Data Science: Finding the ROI

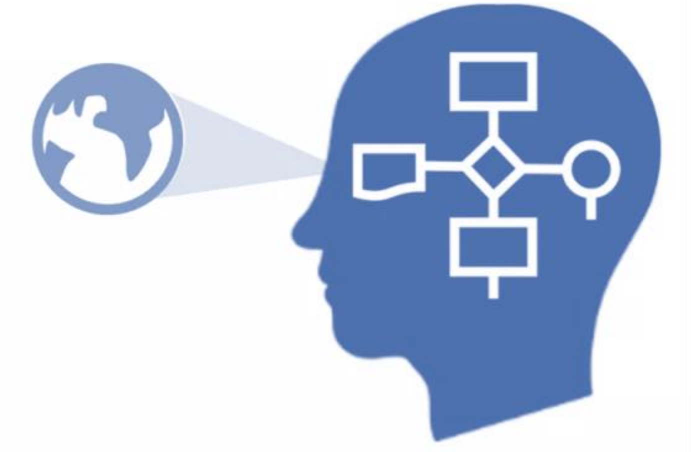
Armin Kakas & Mike Milanowski

Data Science | Finding the ROI

1	Intro Game
2	Basics of Data
3	Ecosystems
4	Data Science
5	Machine Learning Fundamentals

USING MENTAL MODELS FOR A JOURNEY

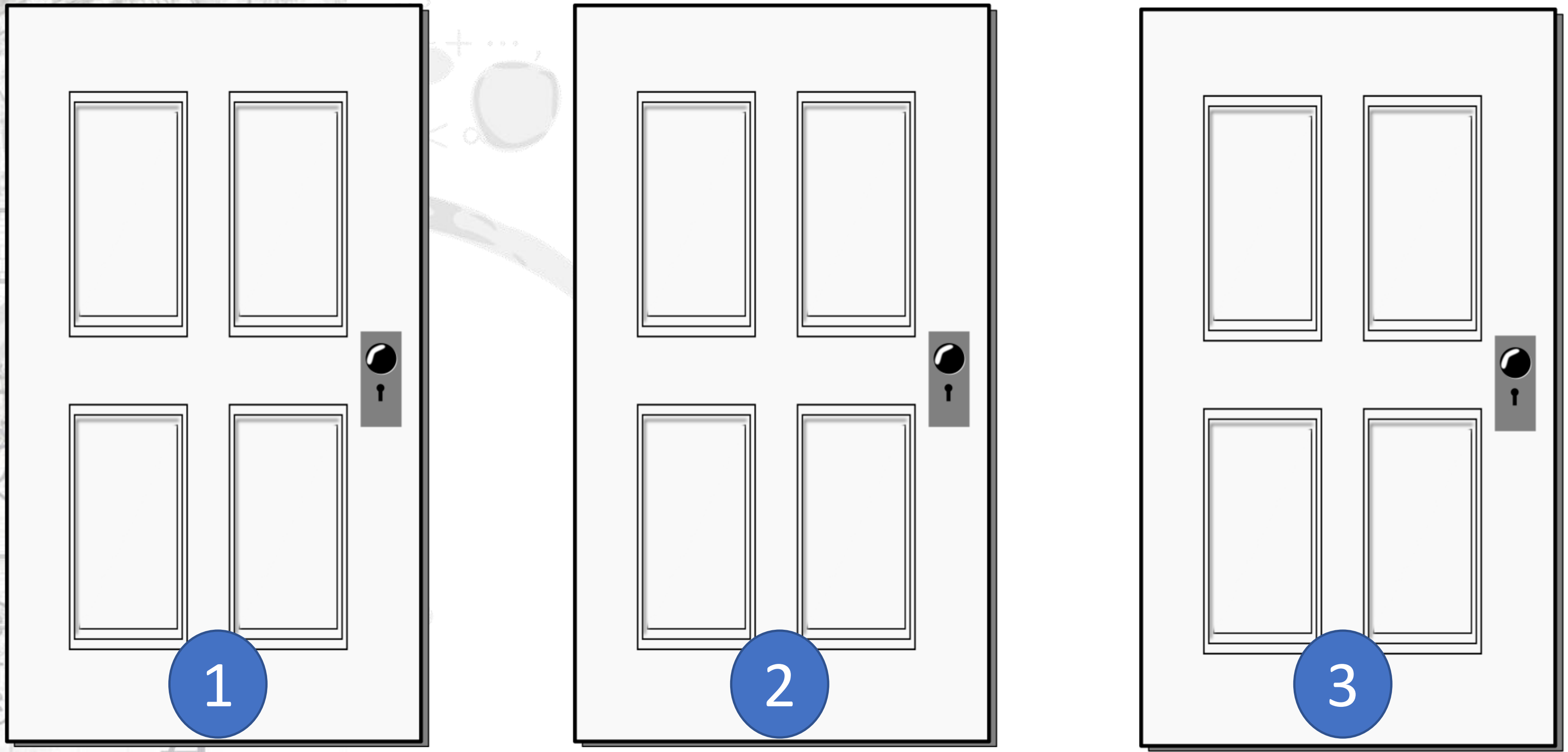
DATA BASICS → MACHINE LEARNING

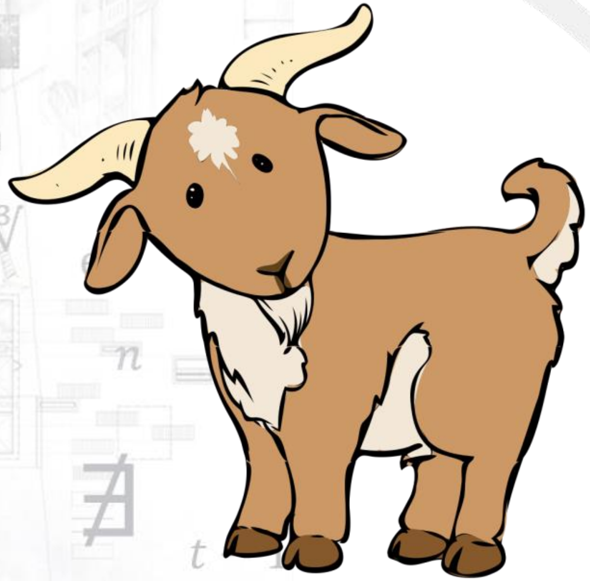


"When we try to pick out anything by itself, we find it hitched to everything else in the Universe." -Muir

