



THE NEW DEAL

COMMENT LA DONNÉE TRANSFORME LE MÉTIER DES ACTUAIRES ?
("CODO ERGO SUM" ?)

Data Science pour les actuaires
2^{ème} promotion 7 mars 2016
Leçon inaugurale

And then he whispered the three
words every woman wants to hear...
"I'm an actuary."



The Unreasonable Effectiveness of Data

Alon Halevy, Peter Norvig, and Fernando Pereira, *Google*

“Invariably, simple models and a lot of data trump more elaborate models based on less data”



A-t-on encore besoin d'actuaire?

BIG DATA WILL BROADEN OUR HORIZONS



New opportunities & challenges



New technologies and profiles



Data science is a process!

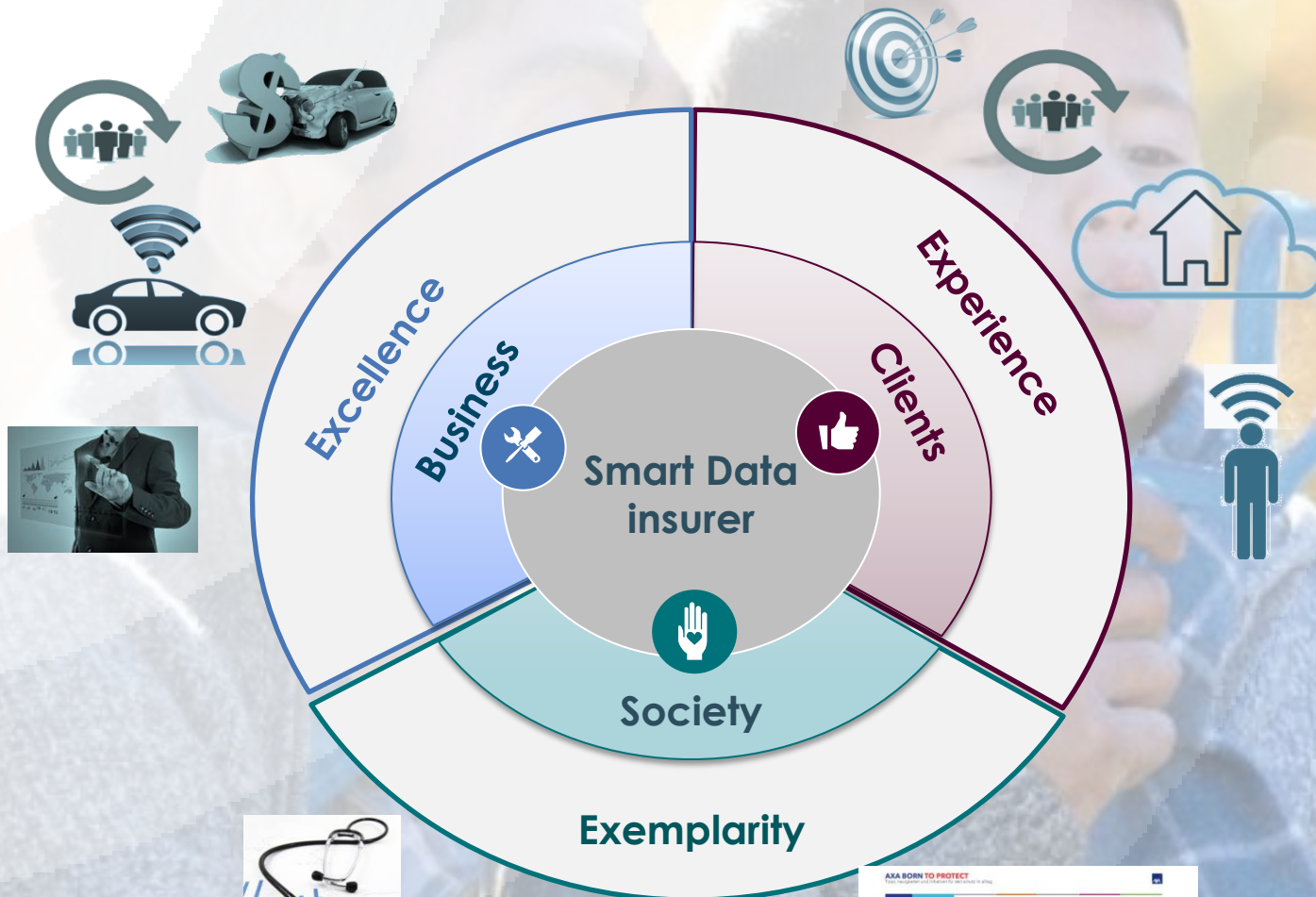
A photograph of Henri de Castries, CEO of AXA, speaking at a podium. He is wearing a grey suit, a white shirt, and a blue patterned tie. He has a microphone clipped to his lapel and is gesturing with his hands. The background is a purple screen with the text "Henri de CASTRIES" and "Directeur Général" visible.

"Big Data is an economical and technological revolution...

...being defensive is a waste of time as it is unavoidable and lethal"

