published positive results.

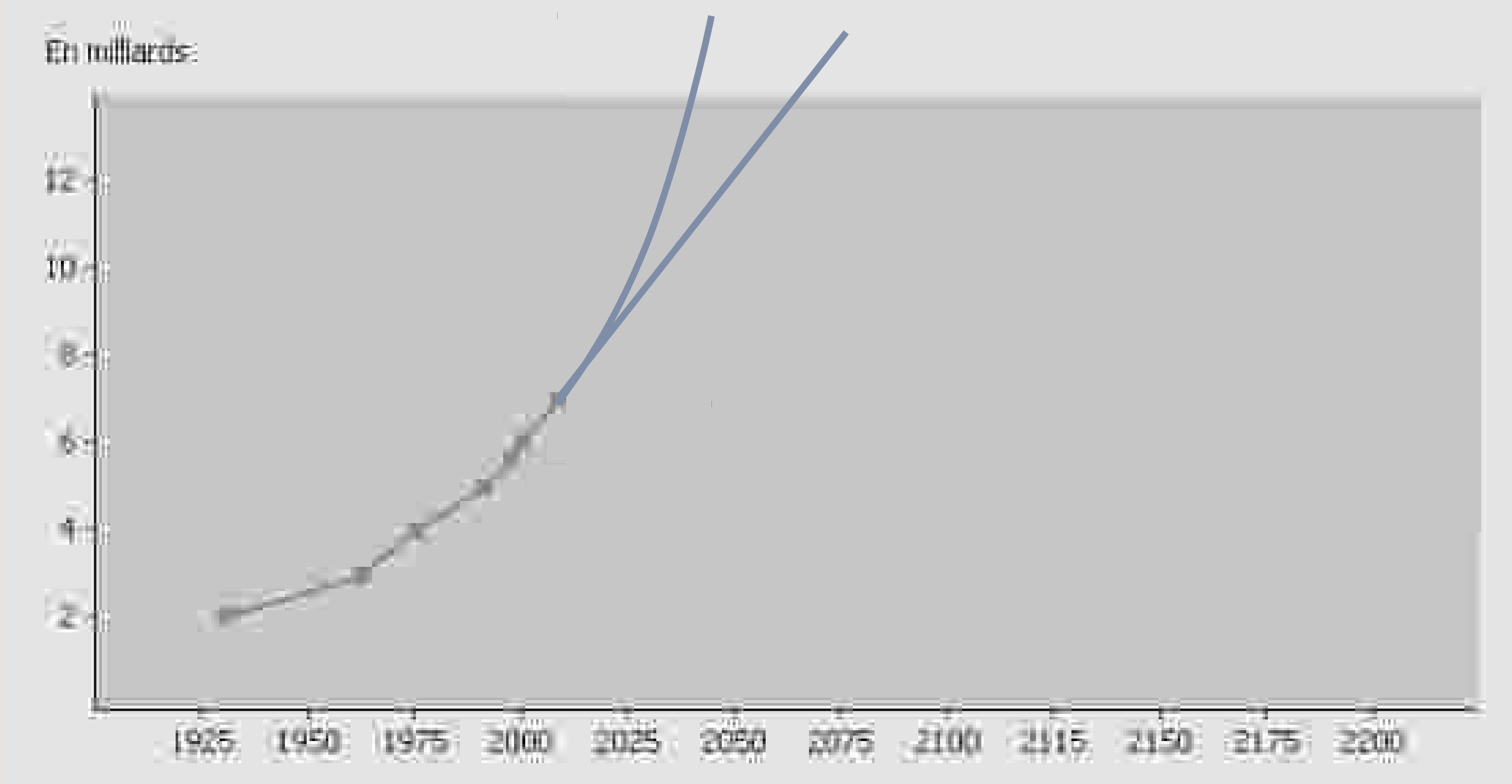
Monya Baker

27 August 2015

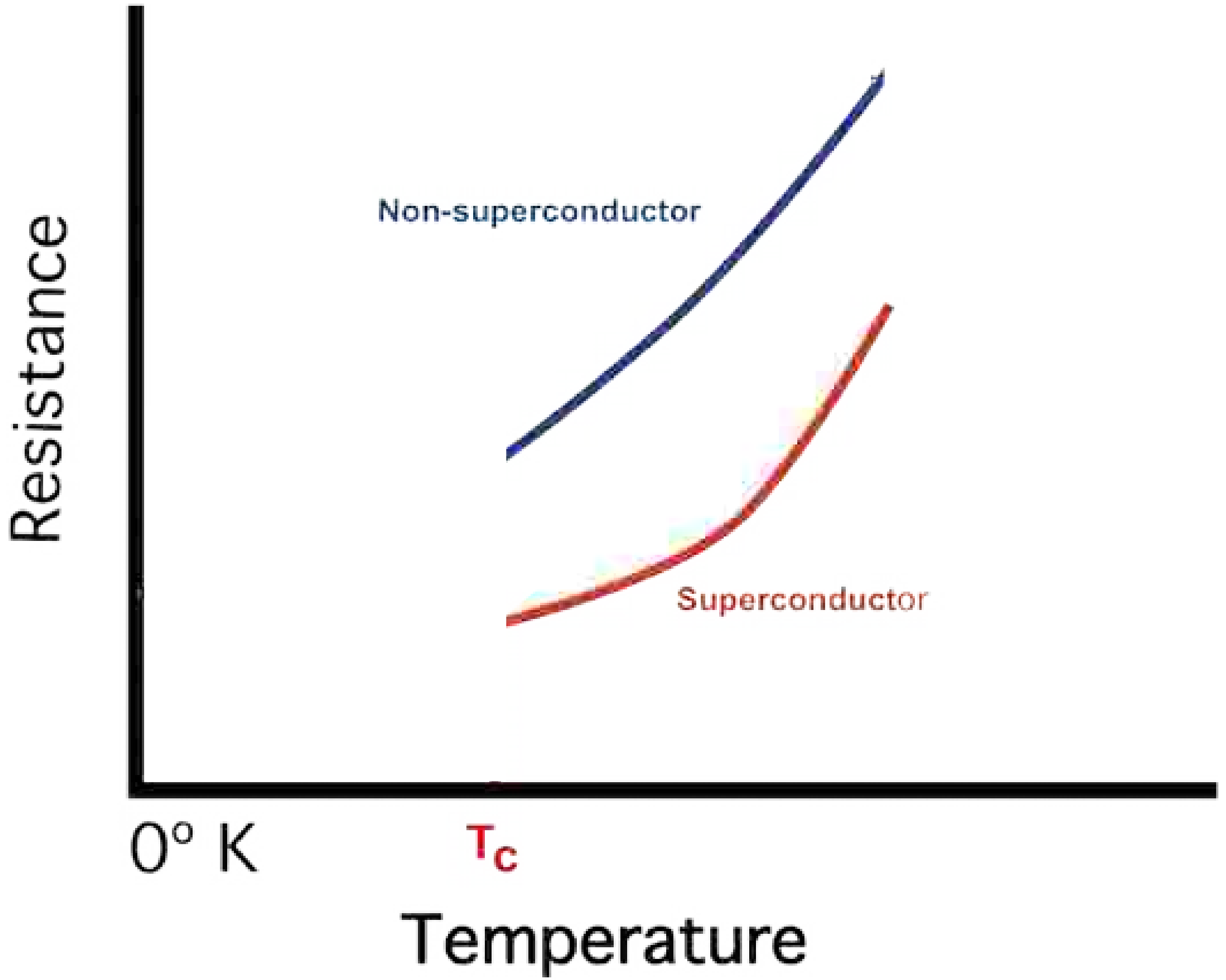
Rights & Permissions

nature briefing

interpolation versus projection



interpolation versus projection



Denying the antecedent

Falsely conclude that A implies B B implies A. Ex. Judging the

evidence on the validity of the answer.

The **through rétrospectf (** *Hindsight bias)* is a miscalculation cognitf designating the tendency of people to surestmer rétrospectvement that the events could have been antcipés means more foresight or clairvoyance. Ex. Numerous lawsuits against doctors.

Ex. The media ofer a explicaton fellows events without ever being able to predict.

The **Through result (** *Outcome bias)* is an error in Evaluaton the quality of a decision when the result of cete decision is already known. Ex. Justce, murder tentatve vs murder

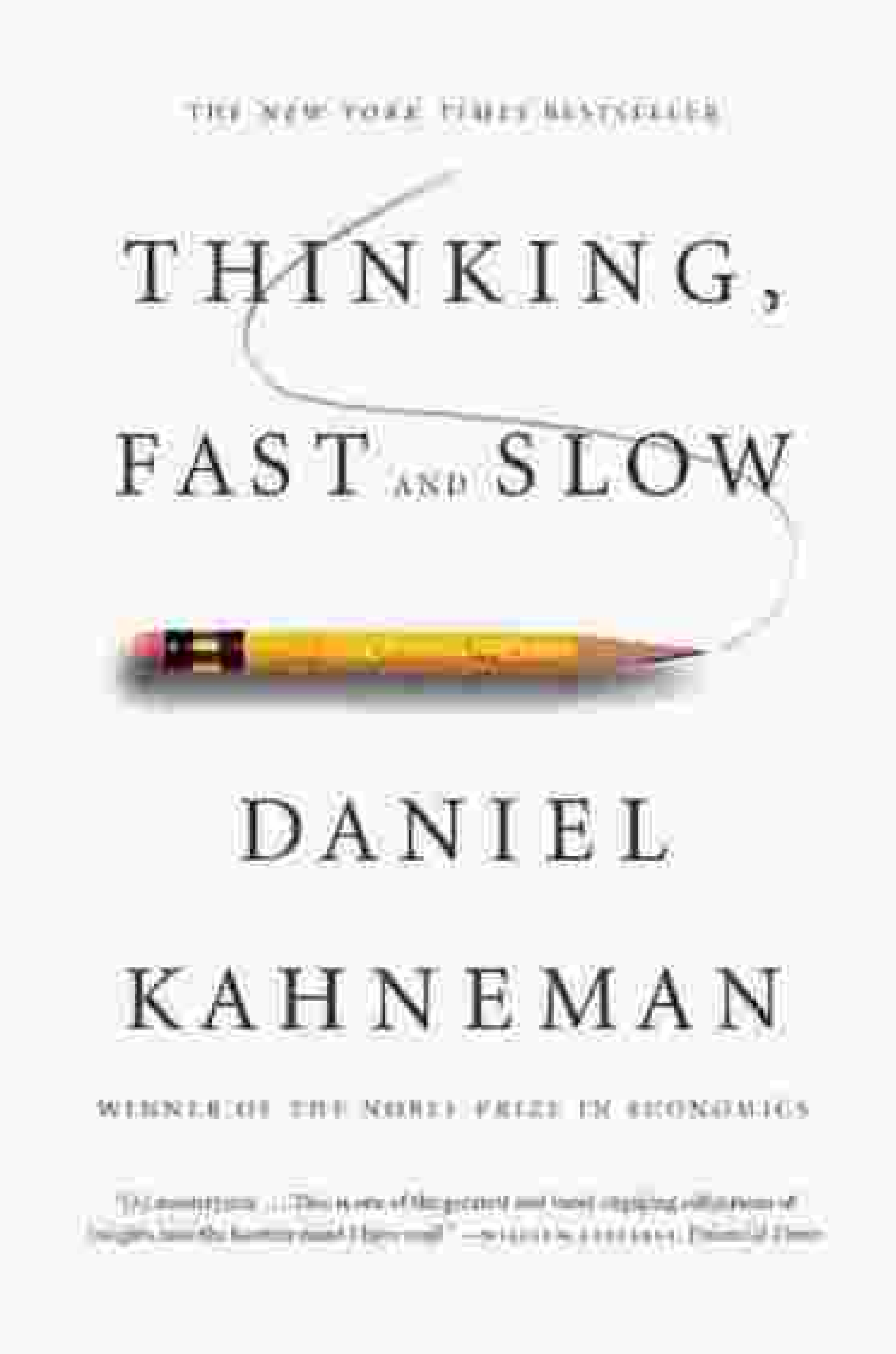
Decision-making, belief, and behavioral Biases

Ambiguity efect
Anchoring gold focalism
anthropocentric thinking
Anthropomorphism gold personifcaton
Atentonal bias
Automaton bias heuristc
Availability Availability
waterfall Backfre efect
Bandwagon efect
Base rate fallacy or base rate neglect
Belief bias
Ben Franklin efect
Berkson's paradox Bias
blind spot Bystander
efect
Choice-supportve bias
clustering illusion
Confrmaton bias congruence
bias Conjuncton fallacy
Conservatsm (belief revision)
Contnued infuence Contrast efect
efect Courtesy bias
Curse of Knowledge
Declinism Decoy efect

