

Why recording No Findings ? A common sample selection bias

5th Annual BayesiaLab Conference in Paris

HIGHER TOGETHER™



5th Annual BayesiaLab Conference in Paris

+33 1 47 11 32 00

gilles.debache@dassault-aviation.com



Presentation Plan

Context

- Machine learning and Screening
- Efficiency for business et technical analysis

No Findings

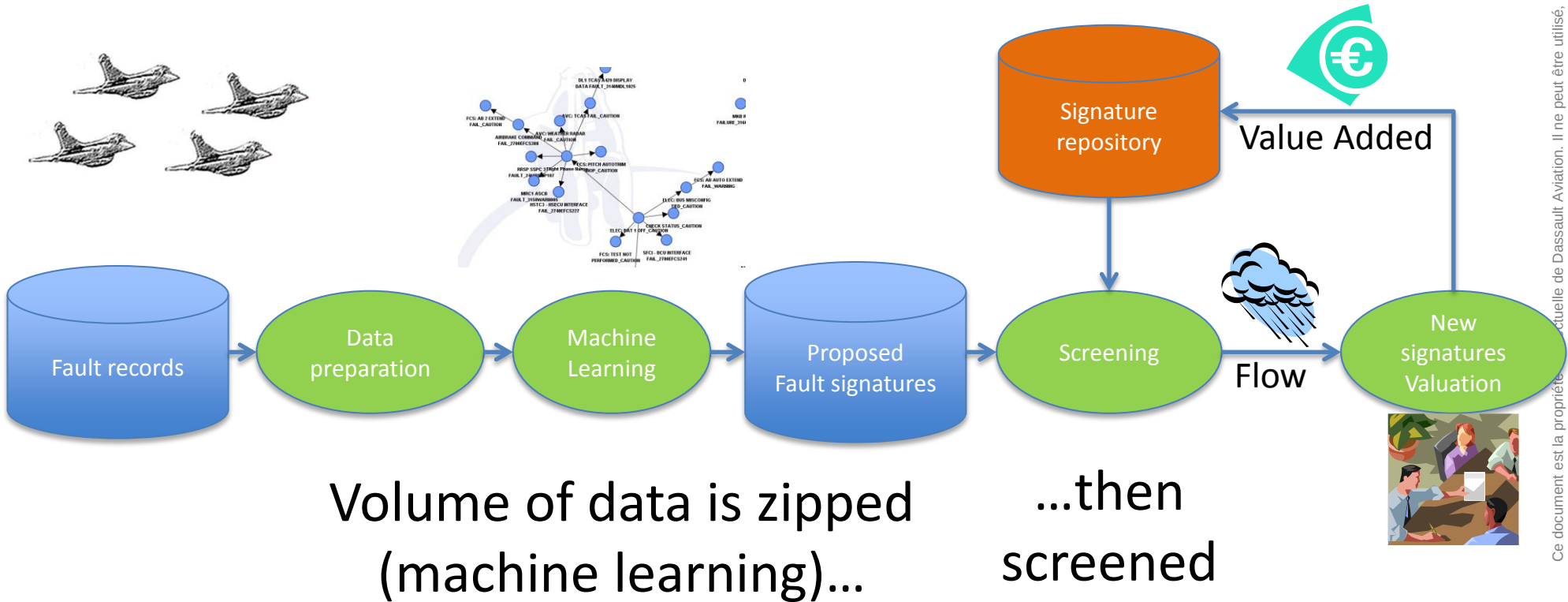
- A reminder about statistic dependency
- The value of No Findings data

Conclusions

- No Findings are required to sign Findings
- ML and Screening for human analysis efficiency

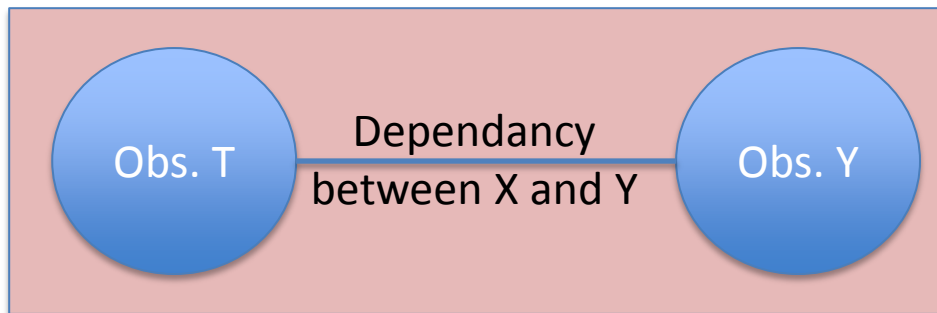
- Machine learning and Screening
- Efficiency for business et technical analysis