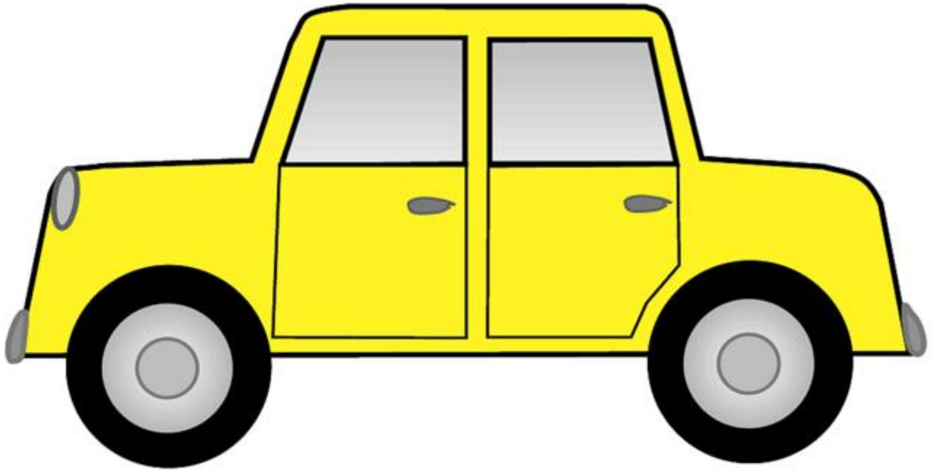
Let's play a game

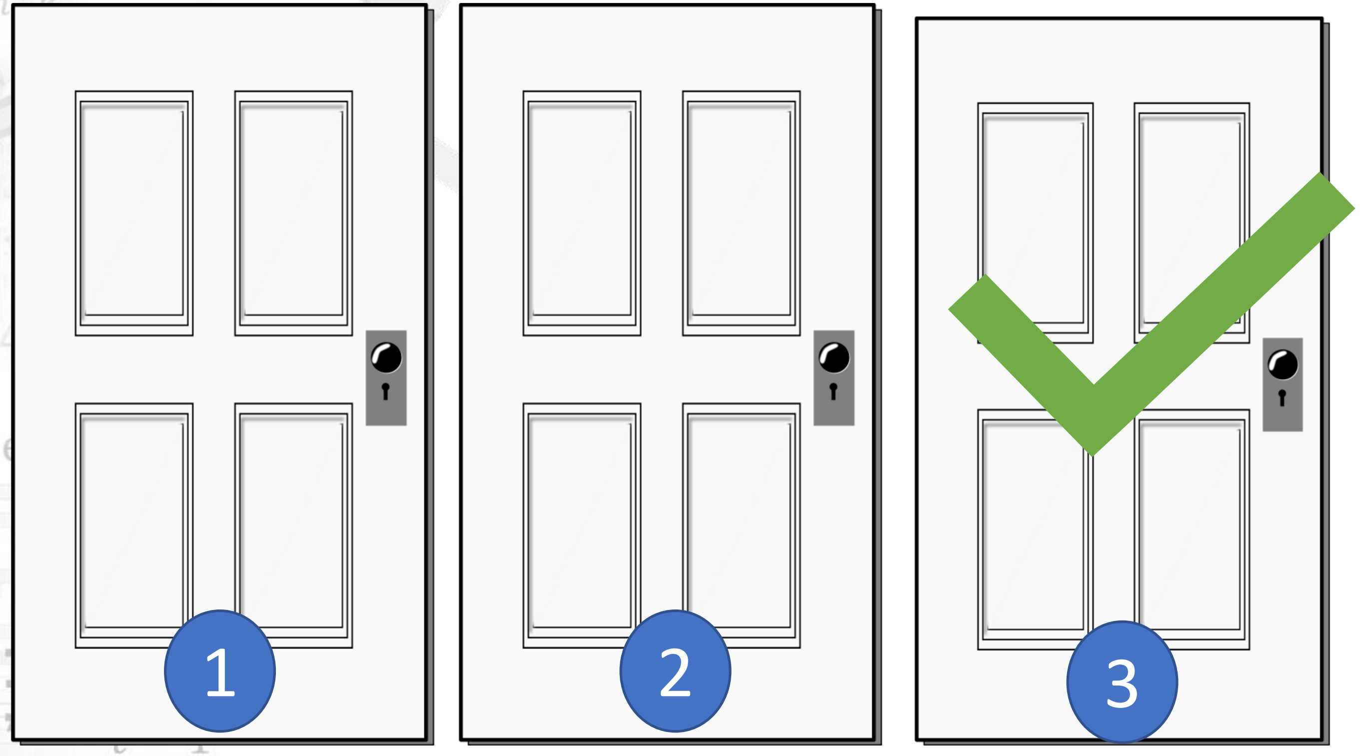
The Monty Hall problem


1
2
3

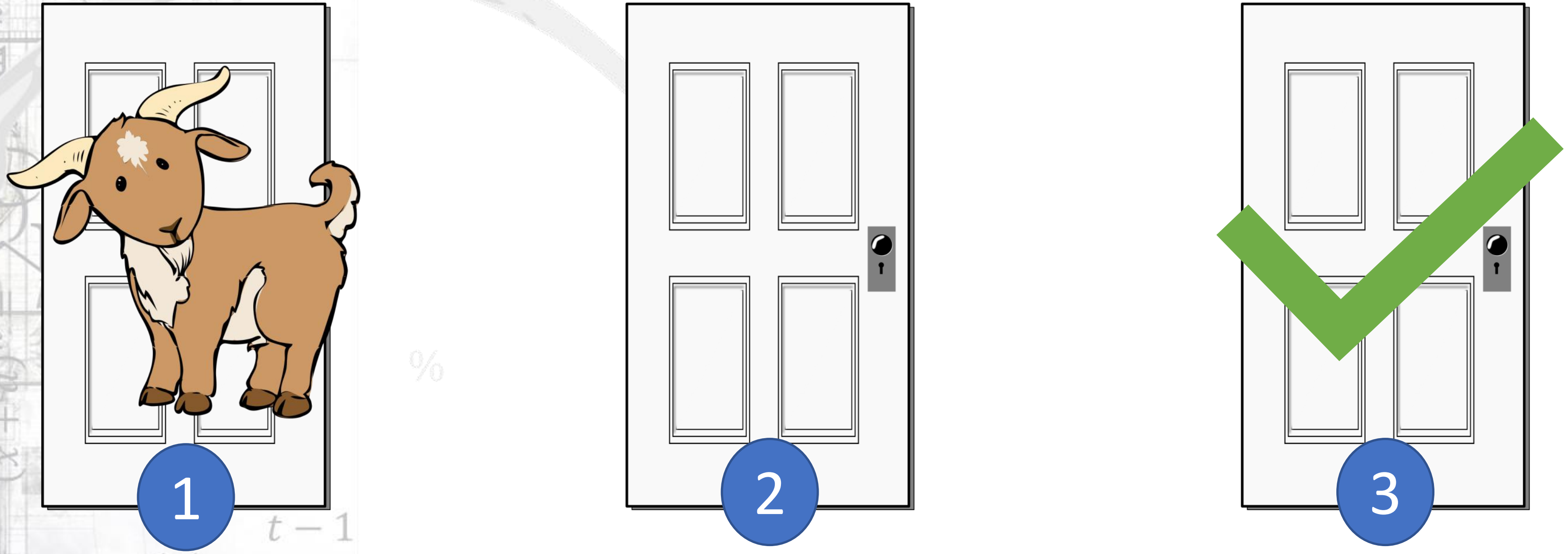


OR

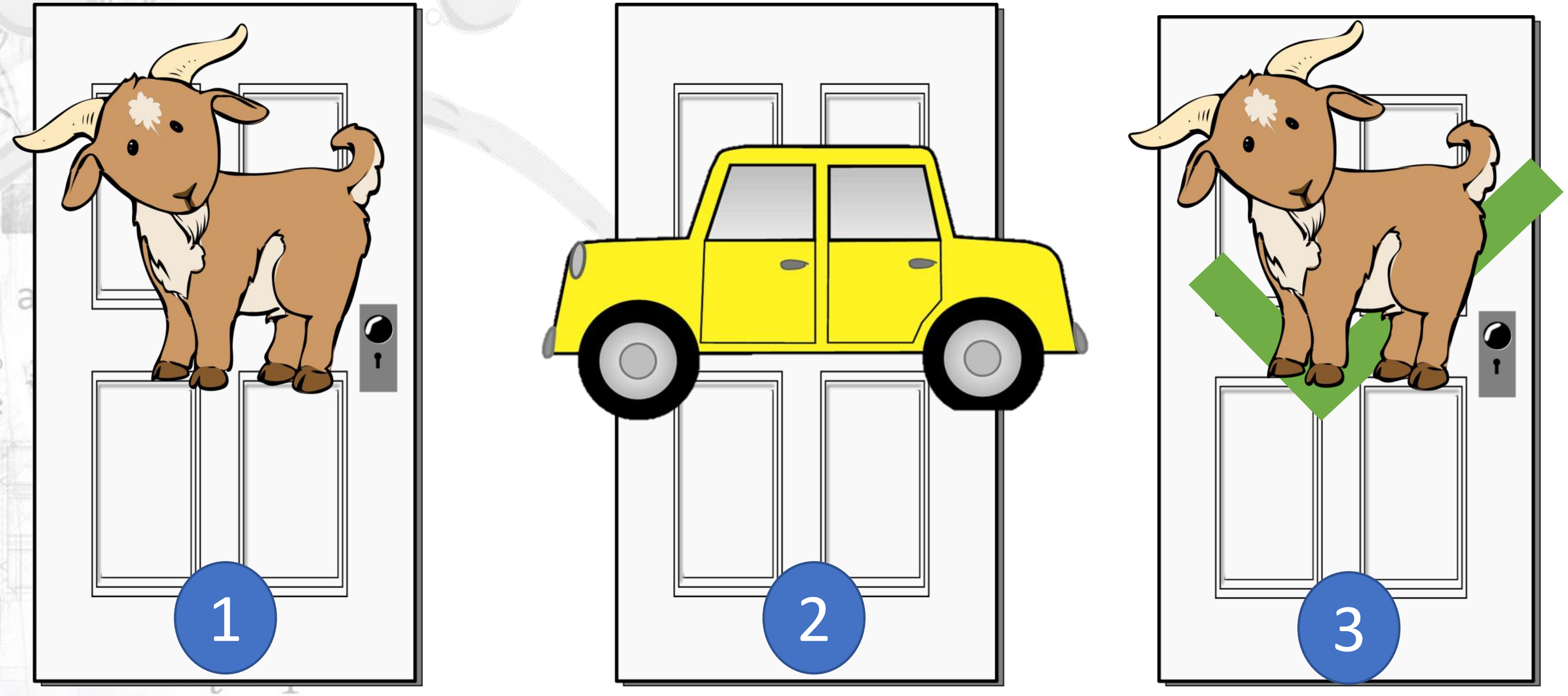




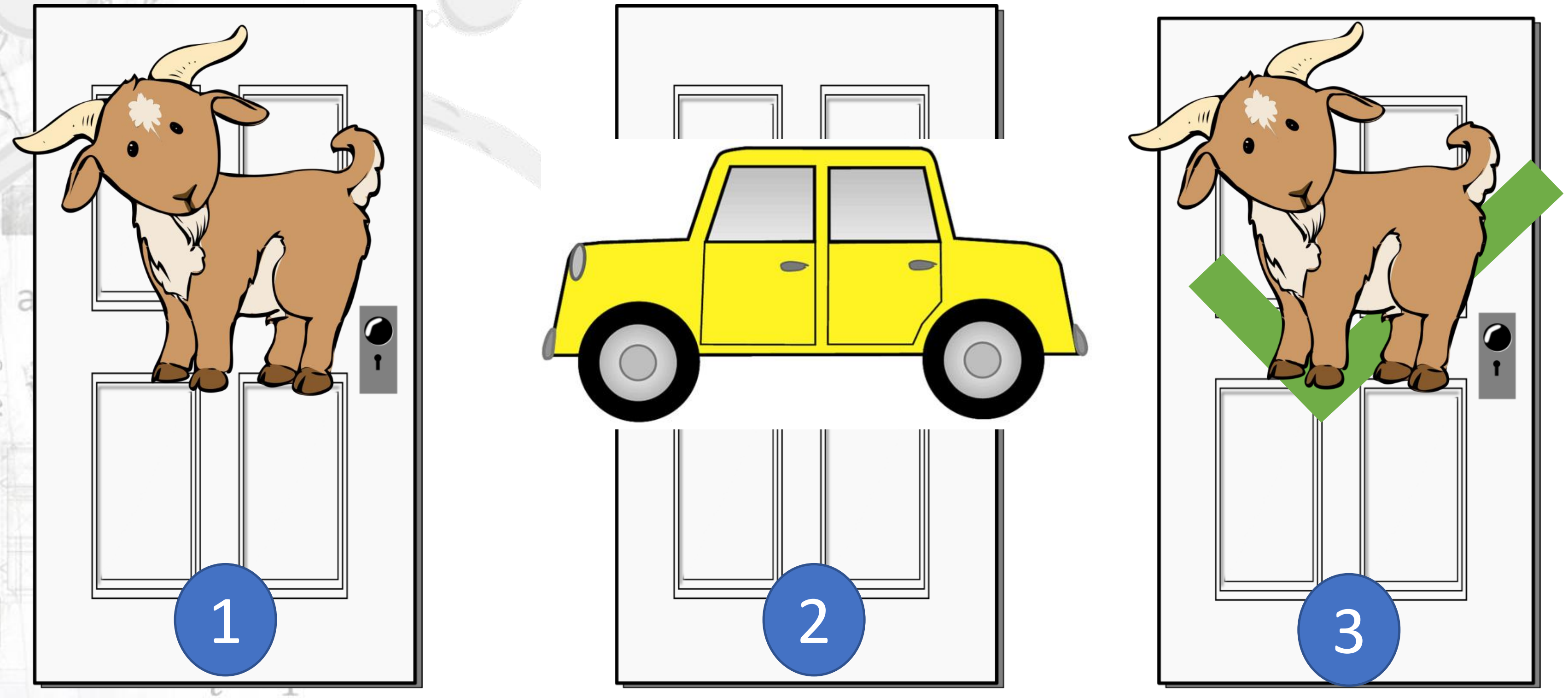
You pick door #3



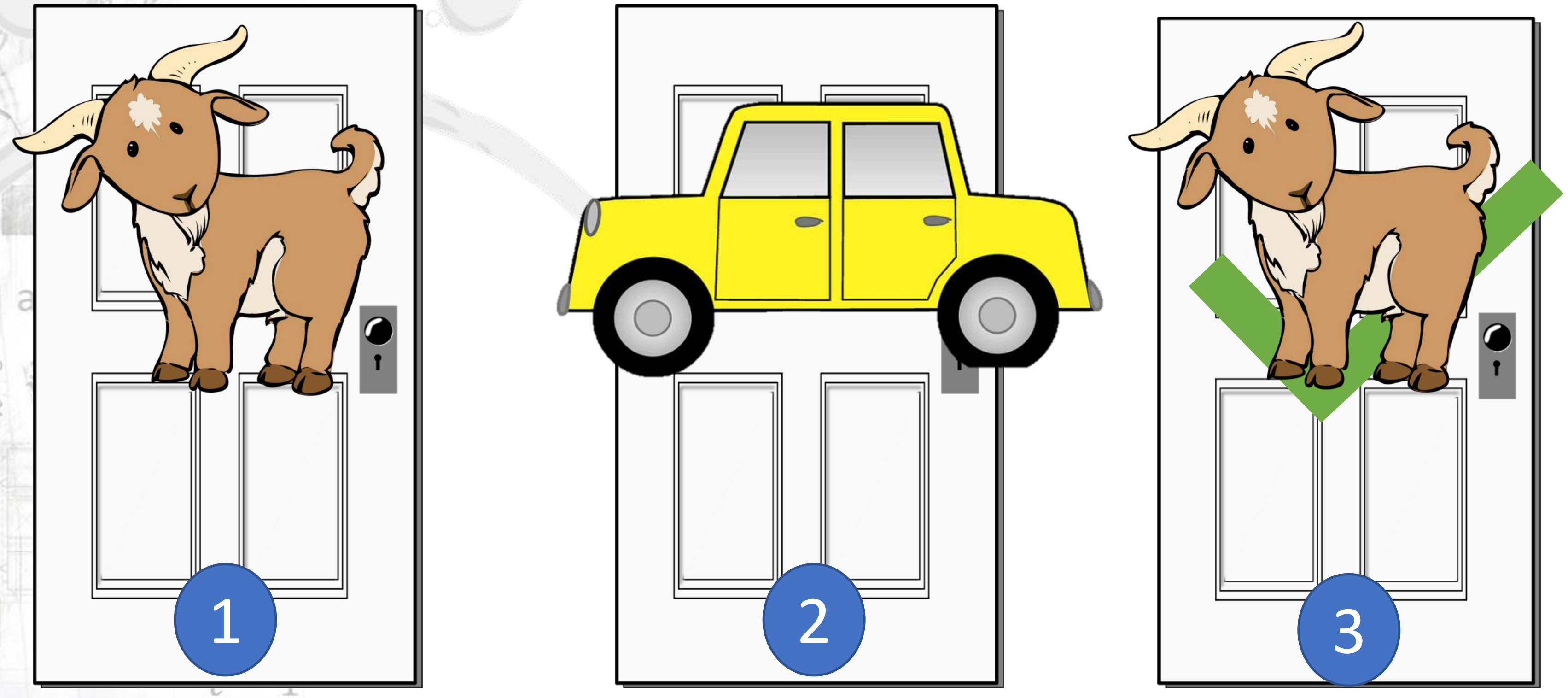
This is a game of choice



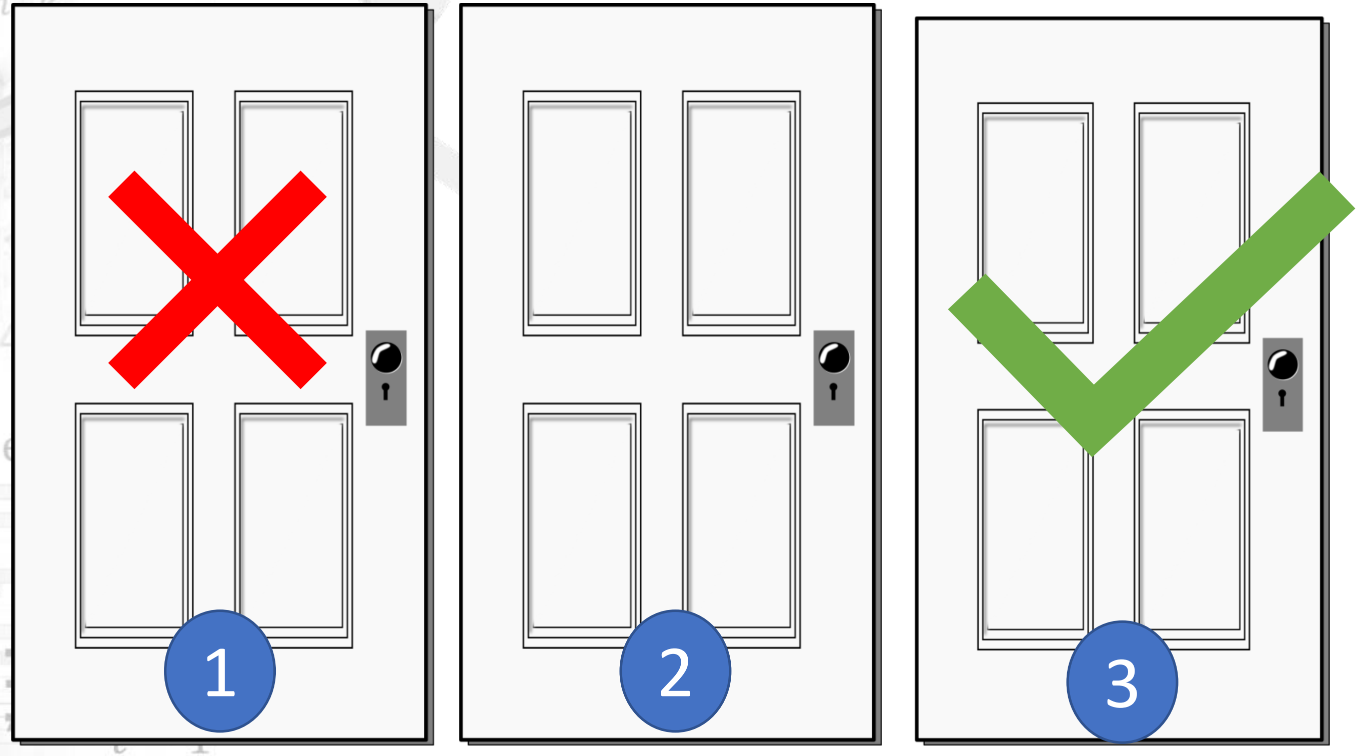
Switching to door #2 is the right choice…



When given the choice, you should always switch doors!

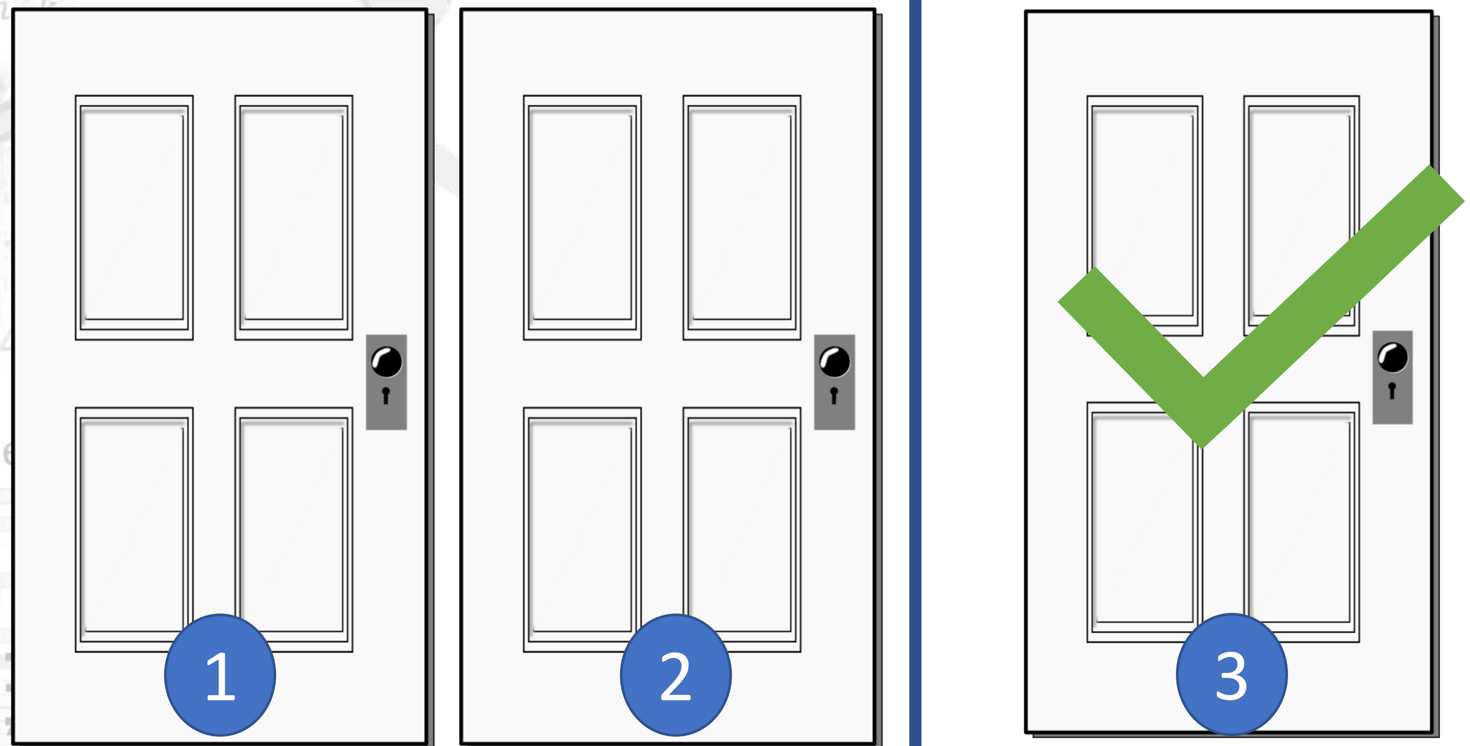


But why?



You pick door #3

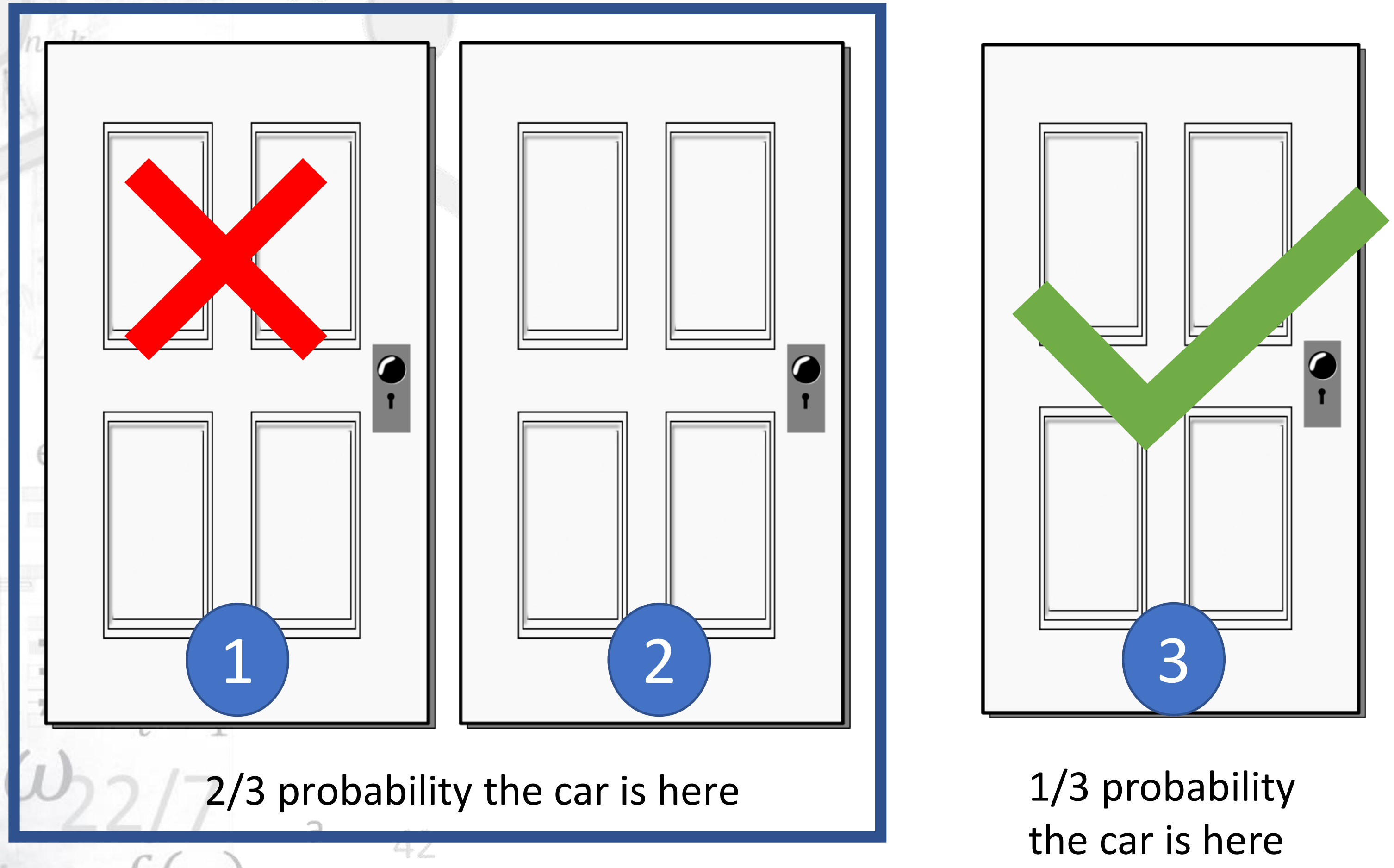
Door #1 is eliminated, which leaves us with 2 doors that can potentially have the car. 50/50 probability, right?



2/3 probability the car is here

1/3 probability
the car is here

After choosing door #3, there is a 2/3 probability of being wrong



Door #1 is eliminated (goat), which leaves a 2/3 probability on door #2… 2/3 > 1/3, thus always switch!

Intuition

?	?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?	?
?	?	?	?	✓	?	?	?	?	?
?	?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?	?
?	?	?	?	?	?	?	?	?	?

If 98 doors are revealed to have the goat, do you think there is still a 50/50 chance that our original choice is correct?

X	X	X	X	X	X	X	X	X	X
X	X	X	X	X	X	X	X	X	X
X	X	X	X	✓	X	X	X	X	X
X	X	X	X	X	X	X	X	X	X
X	X	X	X	X	X	X	X	X	X
X	X	X	X	X	X	X	X	X	X
X	X	X	X	X	X	X	X	X	X
X	X	X	?	X	X	X	X	X	X
X	X	X	X	X	X	X	X	X	X
X	X	X	X	X	X	X	X	X	X

Monty Hall implications for CPG

- Knee-jerk reactions are dominant in business decisions
- We are not good at forecasting, especially new products
- Tendency to stay the course and follow company lore
- We continue to overspend on dilutive customers
- Teams fit analysis to our sales/marketing narrative

Data science is not about confirming our prior beliefs –
it is about uncovering hidden truths or enhancing our knowledge.

Data Basics | Finding the ROI

Data Basics| Finding the ROI

1 **Key Concepts**

2 **Data Strategy**

3 **Data Ecosystems**

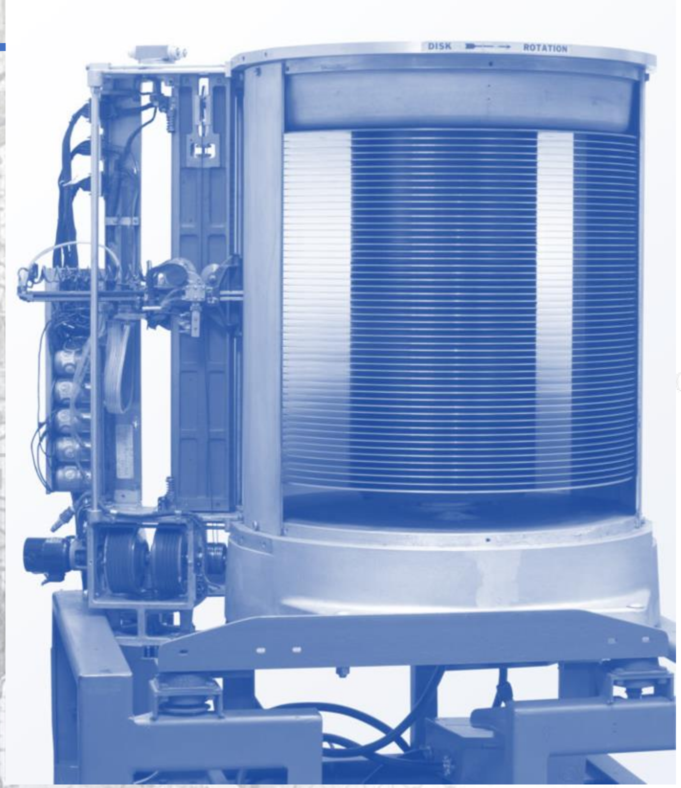
4 **Using the Data**

“A [person] will be imprisoned in a room with a door that's unlocked and opens inwards; as long as it does not occur to him to pull rather than push.” Wittgenstein