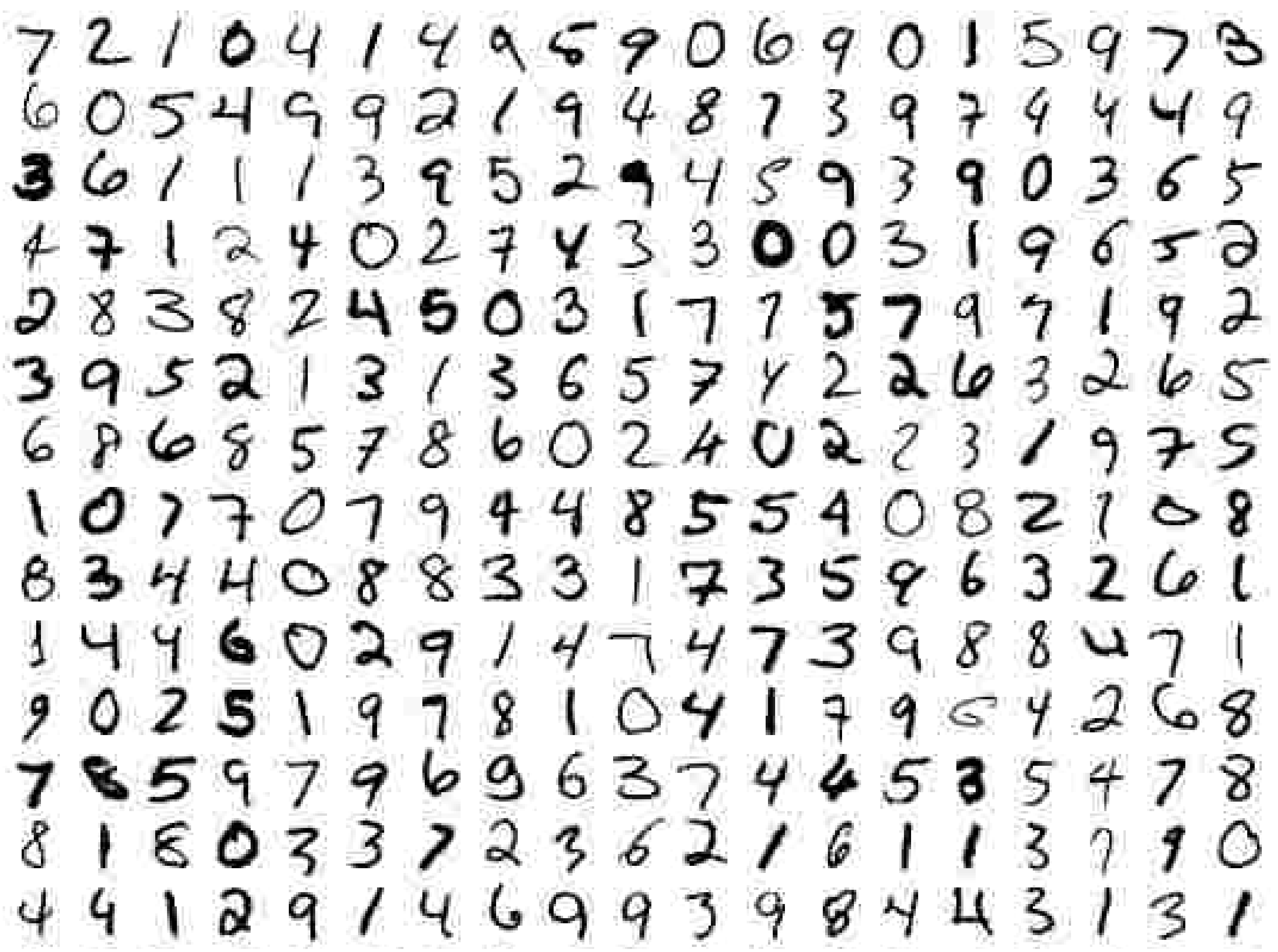
Default Denominaton efect
efect efect dispositon
Distncton bias
Dunning-Kruger efect
Duraton neglect Empathy
Gap Endowment efect
exaggerated expectaton
experimenter's gold expectaton bias
Focusing Drill efect efect gold Barnum efect
Form functon atributon bias Framing efect
Functonal fxedness Gambler's fallacy
Hard-easy efect Hindsight Bias Bias Hostle
atributon hot-hand fallacy Hyperbolic
discountng Identfable victm efect efect
IKEA Illicit transference of control Illusion
Illusion of validity Illusory correlaton Illusory
truth efect

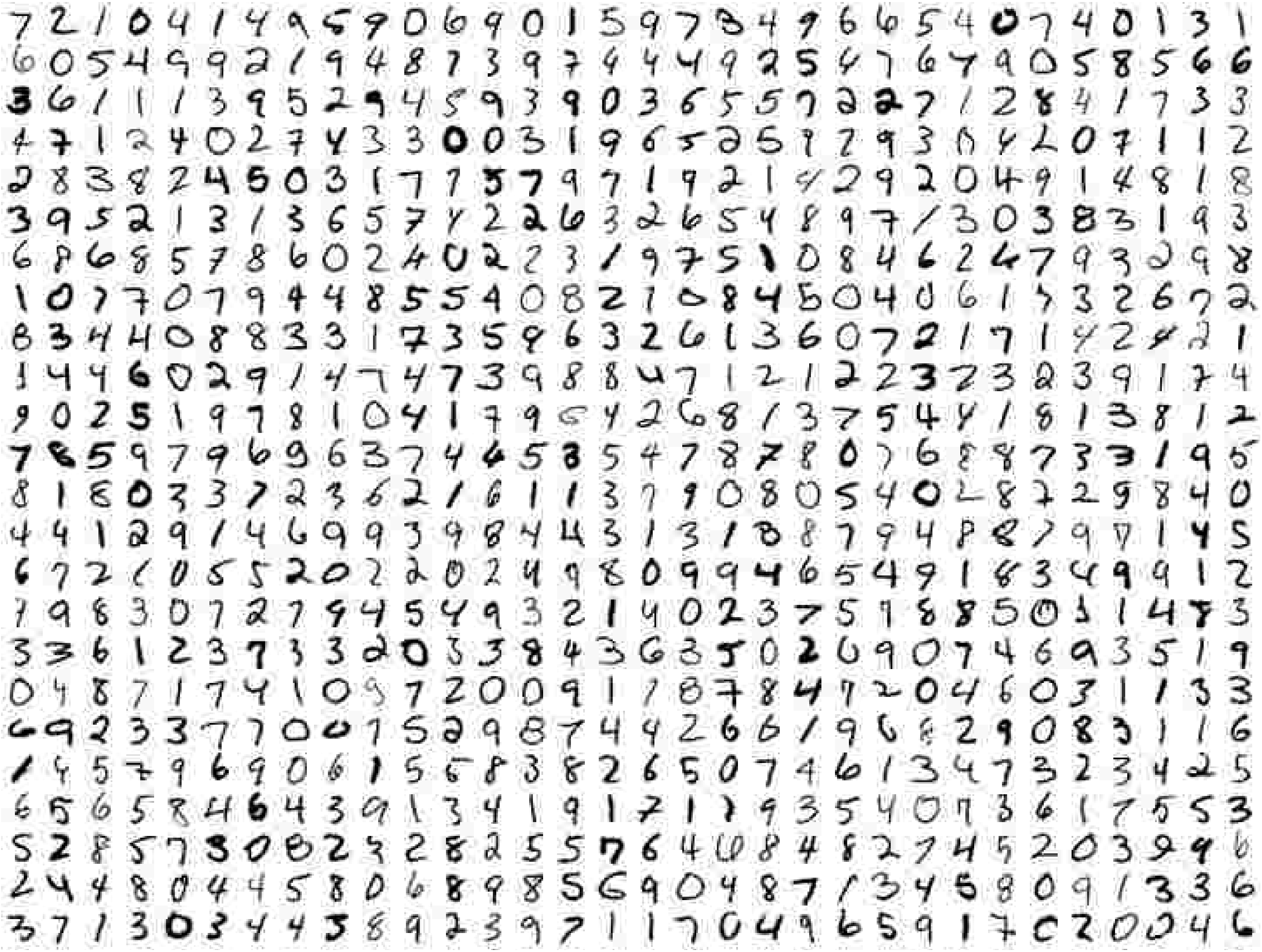
Impact Bias Bias informaton
Insensitvity to sample size
Irratonal escalaton Law of the
instrument Less-is-beter efect
efect Look elsewhere-Loss
aversion Mere exposure efect
Money illusion Moral credental
efect
Negatvity bias gold Negatvity efect
Neglect of probability Normalcy Bias
Not invented here
Observer-expectancy efect Omission
Bias Bias Optmism Ostrich efect
efect Outcome bias Overconfdence
Pareidolia Pessimism bias Placebo
efect Planning fallacy Post-purchase
ratonalizaton innovaton Pro-bias bias
Projecton

Pseudocertainty efect
Reactance Reactve
devaluaton recency illusion
regressive bias
Restraint Bias Rhyme have
reason efect
Risk compensaton / Peltzman efect
Selecton bias Selectve percepton
Semmelweis Refex Sexual overpercepton Bias
/ sexual
Social comparison bias social
desirability bias Status quo bias
Stereotyping Subadditvity efect
Subjectve validaton Surrogaton
Survivorship Bias Time-saving bias
Third-person efect Parkinson's law of
triviality Weber-Fechner law well
traveled road efect Women are
wonderful efect Zero-risk bias
Zero-Sum bias

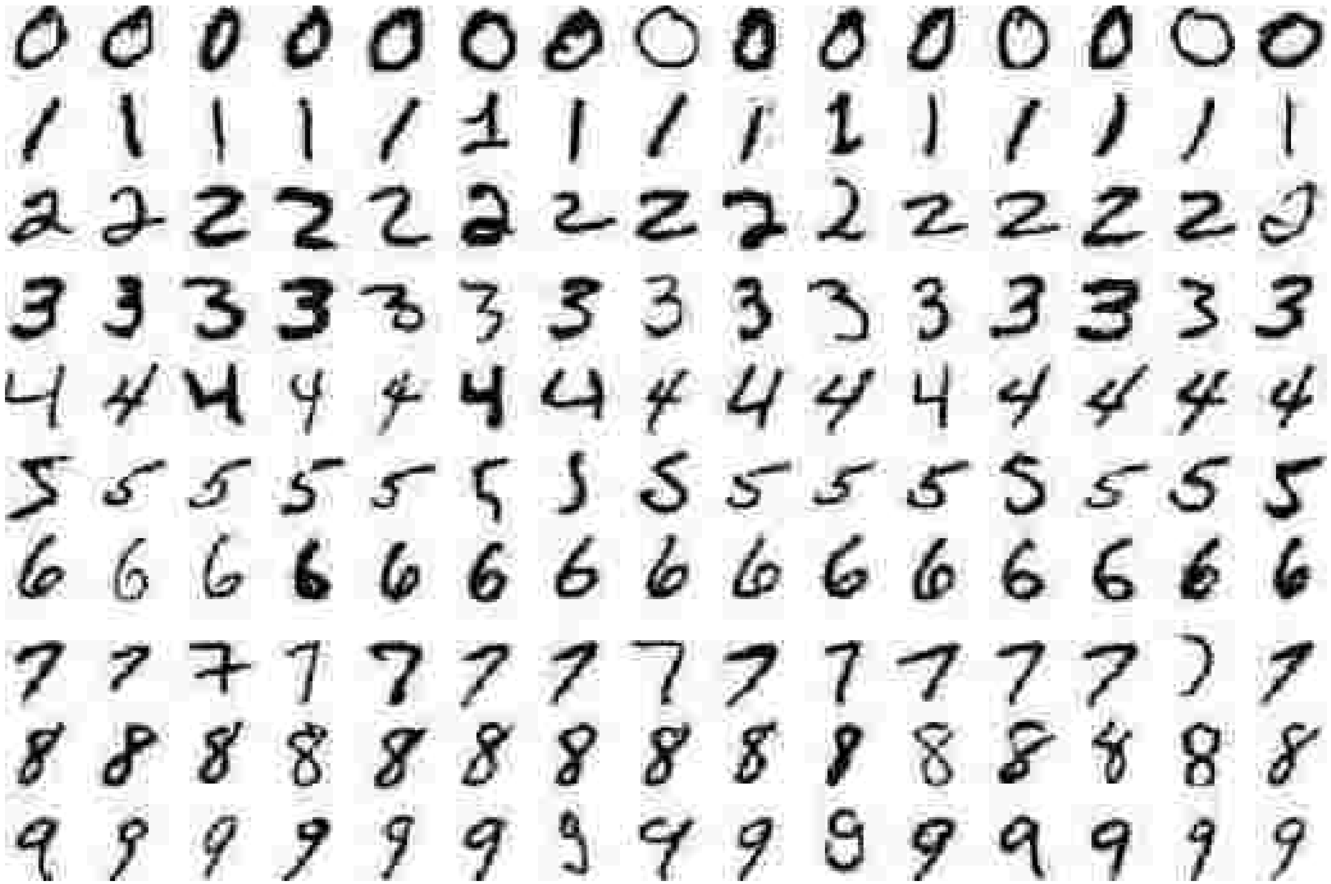

THE NEW YORK TIMES BESTSELLER
THINKING,
FAST AND SLOW
DANIEL
KAHNEMAN
WINNER OF THE NOBEL PRIZE IN ECONOMICS

Datasets



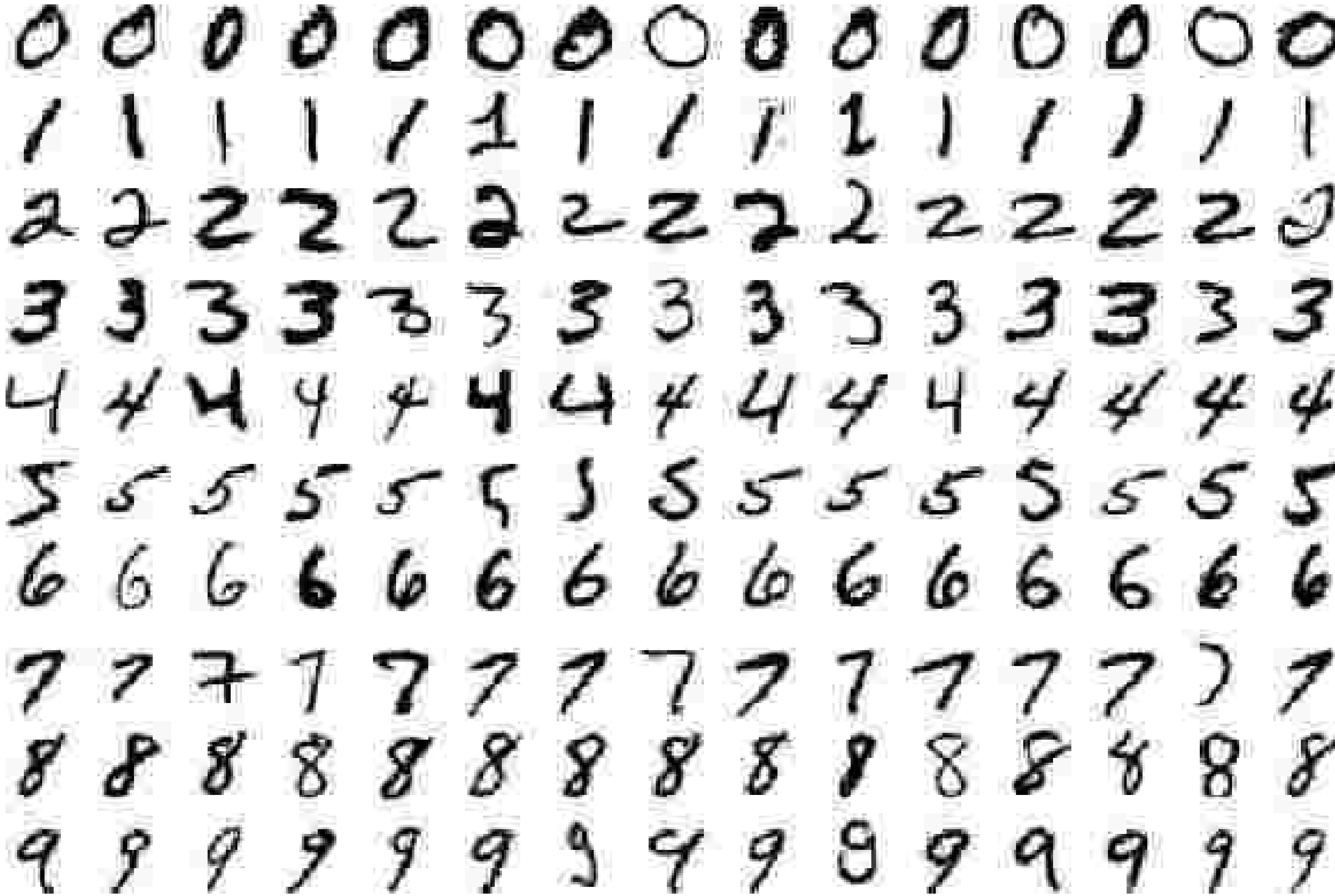


MNIST



70000 examples with good étquete.