Machine learning and screening

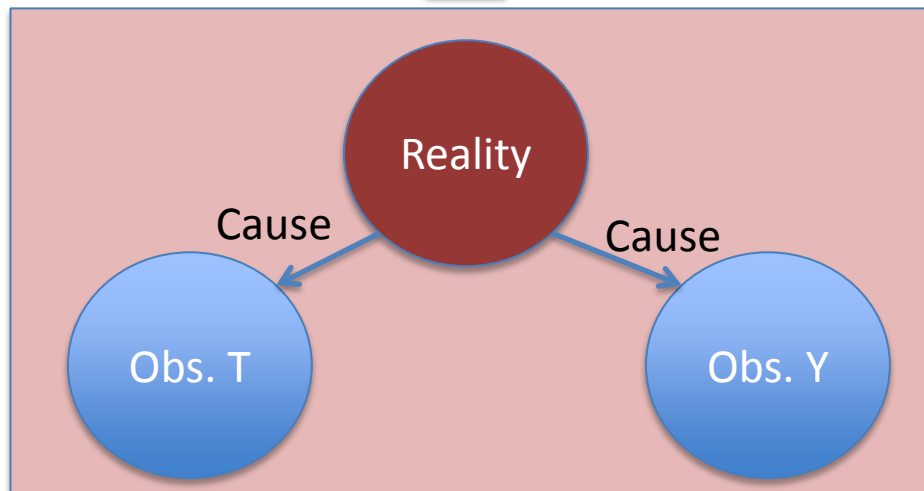


Ce document est la propriété exclusive de Dassault Aviation. Il ne peut être utilisé, reproduit, modifié ou communiqué sans son autorisation. Dassault Aviation Proprietary Data.

Phenomenons as sign of an ultimate reality

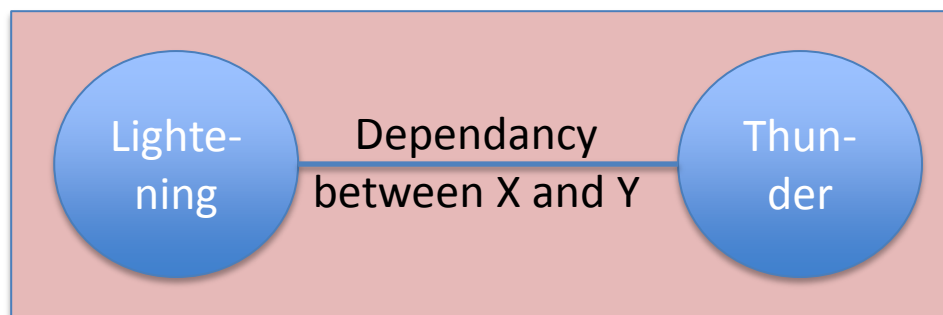


1* Dependency is not causality

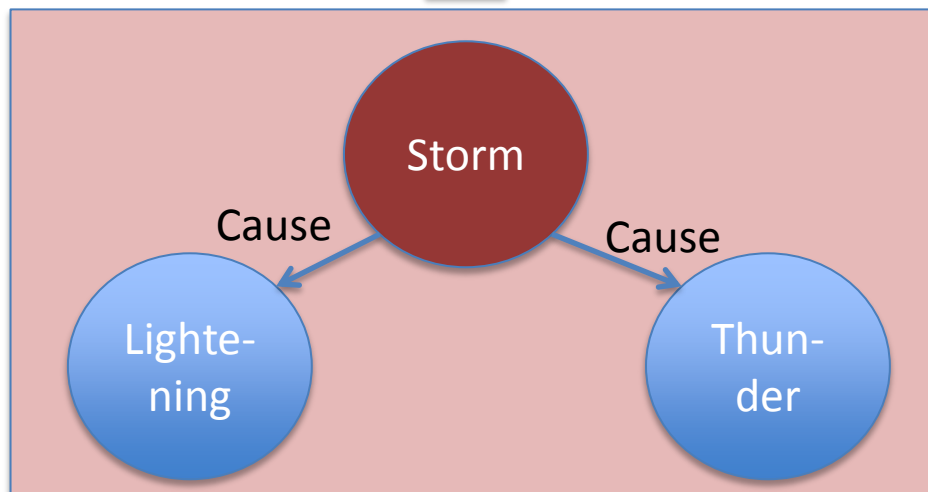


2* Ultimate reality could be hidden (not observed)