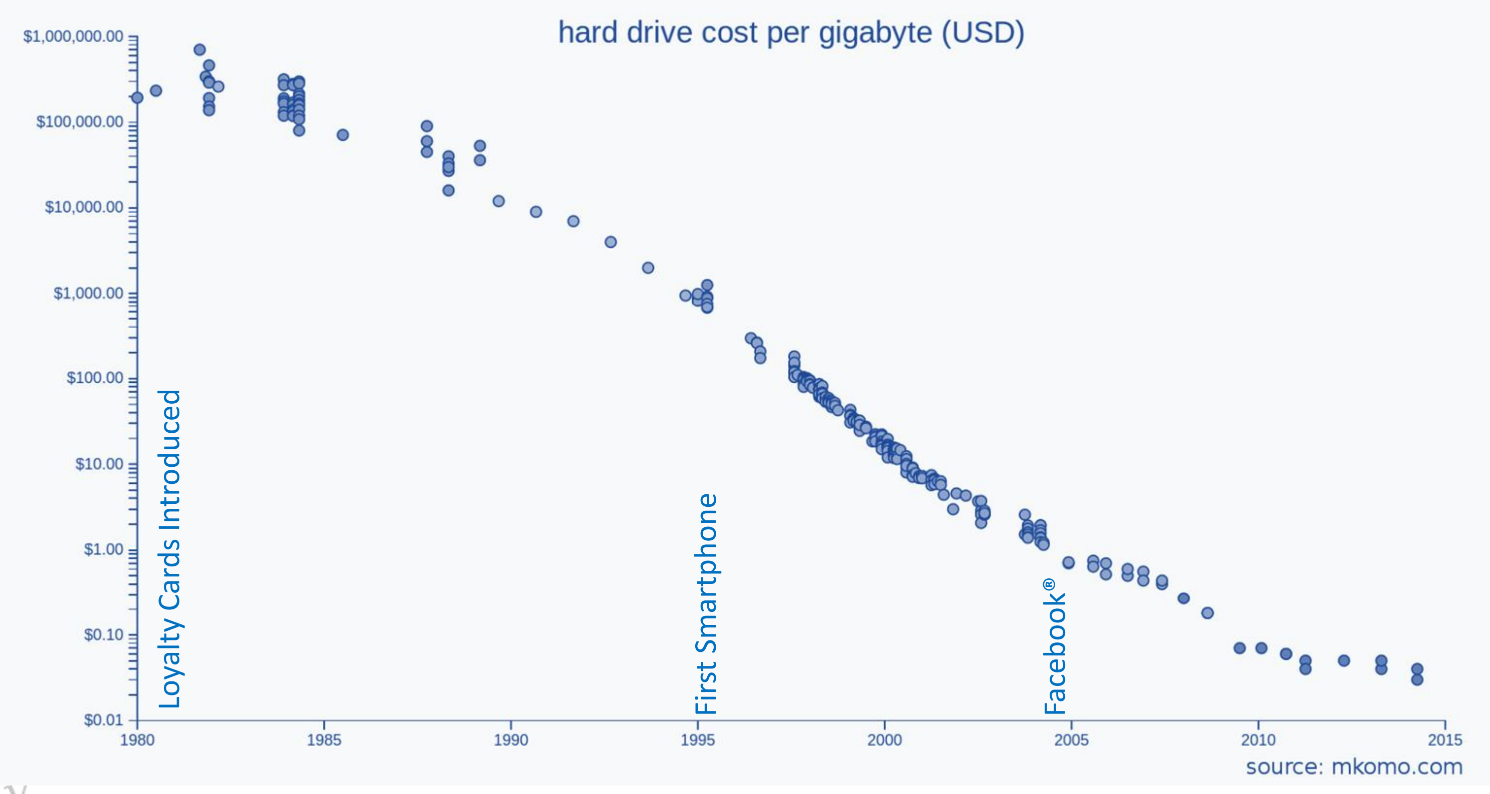
© Larson

Base Setting | Cost Changes

60 Inches | 5MB



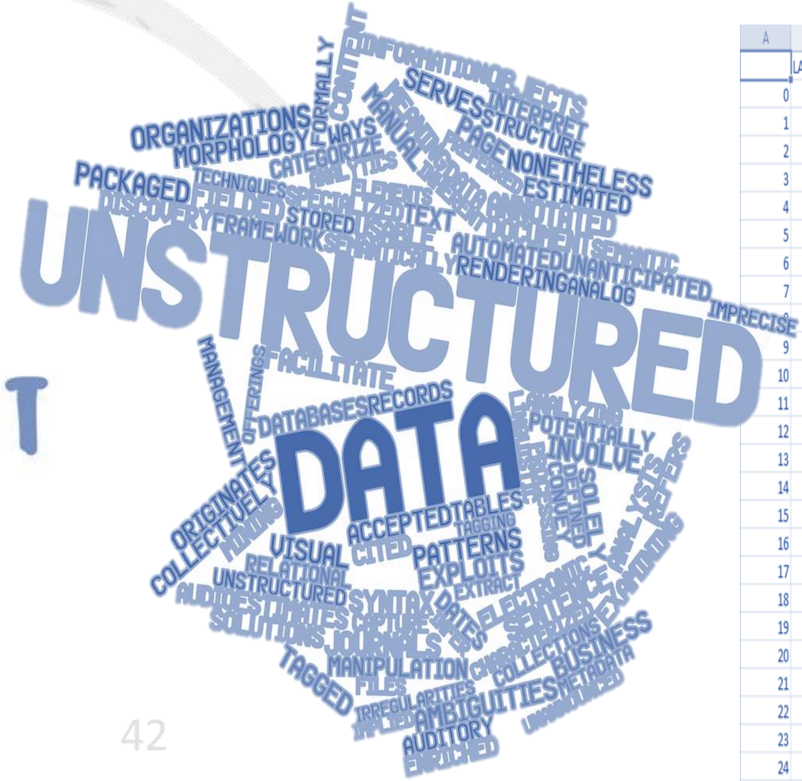
IBM 350 Disk Storage system ~1970



Base Setting | Data Basics

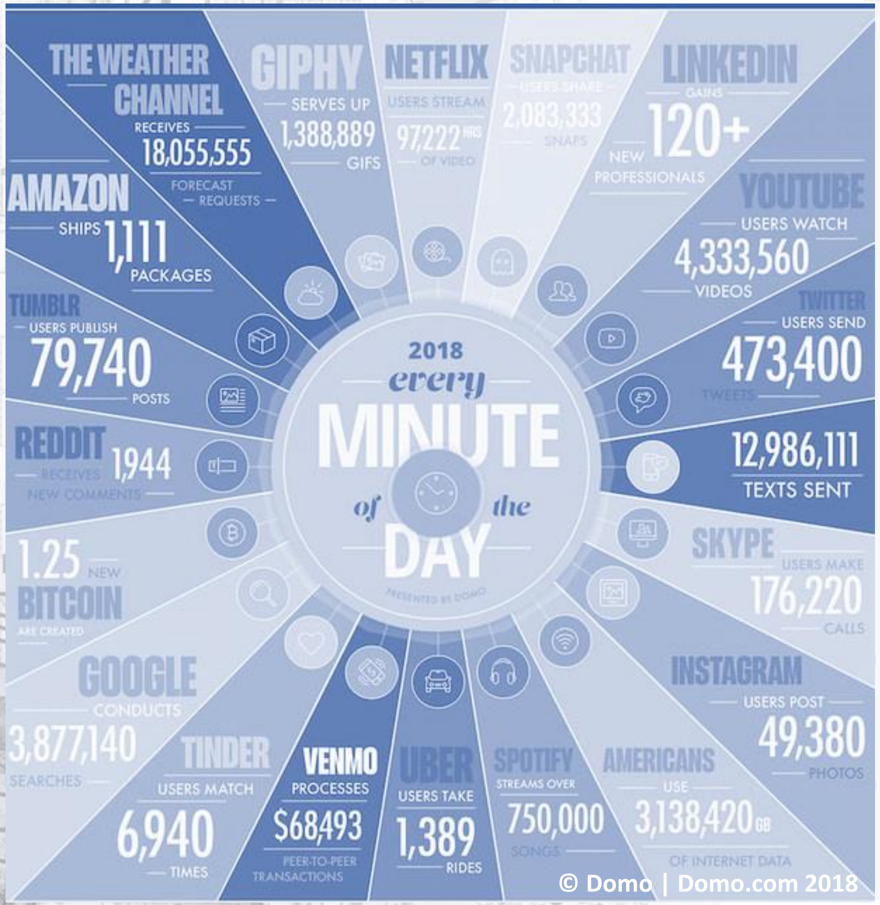
"IN GOD WE TRUST,
ALL OTHERS MUST BRING DATA"

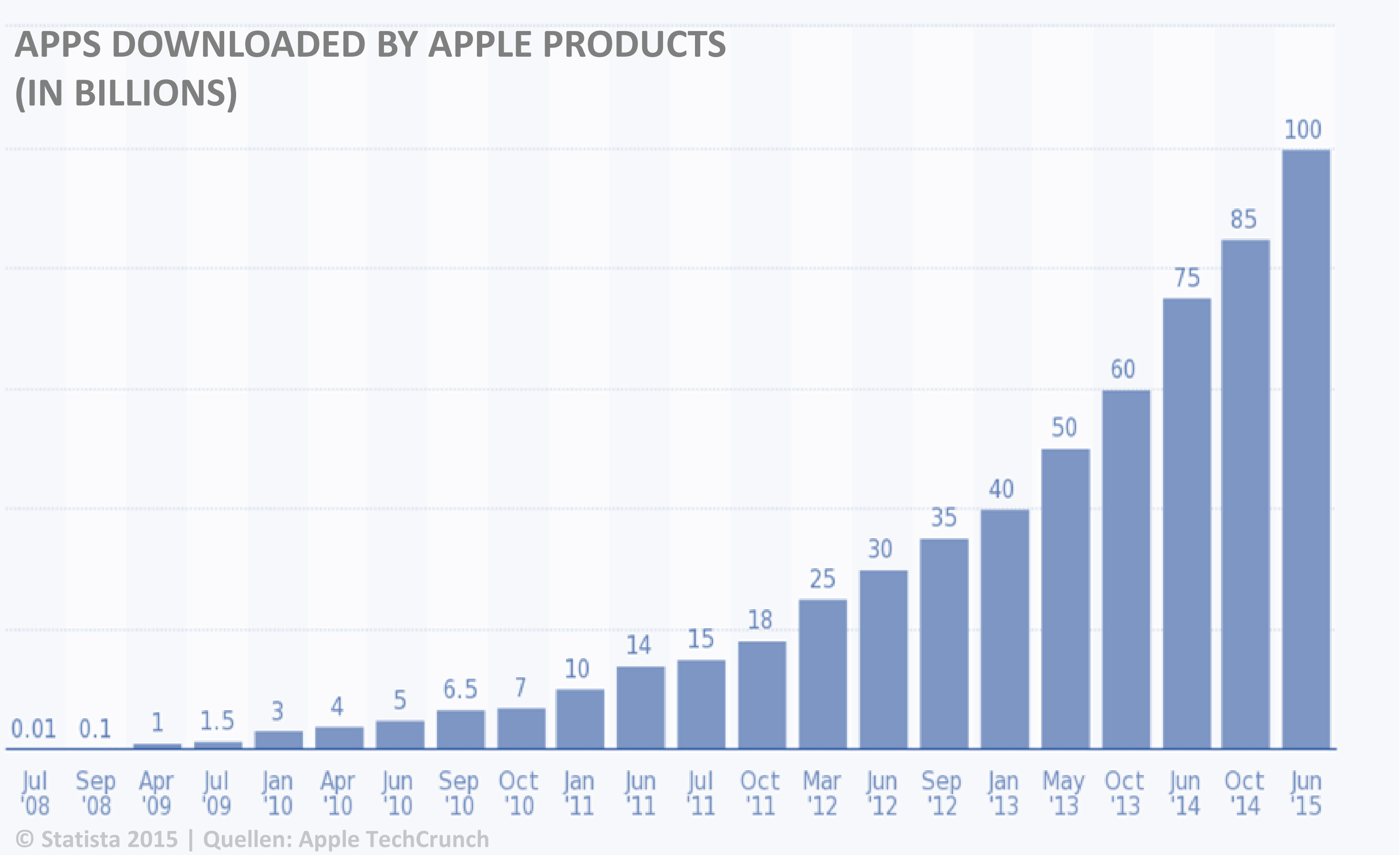
W. EDWARDS DEMING



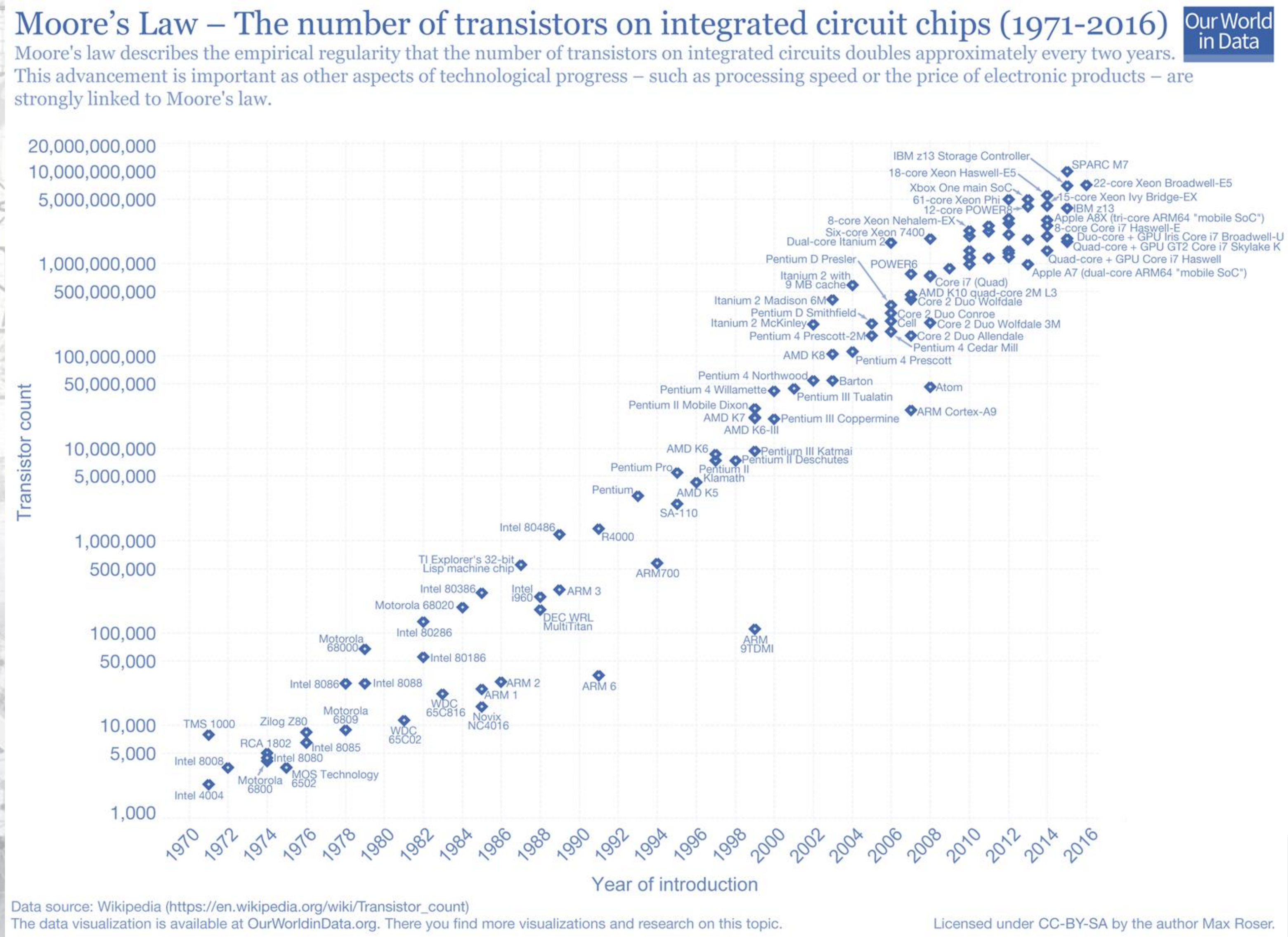
A	B	C	D	E	F	G	H	I	J	K	L	M	N
	LAT	LON	RECORD_DATE	MAX_TEM	MIN_TEMI	CREATED_DATETIME	INTEGRAT	DASK_ID	LAT	LON	LON_ADJU	LON_NOA	LON_NOA
0	13.25	272.75	1/1/2014	93.549	77.37	11/14/2018 1:49	13.25~272	6728	13.25	-87.25	92.75	92.75	272.75
1	13.25	272.75	1/1/2014	93.549	77.37	11/14/2018 1:49	13.25~272	6728	13.25	-87.25	92.75	92.75	272.75
2	13.25	272.75	1/2/2014	93.9465	76.8217	11/14/2018 1:49	13.25~272	9675	13.25	-87.25	92.75	92.75	272.75
3	13.25	272.75	1/2/2014	93.9465	76.8217	11/14/2018 1:49	13.25~272	9675	13.25	-87.25	92.75	92.75	272.75
4	13.25	272.75	1/3/2014	92.8899	77.885	11/14/2018 1:49	13.25~272	1061	13.25	-87.25	92.75	92.75	272.75
5	13.25	272.75	1/3/2014	92.8899	77.885	11/14/2018 1:49	13.25~272	1061	13.25	-87.25	92.75	92.75	272.75
6	13.25	272.75	1/4/2014	90.2433	77.3921	11/14/2018 1:49	13.25~272	6409	13.25	-87.25	92.75	92.75	272.75
7	13.25	272.75	1/4/2014	90.2433	77.3921	11/14/2018 1:49	13.25~272	6409	13.25	-87.25	92.75	92.75	272.75
8	13.25	272.75	1/5/2014	94.2897	77.8821	11/14/2018 1:50	13.25~272	2995	13.25	-87.25	92.75	92.75	272.75
9	13.25	272.75	1/5/2014	94.2897	77.8821	11/14/2018 1:50	13.25~272	2995	13.25	-87.25	92.75	92.75	272.75
10	13.25	272.75	1/6/2014	93.151	75.6971	11/14/2018 1:50	13.25~272	5647	13.25	-87.25	92.75	92.75	272.75
11	13.25	272.75	1/6/2014	93.151	75.6971	11/14/2018 1:50	13.25~272	5647	13.25	-87.25	92.75	92.75	272.75
12	13.25	272.75	1/7/2014	90.0572	75.9713	11/14/2018 1:50	13.25~272	43	13.25	-87.25	92.75	92.75	272.75
13	13.25	272.75	1/7/2014	90.0572	75.9713	11/14/2018 1:50	13.25~272	43	13.25	-87.25	92.75	92.75	272.75
14	13.25	272.75	1/8/2014	89.0187	74.2483	11/14/2018 1:50	13.25~272	2	13.25	-87.25	92.75	92.75	272.75
15	13.25	272.75	1/8/2014	89.0187	74.2483	11/14/2018 1:50	13.25~272	2	13.25	-87.25	92.75	92.75	272.75
16	13.25	272.75	1/9/2014	91.1774	75.1632	11/14/2018 1:50	13.25~272	5217	13.25	-87.25	92.75	92.75	272.75
17	13.25	272.75	1/9/2014	91.1774	75.1632	11/14/2018 1:50	13.25~272	5217	13.25	-87.25	92.75	92.75	272.75
18	13.25	272.75	1/10/2014	94.0385	75.806	11/14/2018 1:51	13.25~272	1812	13.25	-87.25	92.75	92.75	272.75
19	13.25	272.75	1/10/2014	94.0385	75.806	11/14/2018 1:51	13.25~272	1812	13.25	-87.25	92.75	92.75	272.75
20	13.25	272.75	1/11/2014	92.8267	74.7385	11/14/2018 1:51	13.25~272	5969	13.25	-87.25	92.75	92.75	272.75
21	13.25	272.75	1/11/2014	92.8267	74.7385	11/14/2018 1:51	13.25~272	5969	13.25	-87.25	92.75	92.75	272.75
22	13.25	272.75	1/12/2014	94.1406	77.9738	11/14/2018 1:51	13.25~272	4382	13.25	-87.25	92.75	92.75	272.75
23	13.25	272.75	1/12/2014	94.1406	77.9738	11/14/2018 1:51	13.25~272	4382	13.25	-87.25	92.75	92.75	272.75
24	13.25	272.75	1/13/2014	94.0712	71.4597	11/14/2018 1:51	13.25~272	8619	13.25	-87.25	92.75	92.75	272.75
25	13.25	272.75	1/13/2014	94.0712	71.4597	11/14/2018 1:51	13.25~272	8619	13.25	-87.25	92.75	92.75	272.75

Base Setting | Data Generation





Base Setting | Compute Speed



We need to reset our reference point

- It is no longer a question of is there data.
- We no longer need to discuss cost of storage.
- The question isn't computational power.