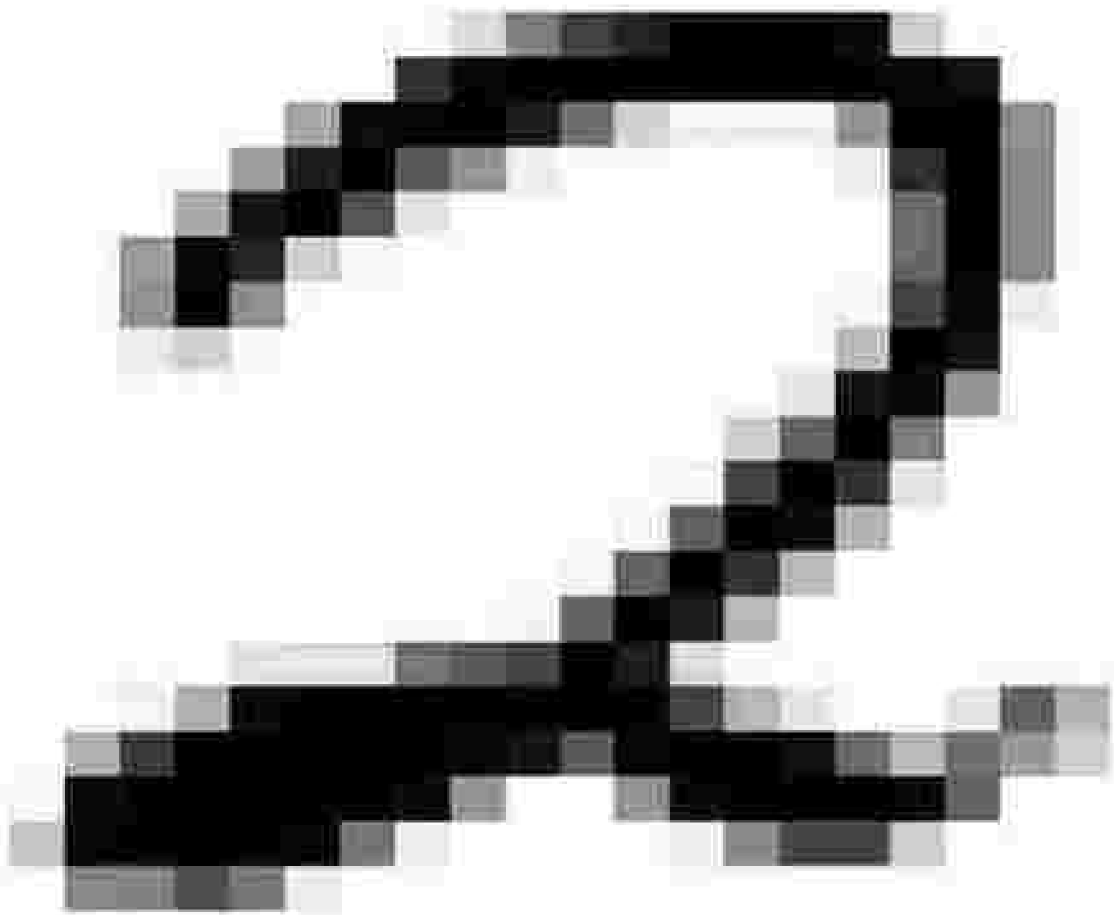
0 =
1 =
2 =
3 =
5 =
4 =
6 =
7 =
8 =
9 =



MNIST CVS

- mnist_train.csv, 60000 x 785, 18KB
- mnist_test.csv, 10000 x 785, 107KB

label 1x1,1x2,1x3,1x0,1x5,1x6,1x7, ..., 1x28,2x1,2x2, 2x3 ...
28x20,28x25,28x26,28x27,28x28



2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 36,125,191,218,255,254,254,241, 51, 0, 0, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,204,249,253,253,253,253,253,253,253,250,235, 0, 0, 0, 0, 0,

0

0, 0, 0, 0, 0, 0, 0, 0, 0, 88,241,251,253,225,142, 49, 12, 12, 12,105,253,253,111, 0, 0, 0, 0,

0

0, 0, 0, 0, 0, 0, 0, 0, 95,225,253,167,113, 14, 0, 0, 0, 0, 0, 16,211,253,117, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 0, 91,238,253,170, 28, 0, 0, 0, 0, 0, 0, 0,150,253,117, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 98,251,218, 48, 5, 0, 0, 0, 0, 0, 0, 0, 0,150,253,117, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0,112,253,112, 0, 0, 0, 0, 0, 0, 0, 0, 0, 9184242, 18, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 20, 45, 5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 67253240, 0, 0, 0, 0, 0, 0

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 27,234,248,105, 0, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 44,157,253,132, 0, 0, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 11,189,253,203, 27, 0, 0, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 11,156,253,246, 77, 0, 0, 0, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 11,156,253,202, 68, 0, 0, 0, 0, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 11,156,253,226, 71, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 0, 0, 30, 33, 33,140,163,186,253,226, 38, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 12, 81,244,253,253,253,253,253,253,253,186, 70, 23, 0, 0, 22.156, 77, 0, 0, 0,

0

0, 0, 0, 0, 0, 80,195,253,253,253,253,253,248,234,166,248,253,253,240,150, 73,144,104, 51, 0,

0, 0, 0,

0, 0, 0, 0, 7,248,253,253,253,253,253,242,105, 0, 0,107,242,253,253,253,245,160, 0, 0, 0, 0,

0, 0,

0, 0, 0, 0, 57,250,253,253,253,247,135, 21, 0, 0, 0, 0, 21,117,183,183, 48, 0, 0, 0, 0, 0, 0,

0

ADULT

08502 individuals (32261 + 16281) 10

caractéristques

age, workclass, fnlwgt, educaton, educaton-num, occupaton, relatonship, race, sex, capital gain, capital loss, hours-per-week, natve-country incomes

htp: //mlr.cs.umass.edu/ml/datasets/Adult

age, workclass, fnlwgt, educaton, educatonnnum, occupaton, relatonship, race, sex, capitalngain, capitalnloss, hoursnpernweek, natvencountry, incomes

39 State-gov, 77516, Bachelors, 13, Never-married, ADM-clerical, Not-in-family, White, Male, 2170, 0, 00, United-States, <= 50K

50, Self-emp-not-inc, 83311, Bachelors, 13, Married CIV-spouse, Exec-managerial, Husband, White, Male, 0, 0, 13, United-States, <= 50K

38, Private, 215606, HS-grad, 9, Divorced, Handlers-cleaners, Not-in-family, White, Male, 0, 0, 00, United-States, <= 50K
53, Private, 230721, 11th, 7, civ-Married-spouse, Handlers- cleaners, Husband, Black, Boy, 0, 0, 00, United-States, <= 50K

Adult

age: contnuous
workclass: Private, Self-emp-not-inc, Self-emp-Inc, Federal-gov, gov-Local, State-gov, Without-Pay, Never- Worked

fnlwgt: contnuous
educaton: Bachelor, Some-college, 11th, HS-grad, Prof-school, Assoc-MDC, Assoc-voc, 9th, 7th, 8th, 12th, Masters, 1st-0th, 10th, Doctorate, 5th-6th, Preschool
educatonnnum: contnuous.marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent Married-AF-spouse
occupaton: Tech-support, Craf-repair, Other Service, Sales Exec-managerial, Prof-specialty, Handlers- cleaners, machine-op-inspct, Adm-clerical, Farming-fshing, Transport-moving, Priv-house-serv, Protectve-serv, Armed Forces,

relatonship: Wife, Own-child, Husband, Not-in-family, Other-relatve, Unmarried
race: White, Asian-Pac-Islander Amer-Indian-Eskimo, Other, Black
sex: Female, Male
capitalngain: contnuous
capitalnloss: contnuous
hoursnpernweek: contnuous
natvencountry: United-States, Cambodia, England, Puerto Rico, Canada, Germany, Outlying-US (USVI-Guam- etc.), India, Japan, Greece, South, China, Cuba, Iran, Honduras, the Philippines, Italy, Poland, Jamaica , Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Hait, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El Salvador, Trinadad & Tobago, Peru, Hong, Holand-Netherlands

incomes: > 50K <= 50K

modeling

modeling

- modeling
- Explain
- Simplifer
- Compress
- Reduce
- Schématser
- mapping
- A theory

modeling

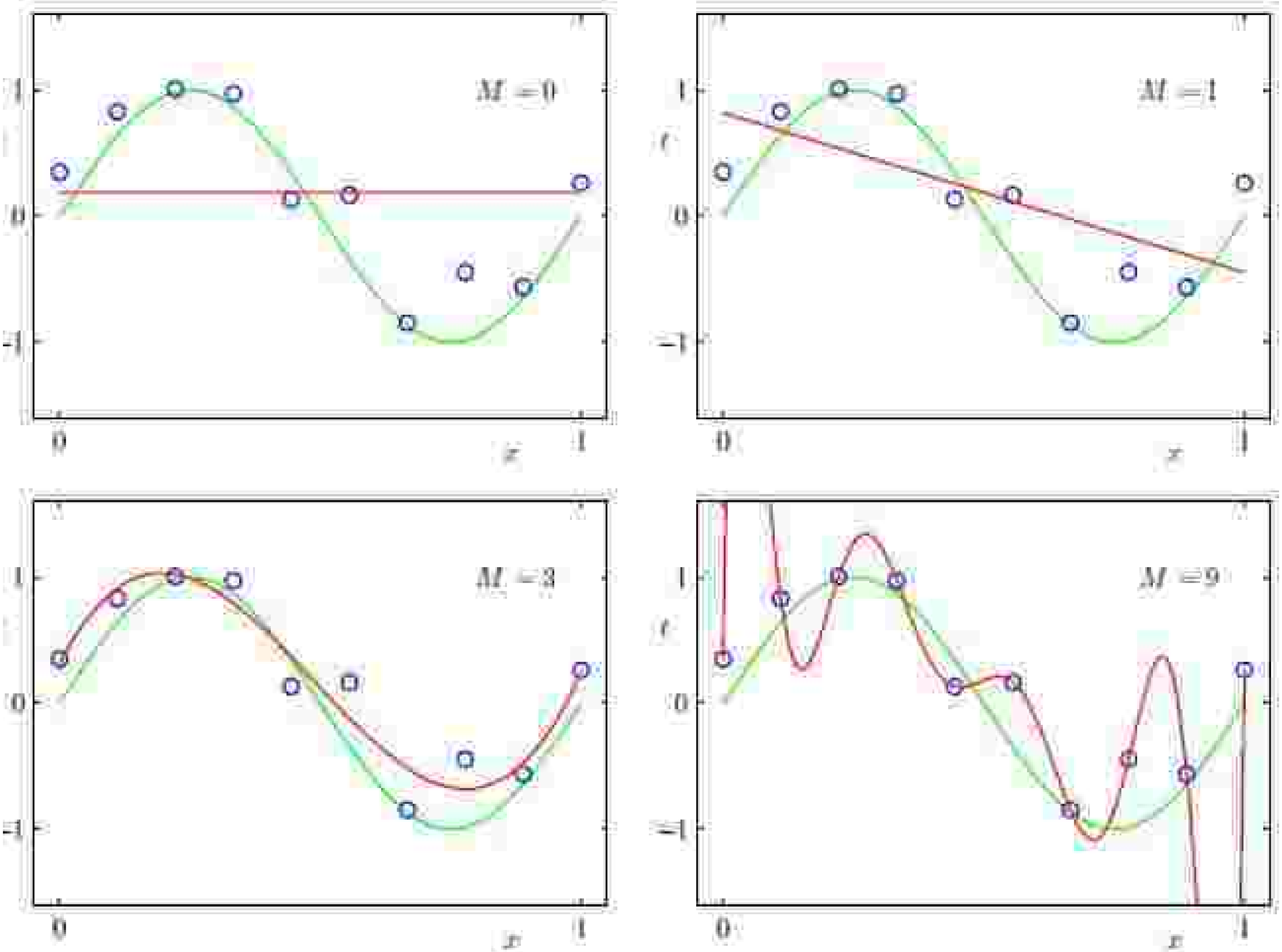
Given: a Characteristc table (dataset)

Model: pressing a formal relaton between caractéristques

Cost: the price paid if utlise model rather than the observed values.

