PRMIA
The Professional Risk Managers' International Association

Data Science for Cybersecurity Risk

Scott Mongeau
Data Scientist – Cybersecurity
SAS Institute

June 22nd 2017
10:00-11:00 am EDT / 3:00-4:00 pm BST / 4:00-5:00 pm CST

PRMIA WEBINAR

SCOTT MONGEAU

Data Scientist
Cyber Security

scott.mongeau@sas.com

+31 68 370 3097

LinkedIn
Twitter
Blog

Experience

- **SAS Institute**
  Data Scientist (cyber/fraud/security)
- **Deloitte**
  Mgr. Analytics (fraud/fin.crime/cyber)
- **Nyenrode University**
  Lecturer Analytics
- **SARK7 Analytics**
  Owner / Consultant (risk/finance)
- **Genentech Inc. / Roche (biotech)**
  Principal Analyst / Sr. Manager
- **Atradius (insurance)**
  Senior R&D Engineer
- **CFSI (credit collateralization)**
  CIO / Head of IT

Academic

- **PhD (ABD)**
- **MBA (OneMBA)**
- **MA Financial Management**
- **Certificate Finance**
- **GD IT Management**
- **MA Computer & Communications Technology**

YouTube

- *Introduction to Advanced Analytics*
- *Introduction to Cognitive Analytics*
- *TedX RSM: Data Analytics*

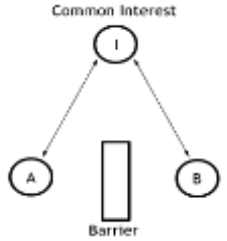
AGENDA

Data Science for Cybersecurity Risk

- Why?
- Cybersecurity risk management
  - Challenges
  - Data-driven approach
- Data Science (DS)
- DS for cybersecurity risk
  - Focused examples
  - Learnings from the field
  - Risk management solutions