Phenomenons as sign of an ultimate reality

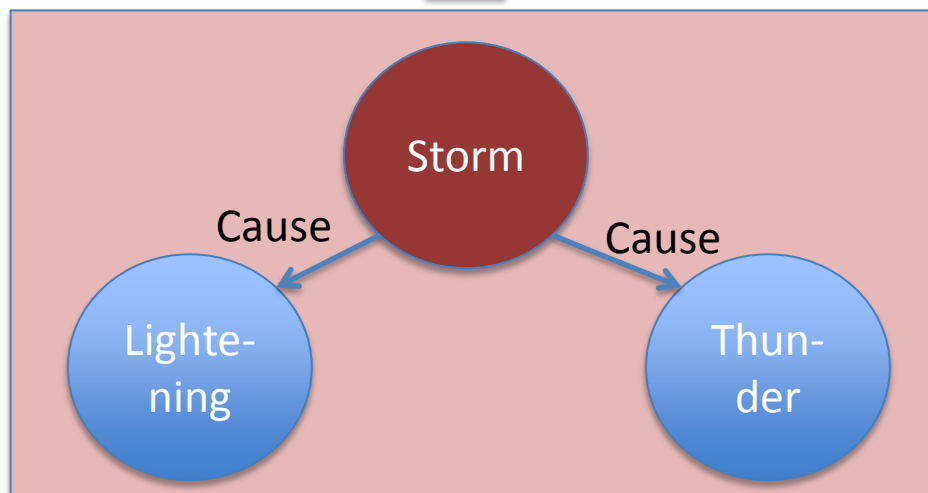
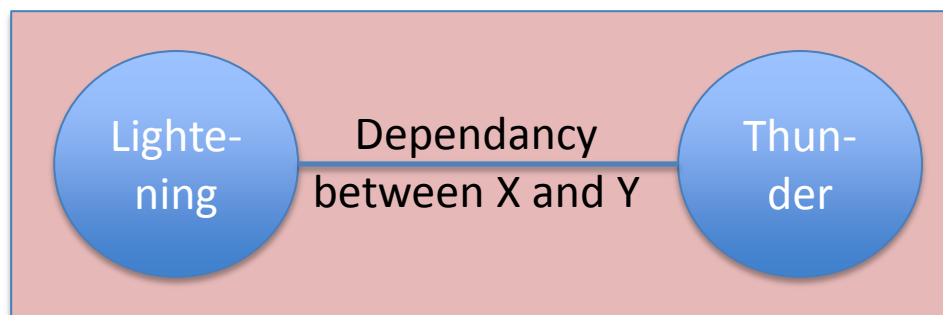


1* Dependency is not causality



2* Ultimate reality could be hidden (not observed)

Let model a storm and its phenomenons



IF(ALEA()>0,99;« Lightning";""))

- Lightning is observed on 1%

IF(B2=" Lightning
";IF(ALEA()>0,95;"";« Thunder");
IF(ALEA()>0,99;« Thunder";""))

- If Lightning then Thunder appears on 95%.
- And if no Lightning, Thunder appear 1%

Let add independant observations :

IF(ALEA()>0,5;« Day »;« Night »)

- And its daylight 50%

IF(ALEA()>0,8;« Sea »;« Land »)

- 80% of experiments are on sea.