Choice of model

- **Family:** choose
- **hyperparameter:** selected or learned
- **Settings:** learned



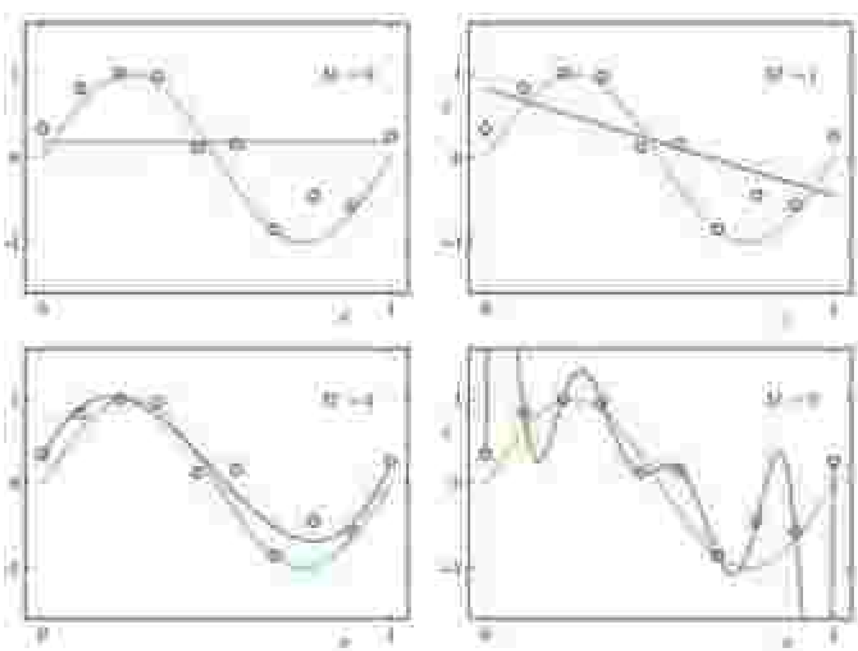
$M = 0$
1
0
−1
0
x
1
$M = 1$
t
1
0
−1
0
x
1
$M = 3$
1
0
−1
0
x
1
$M = 9$
t
1
0
−1
0
x
1

polynomial regression

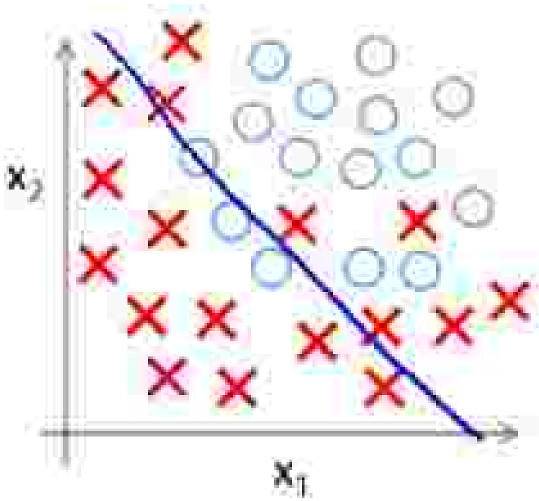
Family: polynomial

Hyperparameter: Settings degree polynomial: Coefcient

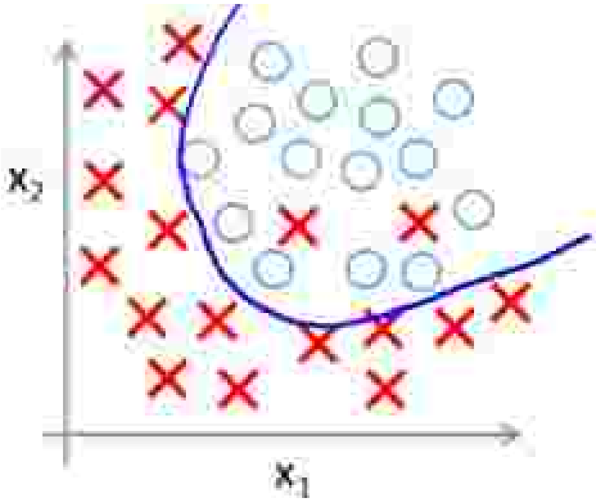
polynomial (degree + 1) cost: Minimizing squared error



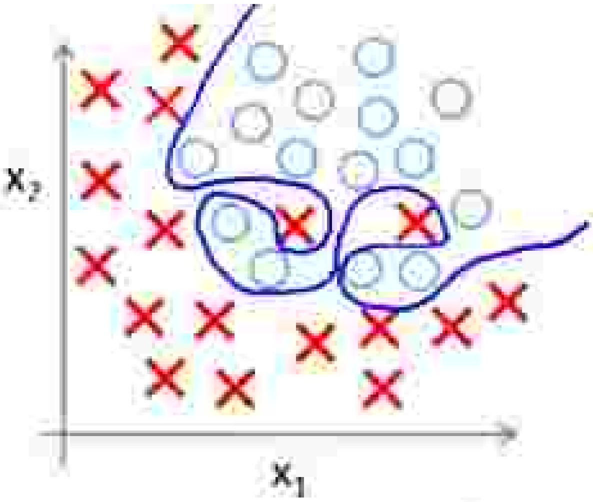
Overtraining?



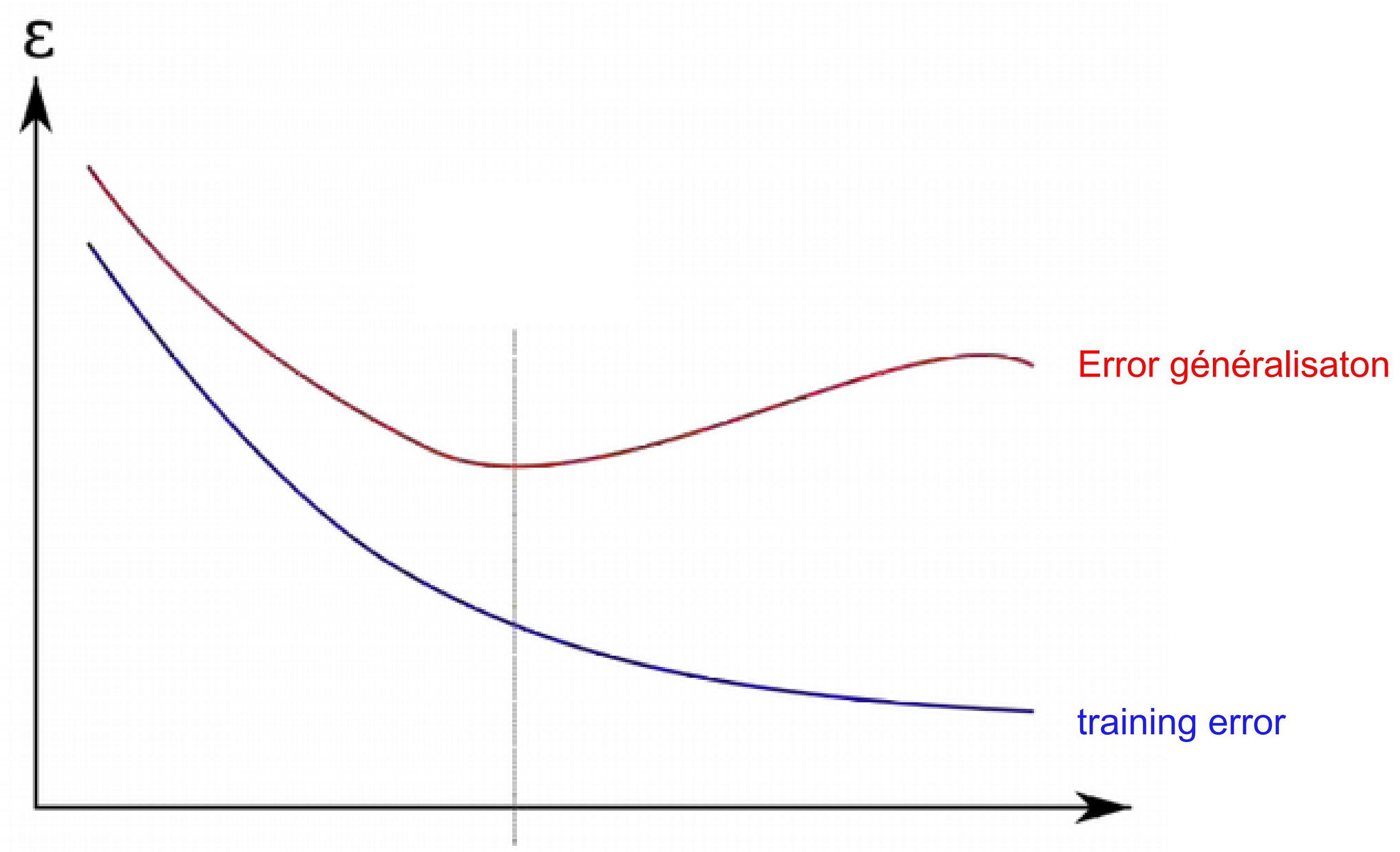
Deputy apprentssage



good généralisaton

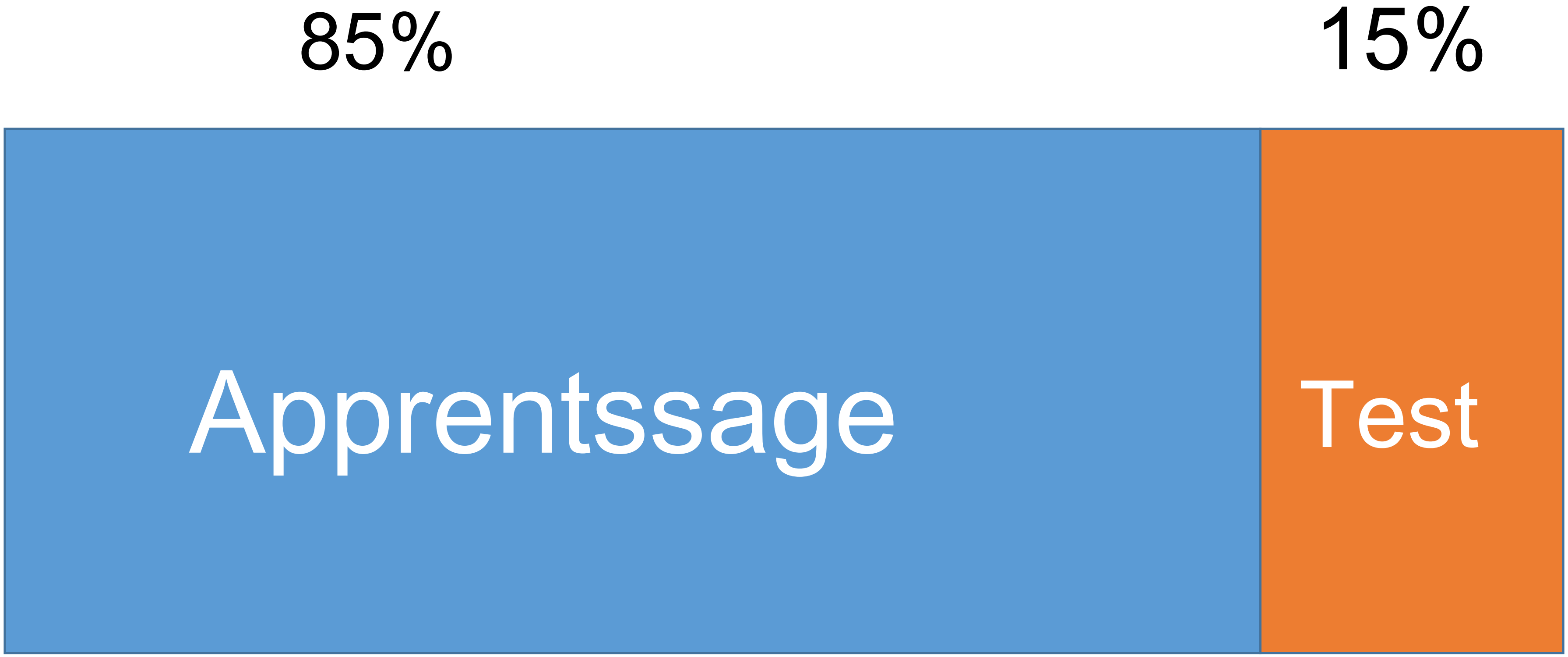


Surapprentssage

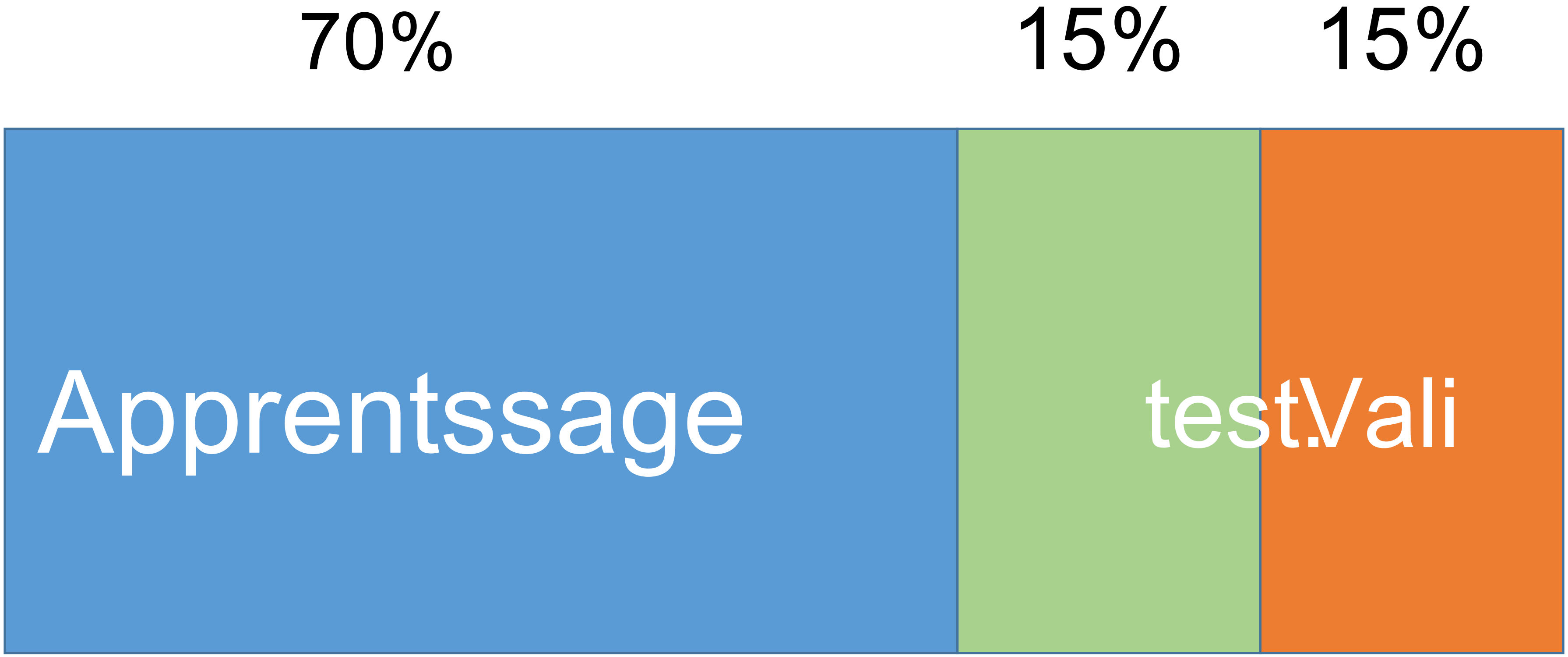

ε
Error généralisaton
training error

100%

Examples



It is learned settings then tested with fresh data.



the parameters and hyper parameters is learned then tested with fresh data.
The hyper parameters are learned Foncton all validaton.