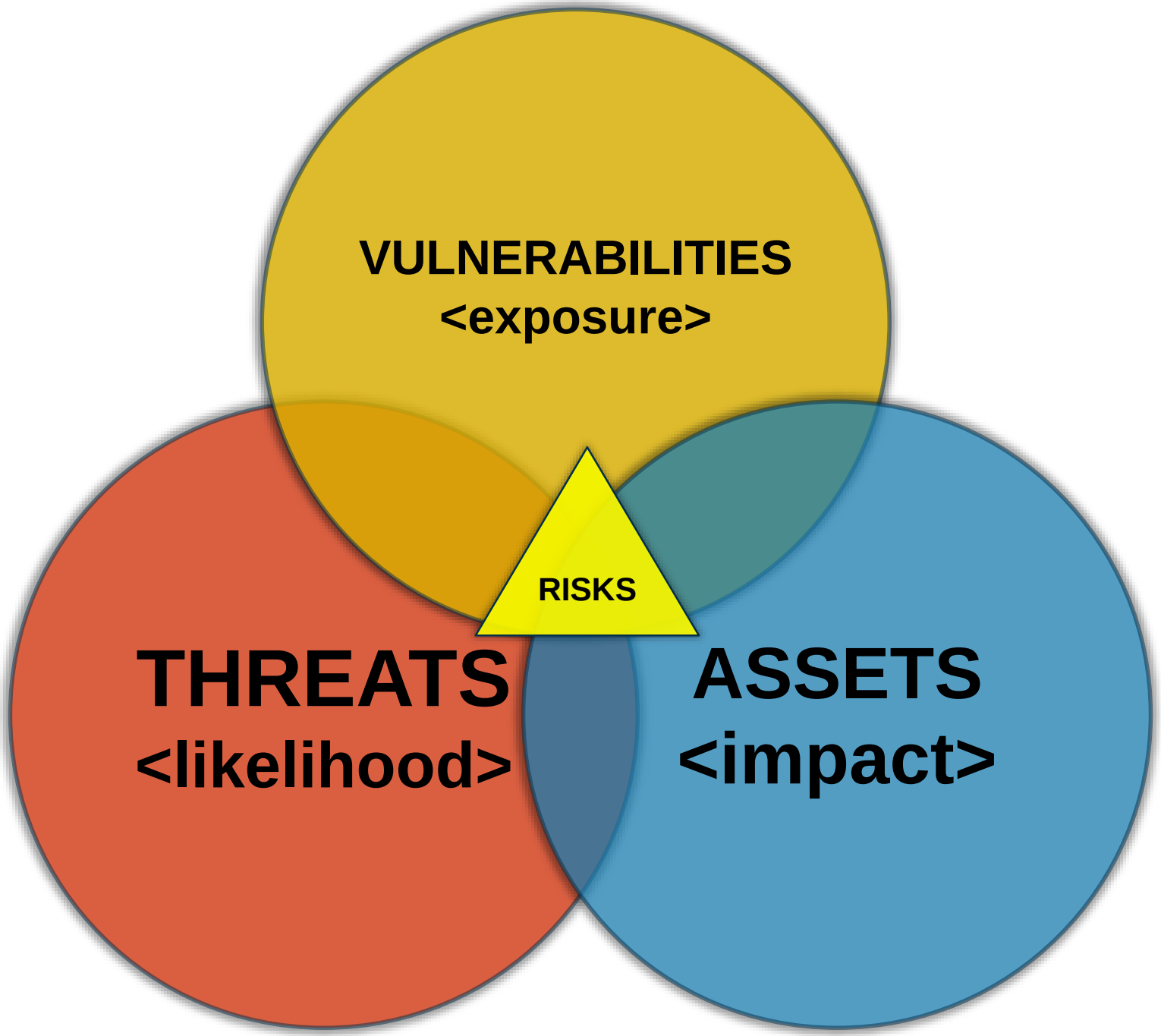
Multidisciplinary Merge

Challenges bringing common interests together

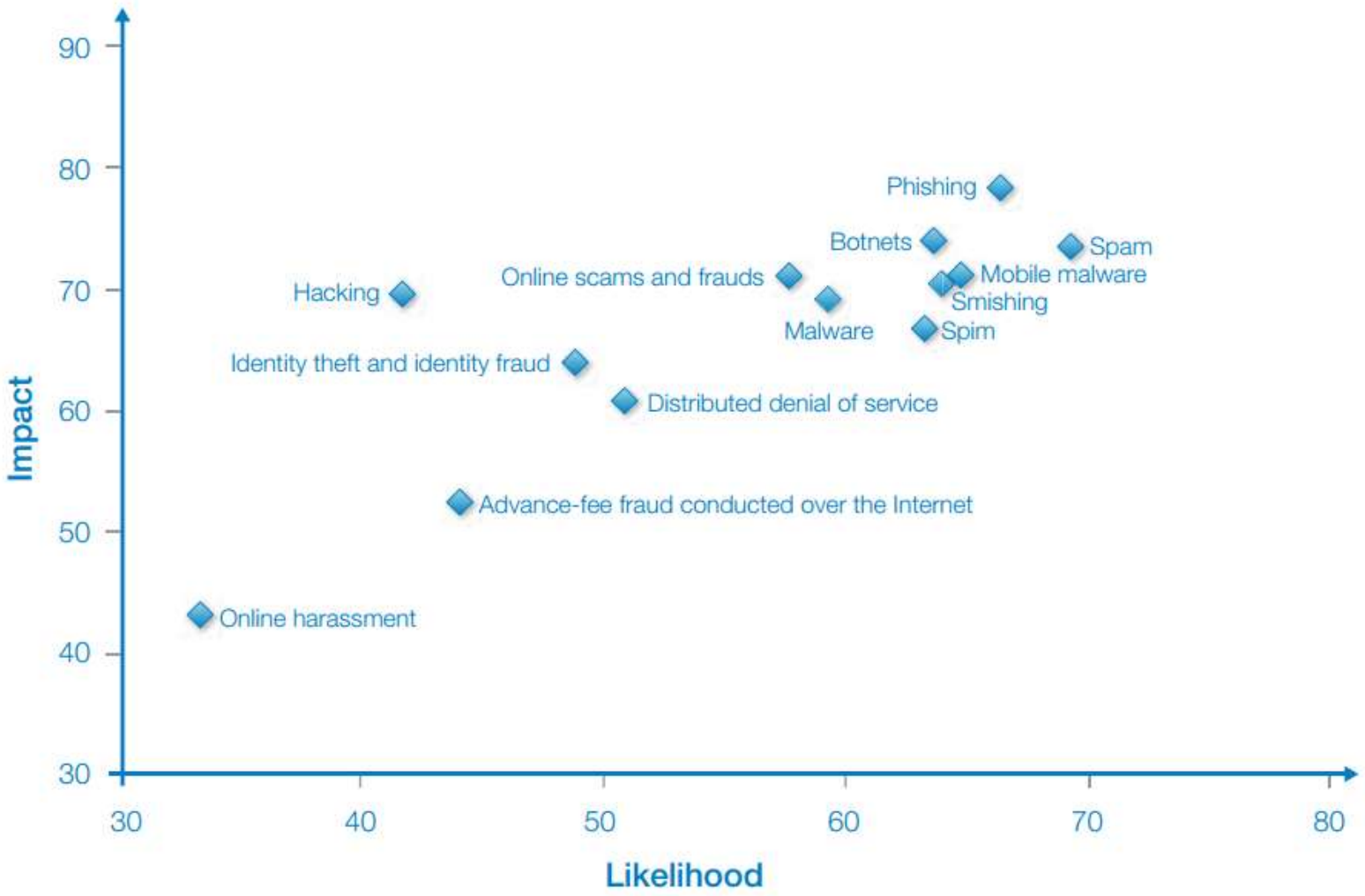


DOMAIN	FOCUS	CHALLENGE
Cybersecurity	Protect network infrastructure and resources	Many unknown-unknowns
Risk Management	Identify probability and impact to control risks/opportunities	Uncertainty gap in cybersecurity domain (avoid, reduce, share, or retain risks?)
Data Science	Application of a range of methods to extract insights from data	Clarifying best practices to support cyber risk management