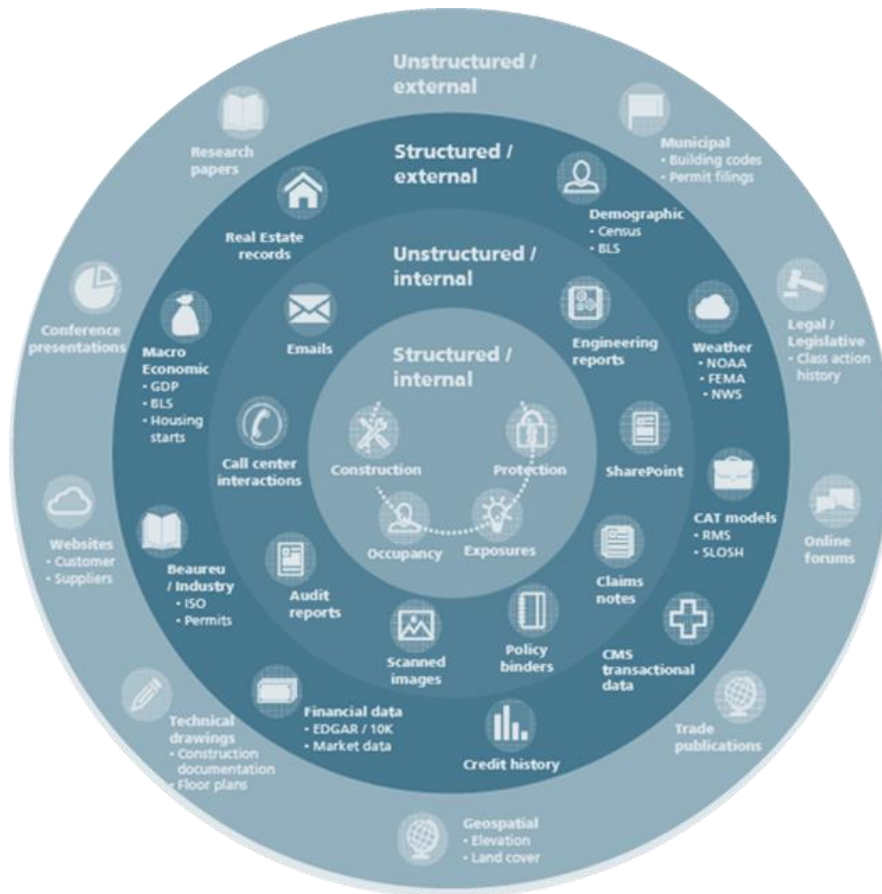
*- Henri de Castries
AXA CEO*

Our conviction: Big Data is an opportunity for our business, clients and society



The challenges of Big Data

> The frenzy trend of data; the 3 V's



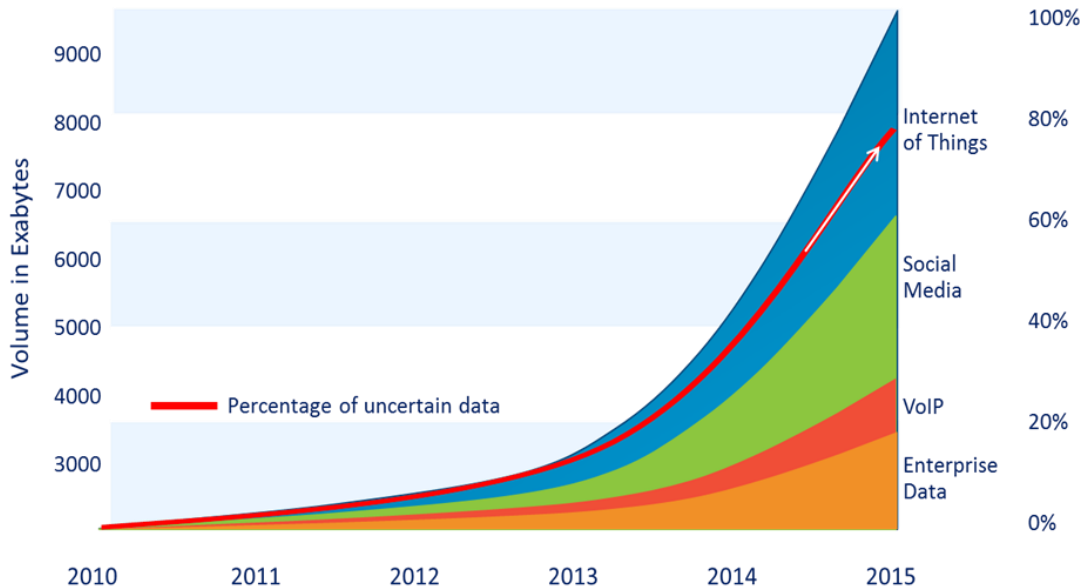
VOLUME

VARIETY

VELOCITY

Big Data is exponential...

> Still a goldmine to exploit



EXPONENTIAL RISE OF DATA

GROWING IMPORTANCE OF UNSTRUCTURED DATA

QUALITY (VERACITY & VALIDITY) IS A GROWING CHALLENGE



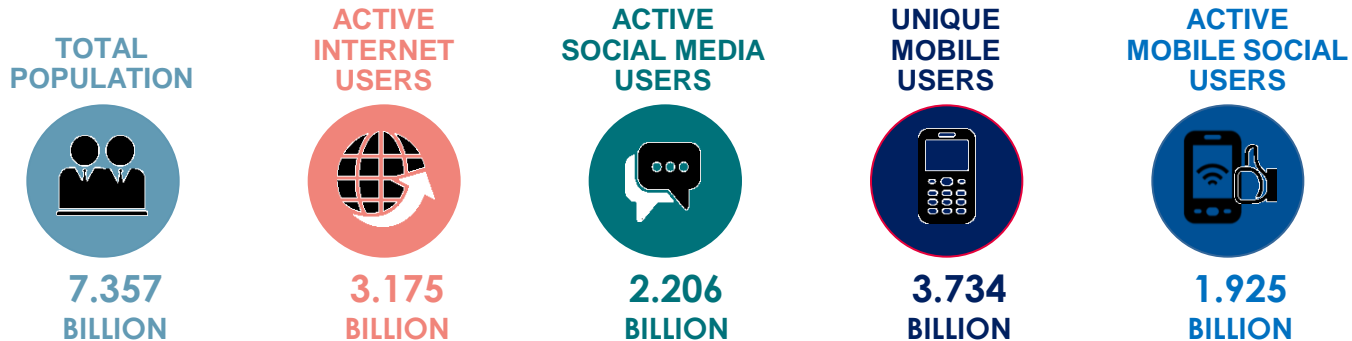
WE TAG AROUND 20% OF THE USEFUL DATA AND ANALYZE ONLY 5%

Sources: IBM Global Technology Outlook – 2012

http://www.progressivepolicy.org/wp-content/uploads/2013/09/09.2013-Mandel_Can-the-Internet-of-Everything-Bring-Back-the-High-Growth-Economy-1.pdf

Data is transforming people's lives

> Internet of people: new interactions, new behaviors, new usages



- Sharing economy: usage vs. ownership
- Solomo [Social – Local – Mobile]: real life in real time



4.9 billion connected things will be in use in 2015 and will reach **25 billion** by 2020**.

*Data wearesocial, August 2015

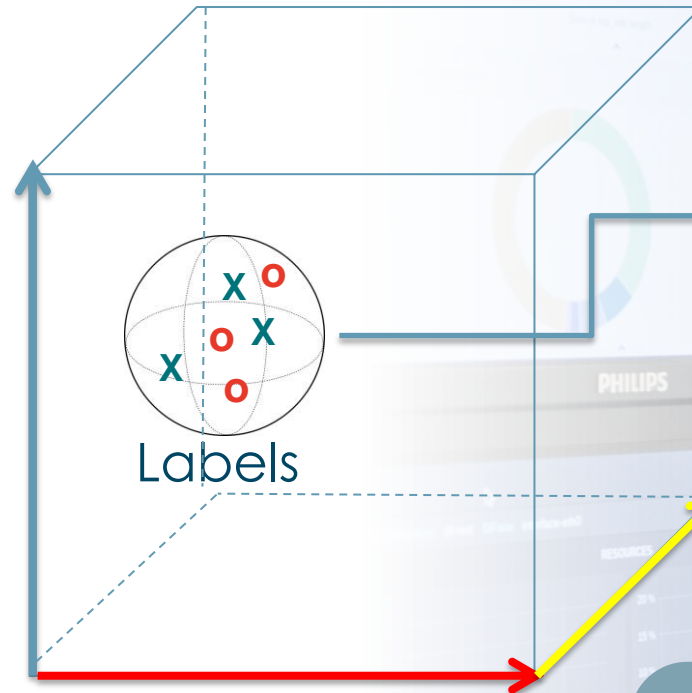
**Data Gartner Inc, 2014

Learning in the data cube*

> An industry perspective

n observations

Biased
Redundancy
Growing volume
Real-time
Low Meta data
management Maturity



Biased
Rare
Imbalanced
Noisy

d dimensions

Access to data
Data quality (format, missing data, noise...)
Historic duration
Unstructured data
Curse of dimensionality (generalization challenge)