Settings and hyper parameters

The apprentssage is an algorithmic mechanism permetant get good parameters for our model.

The **ability of a model** refers to its ability to mémorisaton.

- Capacity
- Number of parameters
- Complexity

The training time refers to the time when we voluntarily decided, at a optmisaton, terminate the progression of the algorithm to avoid over apprentssage. By **régularisaton** of references has potentellement harmful stress has optmisaton, but may have an impact on the généralisaton.

The Bayesian approach

Probability

Definition: events form a universe

Définition: les événements forment un univers Ω

axiom1

Axiome 1

For any event in

Pour tout événement A dans Ω

$$0 \leq P(A) \leq 1$$

axiom2

Axiome 2

$$P(\Omega) = 1$$

axiom3

Axiome 3

Any two events Family disjoint (incompatible)

Toute famille d'événements deux à deux disjoints (incompatibles)

$$P(A_1 \cup A_2 \cup \cdots) = \sum_i P(A_i)$$

Theorem

Théorème

$$P(\emptyset) = 0$$

unsolicited Medical Test

Amyotrophic lateral sclerosis

valid test 95%

False Positive feedback ratio and false négatf

We test you by mistake.

The test is Positive feedback ratio.

Worried?

- valid test was 95% (False Positive feedback ratio and false négatf)
- Populaton 10 million.
- Prevalence in populaton 1/100 000.

If one tests the entire populaton, only 0.019% of positve results are in fact carrier.

	Total	Sick	Healthy
Total	10000000	100	9999900
Positive feedback ratio	50 0095	95	099995
Négatf	9099905	5	9099905

A Positive feedback ratio test indicates augmentaton a factor of 20 risk (the test is wrong again on 20) is signifcatf, but we must not neglect the basic rate.

Bayes Theorem

$P(C)$ = Probabilité de la Classe a priori.

= Probability of Class priori. = Probability of Observaton priori.

$P(O)$ = Probabilité de l'Observation a priori.

$P(O|C)$ = La vraisemblance de l'Observation étant donné la Classe.

= The likelihood of Observaton given the class.

$P(C|O)$ = Probabilité conditionnelle de la Classe étant donné l'Observation.

= Probability of Class conditonnelle given the Observaton.

$$P(C|O) = \frac{P(O|C)P(C)}{P(O)}$$

The Bayesian approach

$P(C)$ = Probability of Class priori. (1/100 0000)

$P(O)$ = Probability of Observaton priori. (P (patient) * 95% + P (healthy) * 5%) = The probability of Observaton given Class. (95%)

$P(C|O)$ = Probability of Class conditionnelle given the Observaton.

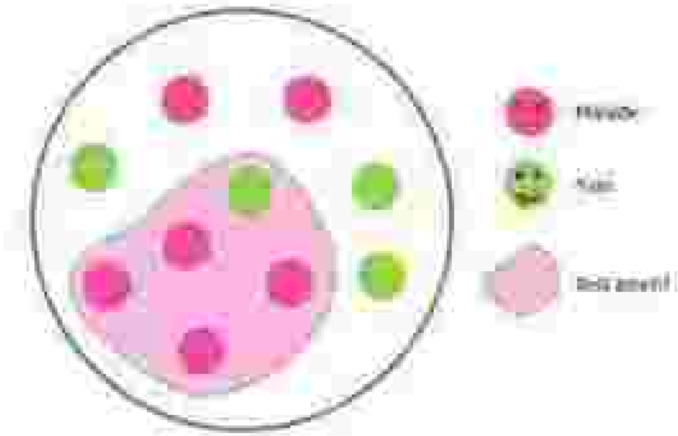
$$P(C|O) = \frac{P(O|C)P(C)}{P(O)}$$

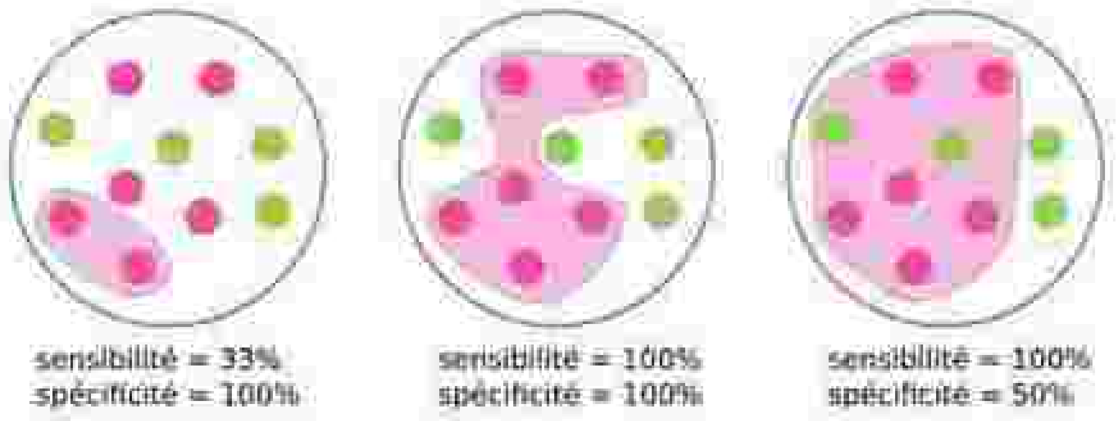
$$P(C|O) = \frac{\frac{19}{20} * \frac{1}{100000}}{\frac{1}{100000} * \frac{19}{20} + \frac{99999}{100000} * \frac{1}{20}}$$

0.0001899658

$$P(C|O) = 0.0001899658$$

Example





Prevalence: 10% of sick

Prévalence : 10% de malade

$$P(M|P) = \frac{P(P|M)P(M)}{P(P)}$$

$$P(S|N) = \frac{P(N|S)P(S)}{P(N)}$$

$$100\% = \frac{0.33 * 0.1}{0.1 \; * 0.33 + 0.9 * 0}$$

$$96\% = \frac{1.0 * 0.9}{0.1 \; * 0.33 + 0.9 * 1.0}$$

100%

100%

$$14\% = \frac{0.75 \; * 0.1}{0.1 \; * 1.0 + 0.9 * 0.5}$$

$$100\% = \frac{0.5 * 0.9}{0.1 \; * 0.0 + 0.9 * 0.5}$$

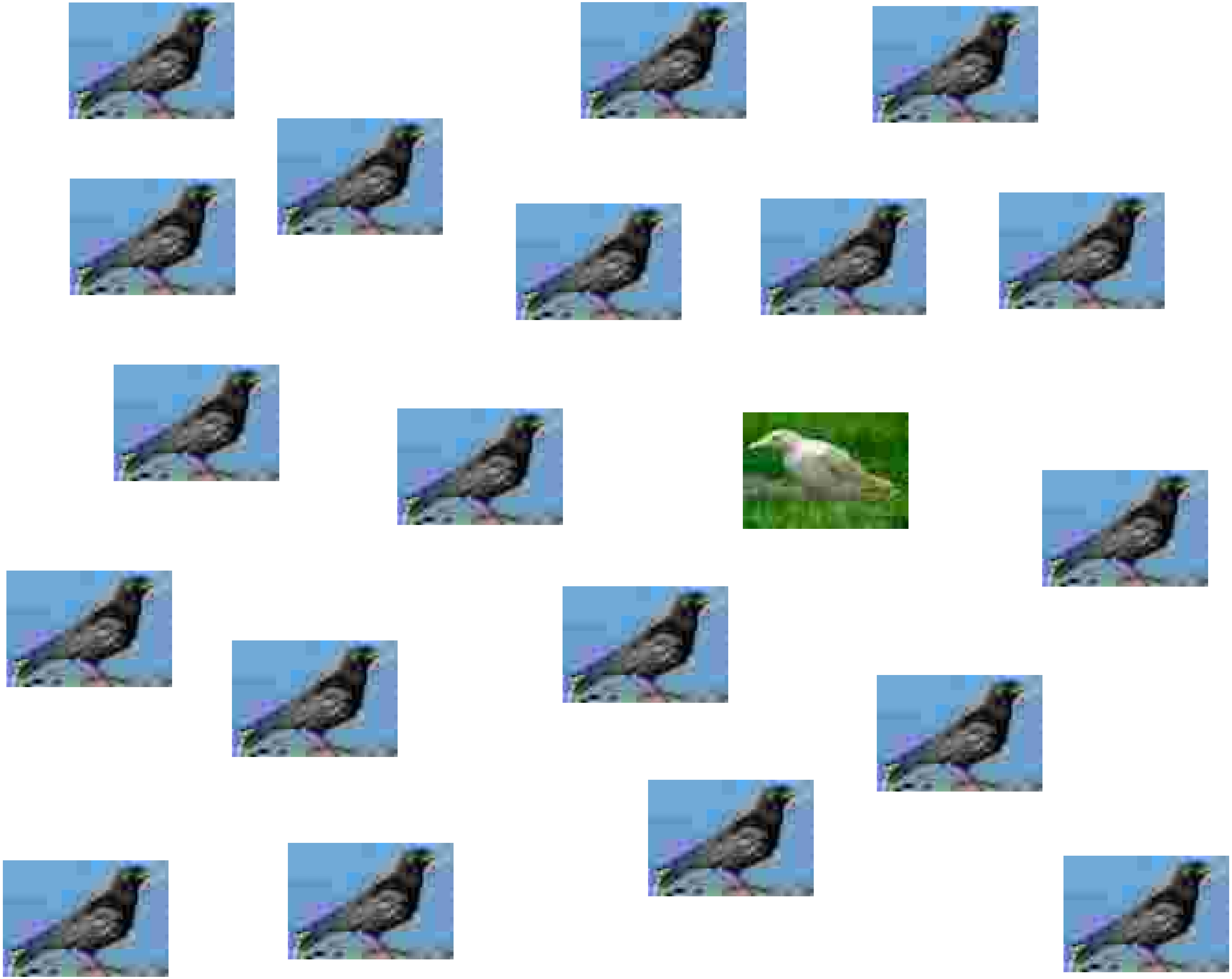
http://dridk.me/le-theoreme-de-bayes.html

The problem of induction

All crows are black?

The inducton (philosophy) is the rule that generalizes partr a number of examples inf has all the elements. This is one of the most problématques concepts in philosophy.

For us the soluton is simple, consider a set of elements inf, uniform échantIlonnage and applying Bayes.



Raven paradox and induction

Black Raven paradox

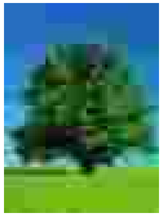
The paradox of Hempel was proposed by the German logician Carl Gustav Hempel in the 1900s to illustrate that the logic inductve could violate intuiton. All A are B

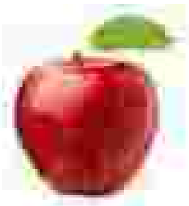
All non-B are not A

Each object (not crow) that is not black confrmed cete hypothesis.

3 despicable **crows**

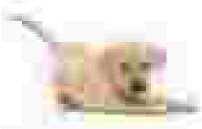
- Black Crow
- **raven nonnnoir**

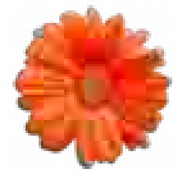


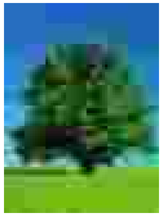


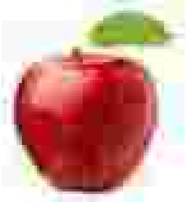
11 items **non-black**

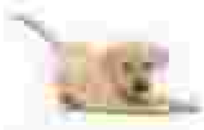
- **non-black raven**
- Other non-black











hypothesis test

hypothesis test

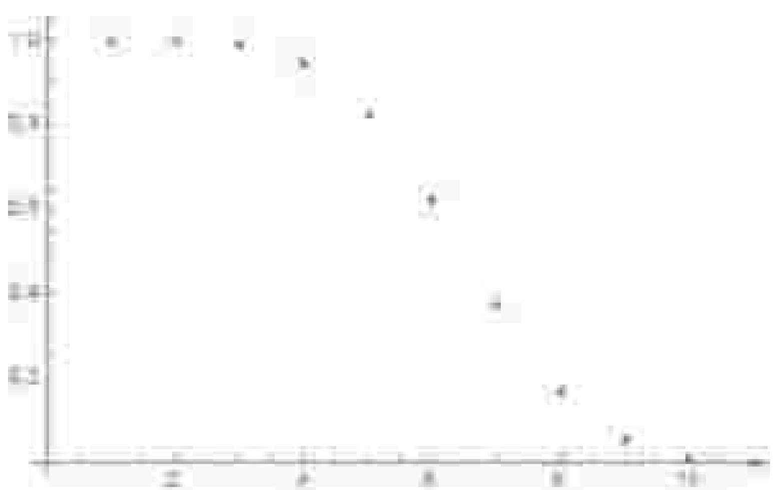
Hypothesis: confrmer what we want to show

Null hypothesis: no confrmaton statstque In a test, the p-value (*p-value)* is the probability that a given statstque model under the null hypothesis of obtaining the same value or a value more extreme than observed.