Cybersecurity Risk Management

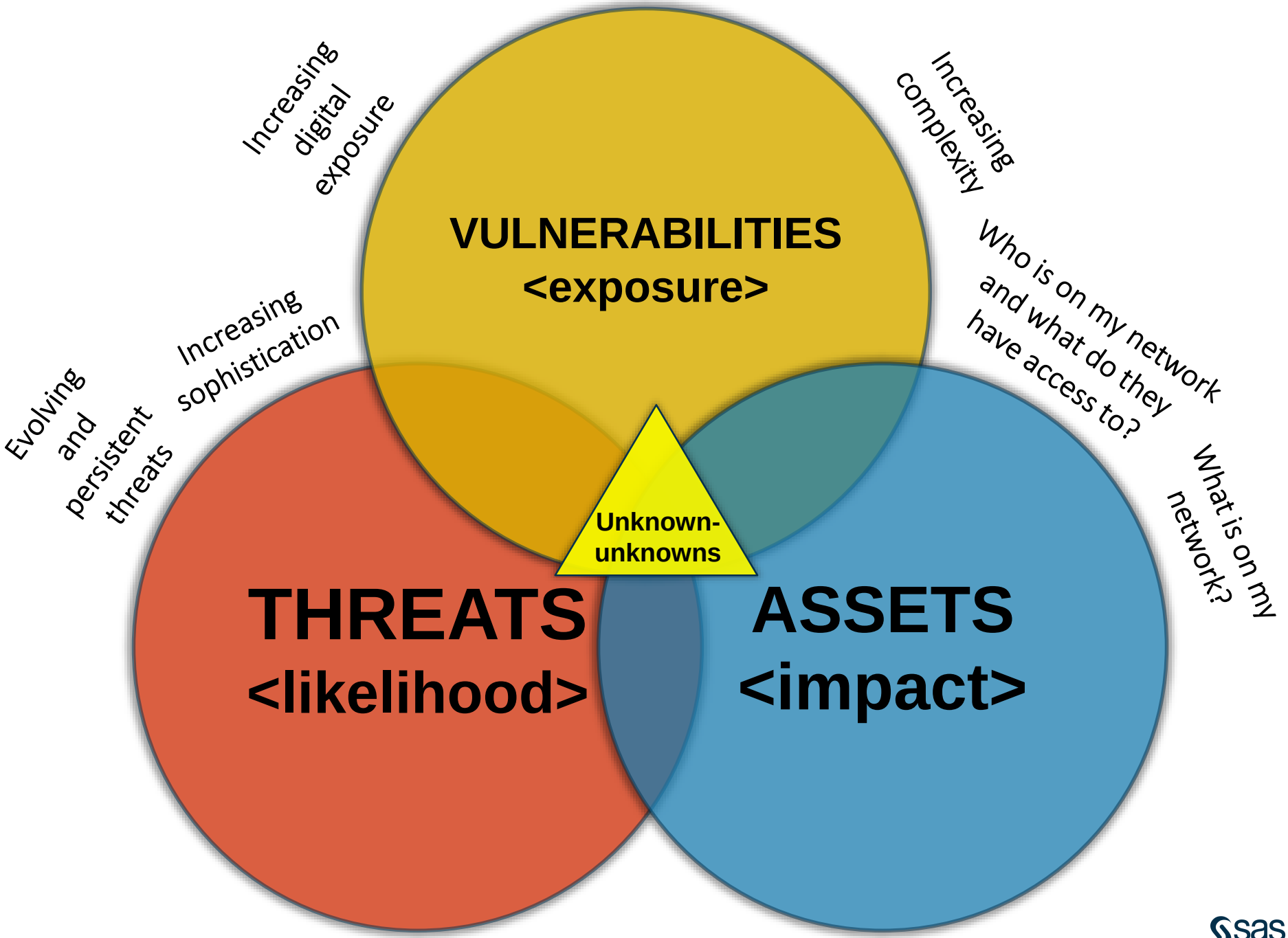


Cyber Incidents: Likelihood and Impact



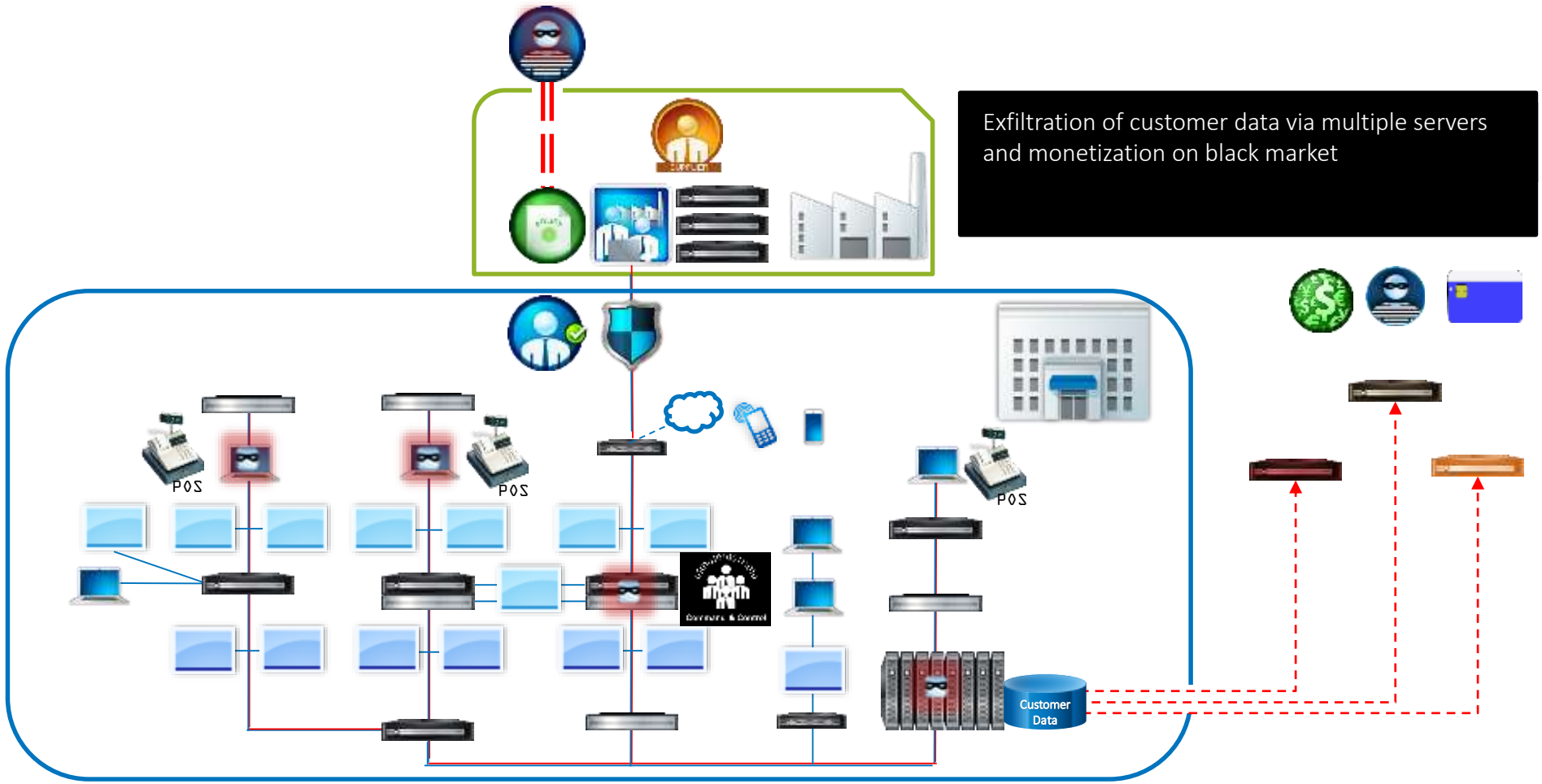
SOURCE

Longitude Research. (2014). "Cyberrisk in banking." Available at http://www.longitude.co.uk/wp-content/uploads/2015/02/cyberrisk-in-banking-106605.pdf



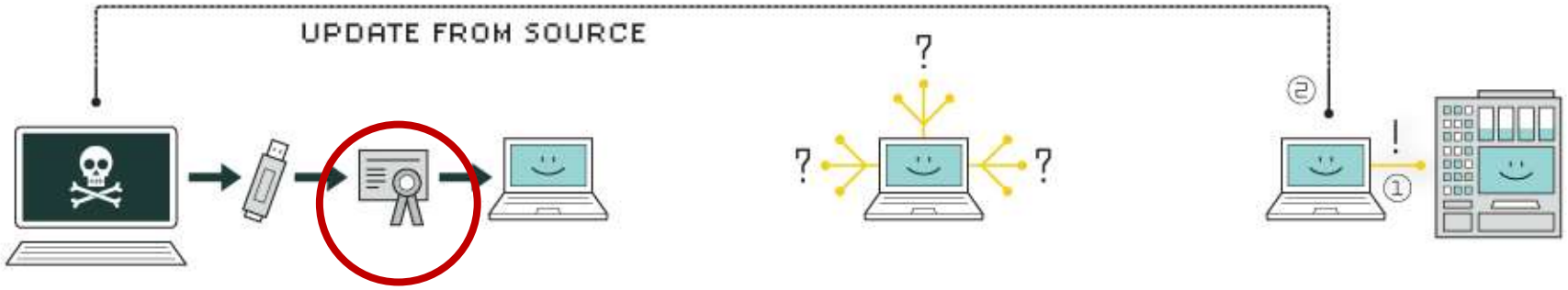
Anatomy of a Sophisticated Attack: Target retailer POS breach (2013)

Hybrid social engineering multi-layered, multi-phased attack



Deep-incursion: STUXNET (Duqu/Flame) (2010)

State-sponsored sophisticated multi-phased worm attack



1. infection
Stuxnet enters a system via a USB stick and proceeds to infect all machines running Microsoft Windows. By brandishing a digital certificate that seems to show that it comes from a reliable company, the worm is able to evade automated-detection systems.

2. search
Stuxnet then checks whether a given machine is part of the targeted industrial control system made by Siemens. Such systems are deployed in Iran to run high-speed centrifuges that help to enrich nuclear fuel.

3. update
If the system isn't a target, Stuxnet does nothing; if it is, the worm attempts to access the Internet and download a more recent version of itself.

4. compromise
The worm then compromises the target system's logic controllers, exploiting "zero day" vulnerabilities-software weaknesses that haven't been identified by security experts.

5. control
In the beginning, Stuxnet spies on the operations of the targeted system. Then it uses the information it has gathered to take control of the centrifuges, making them spin themselves to failure.