The real questions are more fundamental…

The Ask is *Everything*

- What question do I want to answer?
- What do I need to answer the question?
- How will I get to the answer?
- How do I know I can trust the answer?

Data Strategy | Getting the Ask right

What is the elasticity
for *X* item?

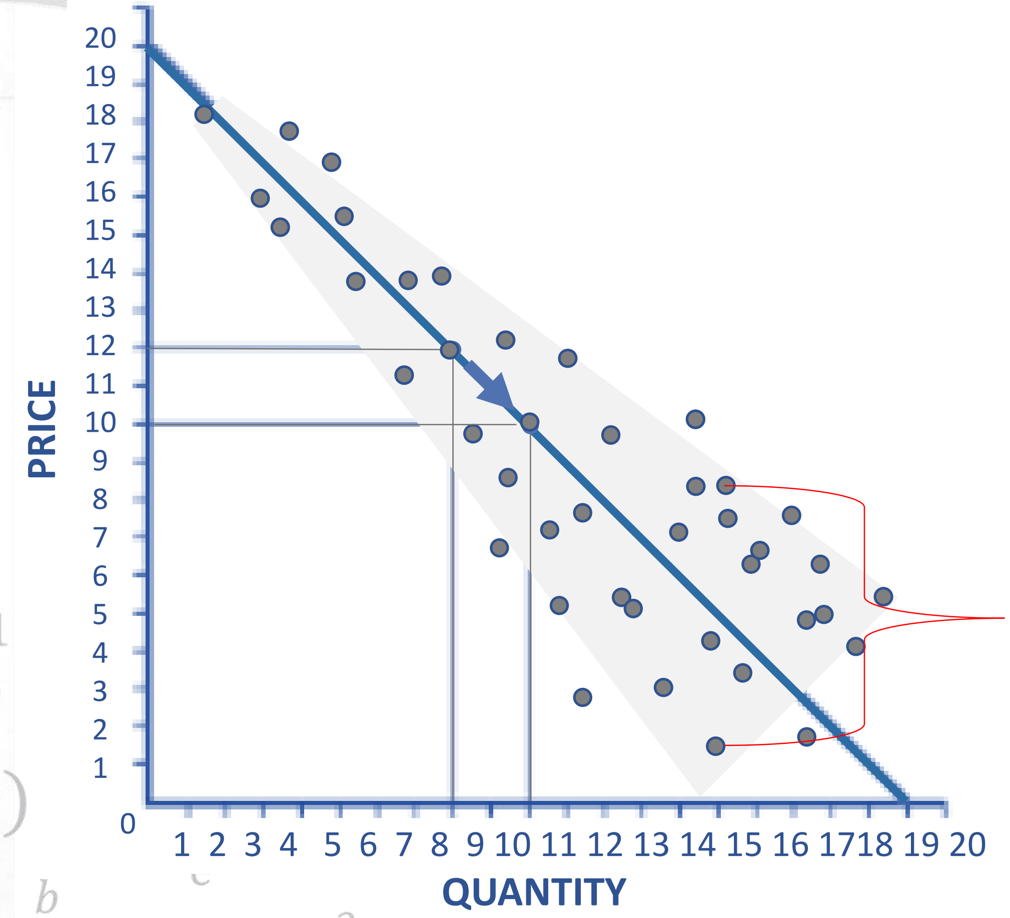
For Discussion Purposes Only | No comments or slides infer legal advice or market actions. All decisions are at the sole discretion of your organization.

"An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem."

-- John Tukey

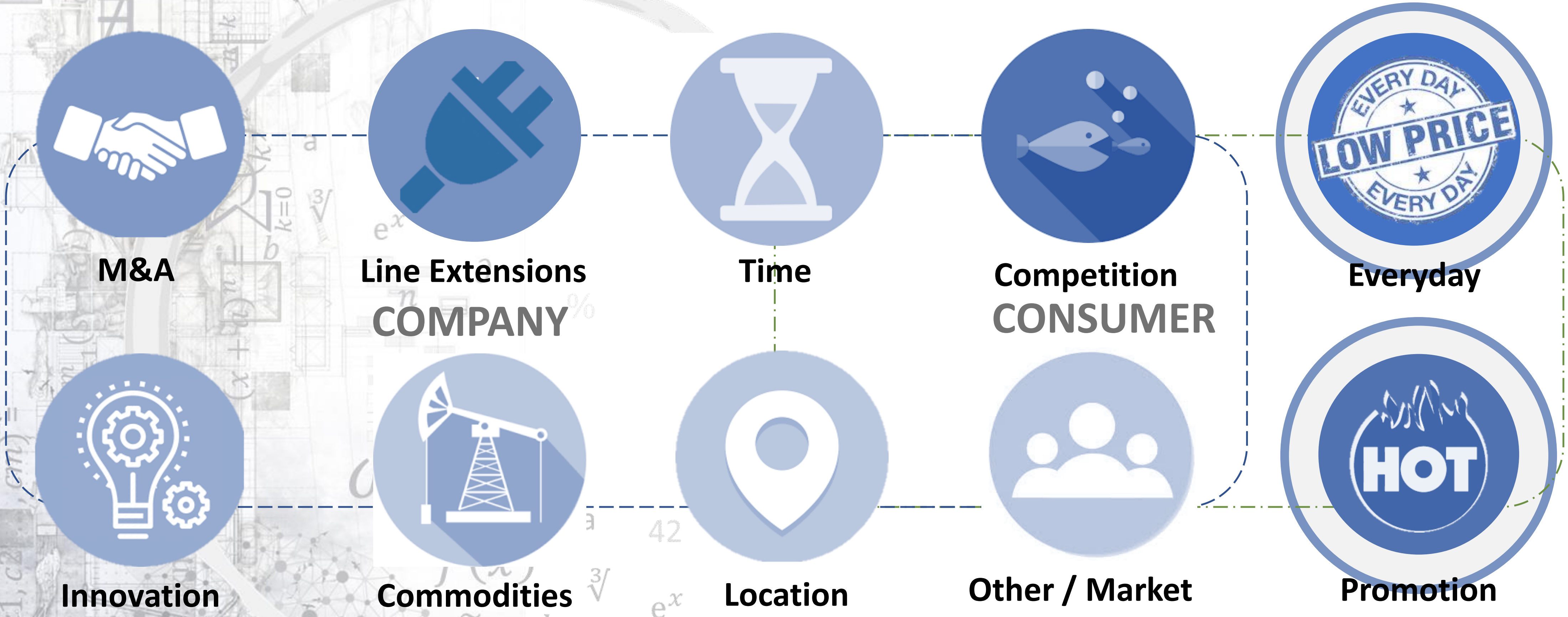
For Discussion Purposes Only | No comments or slides infer legal advice or market actions. All decisions are at the sole discretion of your organization.

Dissonance leads to different requirements



Heteroskedastic:
What else could be causing this?

The world is more complicated



Consumers Have Options

Surveys, Syndicated Sources, Digital Media

Digital Media, Third Party Providers, Marketing Teams

Panel Data, POS Data, Financial Institutions

GIS, Syndicated Sources, POS data

Internal Data, POS, Syndicated

With more information comes more sophisticated approaches

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

Bayes

$$V_n^{\omega}(z, \alpha, c1, c2, \dots, cm) = \inf_{\substack{y_i \geq 0 \\ y \geq q - z}} \Big[\Big(\textstyle\sum_{i=1}^{m} (s_i + ci) y_i + h(y + z - \bar{q})\Big) + \Big(\alpha \int_0^{\infty} \cdots \int_0^{\infty} v_{n+1}(y + z - \bar{q}, \alpha, x_1, x_2, \dots, xm)\, dF_1(x_1)\, dF_2(x_2), \dots, dF_m(x_m) \Big] \text{ where}$$

$$z = \sum_{i=1}^{m} z_i,\ \bar{q} = \sum_{i=1}^{m} \bar{q}_i,\ y = \sum_{i=1}^{m} y_i$$

$(s_i + ci)$ = effective price of brand i, y = units purchased, v_n is the time period, z = inventory

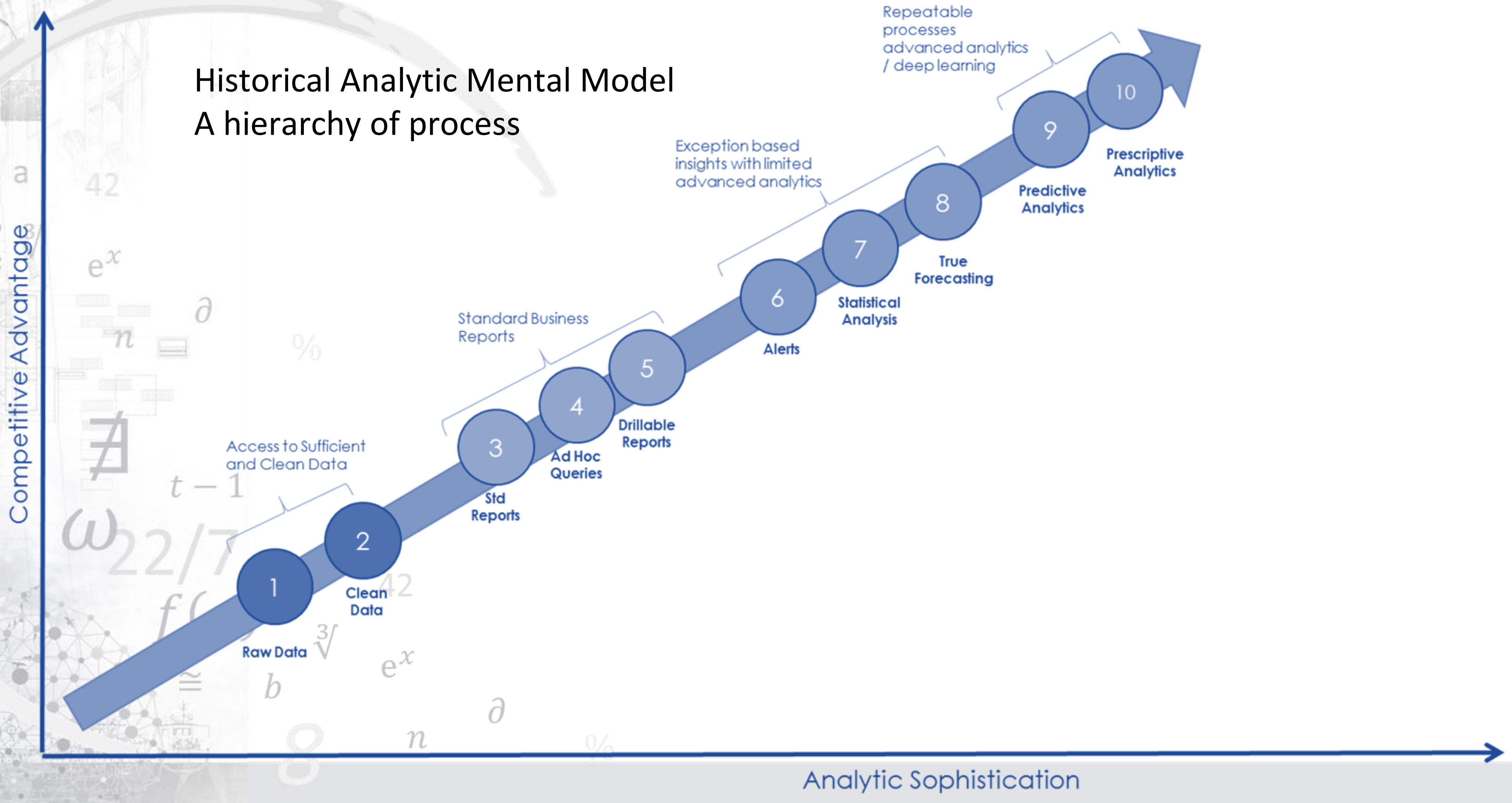
$V_n^{\omega}(z, \alpha, c1, c2, \dots, cm)$ = min expected cost, α = rate of discount per $v_{n+I,}$ $\bar{q}$ = constant consumption rate

Bellman

Bellman's Optimality Equation (1957): Krishna, Aradhna, "The Normative Impact of Consumer Price Expectations for Multiple Brands on Consumer Purchase Behavior," Marketing Science Vol 11, No 3, Summer 1992.

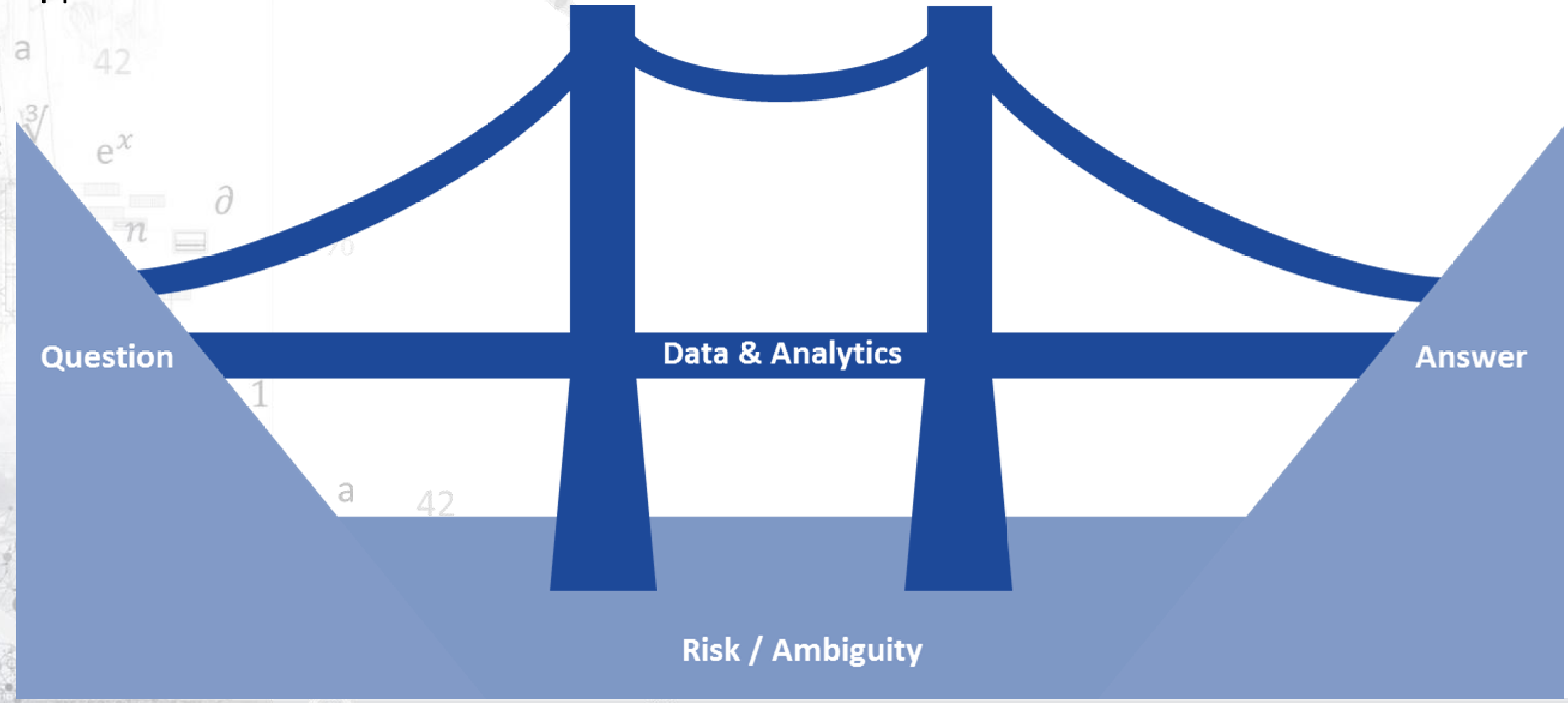
Analytics mindset is shifting

Historical Analytic Mental Model
A hierarchy of process



It’s the journey

New Mental Model
A support driven view.



For Discussion Purposes Only | No comments or slides infer legal advice or market actions. All decisions are at the sole discretion of your organization.

Machine Learning
Fundamentals

ML consists of 3 general areas

Supervised

Apply what has been learned on historical data to new data in order to predict future events.

Unsupervised

Explore the data to discover hidden patterns and structures.

Reinforcement learning

Learn decision-making heuristics from the environment by taking actions, analyzing and optimizing consequences (minimizing errors and maximizing rewards).