Referendum

- Populaton 10 million
- Experience ÉchantIlon 10 (random uniform, honest and perfect)
- Hypothesis: populaton favors
- P value = 0.05 (5%)
- We observe 8 yes and 2 no
- It rejected! (P-value, 0.171875, 17%)

p-value

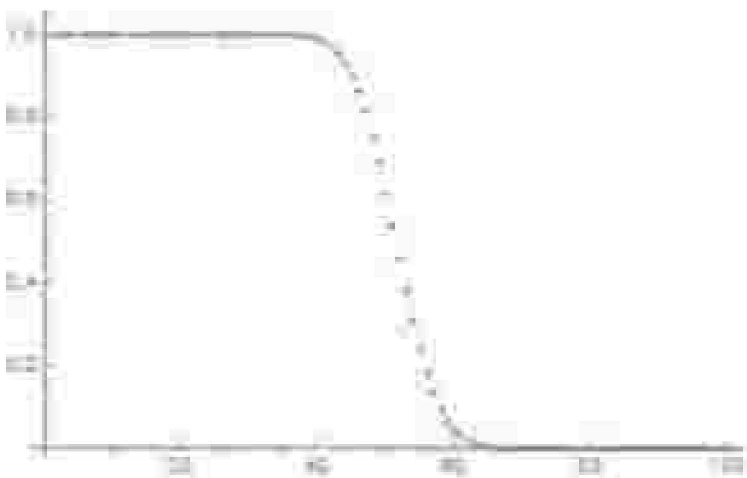


#yes on 10

Referendum

- Populaton 10 million
- Experience ÉchantIlon 100 (random uniform, honest and perfect)
- Hypothesis: populaton favors
- We observe 58 yes and 02 no
- It rejected! (P = 10%)

p-value

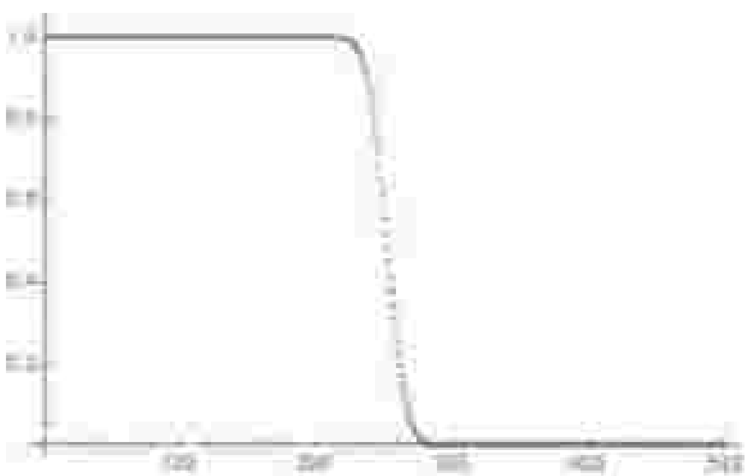


#Yes on 100

Referendum

- Populaton 10 million
- Experience Échantllon 500 (random uniform, honest and perfect)
- Hypothesis: populaton favors
- P value = 0.05 (5%)
- 270 yes we observe, not 230
- We accept. (P-value 0.009)

p-value



 # Yes on 500

Polls

3%, 19 times out of 20 ... yes, but

- Échantllon not uniform at all.
- How to ask the queston has a very big impact.
- Assumes that answer the phone an unknown causes even réacton the confdentalité urn

- People do not know what they did and they do.

correlation and causation

statistical Reminder

the data

$$X = \{x_1, x_2, \dots, x_n\} \in \mathbb{R}^n$$
$$Y = \{y_1, y_2, \dots, y_n\} \in \mathbb{R}^n$$

The average

$$\mu_X = \mathrm{E}(X) = \frac{1}{n}\sum_{I=1}^{n} x_i$$

The standard deviation

$$\sigma_X = \sqrt{\frac{1}{n}\sum_{I=1}^{n}(x_i - \mu)^2}$$

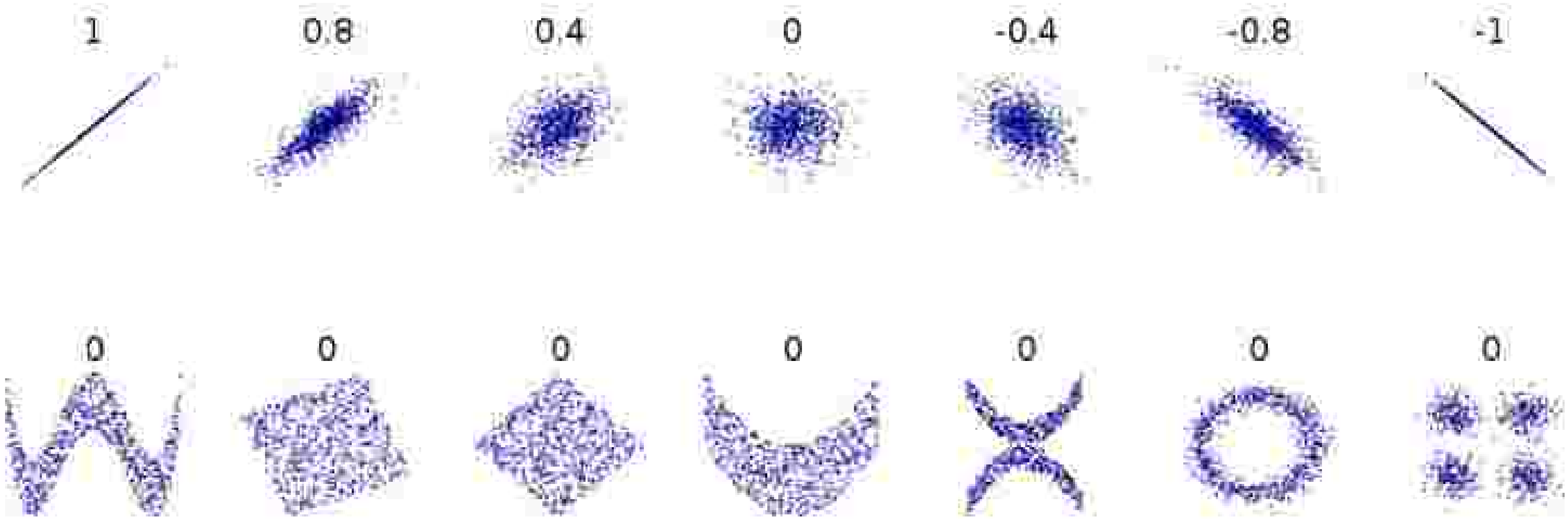
Coefcient of corrélaton

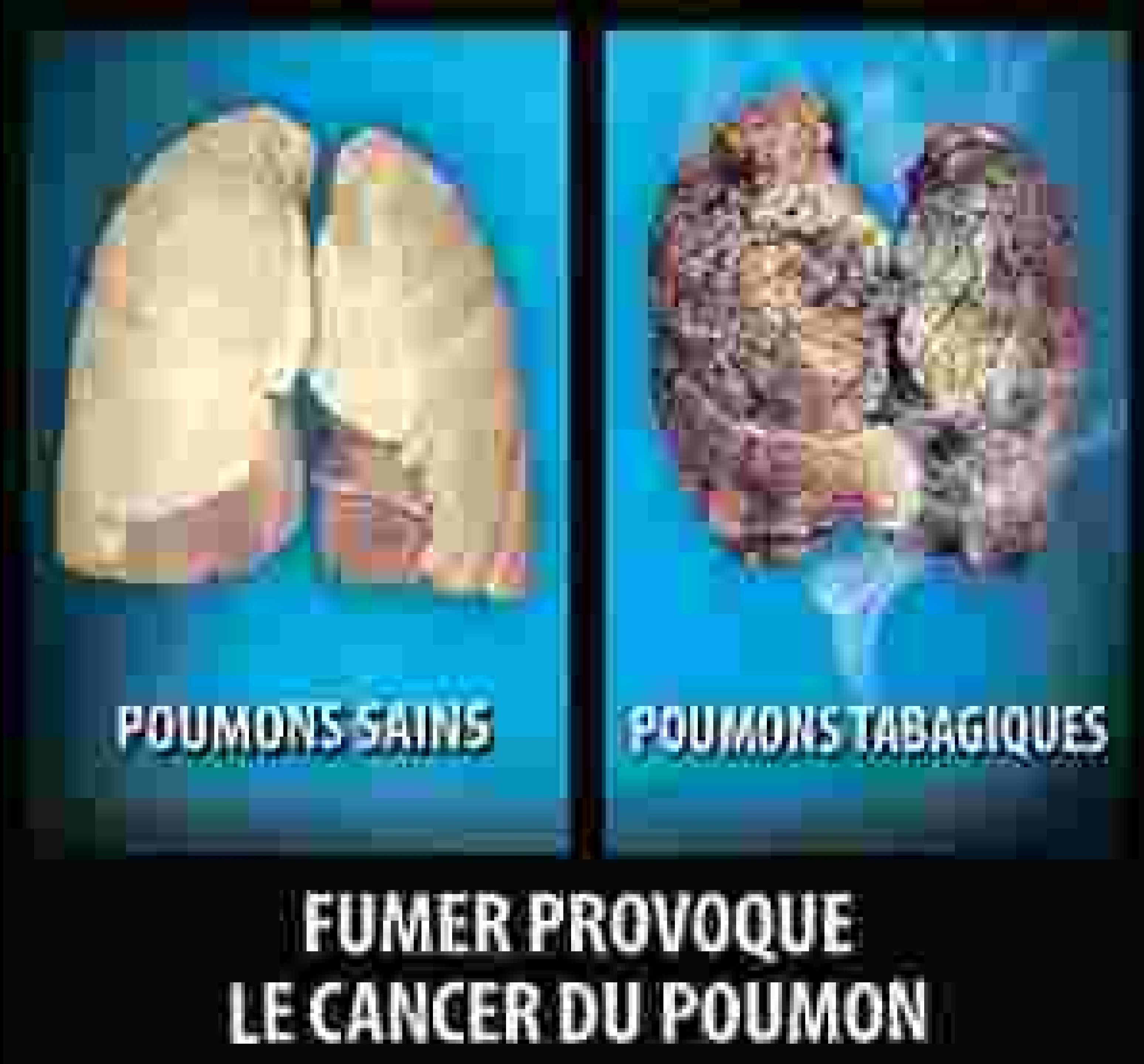
$$r_{X,Y} = \frac{\frac{1}{n}\sum_{I=1}^{n}(x_i - \mu_X)(y_i - \mu_Y)}{\sigma_X \sigma_y}$$

Correlation Coefcient

$$r_{X,Y} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)(y_i - \mu_Y)}{\sigma_X \sigma_Y}$$

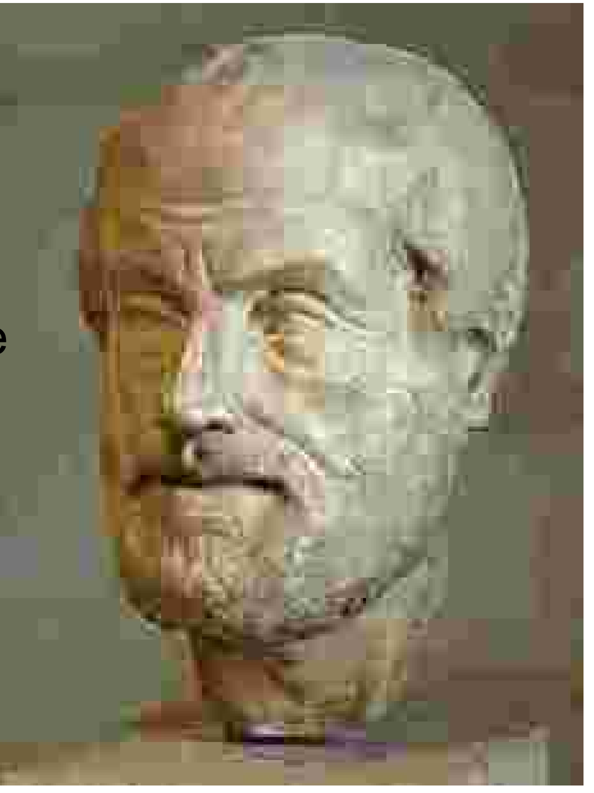
$$-1 \leq r_{X,Y} \leq 1$$




POUMONS SAINS
POUMONS TABAGIQUES
FUMER PROVOQUE
LE CANCER DU POUMON

Aristotle

"If so everything moved is necessarily moved by something [...] there must be a first mover that is not moved by anything [...] In efect, it is impossible that the series of engines that are themselves moved by something else go to infni, since in infnies series there is nothing that is first



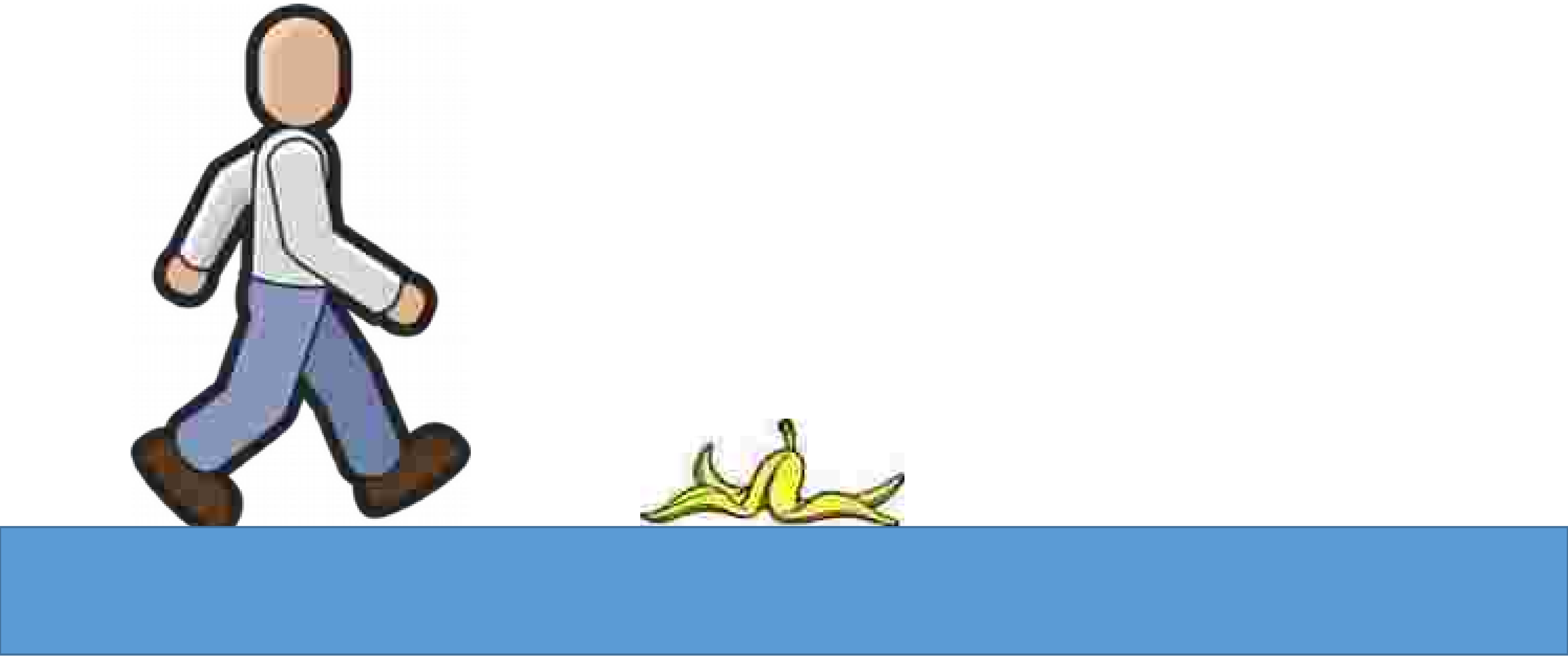
Aristotle: The nature of the apple.