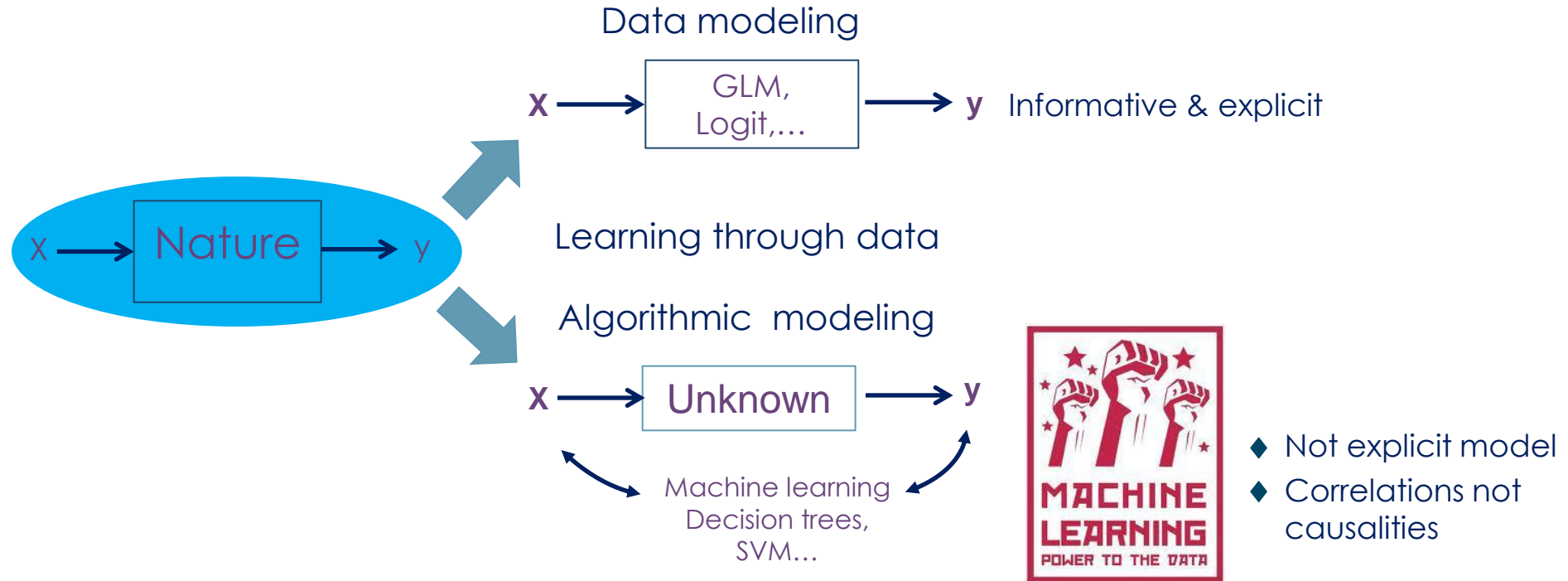
k actions

Personalized treatment learning (causal inference)
Not randomized treatment
Interpretability
Reality
Performance monitoring and causality (e.g. homophily vs influence, true lift)

* From an idea of F. Bach

... and steers the development of an algorithmic modeling culture*

> The emergence of Machine Learning: here is the age of algorithms

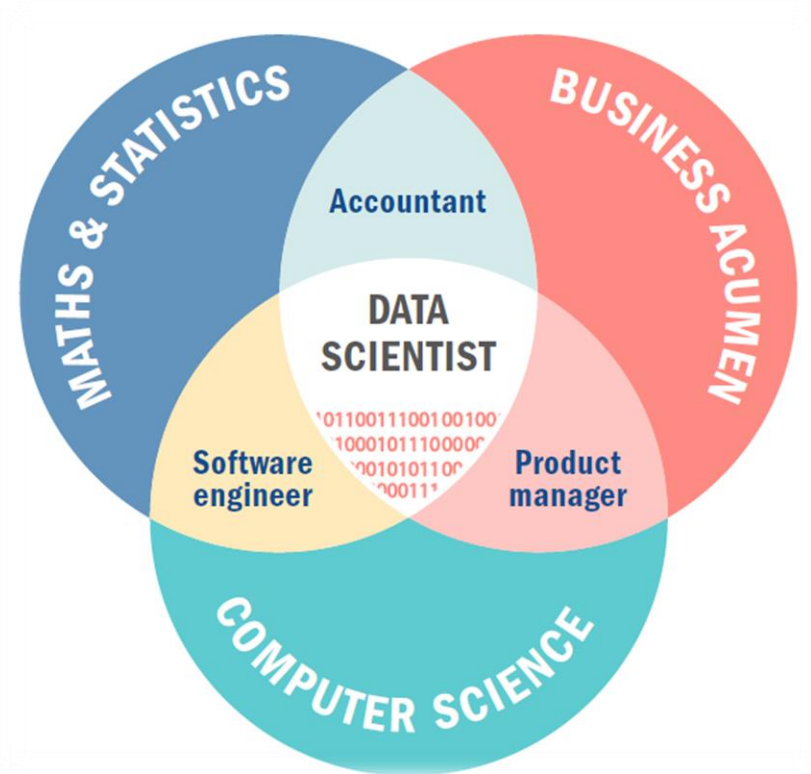


From static approach to more Iterative and adaptive process
New kind of ecosystem

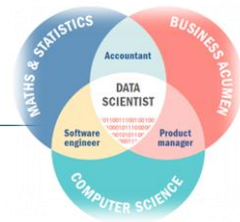
* cf. "Statistical modeling: the two cultures" of Léo Breiman

... the emergence of data scientists...

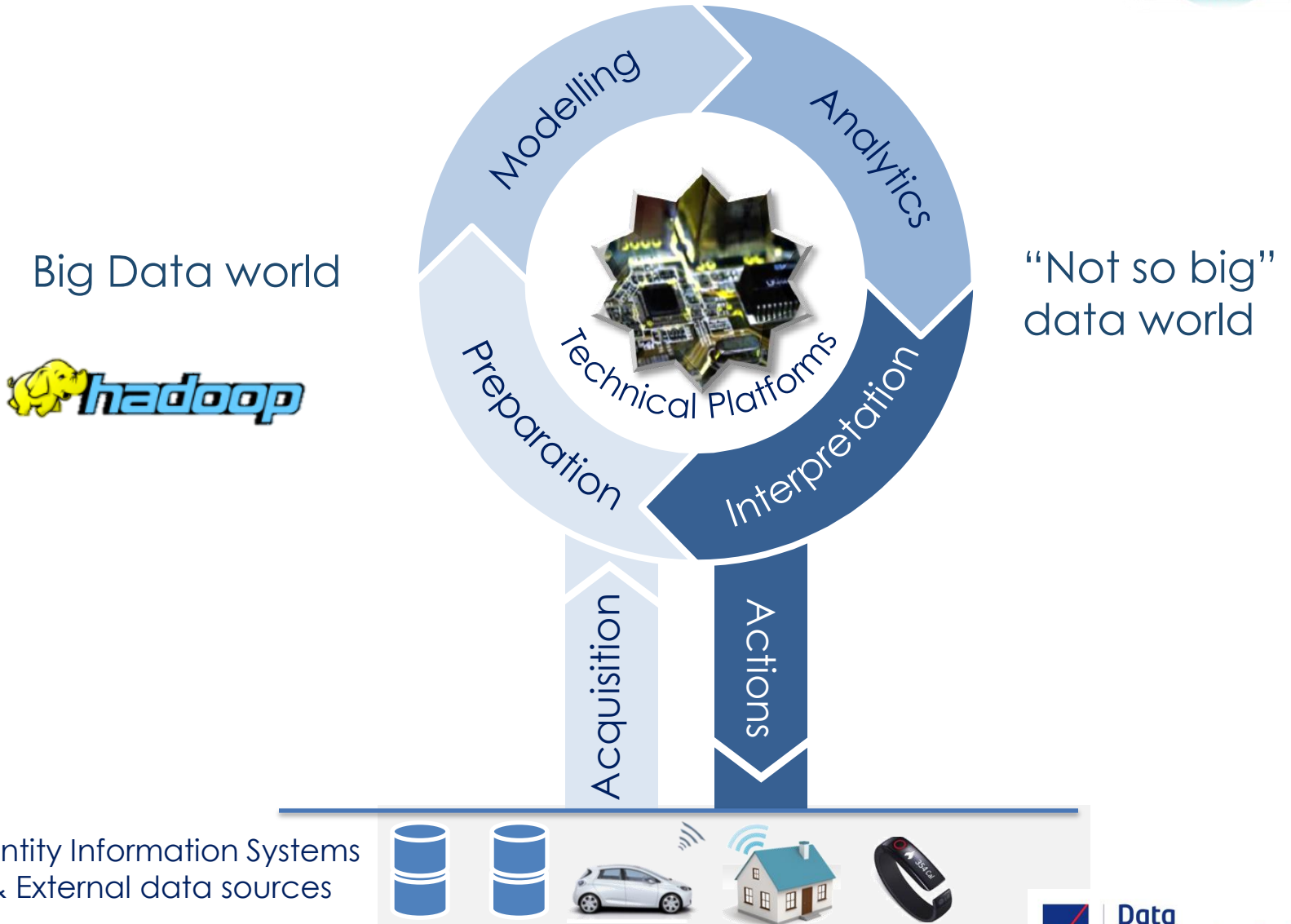
> The Data scientist definition



...and data science



> Data science is a cross-disciplinary and iterative process



Big Data world



"Not so big" data world

Entity Information Systems & External data sources



Illustration Telematics

The Unreasonable Effectiveness of Data

Alon Halevy, Peter Norvig, and Fernando Pereira, *Google*

“Invariably, simple models and a lot of data trump more elaborate models based on less data”