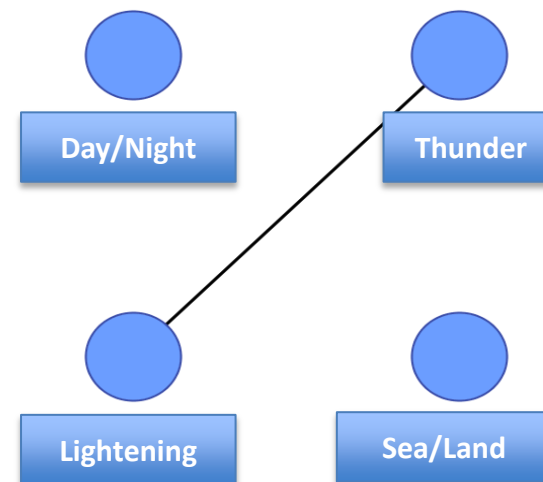
- A reminder about statistic dependency
- The value of No Findings data

True dependancies are found and no False dependancies appear

Day/Night	Lightening	Thunder	Sea/Land
Jour	-	Tonnerre	Mer
Nuit	-	Tonnerre	Terre
Jour	Foudre	-	Mer
Nuit	-	Tonnerre	Mer
Jour	-	Tonnerre	Mer
Jour	Foudre	Tonnerre	Mer
Nuit	Foudre	-	Terre
Nuit	Foudre	Tonnerre	Mer
Jour	-	Tonnerre	Mer
Nuit	Foudre	Tonnerre	Mer
Jour	Foudre	-	Mer
Jour	-	-	Mer
Nuit	-	-	Mer
Nuit	-	-	Mer
Nuit	-	-	Mer
Jour	-	-	Mer

Findings
11 lignes

« No Findings »
352 lignes



5th Annual BayesiaLab Conference in Paris

- A reminder about statistic dependency
- The value of No Findings data

But without « No Findings » :

Day/Night	Lightening	Thuneder	Sea/Land
Jour	-	Tonnerre	Mer
Nuit	-	Tonnerre	Terre
Jour	Foudre	-	Mer
Nuit	-	Tonnerre	Mer
Jour	-	Tonnerre	Mer
Jour	Foudre	Tonnerre	Mer
Nuit	Foudre	-	Terre
Nuit	Foudre	Tonnerre	Mer
Jour	-	Tonnerre	Mer
Nuit	Foudre	Tonnerre	Mer
Jour	Foudre	-	Mer
Jour	-	-	Mer
Nuit	-	-	Mer
Nuit	-	-	Mer
Nuit	-	-	Mer
Jour	-	-	Mer

Dependancy between Lightning and Thunder is lost

Findings
11 lignes

Day/Night

Thunder

Lightening

Sea/Land

No dependancy incuding by increasing the number of Findings

« No Findings » are removed

5th Annual BayesiaLab Conference in Paris



- A reminder about statistic dependency
- The value of No Findings data

Findings have value...

Day/Night	Lightening	Thuneder	Sea/Land
Jour	-	Tonnerre	Mer
Nuit	-	Tonnerre	Terre
Jour	Foudre	-	Mer
Nuit	-	Tonnerre	Mer
Jour	-	Tonnerre	Mer
Jour	Foudre	Tonnerre	Mer
Nuit	Foudre	-	Terre
Nuit	Foudre	Tonnerre	Mer
Jour	-	Tonnerre	Mer
Nuit	Foudre	Tonnerre	Mer
Jour	Foudre	-	Mer
Jour	-	-	Mer
Nuit	-	-	Mer
Nuit	-	-	Mer
Nuit	-	-	Mer
Jour	-	-	Mer

...only with « No Findings »

« $\neg q \Rightarrow \neg p$ » is equivalent to « $p \Rightarrow q$ »

If no $\neg q \Rightarrow \neg p$
then no $p \Rightarrow q$





- A reminder about statistic dependency
- The value of No Findings data

The Raven paradox



Carl Gustav
Hempel
(1905-1997)

Day/Night	Lightening	Thuneder	Sea/Land
Jour	-	Tonnerre	Mer
Nuit	-	Tonnerre	Terre
Jour	Foudre	-	Mer
Nuit	-	Tonnerre	Mer
Jour	-	Tonnerre	Mer
Jour	Foudre	Tonnerre	Mer
Nuit	Foudre	-	Terre
Nuit	Foudre	Tonnerre	Mer
Jour	-	Tonnerre	Mer
Nuit	Foudre	Tonnerre	Mer
Jour	Foudre	-	Mer
Jour	-	-	Mer
Nuit	-	-	Mer
Nuit	-	-	Mer
Nuit	-	-	Mer
Jour	-	-	Mer