Newton: gravitatonnelle Force

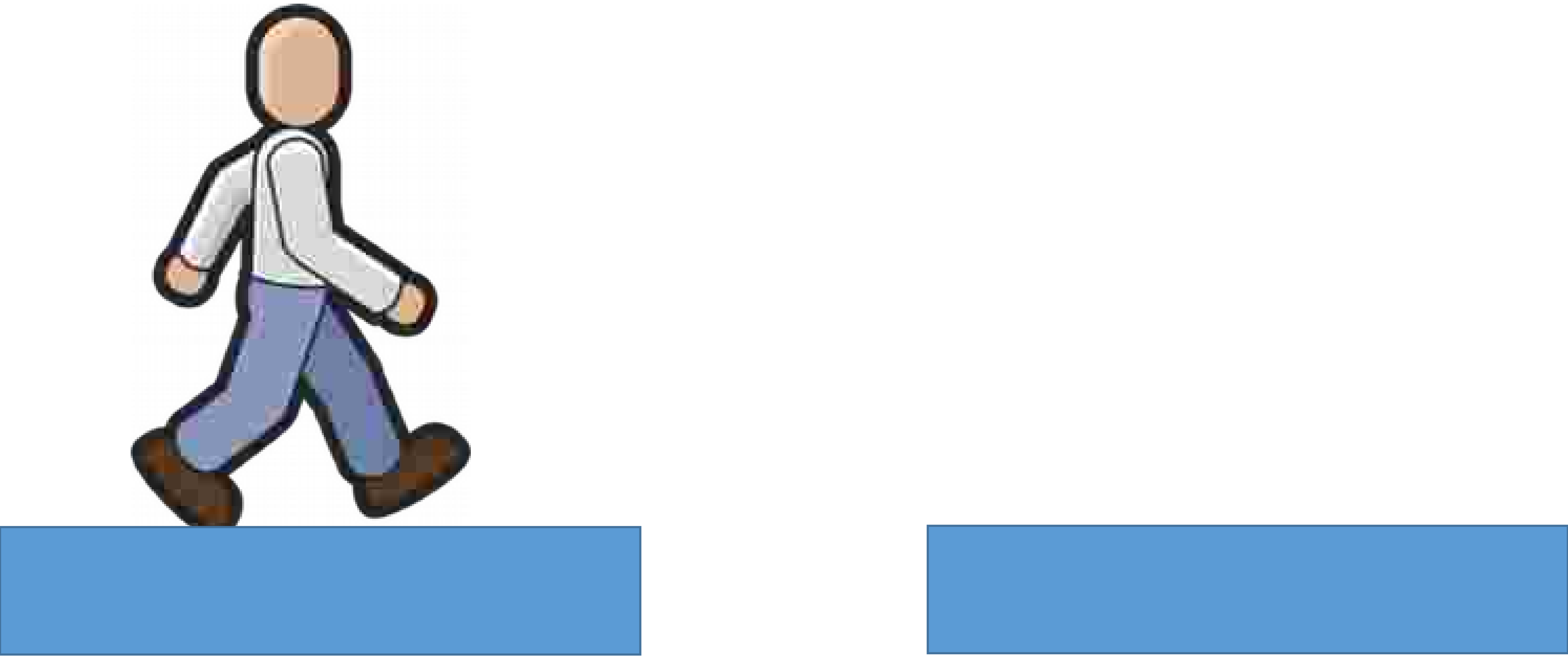
Einstein: Curvature of space

Prédictons quality is improving, but we still have not established **the cause.**

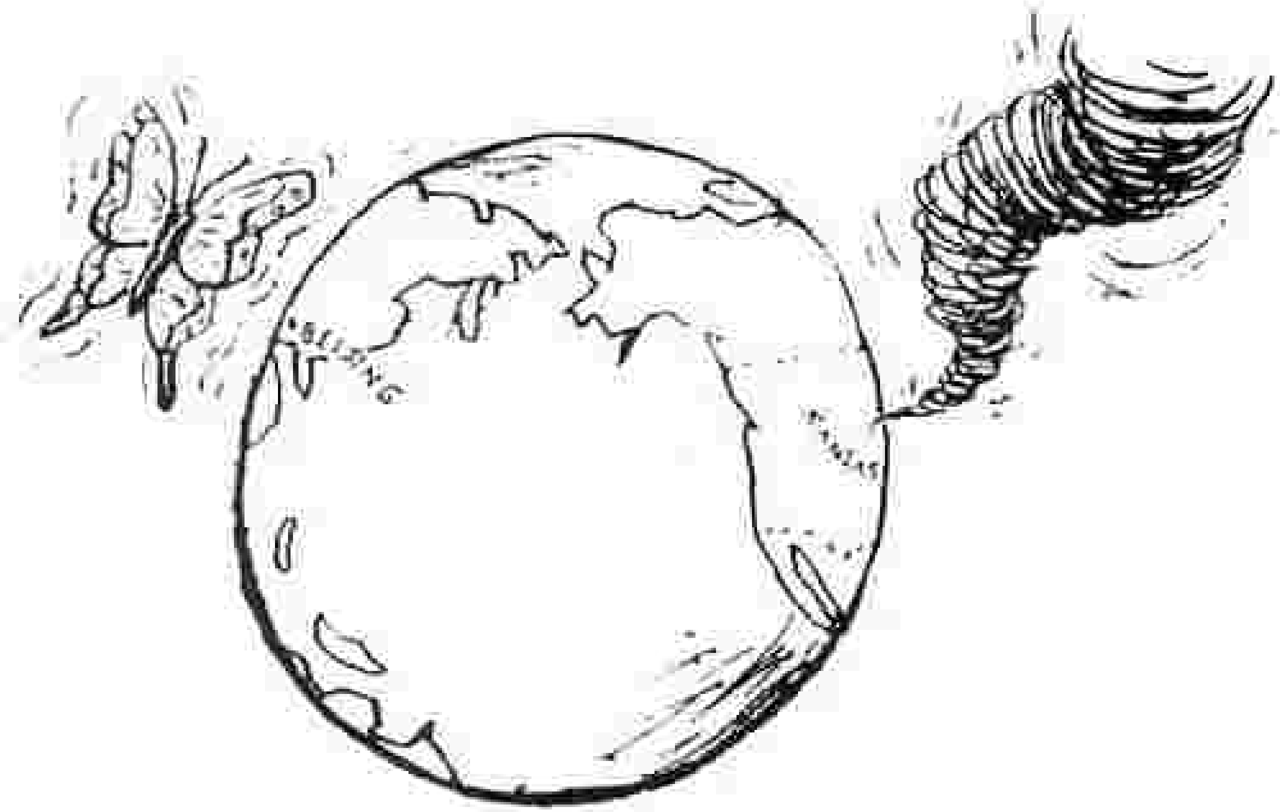
What is the cause of his fall?



What is the cause of his fall?



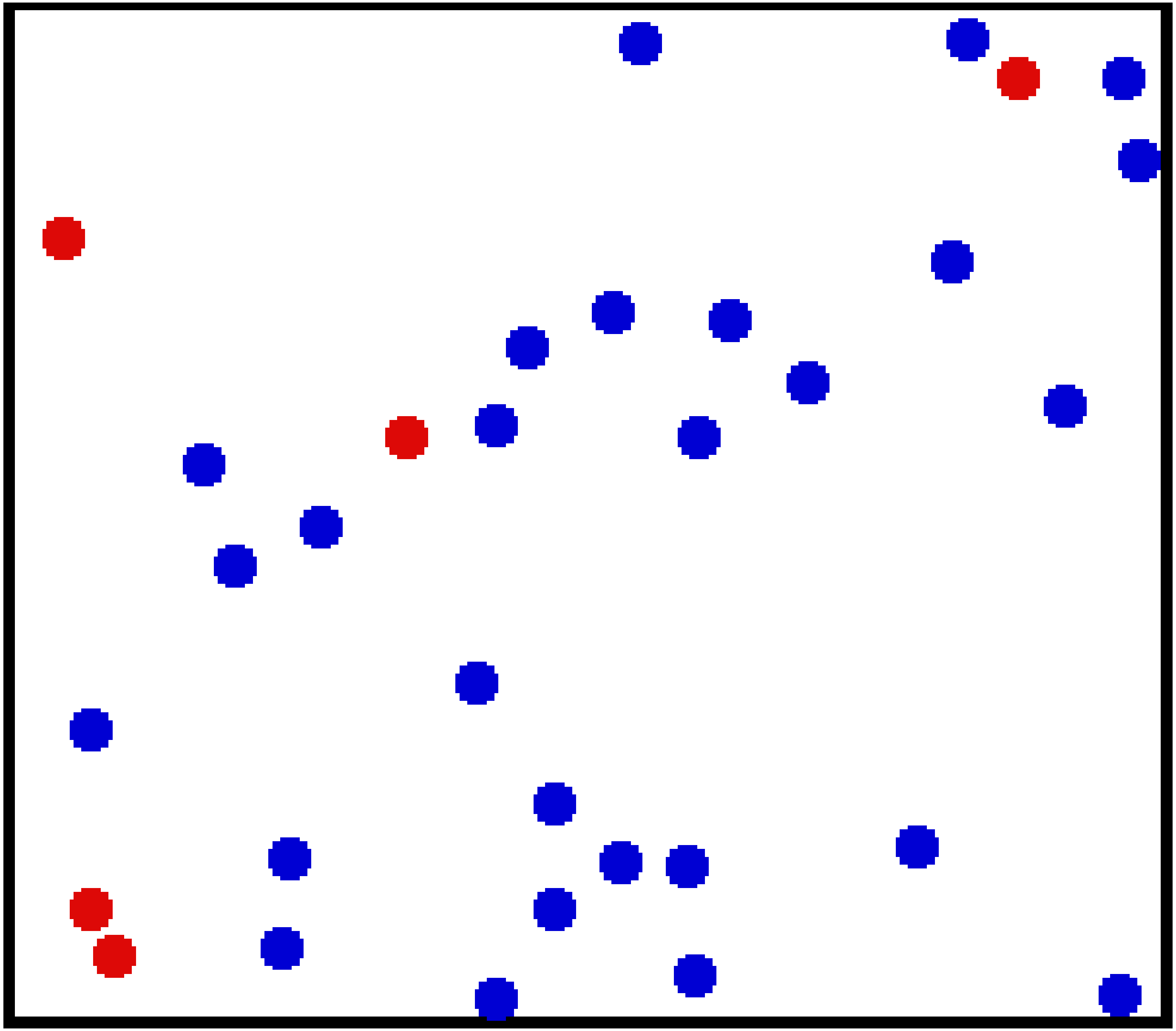
The butterfly efect



What is efet butterfly? It's a **theory** that a butterfly wings in Brazil can batement **provoke**

a storm in Texas. According to the expression coined by meteorologist Edward Lorenz, he suft of MODIFY so infamous a parameter in a weather model for the latter is gradually amplife and causes long-term, colossal change.

HTPS: //www.caminteresse.fr/environnement/quest-ce-que-lefet-papillon-1163580/



Free will

Experimental science

- We **chooses** X and **observed** Y.

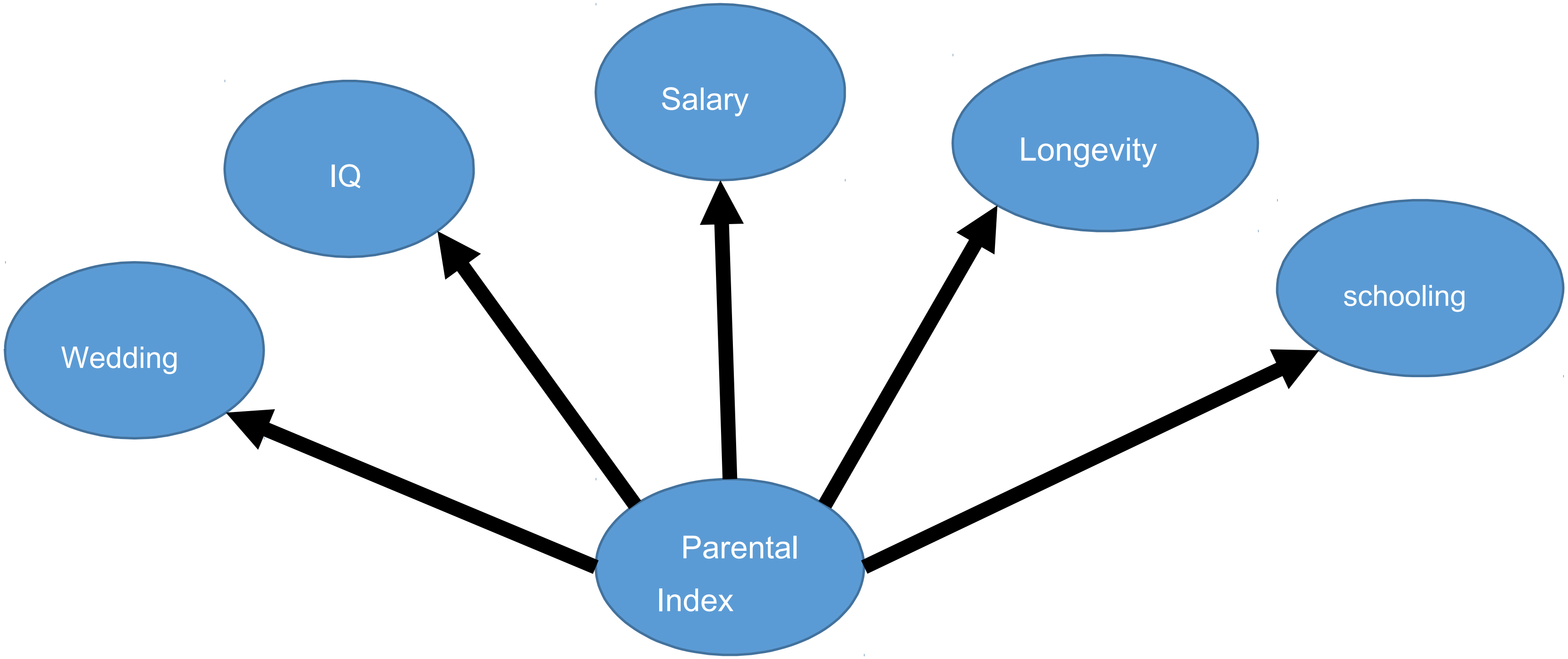
- If X had been diferent then Y would
  also have been diferent

AT **is a cause** B if anyone could change A B to try to avoid and what to produce.