FEATURE ENGINEERING IS BECOMING MORE AND MORE IMPORTANT

Presentation of DIL Telematics solution

1

CONNECT YOUR CAR



2

DRIVE



3

ENJOY OUR SERVICES



Behind the scenes

1

CONNECT



iOS 9

2

COLLECT



21 kms



here



The Free Wiki World Map



3

COMPUTE

TRIP INTERPRETATION

```
{ "timestamp": 1437856905982, "location": { "bearing": 269.296875, "altitude": 94.0, "precision": 5.0, "longitude": 2.577787, "latitude": 49.004018, "speed": 5.166353 }, { "motion": { "acceleration": { "y": 1.101642, "x": 1.361841, "z": 0.549481 }, "gravity": { "y": -5.832105, "x": 1.312946, "z": 0.1778098 }, "rotation_rate": { "y": 0.049503, "x": 0.191346, "z": 0.153111 }, "timestamp": 1437856906243 }, { "timestamp": 1437856906735, "location": { "bearing": 266.132813, "altitude": 91.0, "precision": 5.0, "longitude": 2.577787, "latitude": 49.00401, "speed": 5.168603 }, { "motion": { "acceleration": { "y": 0.50353, "x": 0.59418, "z": -0.366929 }, "gravity": { "y": -5.57418, "x": 1.790774, "z": -7.899332 }, "rotation_rate": { "y": 0.122412, "x": -0.219113, "z": 0.526752 }, "timestamp": 1437856907256 }, { "timestamp": 1437856907693, "location": { "bearing": 247.148438, "altitude": 91.0, "precision": 5.0, "longitude": 2.577639, "latitude": 49.003995, "speed": 5.178486 }, { "motion": { "acceleration": { "y": 0.817697, "x": 1.307687, "z": ...
```

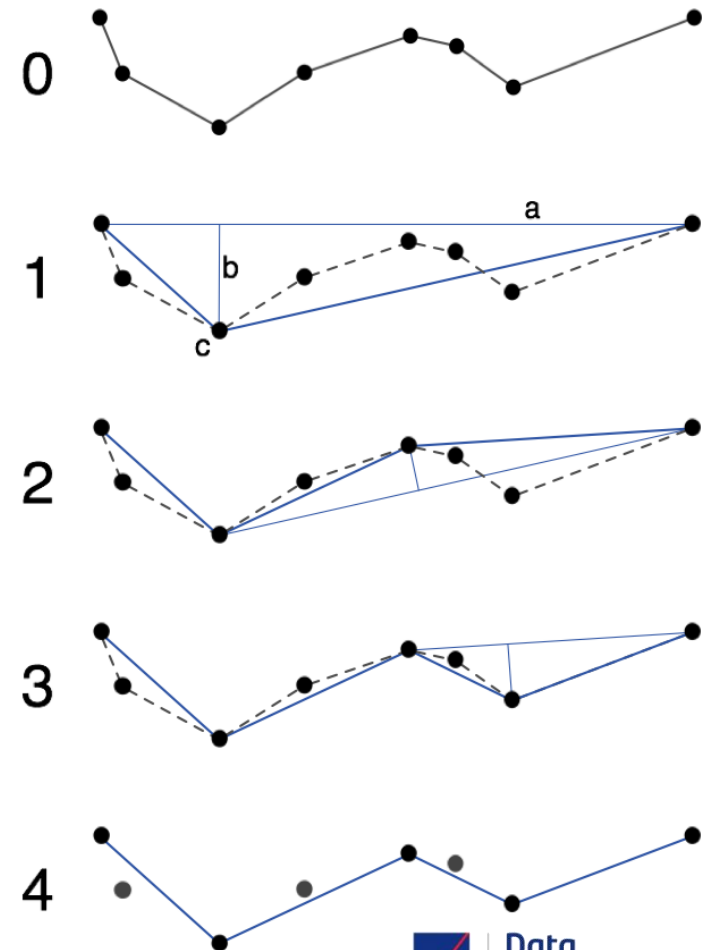
SCORING

You are among the 5% best drivers

How to tag a corner on a trip ?

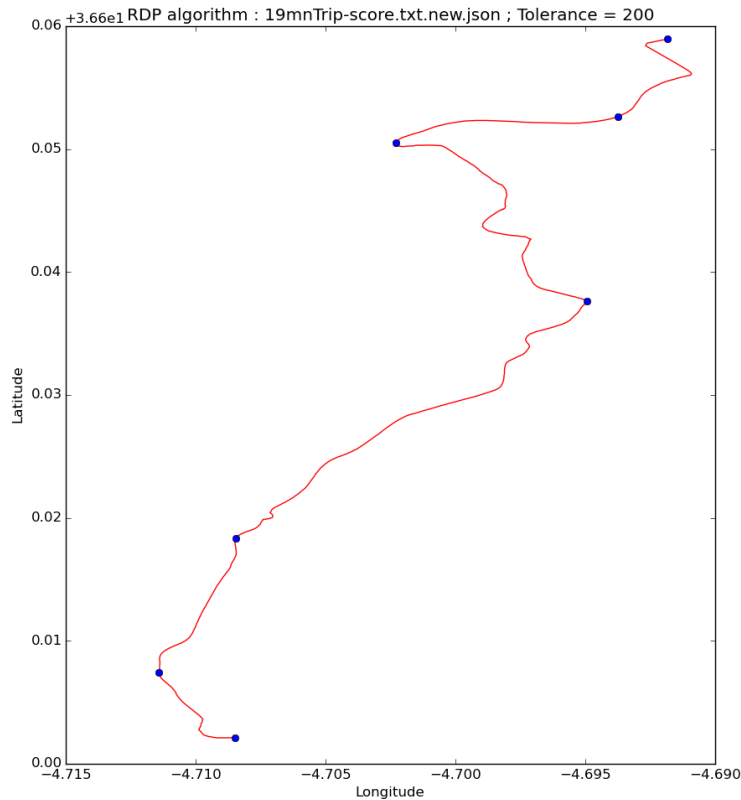
- ➔ **Initial algo: Forward States algorithm (FS) –curvatures sinuosity and angles**
 - ➔ Too many false positives due to noisy GPS data. Tolerance parameters needed for adjustment
- ➔ **Algo needs to be simplified, automated and more accurate**
- ➔ **Tracking trajectory turn – the Ramer-Douglas-Peucker algorithm (RDP)**
 - ➔ Introducing a tolerance parameter as the input
- ➔ **RDP algorithm appears to be efficient in tagging trajectory-shaping corners**

RDP algorithm

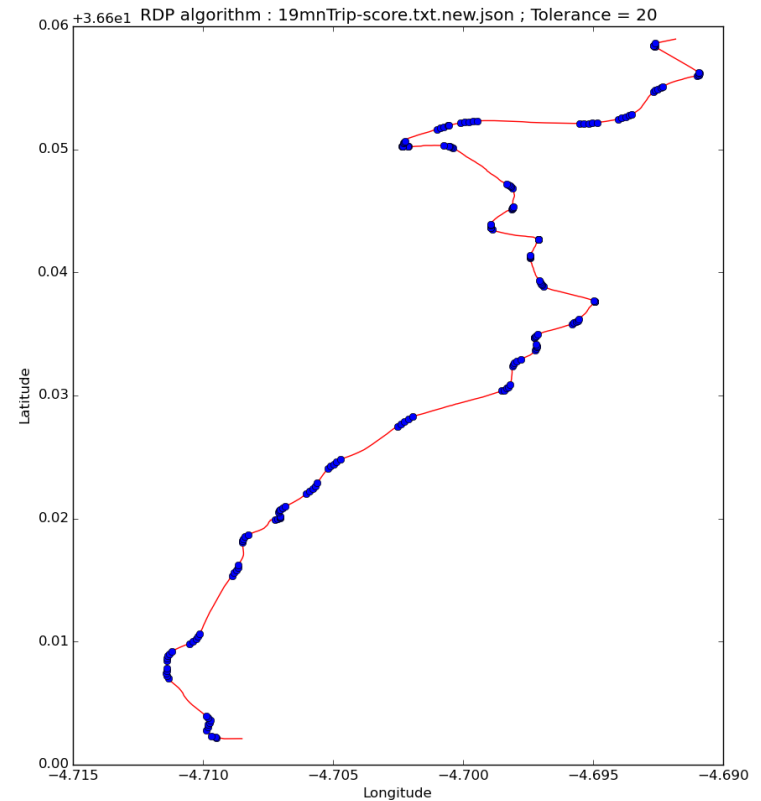


How to tag a corner on a trip ?