Common ML models

Supervised Learning

Linear Regression

Logistic Regression

Nearest Neighbor

Decision Trees

Random Forest

Gradient Boosting Machines (GBM) / xgboost

Neural Networks

Unsupervised Learning

Clustering

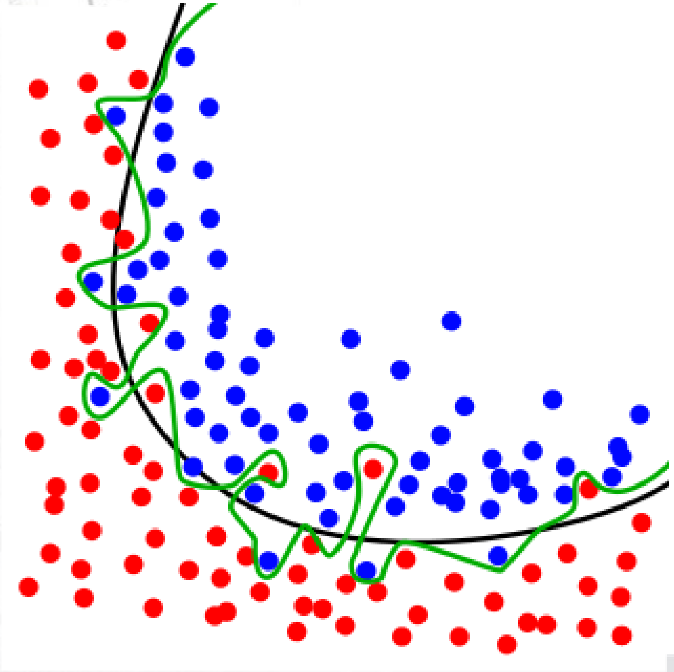
Association analysis

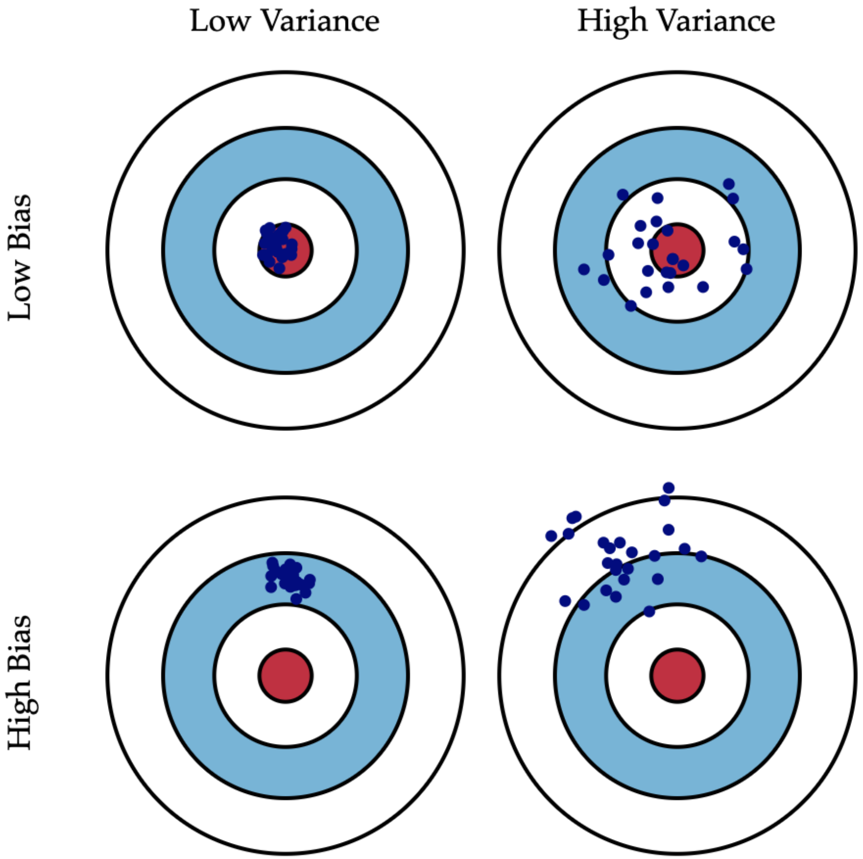
Principal Component Analysis

Neural Networks

Key concepts to know

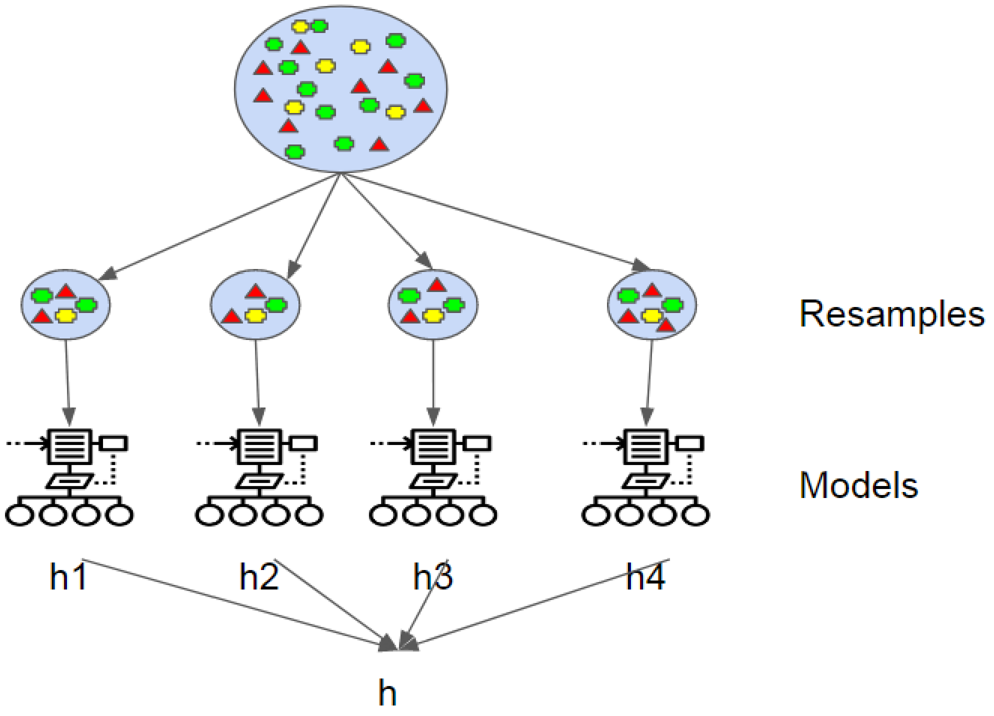
- Training/Validation/Test data sets
- Multi-collinearity (it's a problem!)
- Feature engineering
- Bias vs. Variance
- Overfitting

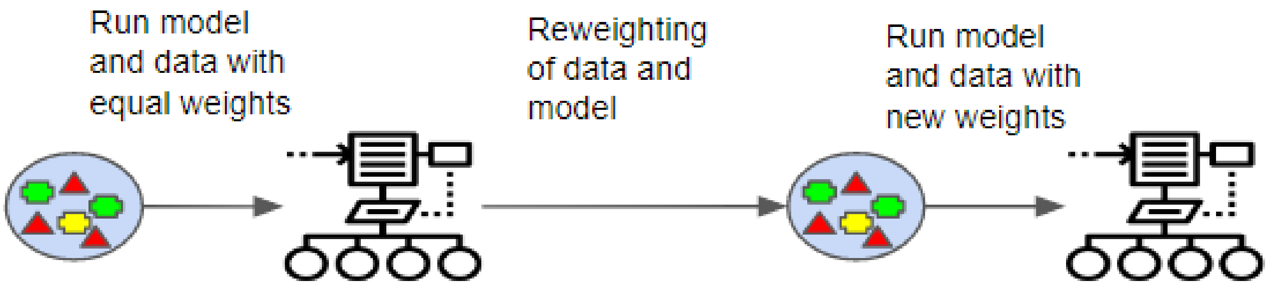




Ensemble methods are typically much more accurate than a standalone model

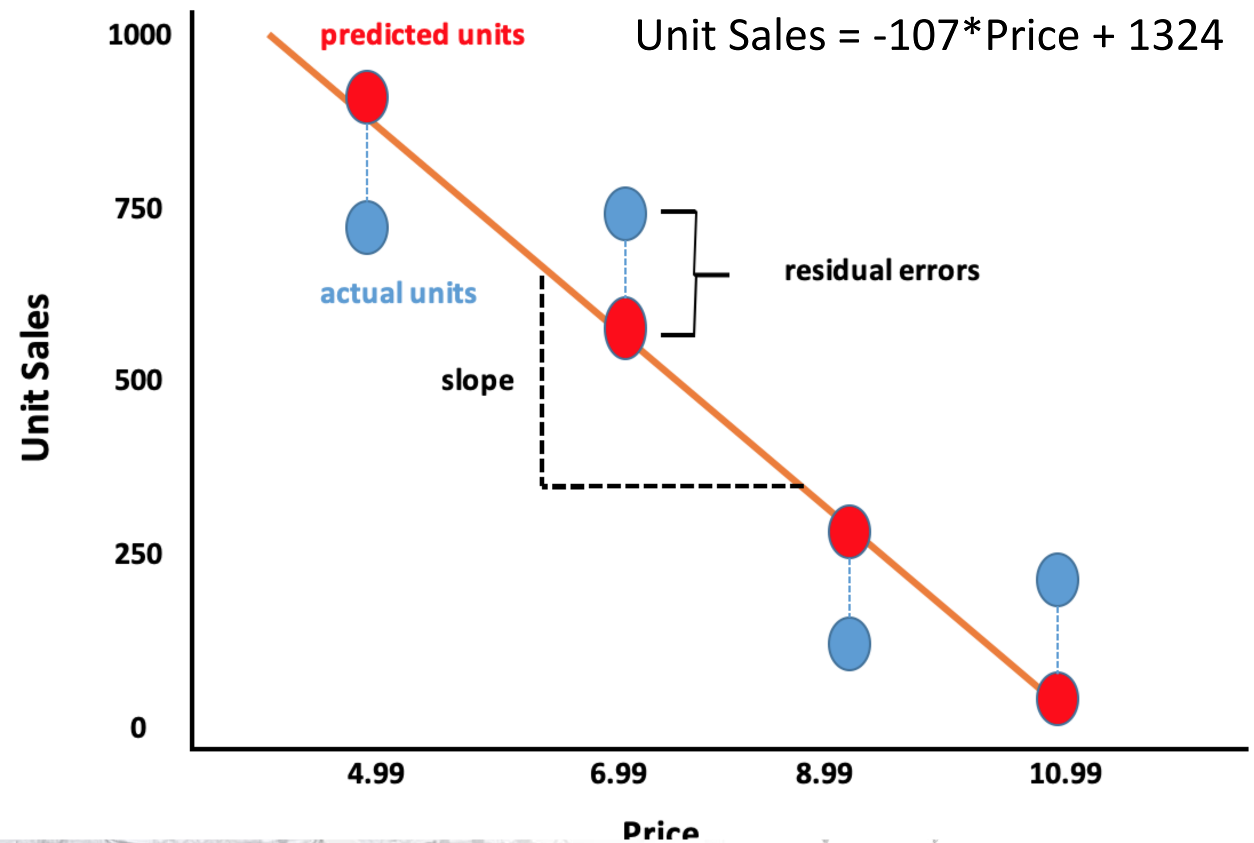
Most common are bagging and boosting





Images sourced from: https://medium.com/@SeattleDataGuy/how-to-develop-a-robust-algorithm-c38e08f32201

Linear Regression

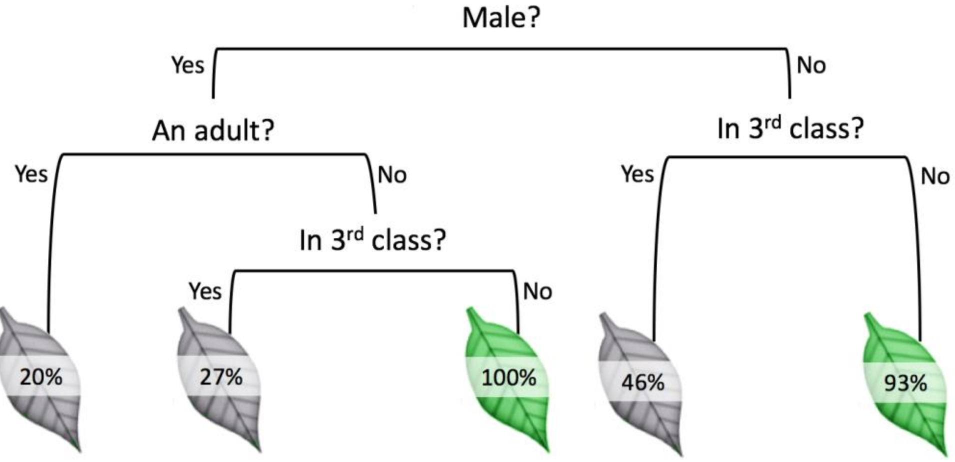


- The basis of modern day machine learning
- Simple; fast; great baseline model before employing more advanced algorithms; easy to explain
- Multi-collinearity and overfitting are a big problem; low accuracy vs. more advanced models; most business problems are complex, non-linear (not suitable for many real-world applications)

Decision Tree

Classification Tree

Predicting Titanic survival rates

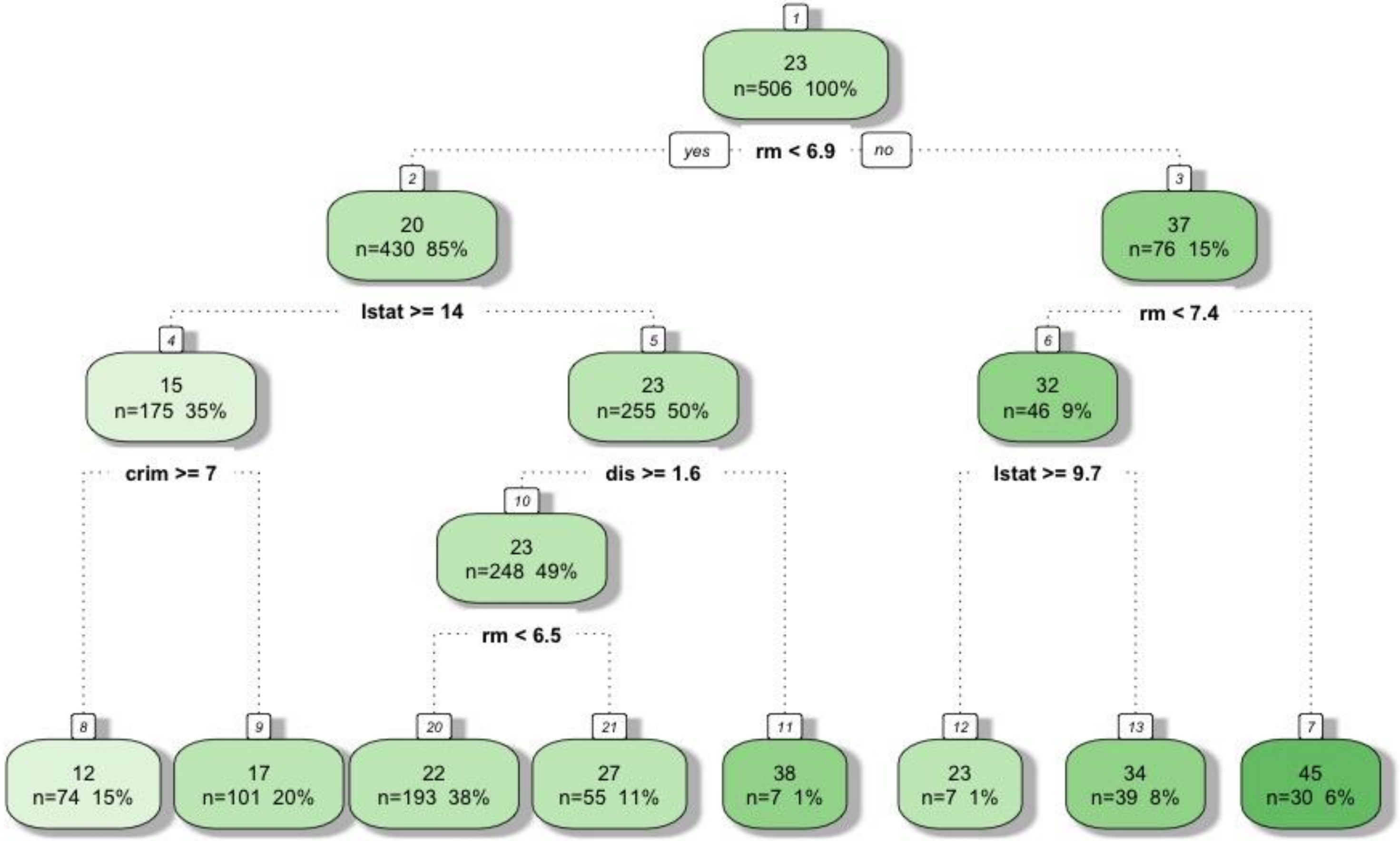


- Simple; fast; deals with multi-collinearity; addresses non-linear nature of business problems; easy to visualize and explain
- Low accuracy vs. more advanced models; prone to overfitting

Regression Tree

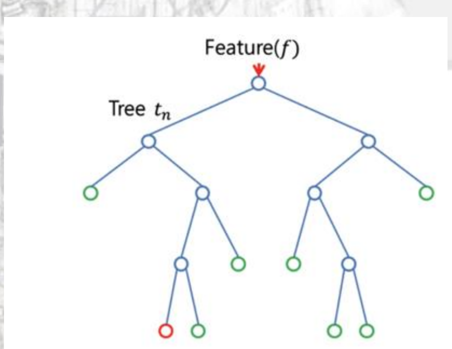
Predicting Median Home Values in Boston in 1978

rm = average number of rooms per dwelling
lstat = percent low income population
crim = per capita crime rate by town
dis = distance to employment centers

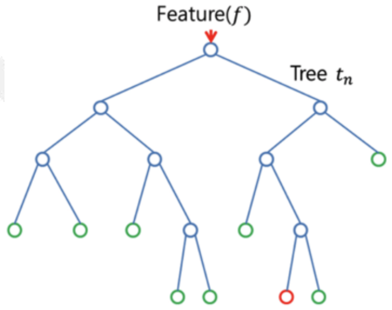


Random Forest

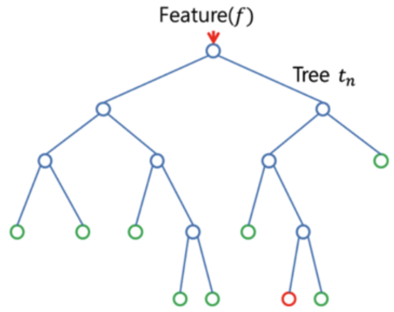
Tree #1



Tree #2



....Tree #500

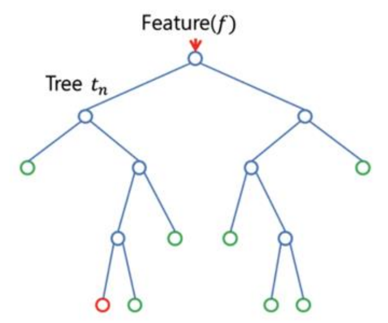


- Independent trees, each with the same level of importance (bagging)
- Classification: majority rules
- Regression: average out the predictions across all the decision trees

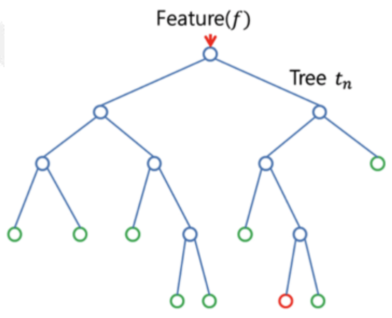
- Simple; fast; few algorithms can produce more accurate results; multi-collinearity and overfitting not an issue; can be used as a baseline for building more complex models (variable importance)
- Hard to explain; can be slow to train

Gradient Boosting Machines

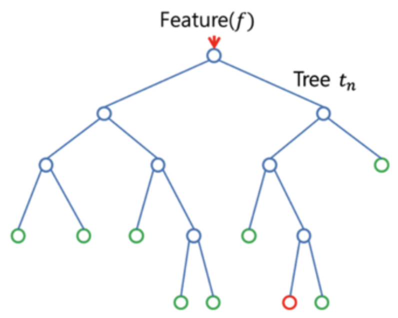
Tree #1



Tree #2



....Tree #500



- Sequential trees. Prediction errors that Tree #1 makes influence how Tree #2 is constructed; errors that Tree #2 makes influence how Tree #3 is made, etc… (boosting)
- Classification: based on weighted importance of all decision trees
- Regression: weighted average of predictions

- GBM and its variations can often produce the most accurate models
- Hard to explain; can be much slower to train than Random Forest; prone to overfitting

Which algorithm should I choose?

Algorithm	Can it handle Regression, Classification or Both?	Model Interpretability	Easy to Explain to your EVP of Sales?	Predictive Accuracy	How fast can you train a model?	How fast can you predict on new data?	How much data do you need to build a model?	Do they handle multi-collinearity well?
Linear regression	Regression	Very Strong	Yes	Lower	Fastest	Fastest	Very little	No
Logistic regression	Classification	Strong	Yes	Lower	Fastest	Fastest	Very little	No
Nearest Neighbor	Both	Very Strong	Yes	Lower	Fast	Slower	More	No
Decision trees	Both	Medium	Yes	Lower	Fast	Fast	Little	Yes
Random Forests	Both	Low	Somewhat	Higher	Slower	Fast	Much more	Yes
GBM	Both	Low	Somewhat	Higher	Slower	Fast	Much more	Yes
Neural networks	Both	Lowest	Much harder	Highest	Slow	Fast	Most	Yes

Data Strategy| Key Components

Data Infrastructure

- Development supporting data acquisition and maintenance.
- Architecture that supports and *usage*.

Governance

- GDPR, PII, and news stories are bringing data governance forward.
- As systems, tools, MDM and data change our views must *evolve*.

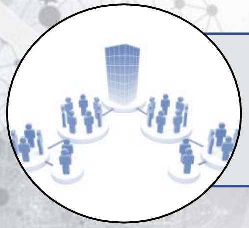
Technology Stack

- Technology that is nimble and scalable.
- Solution orientation focused on continuous *evolution*.