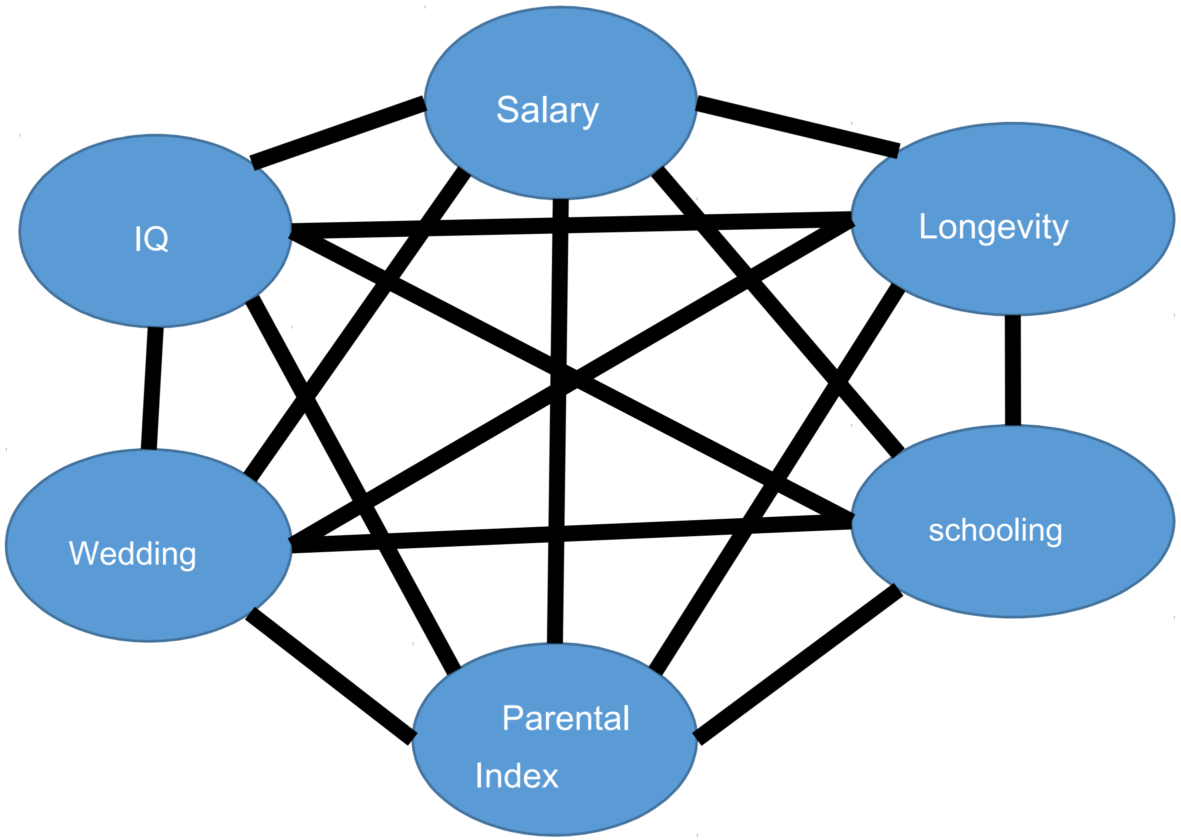
Cete défniton requires free will as a first hypothesis.

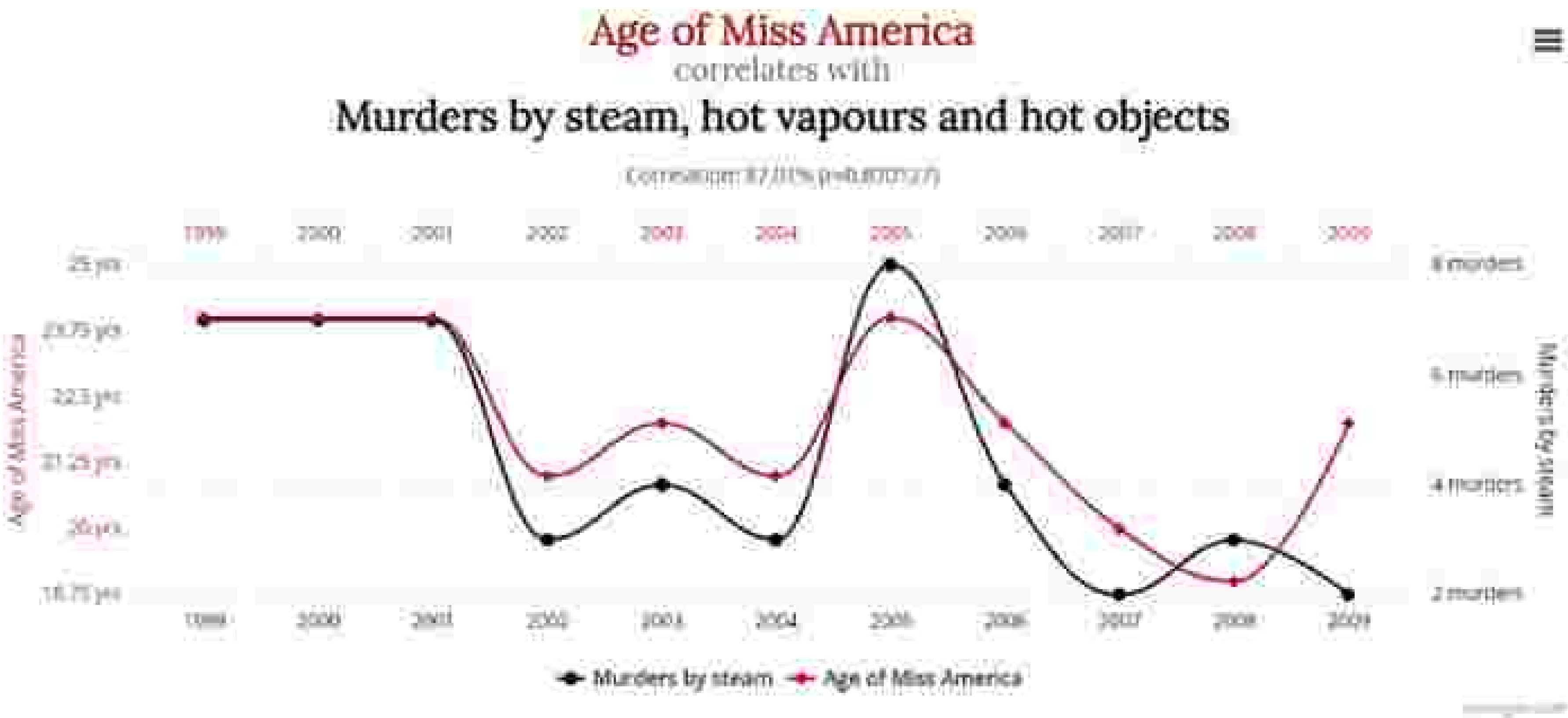
Ask freewill assumption provides a ratonnelle Attude is in that key responsibility, good and justce.

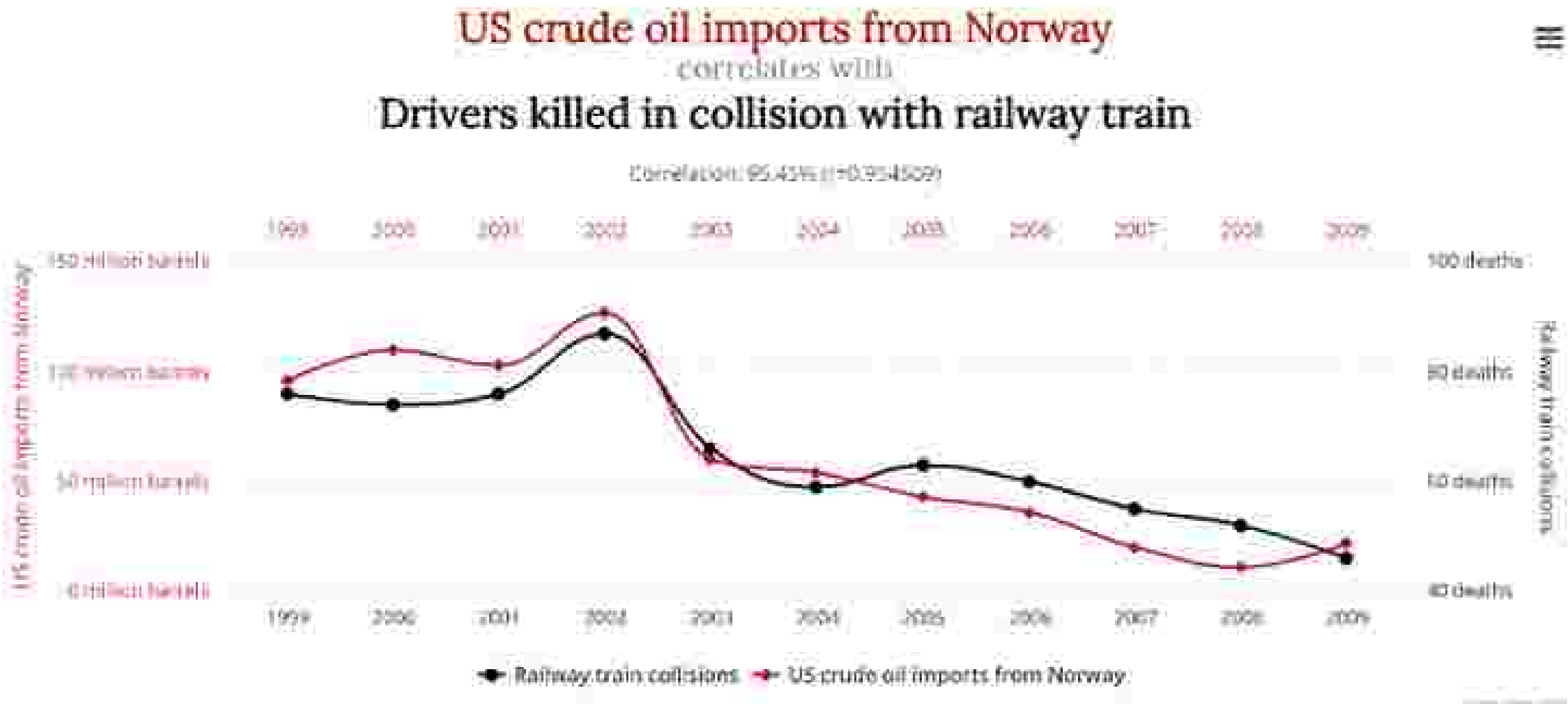
parental Index



30 links but 5 causes!




Age of Miss America
correlates with
Murders by steam, hot vapours and hot objects
Age of Miss America
Murders by steam
Murders by steam
Age of Miss America



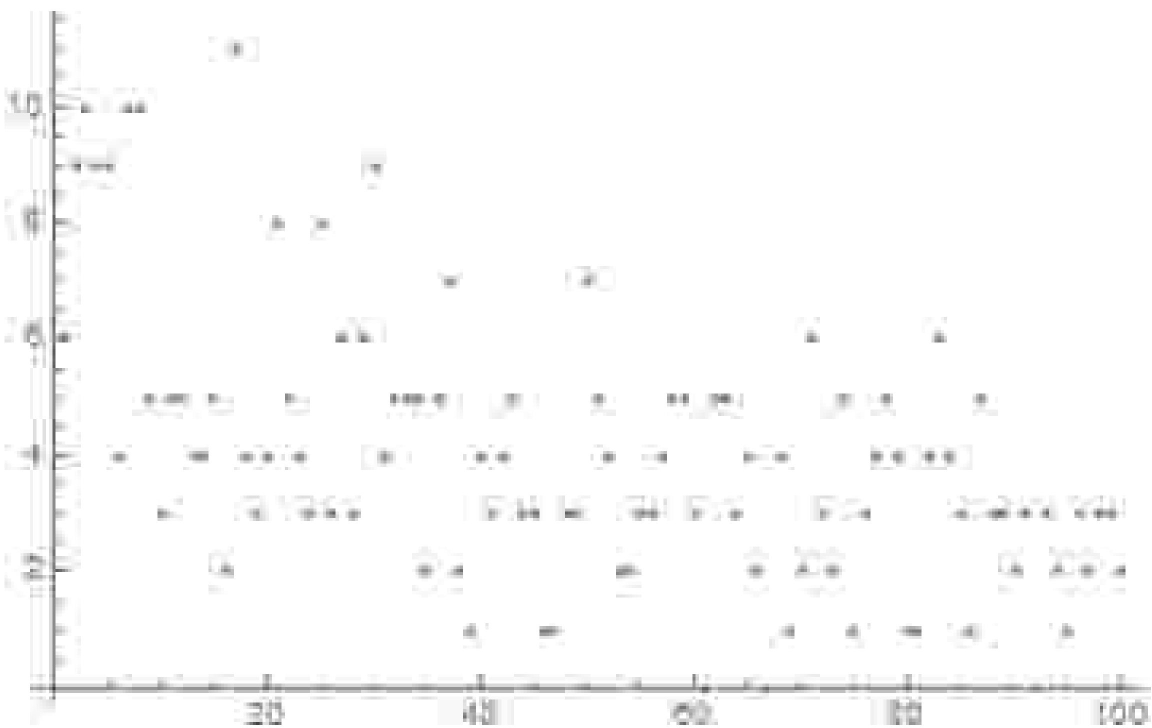
US crude oil imports from Norway
correlates with
Drivers killed in collision with railway train
Correlation: 95.43% (r=0.954509)
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
100 deaths
80 deaths
60 deaths
40 deaths
Railway train collisions
US crude oil imports from Norway

Spurious correlations

http: //tylervigen.com/spurious-correlatons

Warning!

Number of days the record heat is batu per year if we compile the data for 50 years, assuming that the temperature is normally distributed and **the mean is fixed.**



Warning!

If a researcher is studying the impact of coffee on cogniton Foncton of 0
following variables.

- Man and woman
- Child, adolescent, adult and old
- Efet short term and long term
- Afecte IQ, créatvité or émotonnelle intelligence.

There is a 99% chance to discover a négatf Positive feedback ratio or impact on a subgroup even if we replace the coffee with water.