RDP-tagged datapoints on a given trajectory, for different tolerance parameters



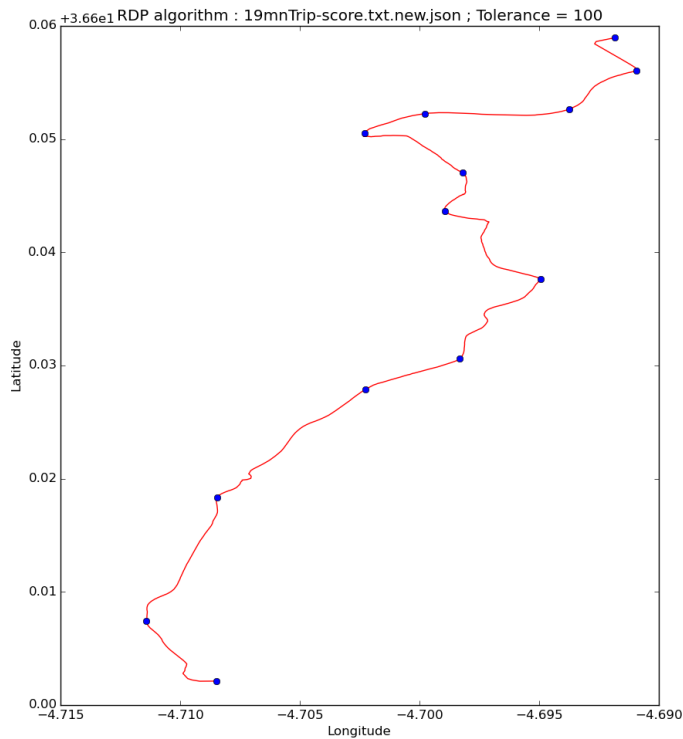
(a) Tolerance : 200 meters



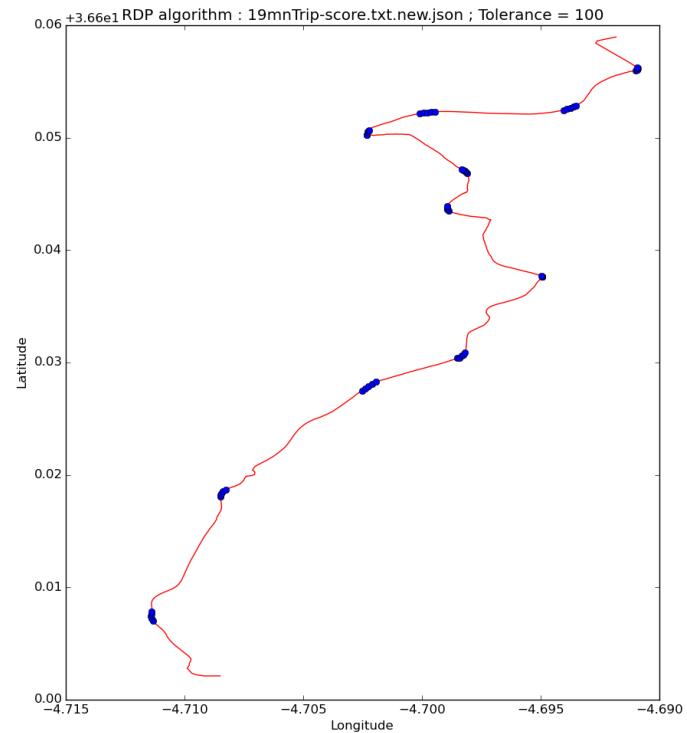
(b) Tolerance : 20 meters

How to tag a corner on a trip ?

Post processing allowed to consider the whole cornering



(a) No post-processing

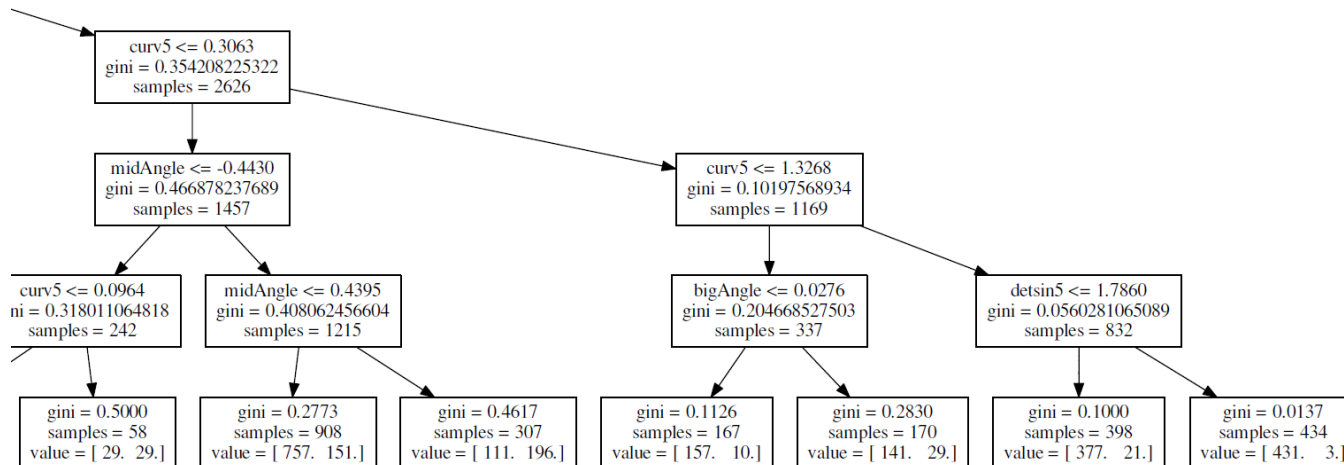


(b) Post-processing

How to tag a corner on a trip ?

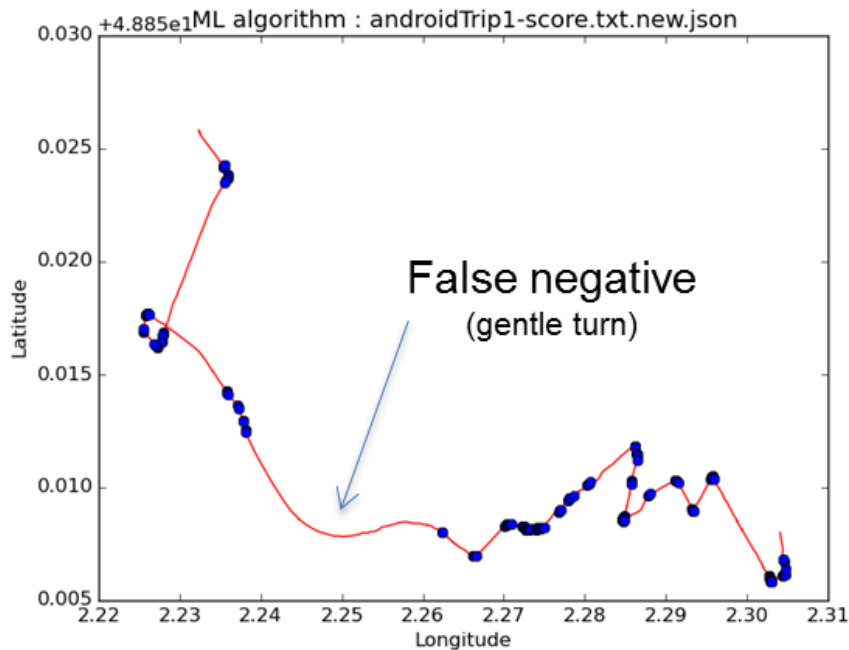
Post processing allowed to consider the whole cornering