Observing objects that are neither black nor ravens **improve** the weight of evidence in favor of the hypothesis that all ravens are black,

Please change improve with **is required** :

If no $\neg q \Rightarrow \neg p$
then no $p \Rightarrow q$

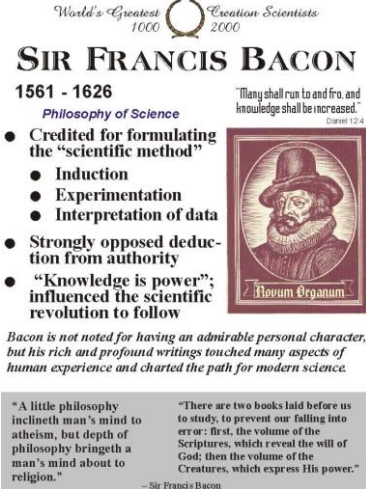


Ce document e
ou communiqué sans son autorisation. Dassault Aviation Proprietary Data.

The 3 tables for Induction

According to Bacon, the search for forms must continue investigating all known instances

- Presence (Presencia) = where are recorded all instances in which the phenomenon occurs.
- Absence (Absencia) = where some of the cases related to previous but where the phenomenon does not occur.
- Gradi (RANK) = in which are recorded all the cases or instances in which the phenomenon presents itself according to a greater or lesser intensity.



The 3 tables for Induction

Day/Night	Lightening	Thuneder	Sea/Land
Jour	-	Tonnerre	Mer
Nuit	-	Tonnerre	Terre
Jour	Foudre	-	Mer
Nuit	-	Tonnerre	Mer
Jour	-	Tonnerre	Mer
Jour	Foudre	Tonnerre	Mer
Nuit	Foudre	-	Terre
Nuit	Foudre	Tonnerre	Mer
Jour	-	Tonnerre	Mer
Nuit	Foudre	Tonnerre	Mer
Jour	Foudre	-	Mer
Jour	-	-	Mer
Nuit	-	-	Mer
Nuit	-	-	Mer
Nuit	-	-	Mer
Jour	-	-	Mer

Presencia Table

Absencia Table

World's Greatest

1000

Creation Scientists

2000

SIR FRANCIS BACON

1561 - 1626

Philosophy of Science

- Credited for formulating the "scientific method"
 - Induction
 - Experimentation
 - Interpretation of data
- Strongly opposed deduction from authority
- "Knowledge is power"; influenced the scientific revolution to follow

Bacon is not noted for having an admirable personal character, but his rich and profound writings touched many aspects of human experience and charted the path for modern science.

"A little philosophy inclineth man's mind to atheism, but depth of philosophy bringeth a man's mind about to religion."

— Sir Francis Bacon

"There are two books laid before us to study, to prevent our falling into error: first, the volume of the Scriptures, which reveal the will of God; then the volume of the Creatures, which express His power."

Ce document est la propriété intellectuelle de Dassault Aviation. Il ne peut être ou communiqué sans son autorisation. Dassault Aviation Proprietary Data.

The Absentia table **is required**

Day/Night	Lightening	Thuneder	Sea/Land
Jour	-	Tonnerre	Mer
Nuit	-	Tonnerre	Terre
Jour	Foudre	-	Mer
Nuit	-	Tonnerre	Mer
Jour	-	Tonnerre	Mer
Jour	Foudre	Tonnerre	Mer
Nuit	Foudre	-	Terre
Nuit	Foudre	Tonnerre	Mer
Jour	-	Tonnerre	Mer
Nuit	Foudre	Tonnerre	Mer
Jour	Foudre	-	Mer
Jour	-	-	Mer
Nuit	-	-	Mer
Nuit	-	-	Mer
Nuit	-	-	Mer
Jour	-	-	Mer

Absencia Table

World's Greatest

1000

Creation Scientists

2000

SIR FRANCIS BACON

1561 - 1626

Philosophy of Science

- Credited for formulating the "scientific method"
 - Induction
 - Experimentation
 - Interpretation of data
- Strongly opposed deduction from authority
- "Knowledge is power"; influenced the scientific revolution to follow

"Many shall run to and fro, and knowledge shall be increased."