Talent

- Proprietary or Third Party, it takes *aptitude* to know how or why.
- Data Science talent isn't one size fits all.

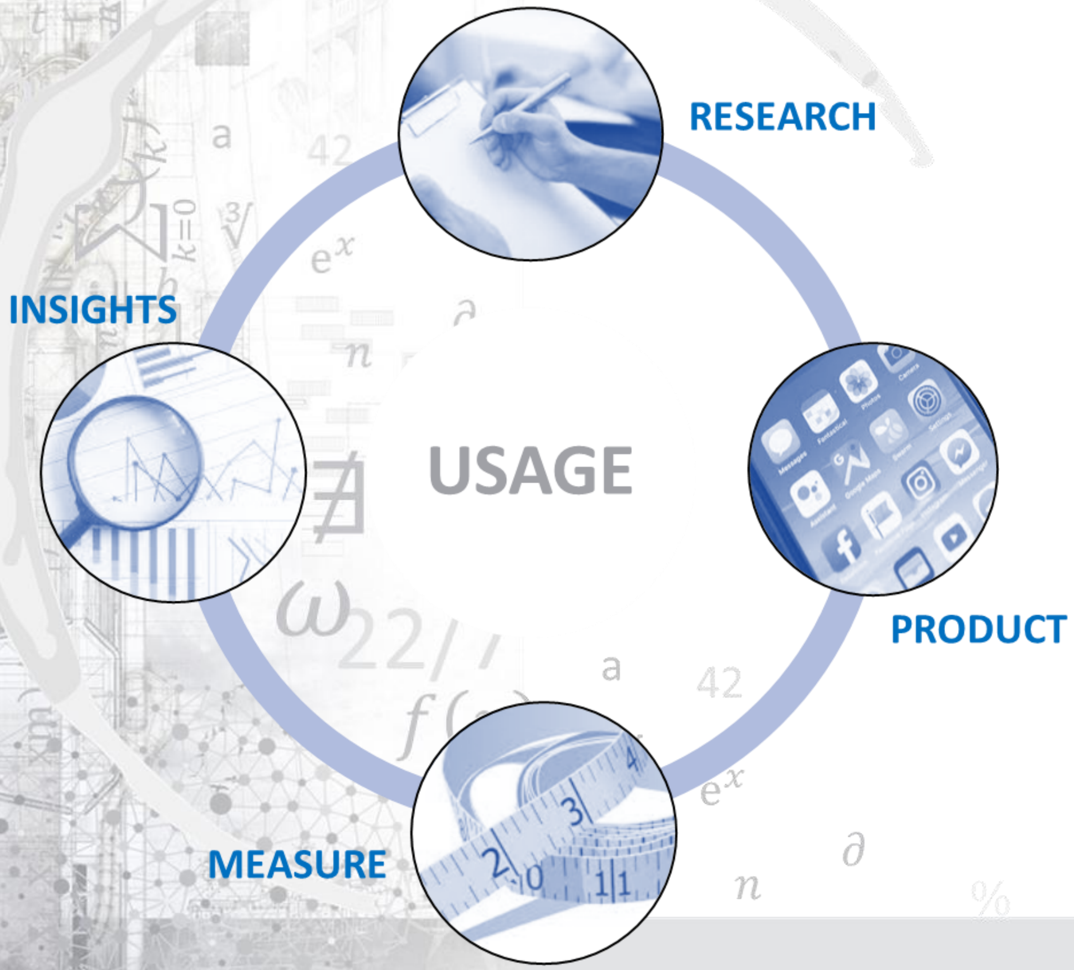
Integration



- Executive sponsorship must be more than verbal.
- Organizational design is *idiosyncratic*.

No one defines data the same

Generalized Data



Master Data

Transactional

Analytical

Reference

Data Governance: *Practice of ensuring access to high-quality data throughout the data life cycle.*

Governance/MDM In Data Science Era

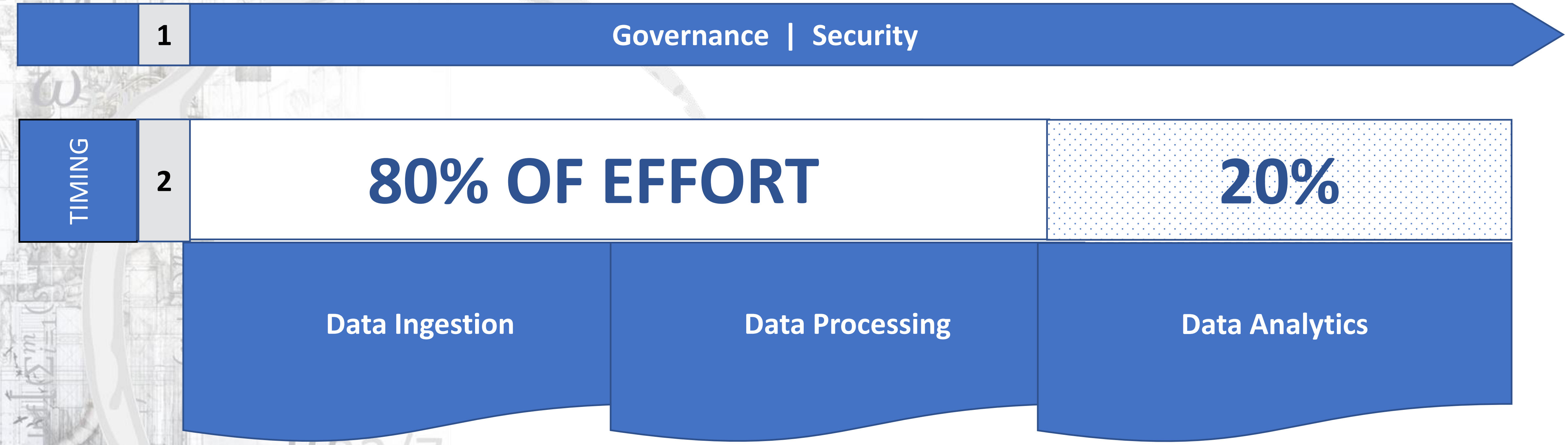
Generalized MDM

- Organization
- Data Origination
- Legal Requirements
- Security

MDM Qualities

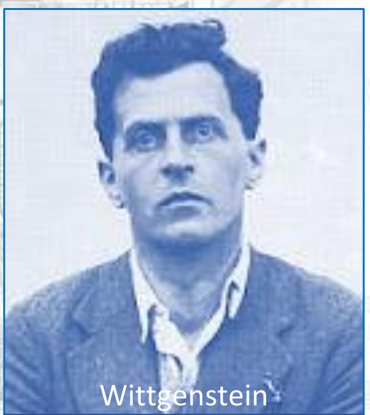
- Relationships
- Quality
- Adaptability

Understanding the Data Ecosystem



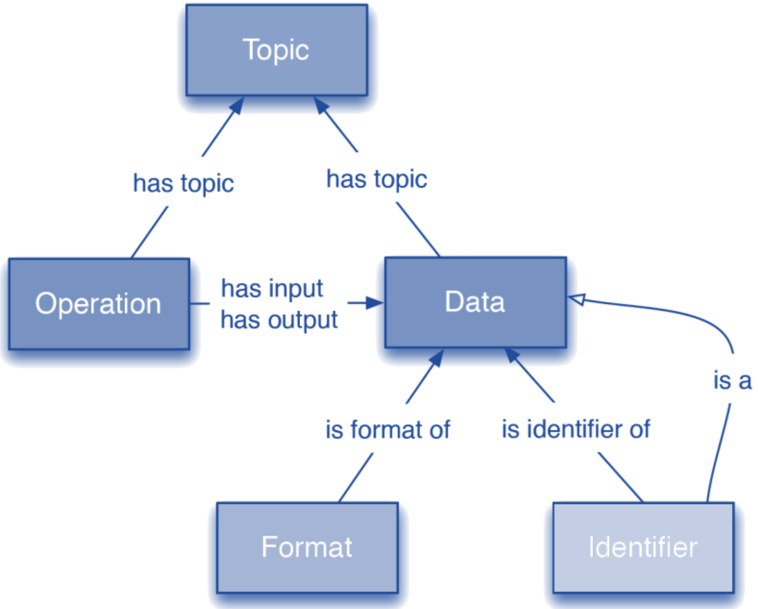
Static No More: Ontologies & Labelling

We have never been in a static world we need adaptive ontologies.



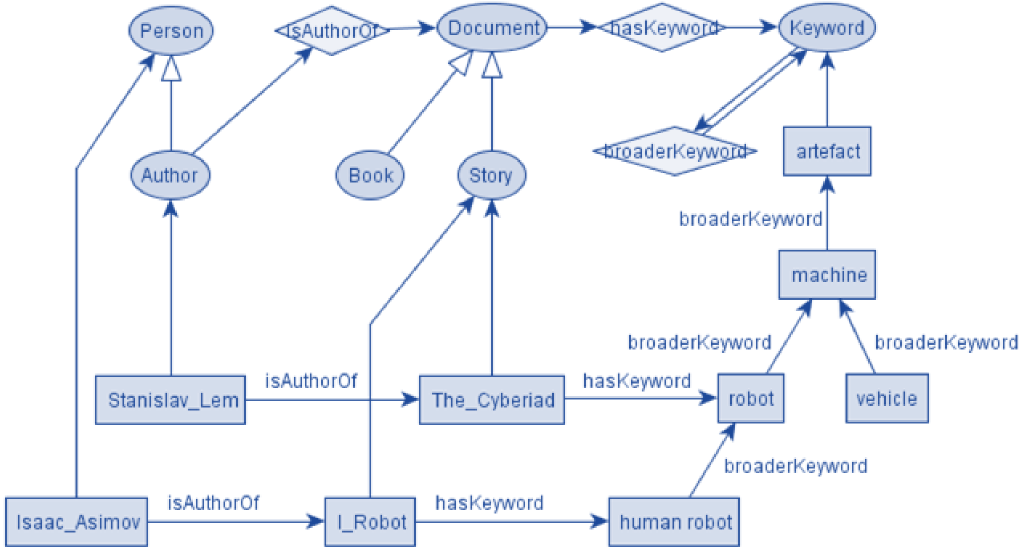
The limits of my language mean the limits of my world.

Global Connections

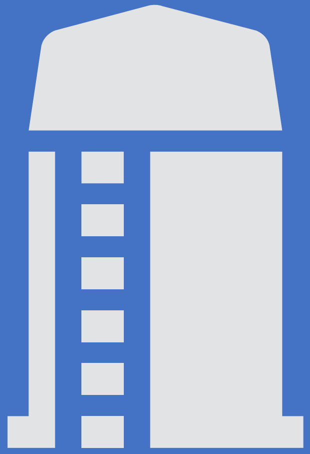


http://edamontology.org

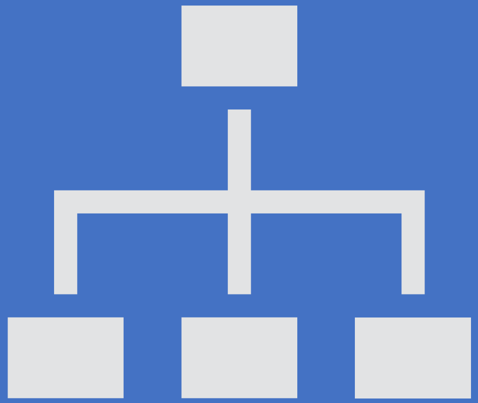
Respective Associations



Changing data means a changing framework



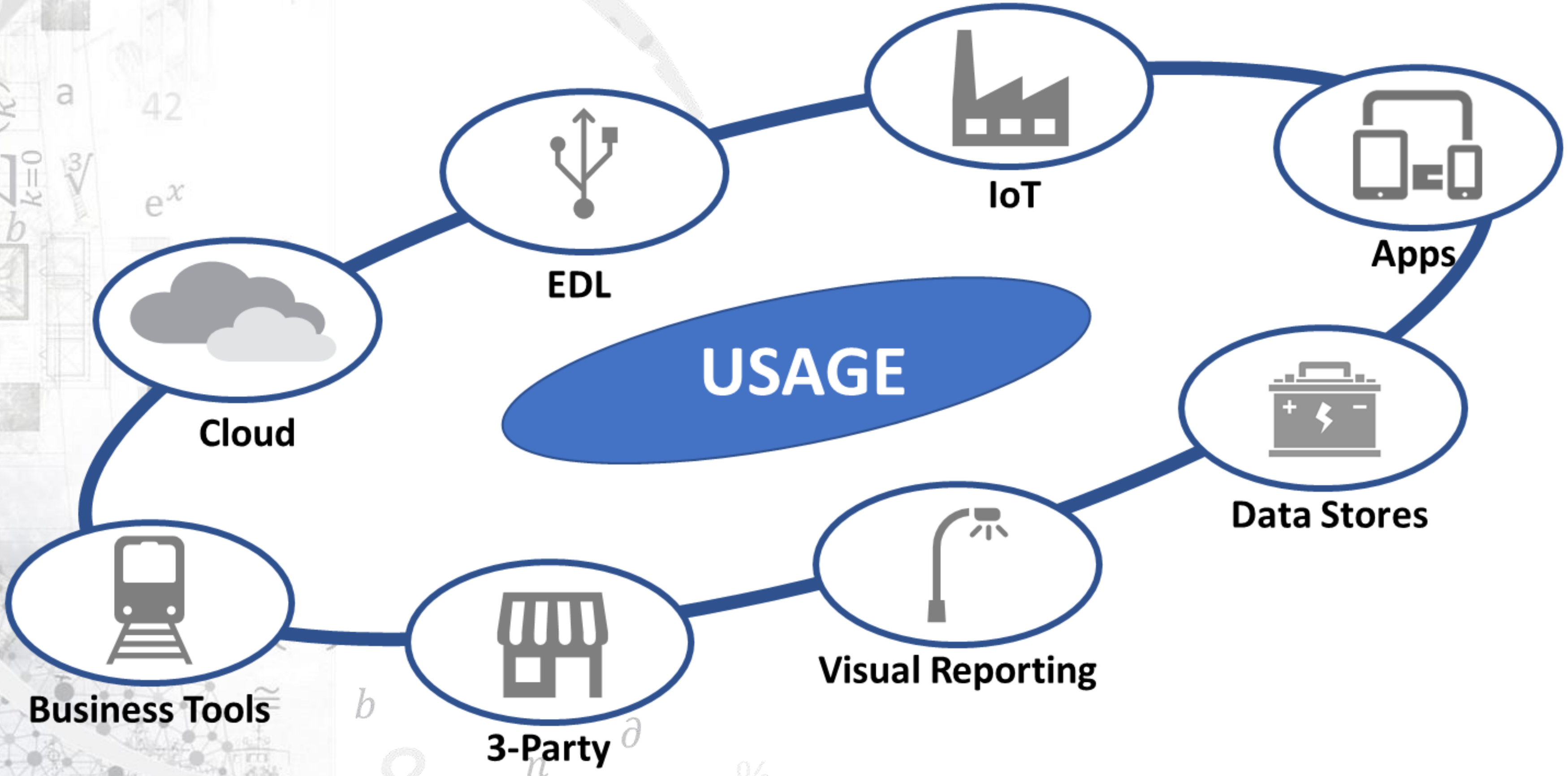
SILO’D



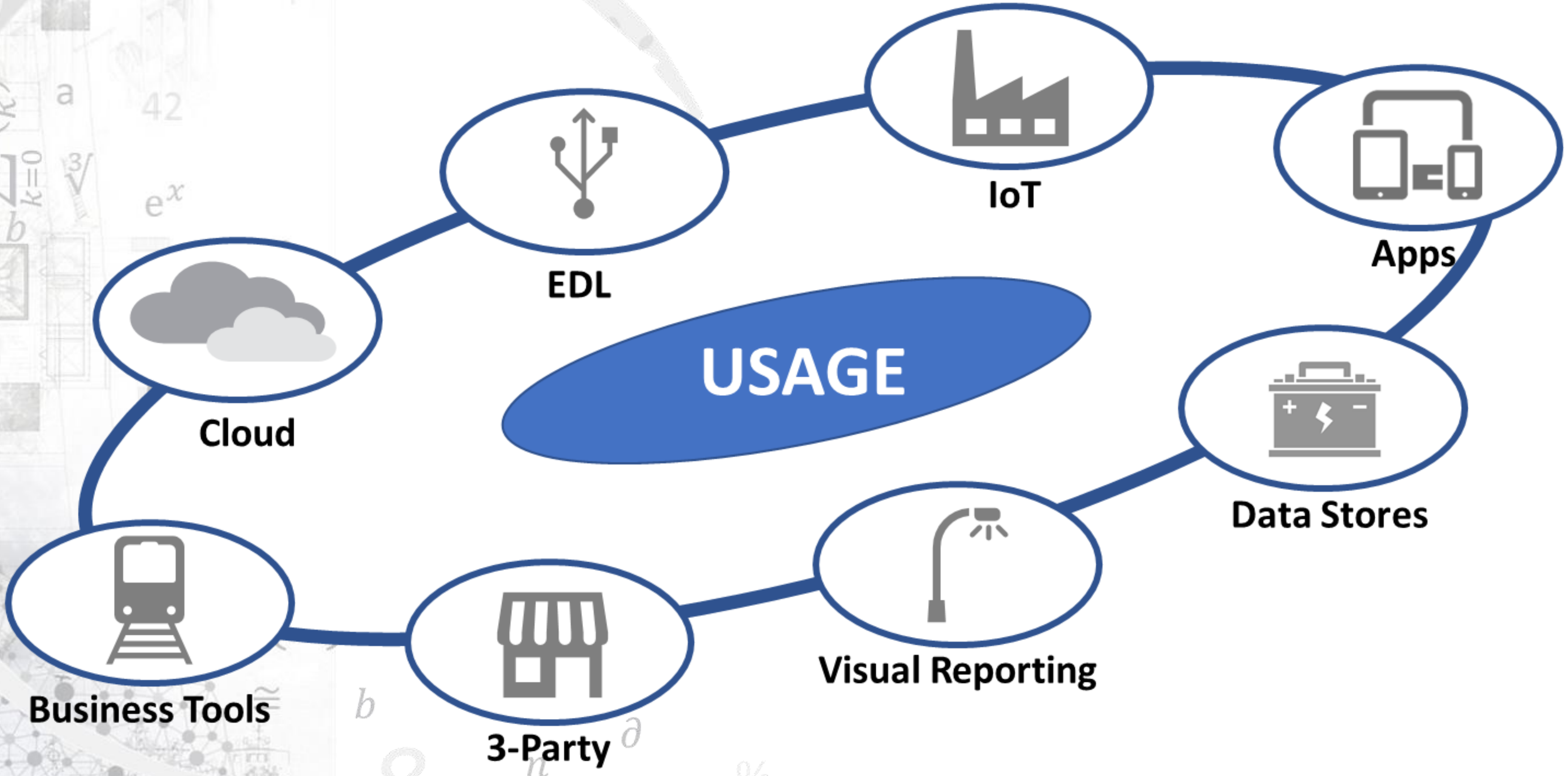
ORG

"Everything should be made as simple as possible, but not simpler." - *Albert Einstein*