- RDP algorithm tags poorly the local turns
- structure of a corner is inherently absorbed in the features of a given datapoint (GPS positions + specific features)
- Learning set: implementation of a user-friendly method to tag corners within a given trajectory
- Training of a Random Forest on tagged trajectories

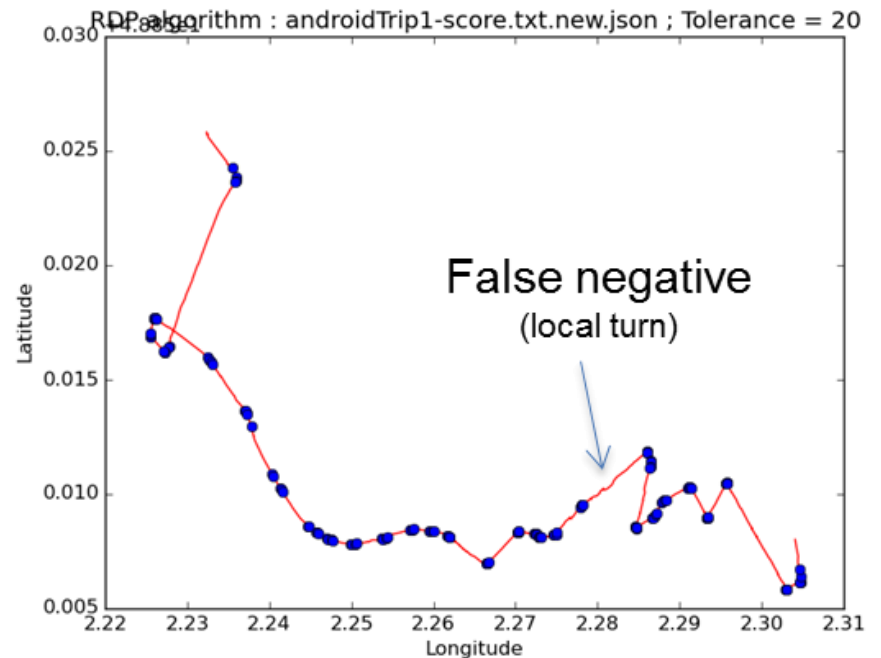


How to tag a corner on a trip ?

Combination of a geometric algorithm and a machine learning algorithm: automation of the cornering process and accurate results

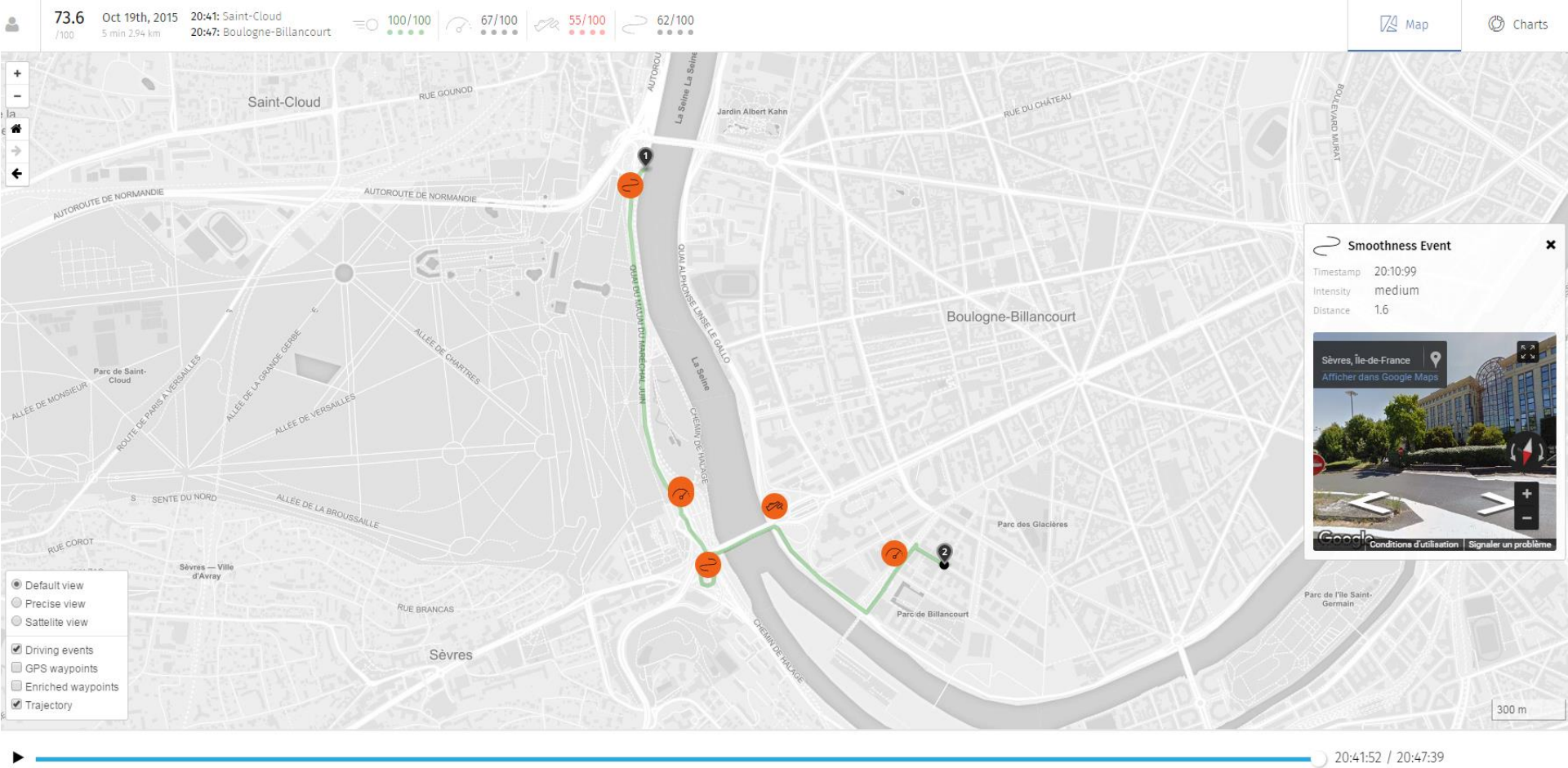


(a) False negatives for Random Forests



(b) False negatives for RDP algorithm

Telematics: Data viz



New ways of working to meet new challenges



Cloud & virtualization

« **Designed for failure** »



Source code management

« **With infrastructure as a code, systems engineers need to become developers** »

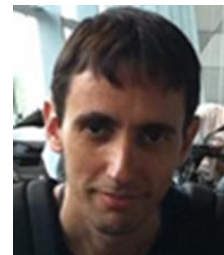


Business monitoring

And end-to end search & analytics platform infinitely versatile.



Collaborative work and Backlog management



Completed • \$30,000 • 1,528 teams

Driver Telematics Analysis

Mon 15 Dec 2014 – Mon 16 Mar 2015 (6 months ago)



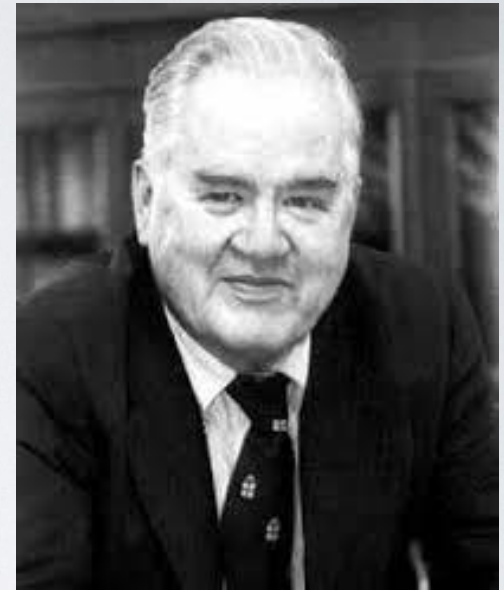
Dev & test and continuous integration



A revolution? You're kidding!