Novum Organum

Bacon is not noted for having an admirable personal character, but his rich and profound writings touched many aspects of human experience and charted the path for modern science.

"A little philosophy inclineth man's mind to atheism, but depth of philosophy bringeth a man's mind about to religion."

"There are two books laid before us to study, to prevent our falling into error: first, the volume of the Scriptures, which reveal the will of God; then the volume of the Creatures, which express His power."



Observing objects that are neither black nor ravens **improve** the weight of evidence in favor of the hypothesis that all ravens are black,

Please change improve with **is required** :

If no $\neg q \Rightarrow \neg p$

then no $p \Rightarrow q$





- A reminder about statistic dependency
- The value of No Findings data

A common sample bias

Challenger
disaster

To ignore « No Findings »

Seal Failure	X	X	X	X	X	X	X	X

Cold
Weather

Warm
Weather

There are the same number of seal in fault
when the weather is cold then
when the weather is warm. So ?

The Flight director concluded that weather has no link
with seal fault.



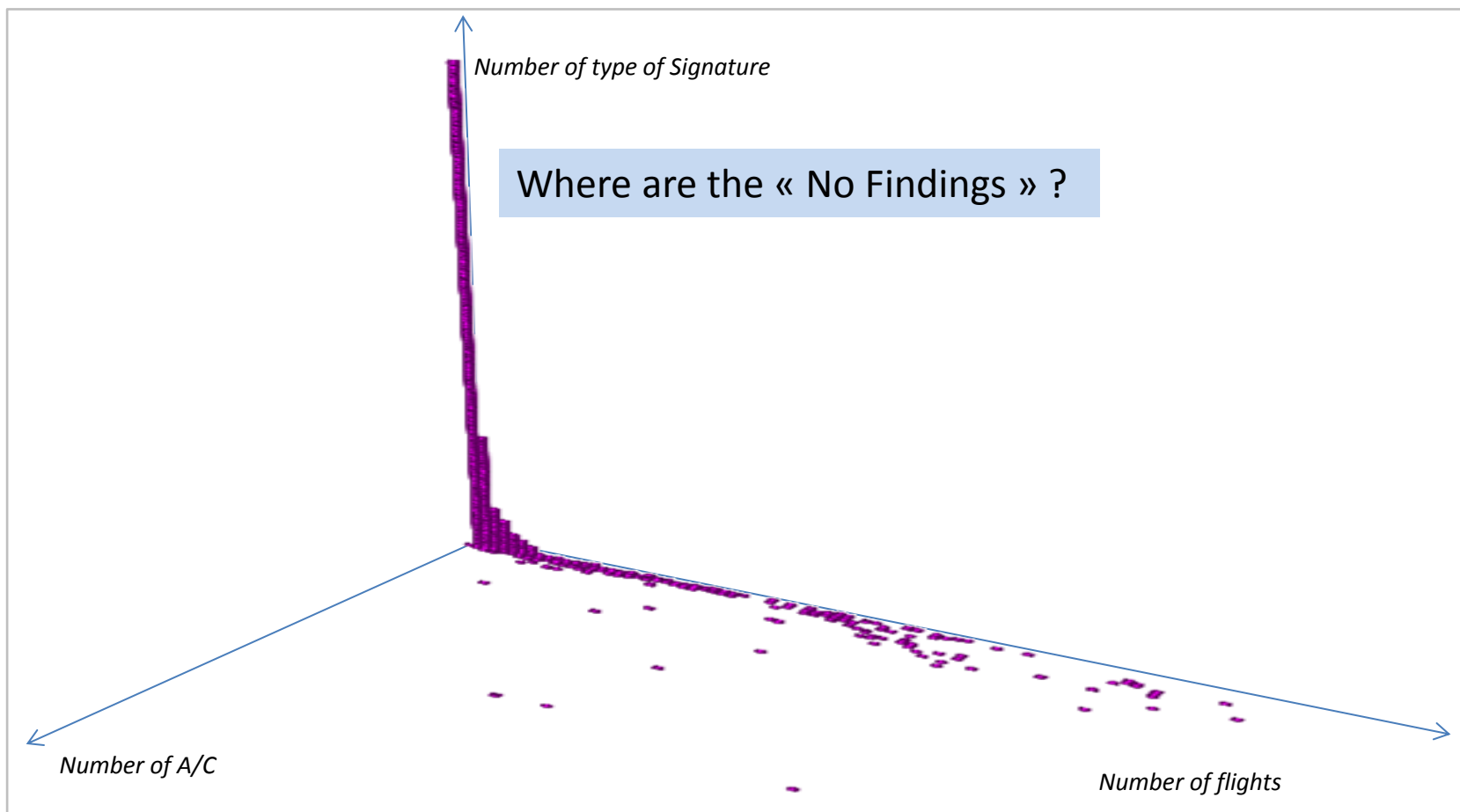
Challenger disaster

Seal Failure	X	X	X	X	X	X	X	X
Seal OK					XXX	XXXX	XXXX	XXXXX
	Cold Weather				Warm Weather			



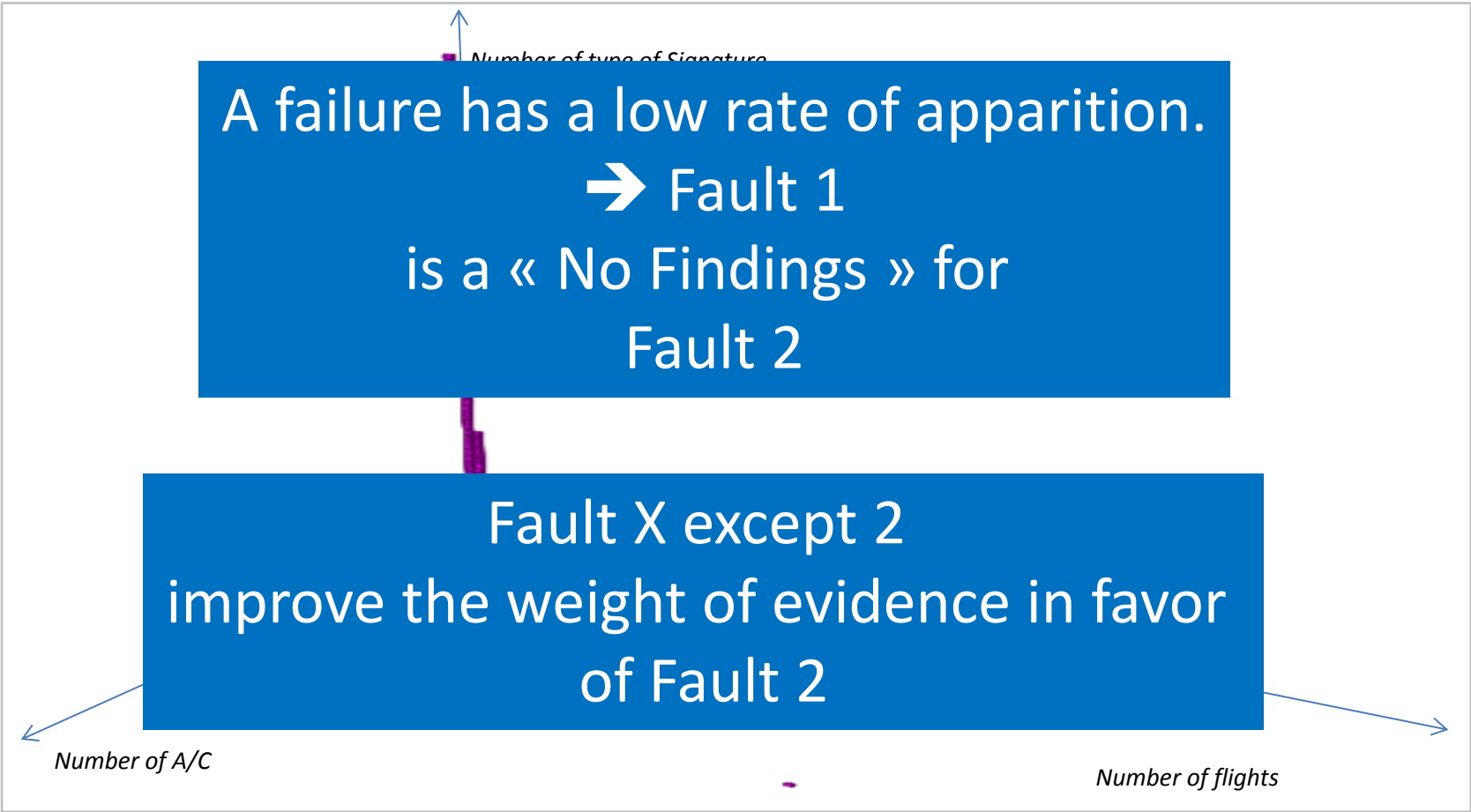
- No Findings are required to sign Findings
- ML and Screening for human analysis efficiency

Mainly failures with low rate of apparition...