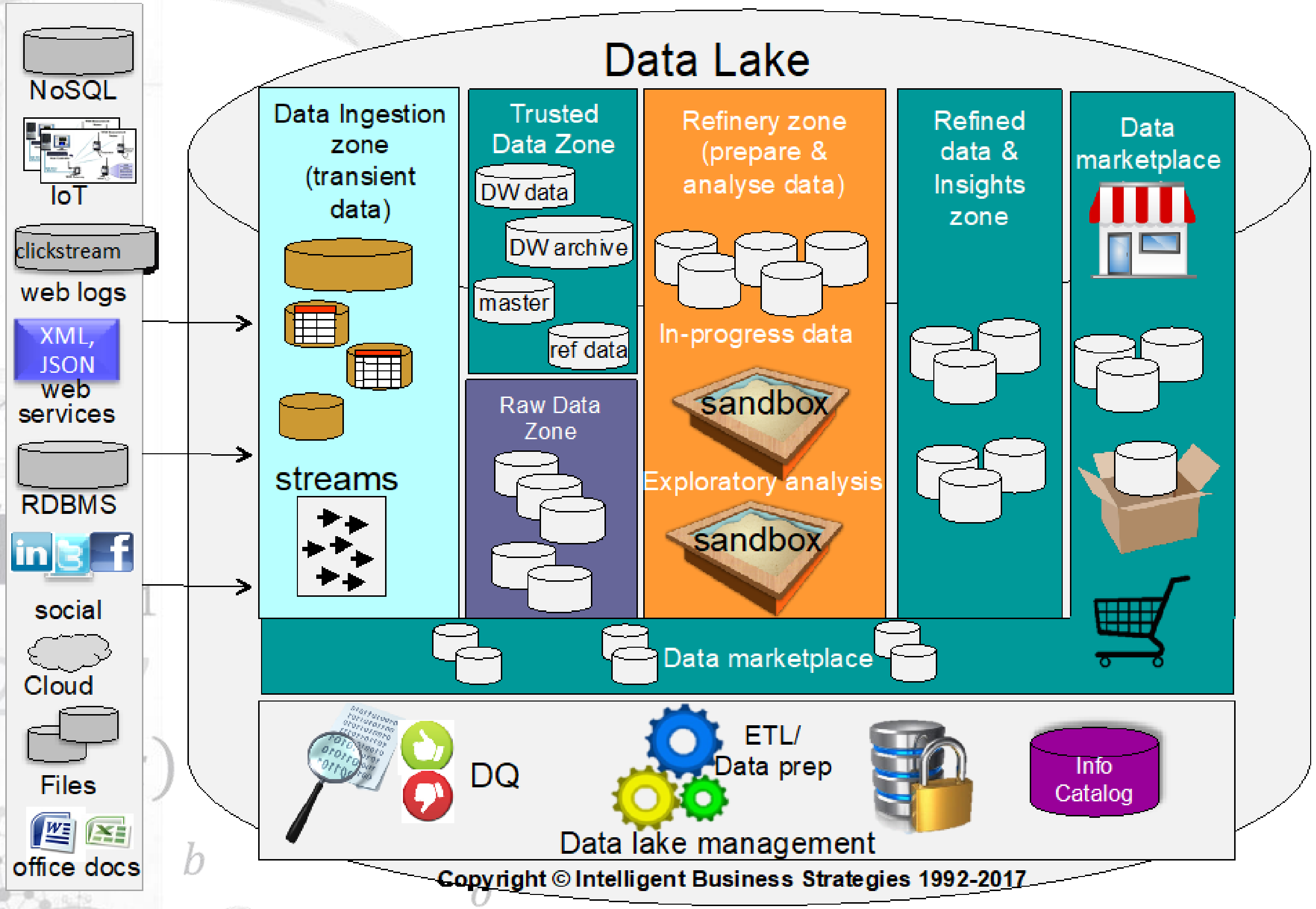
Evolving Data Ecosystem



Bringing this to life



Data Lakes are more than Big Data



“Errors using inadequate data are much less than those using no data at all. ” --Charles Babbage

The Cloud: Everyone is doing it…

83% of enterprises have a Cloud Strategy…^

13,202 views | Jul 18, 2018, 08:11am

Walmart CIO: We Picked Microsoft For Huge Cloud Deal To Accelerate Digital Transformation

Bob Evans Contributor
Enterprise & Cloud

salesforce

Enable smarter, more connected experiences with Salesforce

Salesforce has named Google Cloud a preferred public cloud provider and integrates with G Suite and Google Analytics 360 to help businesses create smarter, richer customer experiences.

LEARN MORE →

Netflix & Amazon Kinesis Streams Case Study

NETFLIX

Amazon Kinesis Streams processes multiple terabytes of log data each day, yet events show up in our analytics in seconds. We can discover and respond to issues in real time, ensuring high availability and a great customer experience.

John Bennett
…gineer, Netflix

Kellogg's

The Kellogg Company

Kellogg's uses AWS to to deliver spend analyses and data simulations in minutes.

More big data stories »

Microsoft Azure

The University of Ottawa Heart Institute

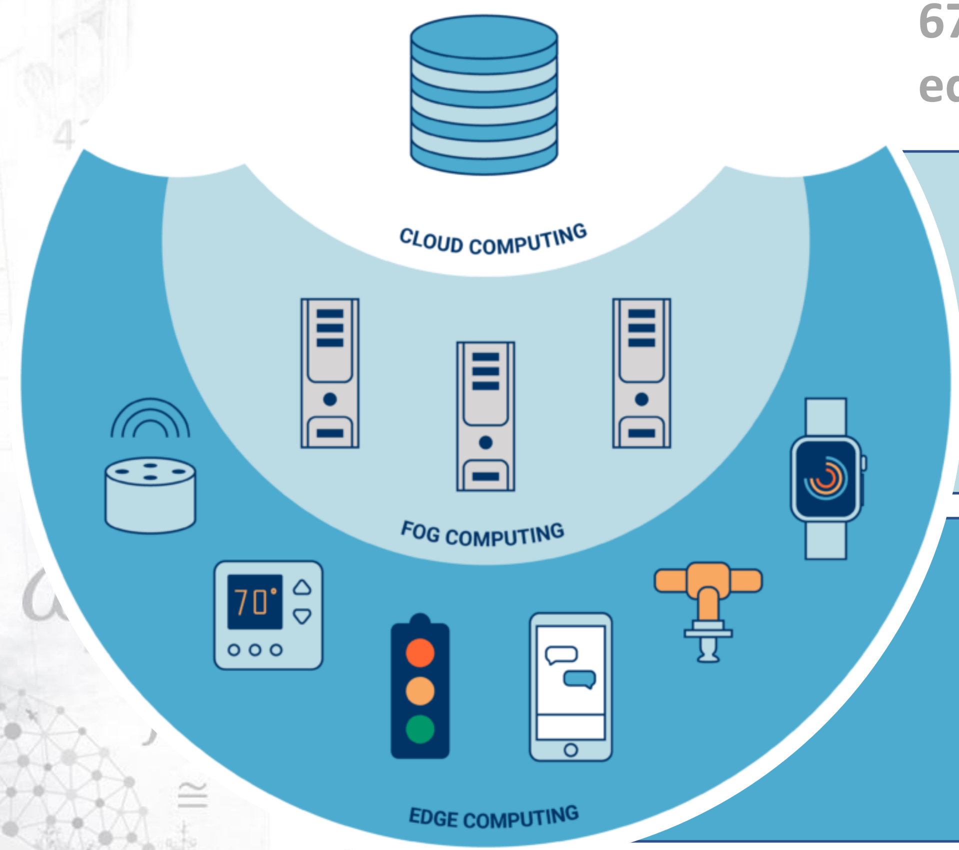
Global cardiac leader gains new insight to improve care, increase efficiency, and facilitate research

nielsen

Google Cloud

^ Forbes.com

Not only the cloud but living on the Edge (IoT)



67% of enterprises increasing edge deployments…^

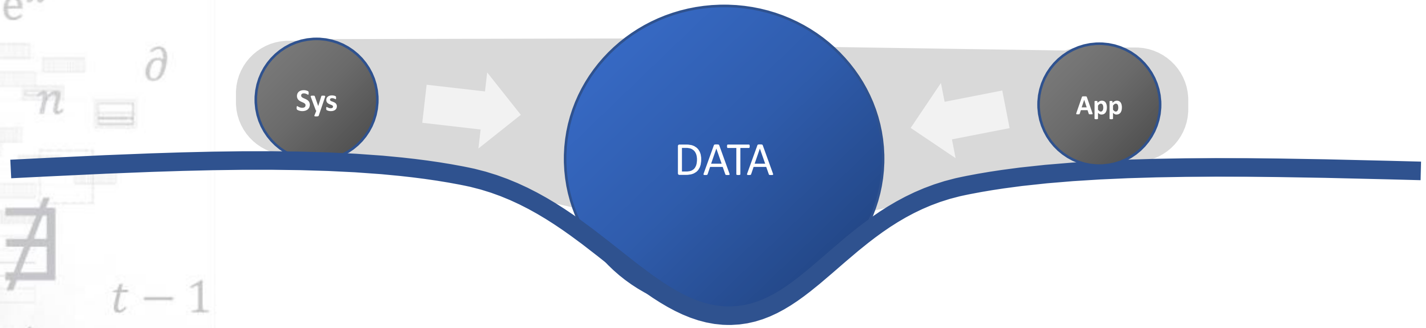
FOG systems exchange data between the edge devices and cloud systems

Edge Computing is IoT. It is your Apple® watch or your Nest® home device.

https://www.cbinsights.com/research/what-is-edge-computing/

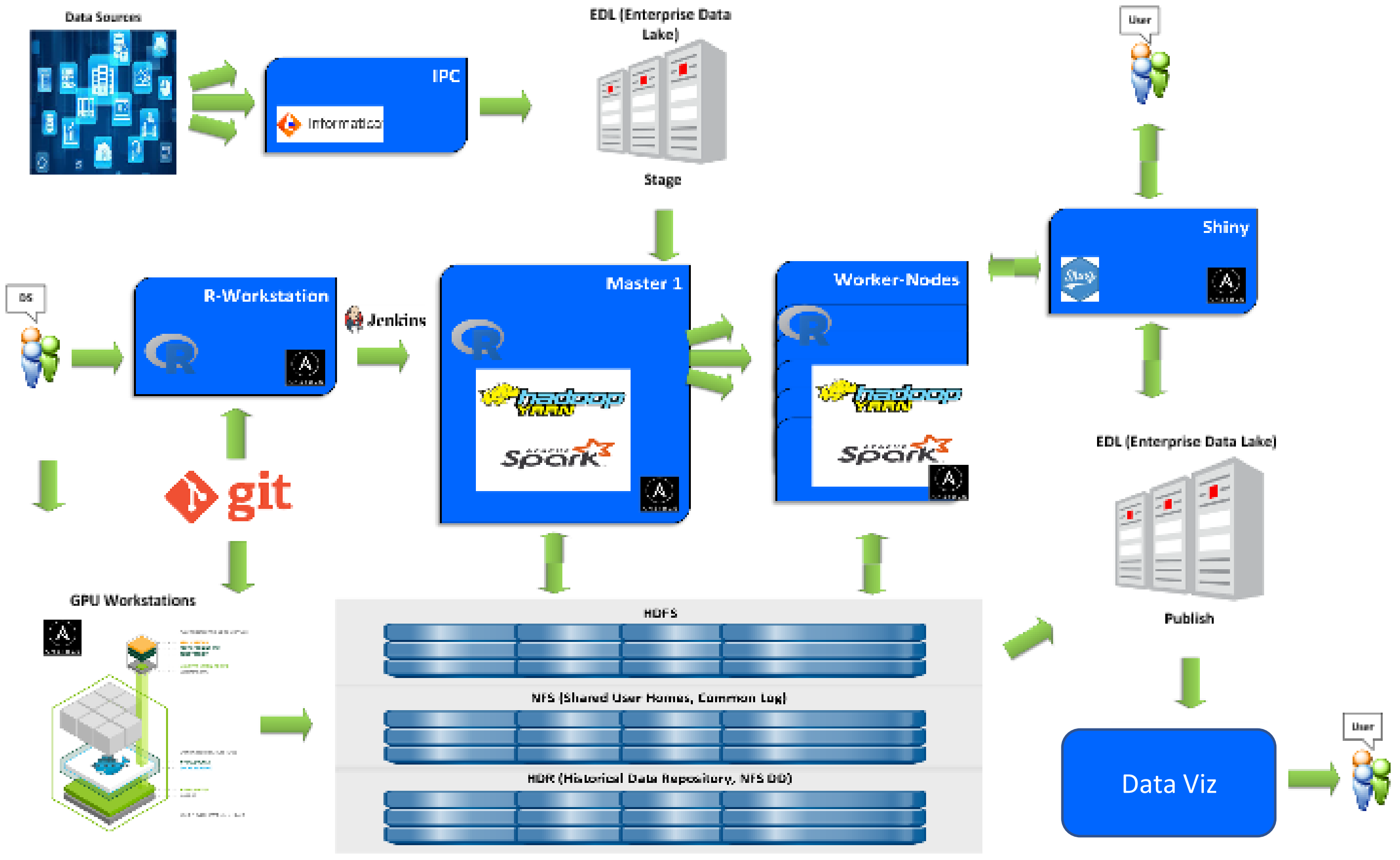
^ Forbes.com

Data Gravity is the new standard



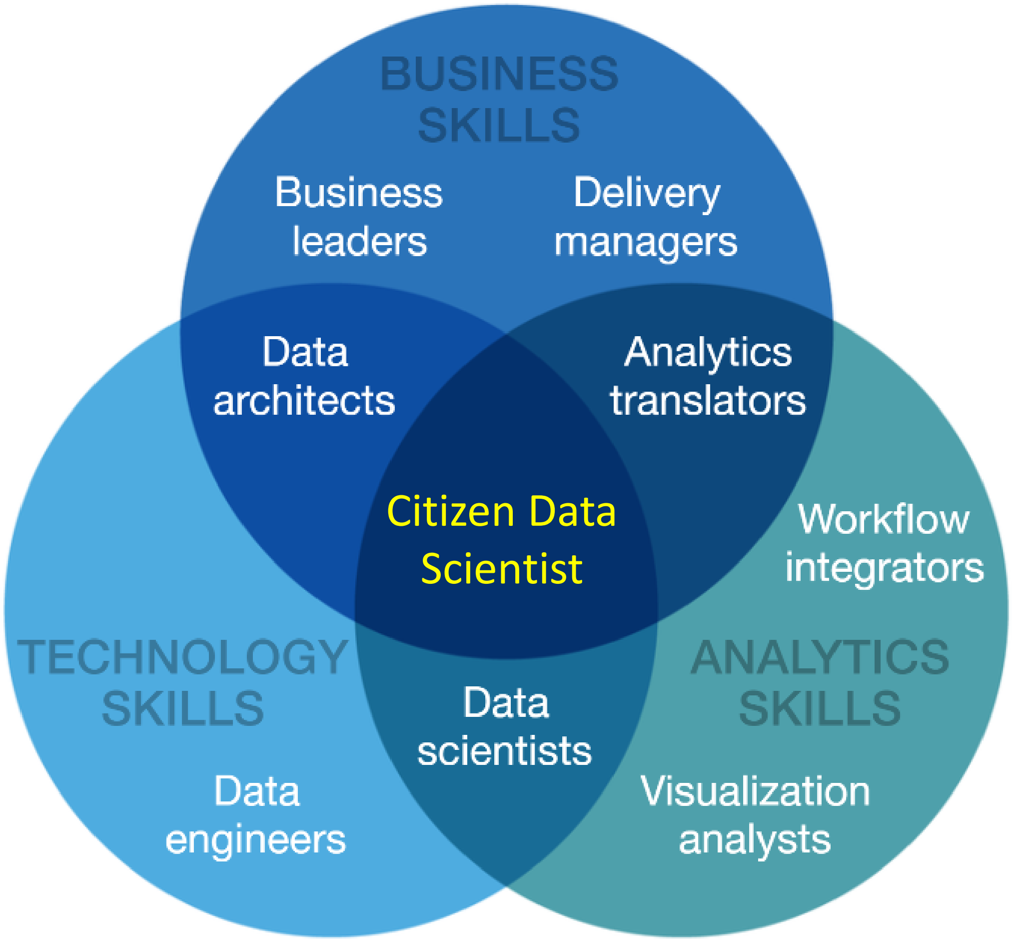
Evolving Data Ecosystem

Example:
Reimagining on-premise and cloud based architecture to more quickly address business questions.



Complex subjects need expert support

Data science translators are critical



Upskilling traditional analysts is perhaps the most important

Enhance skills of traditional analytical roles:

- Coding (R and/or Python)
- Foundational ML algorithms
- Instilling a Statistical acumen
- Utilize quick hit Proof of Concept efforts
- A connected relationship

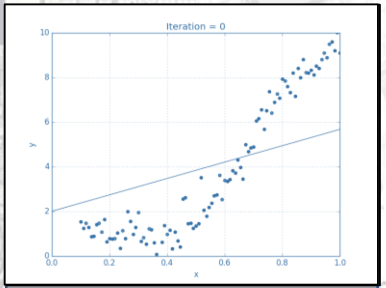
Source: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/analytics-translator

Talent: Single COE or Live within Functions?

Gustavo Dudamel | Los Angeles Philharmonic

The various instruments add color

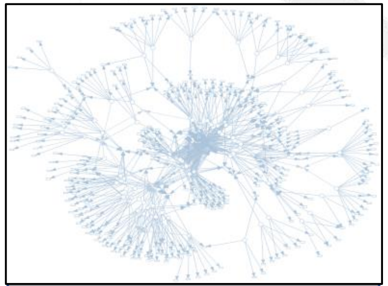
Pedro Domingos’ *Five Tribes*



SYMBOLIST

Consider logical form and linear deduction

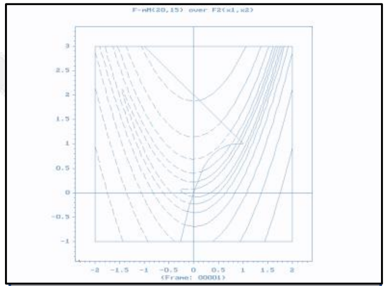
Logit Functions
Avg Elasticity Models



CONNECTIONIST

Bring disparate pieces of data together

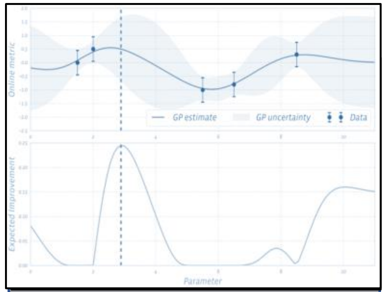
Neural Networks
CNN’s
Squared Error’s



EVOLUTIONARY

Algorithms that learn and develop and learn and…

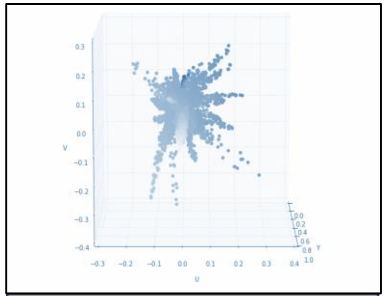
Swarm Algo (Routing)
Search Algorithm



BAYESIAN

Utilize history with snippets of today to inform tomorrow

Probabilistic Inference
Posterior Probability



ANALOGIZER

Efficient optimization to highlight hidden patterns

SVM
Constrained Optim.