> Why we could (wrongly) disregard the Big Data impact ?

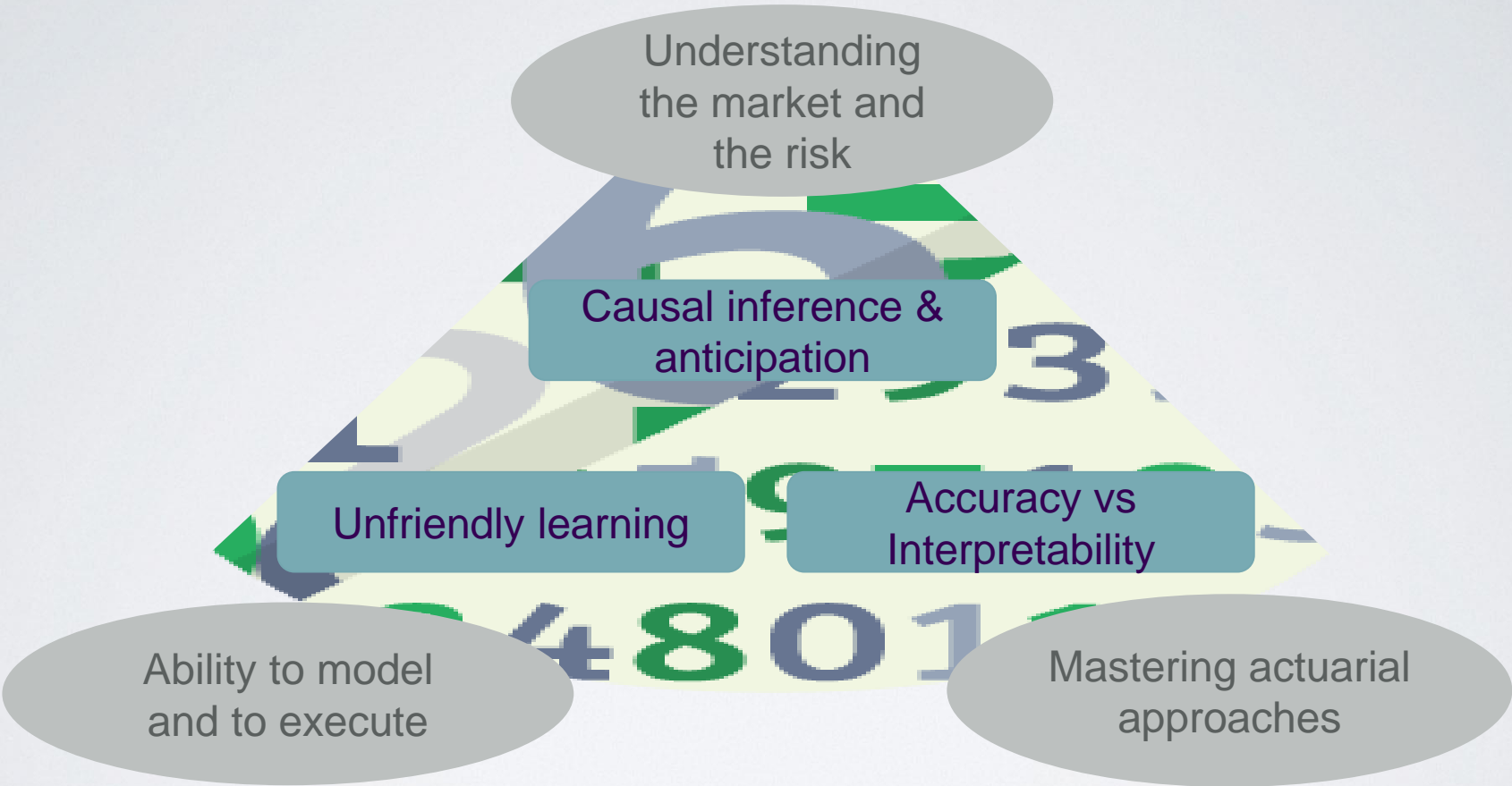
- ➔ « This isn't all that new » (TW)
 - ➔ **Insurance is the only industry (with banks) to have dealt with data in recent years**
- ➔ « Insurers have quasi-data scientist » (TW)
 - ➔ « **DS companies hires actuaries** »
 - ➔ **The Economist 2015 : « Google and Amazon hires micro-economist »**
- ➔ « A huge proportion of big data is irrelevant » (TW)
 - ➔ **relevance of normal data (claims,...)**
 - ➔ **Data Enrichment is nevertheless one of the Strategic axis of technical excellence**



"The future of data analysis"

Academic paper - John W. Tukey 1961

SO WHY THE DATA SCIENTIST HAS NOT REPLACED THE ACTUARY YET?



MAIN TECHNICAL EVOLUTIONS ACTUARIES NEED TO COPE WITH...

NEW CAPABILITIES TO HANDLE DATA

- Automatic data Extraction framework
- Acquisition of unstructured data
- Advanced data preparation (including complex encoding such as SDR*)
- Advanced Feature engineering

ADVANCED MODELING APPROACH

- from cross-section data to longitudinal information (panel data)
- Dependences could be modeled differently (GLM enriched by ML)
- Tracking of insured risks
- Dynamic ratemaking could be reviewed with direct links between the observed statistics and the proposed rates

DEVELOPMENT OF SPECIFIC MODEL IMPLEMENTATION, MONITORING AND MAINTENANCE)

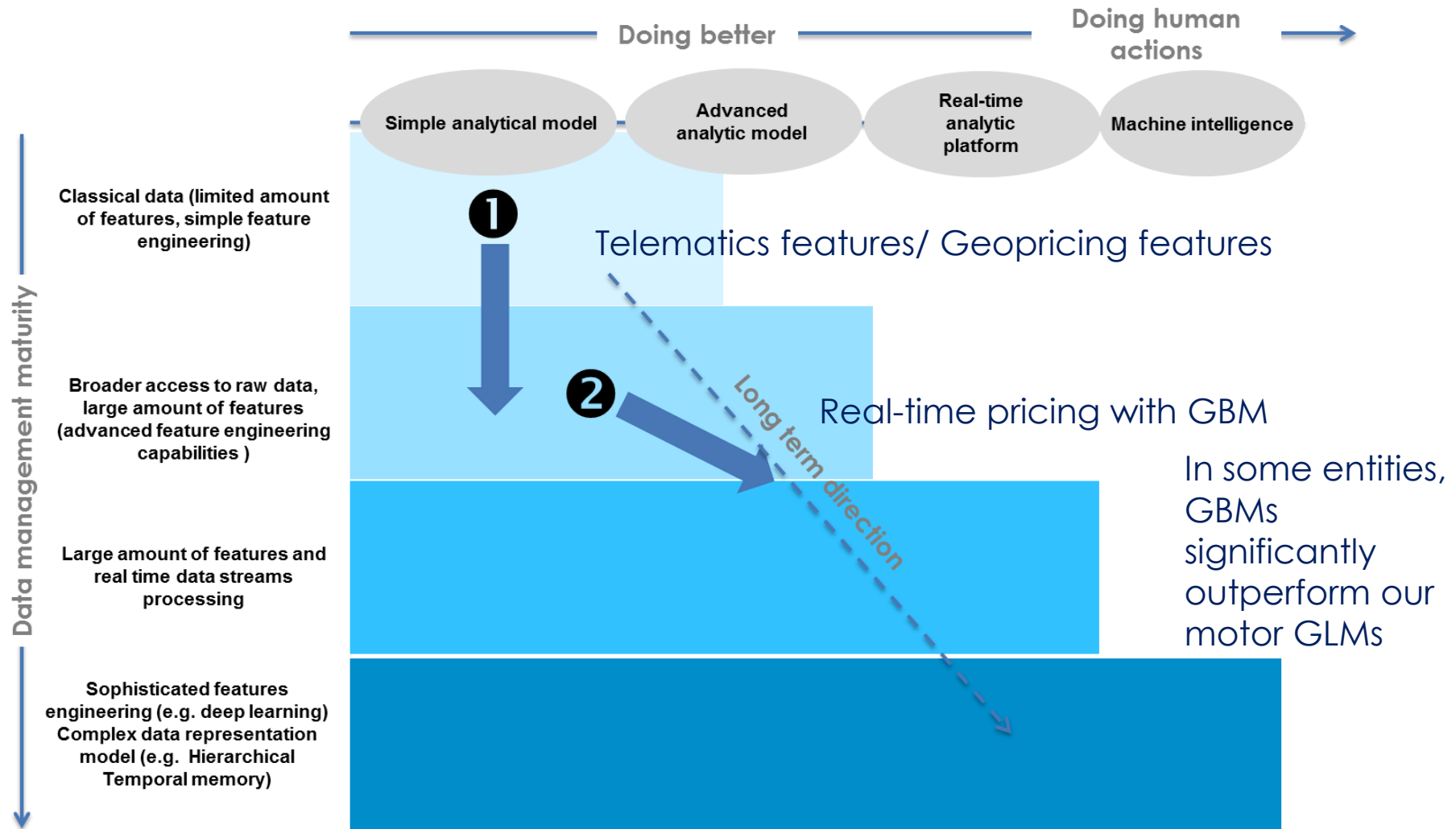
- Automatic checks of model accuracy (incl. Gini curves)
- Technical model deployment
- Real time quotation & optimization
- Training process
- Performance monitoring (A/B testing, True Lift approach...)
- Active learning (Contextual-Bandit approach ...)