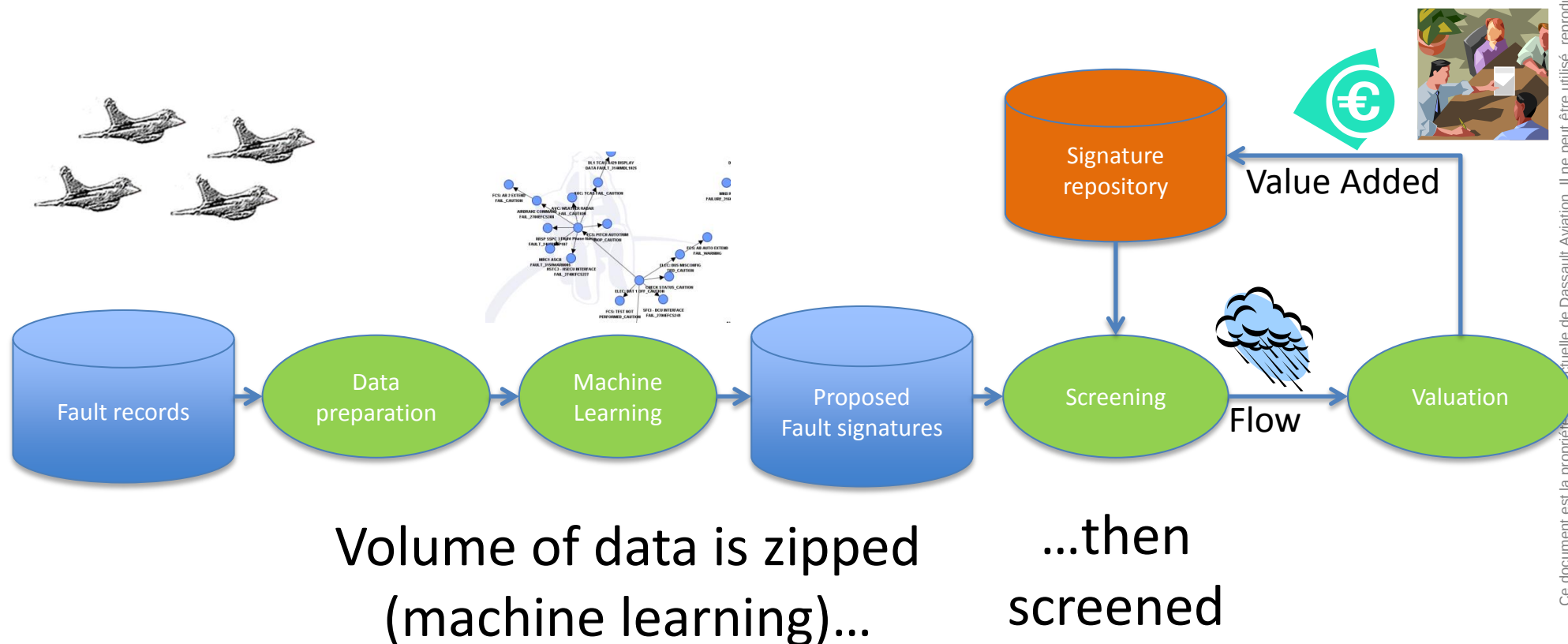
- No Findings are required to sign Findings
- ML and Screening for human analysis efficiency

...which are also « No Findings »
for the other faults.



- No Findings are required to sign Findings
- ML and Screening for human analysis efficiency

Data reduction is for human analysis



- No Findings are required to sign Findings
- ML and Screening for human analysis efficiency

Context

- Machine learning and Screening
- Efficiency for business et technical analysis

No Findings

- A reminder about statistic dependency
- The value of No Findings data

Conclusions

- No Findings are required to sign Findings
- ML and Screening for human analysis efficiency