They represent a breadth of application

Multi-tiered demand forecasting

Trade Promotion Effectiveness & Optimization

Advertisement ROI

Optimized product assortments at the individual store level

Price positioning studies (new product pricing, price-value curves)

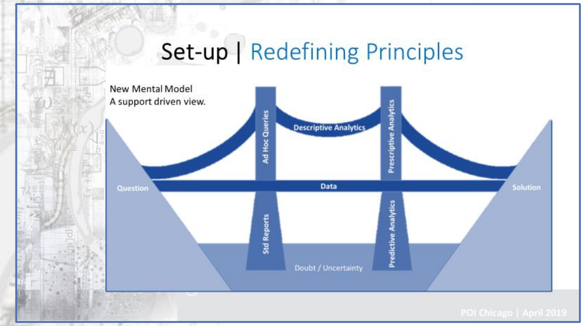
Process automation

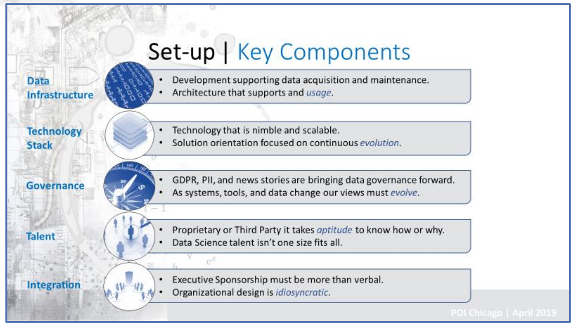
Personalized consumer engagement

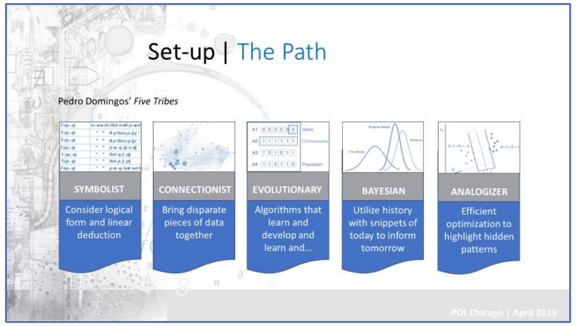
Natural Language Processing areas:

- Trend predictions for product development
- Sentiment analysis for pricing, competitive intel and concept testing
- NLP for descriptive and diagnostic analytics

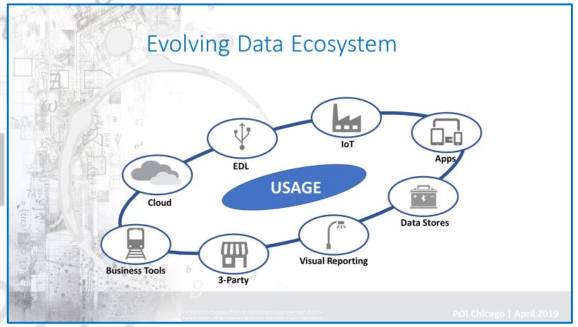
Set-up | Data Foundation Review

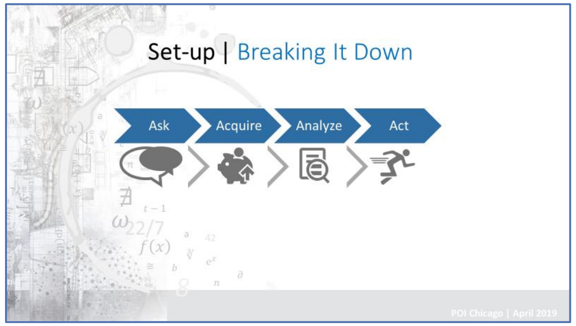






The are no Unicorns but schools of thought.





Take your time |
Get the ask right |
Get the right data |
Then Analyze |
Finally Act on the Facts

Ultimately Change Management Is Critical

Data Science

Realities and Key Learnings

The realities for CPG

Realities

Few companies have seen payoffs from Data Science investments

Few CPGs treat Data as a key strategic asset or Analytics as a core competency

Many prototypes, but few large-scale implementations

Roadblocks

Lack of Vision

Bad Prioritization

Human Capital

Little to no data governance

Estimating Impact

$1M Netflix prize: complex, accurate…negative ROI!

Why Netflix Never Implemented The Algorithm That Won The Netflix $1 Million Challenge

from the *times-change* dept

You probably recall all the excitement that went around when a group finally won the big Netflix $1 million prize in 2009, improving Netflix's recommendation algorithm by 10%. But what you might *not* know, is that **Netflix never implemented that solution itself**. Netflix recently put up a blog post discussing some of the details of its recommendation system, which (as an aside) explains why the winning entry never was used. First, they note that they *did* make use of an earlier bit of code that came out of the contest:

> *A year into the competition, the Korbell team won the first Progress Prize with an 8.43% improvement. They reported more than 2000 hours of work in order to come up with the final combination of 107 algorithms that gave them this prize. And, they gave us the source code. We looked at the two underlying algorithms with the best performance in the ensemble: Matrix Factorization (which the community generally called SVD, Singular Value Decomposition) and Restricted Boltzmann Machines (RBM). SVD by itself provided a 0.8914 RMSE (root mean squared error), while RBM alone provided a competitive but slightly worse 0.8990 RMSE. A linear blend of these two reduced the error to 0.88. To put these algorithms to use, we had to work to overcome some limitations, for instance that they were built to handle 100 million ratings, instead of the more than 5 billion that we have, and that they were not built to adapt as members added more ratings. But once we overcame those challenges, we put the two algorithms into production, where they are still used as part of our recommendation engine.*

Neat. But the winning prize? Eh... just not worth it:

> *We evaluated some of the new methods offline but the additional accuracy gains that we measured did not seem to justify the engineering effort needed to bring them into a production environment.*

It wasn't just that the improvement was marginal, but that Netflix's business had *shifted* and the way customers used its product, and the kinds of recommendations the company had done, had shifted too. Suddenly, the prize winning solution just wasn't that useful -- in part because many people were *streaming* videos rather than renting DVDs -- and it turns out that the recommendation for streaming videos *is different* than for rental viewing a few days later.

Neat. But the winning prize? Eh... just not worth it:

> *We evaluated some of the new methods offline but the additional accuracy gains that we measured did not seem to justify the engineering effort needed to bring them into a production environment.*

It wasn't just that the improvement was marginal, but that Netflix's business had *shifted* and the way customers used its product, and the kinds of recommendations the company had done, had shifted too. Suddenly, the prize winning solution just wasn't that useful -- in part because many people were *streaming* videos rather than renting DVDs -- and it turns out that the recommendation for streaming videos *is different* than for rental viewing a few days later.

Source: https://www.forbes.com/sites/ryanholiday/2012/04/16/what-the-failed-1m-netflix-prize-tells-us-about-business-advice/

Advanced Analytics @ ATD

Key takeaways on building a successful team

- Start at the very top
- Solve real problems
- Don’t sound like Your PhD thesis
- Monetize
- Learn the domain…push the envelope
- Evangelize
- Push back
- Admit failure

Advanced Analytics @ ATD

Sample Capabilities

Pricing insights and processes running on Open Source Systems

Predict customer churn with 85% accuracy 90 days out

Runs+ scenarios daily to optimize 50K+ retail partners' profitability

Forecasts demand based on 1000+ features, not just past sales

Ability to tell retail partners how much demand an inch of snowfall will generate

Train Data Translators and Citizen Data Scientists

ATD
AMERICAN TIRE DISTRIBUTORS

Example: Inventory optimization via drone based image recognition

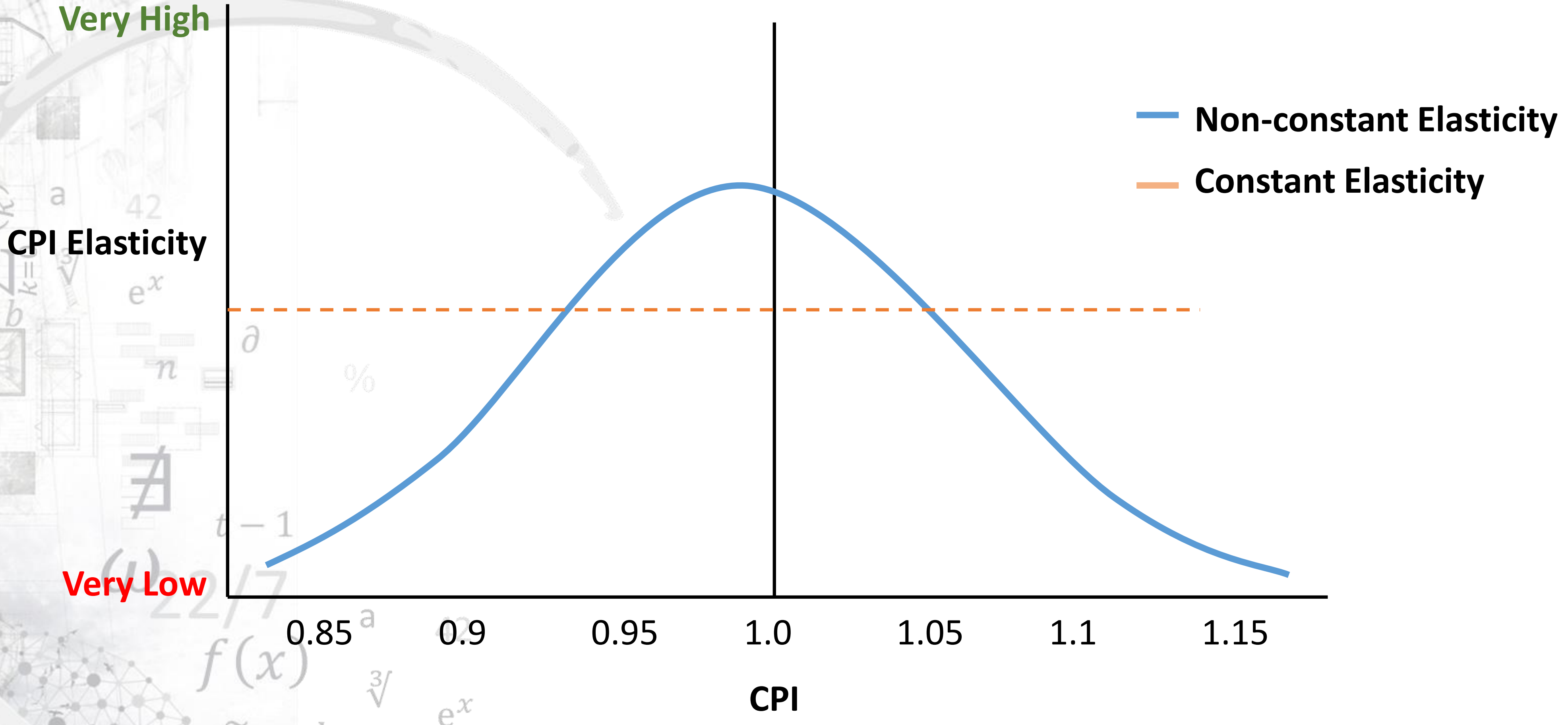
Price Elasticity Modeling

When being perturbed is a good thing

Own Elasticity vs. CPI Elasticity

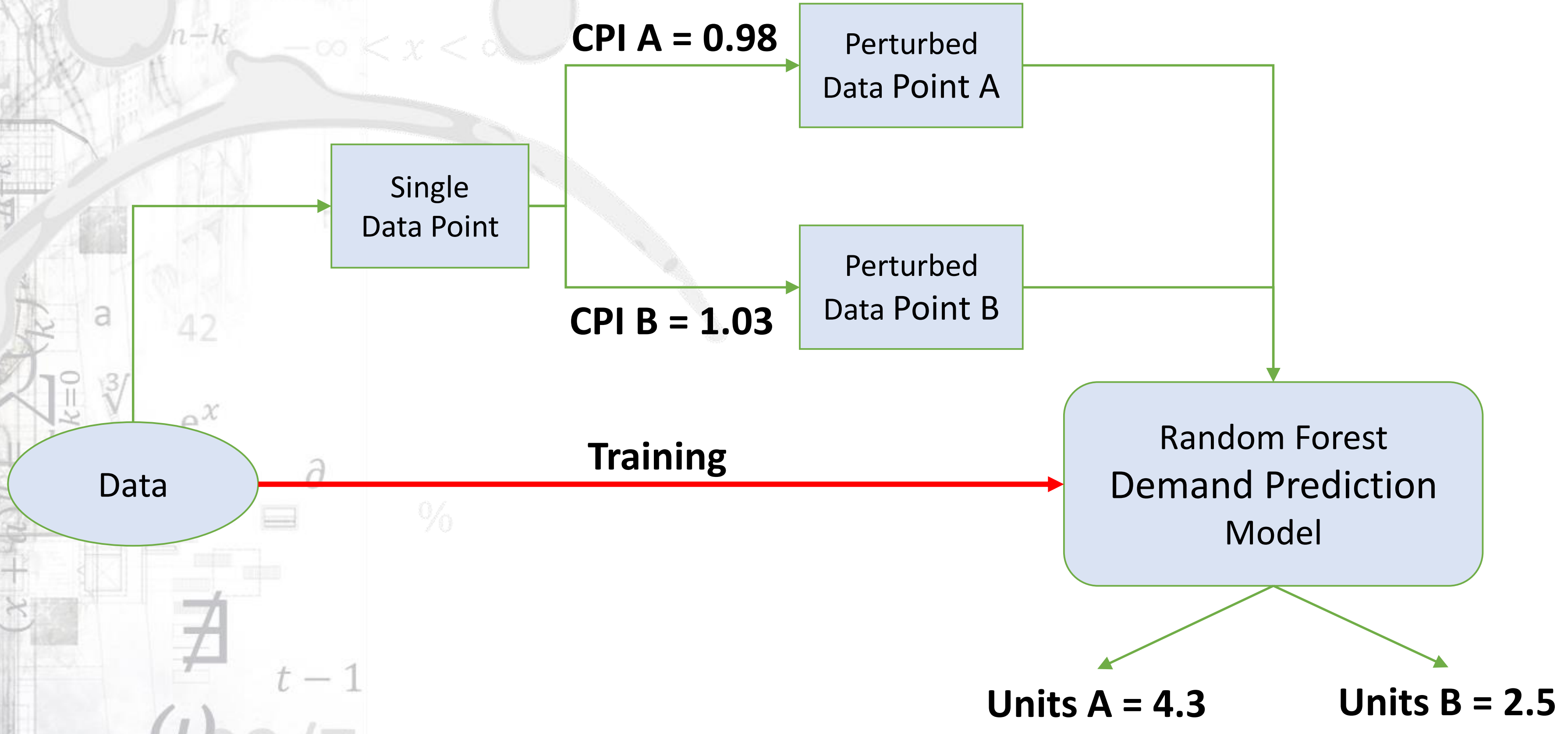
- In Consumer Products / Manufacturing / Distribution type industries, Price Elasticity is typically estimated via linear regression models
- Log(Units Sold) ~ Log(Price) + Past Units Sold + Seasonality + other factors
    - The **Price coefficient** in the model loosely equates to price elasticity
- Instead of measuring unit sales' sensitivity to price changes, we can also measure **sensitivity to CPI changes (aka. CPI Elasticity)**:
    - CPI=Competitive Price Index, aka. Our Price / Competitive Market Price
        - Example 1: Our Price = 90 and Comp Market Price = 100… CPI = 0.9
        - Example 2: Our Price = 105 and Comp Market Price = 100… CPI = 1.05
    - Example: CPI elasticity = -2
        - If **CPI** goes from 1.05 to 1.03 (-2% change), units are expected to increase by 4%
- CPI-elasticity model better aligns with human and marketplace behavior

CPI Elasticity in Reality (Non-Constant)



CPI = Competitive Price Index (Our Price / Competitor Price)

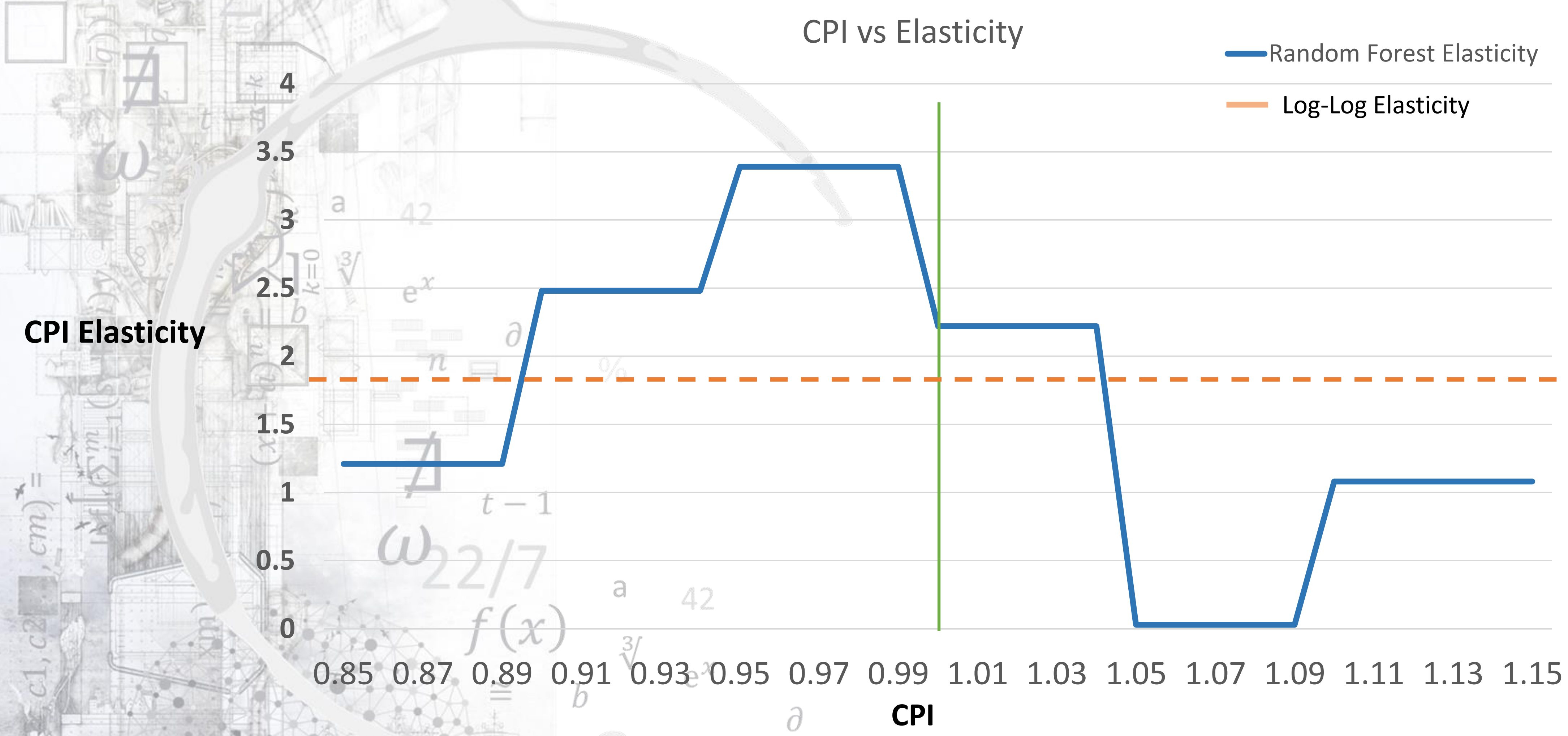
CPI Elasticity simulation using Random Forest model



$$Imputed\ CPI\ Elasticity = \frac{\%\ Units\ change}{\%\ CPI\ change} = \frac{-41.9\%}{5.1\%} = -8.2$$

Elasticity between CPI 0.98 to 1.03 is -8.2

Elasticity curve for a Customer-Product



Organizational Approach is Idiosyncratic

YOUR ORGANIZATION'S
PROJECT MANAGEMENT STYLE

- **Experiment:** Prove the concept
- **Patience:** Everyone will want it now
- **Innovate:** Be willing to fail forward
- **Budget:** Build Accordingly
- **Scalability:** There is no future proof

C-Suite Must Be 100% On Board | Committed