DEVELOPMENT OF ALGORITHMIC CULTURE AND COMPUTER SCIENCE

- Predictive power and generalization vs asymptotic property
- Iterative and learning process
- Scalability and performance optimization (incl. production design)
- New type of data (more diverse...)
- Real time and better responsiveness
- Cross-validation culture

...and what will change with data science

> The biggest challenge however is assembling all this information into a coherent mode (P. Domingos*)



NEW CHALLENGES REQUIRE NEW APPROACHES FOR ACTUARIES

 Scope: new playground

 Tools & capabilities

 Agile & cross-disciplinary approach

Some Big Data business challenges for actuaries



Connected Devices



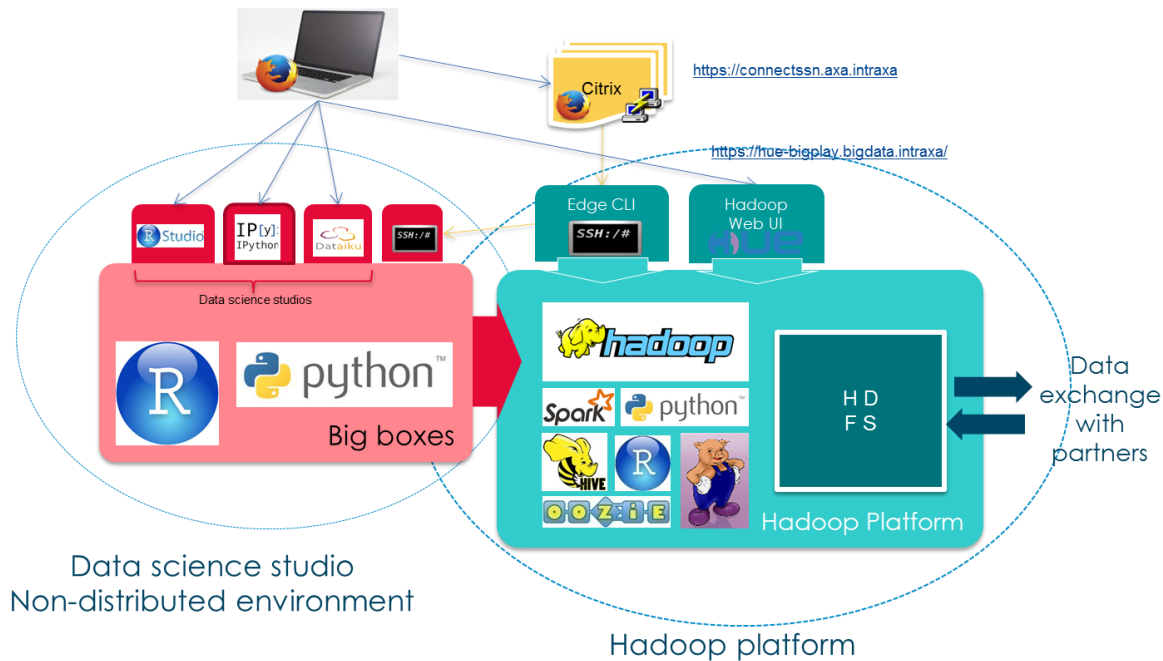
Predictive Behavior



**Risk Management
Advanced Analytics**

New environment and new capabilities needed

Coding!





Big Data - New questions call for new techniques *



- 1st generation
- 2nd generation
- 3rd generation
- 4th génération
- 5th génération

Ruin theory and collective risk model

Credibility and segmentation

GLM & non-linear approach

ERM/finance (DFA, Options, Solvability, Cat modeling, EVT...)

Applied Insurance micro-economy (CMA, price-elasticity modeling , nano segmentation)

ML & CODING



* Paul Embrechts – Astin Colloquium Cannes 1994



Data science process require different profiles

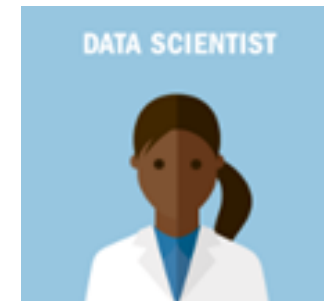


Expert of Big Data and distributed environment

Strong IT profile and mastering of several programming languages



Business background with change management skills and analytical insights



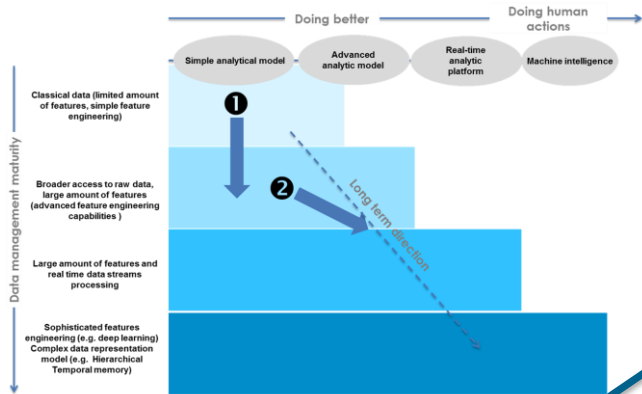
Data-driven problem solver who tries to make discoveries from data

Strong programming and modeling expertise

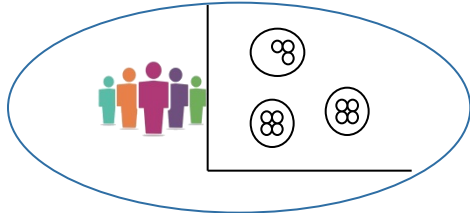
+ Data manager and junior data scientists

How to really become data driven?

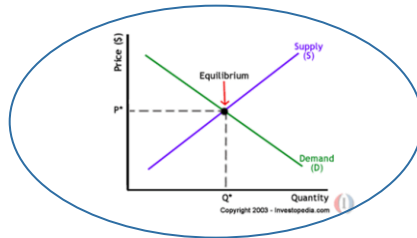
Key challenges to really change the business means to go beyond analytics



New challenges for actuaries



How much will data affect risk pooling?



How will big data modify market dynamic?



Data quality
Privacy & inference
Exclusion & non explicit
Discrimination



Will Information asymmetry disappear?



Will Big Data create new insurance opportunities?

**The future belongs to the companies
and people that turn data into products**

Mike Loukides

ACTUARY, A FUTURE BUSINESS TRANSFORMER ?



THANK YOU!

Philippe.mariejeanne@axa.com