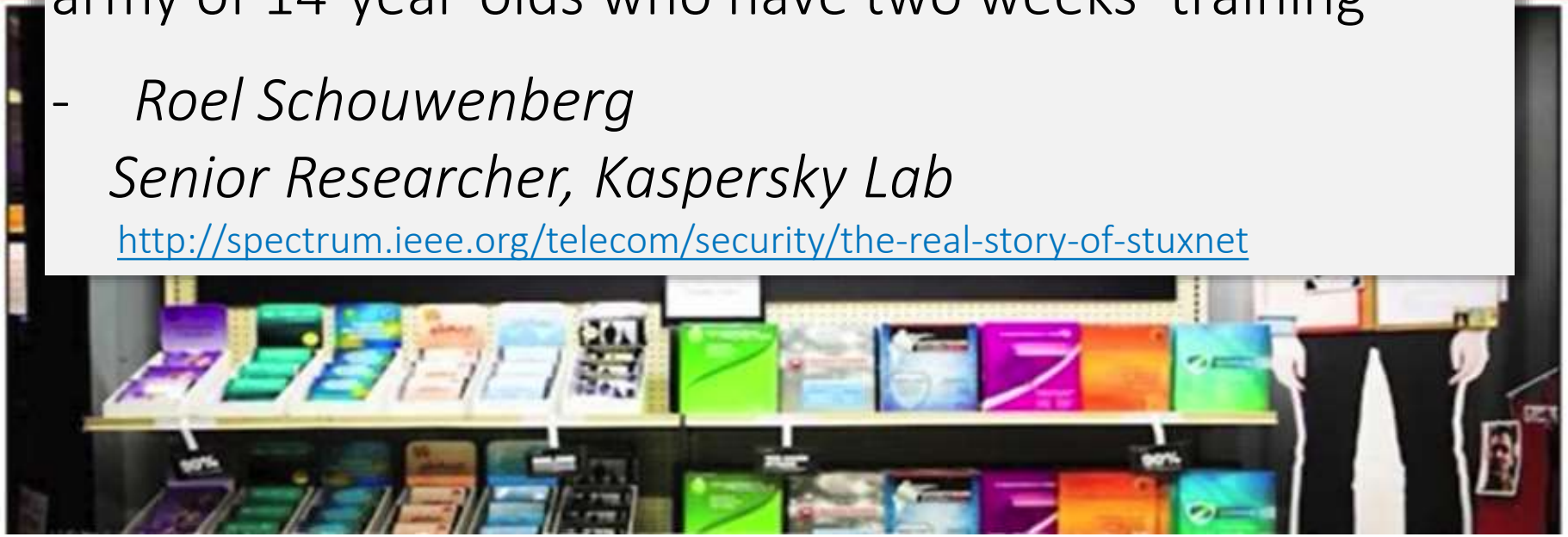
6. deceive and destroy
Meanwhile, it provides false feedback to outside controllers, ensuring that they won't know what's going wrong until it's too late to do anything about it.

SOURCE: IEEE Spectrum http://spectrum.ieee.org/telecom/security/the-real-story-of-stuxnet

DISTRIBUTE, WHOLESALE, RESELLERS……

"There's a lot of talk about nations trying to attack us, but we are in a situation where we are vulnerable to an army of 14-year-olds who have two weeks' training"

- *Roel Schouwenberg*
  Senior Researcher, Kaspersky Lab
  http://spectrum.ieee.org/telecom/security/the-real-story-of-stuxnet




The Internet
(more or less)
Digg
reddit
LiveLeak
tumblr.
random wikis and blogs
Gaia
808chan, 7chan, 420chan
2channe
4chan
sTORage
KindzazaChan
PedoPlanet
Hard Candy
TorChan
Hidden Wiki
Tor P.M.
OnionChan
OnionIB
Internet
Deep Web
Darknet

NOTICE OF EXTORTION

Your business, [redacted], has been targeted for extortion. The selection process is random, and was not triggered by any event under your control.

Should you fail to pay the one-time monetary tribute, by the deadline provided below, your business will be **severely and irreparably damaged**. The following methods are commonly employed in cases of non-compliance:

- Negative Online Reviews
- BBB Complaints
- Harassing Telephone Calls
- Fraudulent Delivery Orders
- Telephone Denial-of-Service
- Bomb Threats
- Vandalism
- Mercury Contamination

Anonymous Reports of:

- Health Code Violations
- OSHA Violations
- Criminal Tax Evasion
- Money Laundering
- Illegal Drug Sales
- Marijuana Grow Operations
- Methamphetamine Production
- Terrorist Training Activity

The tribute price is only One Bitcoin (1 BTC), but must be paid by **August 15, 2014**. Payment is to be made to the Bitcoin Wallet Address listed below.

If payment is not received, our team will begin taking the actions listed above. Once engagement has begun, it can only be stopped for a tribute of Three Bitcoin (3 BTC). Because many of the actions we take are catastrophic and irreversible, is it advised to pay the tribute before the deadline is reached.

Payment Type: Bitcoin

Deadline: August 15, 2014

Amount Due: (If paid before deadline) 1 Bitcoin

Amount Due: (If paid after deadline) 3 Bitcoin

Purchase Bitcoin @
https://www.coinbase.com/

17gt1BancvtnnJwy4BA41VBUH3pfbUvzE

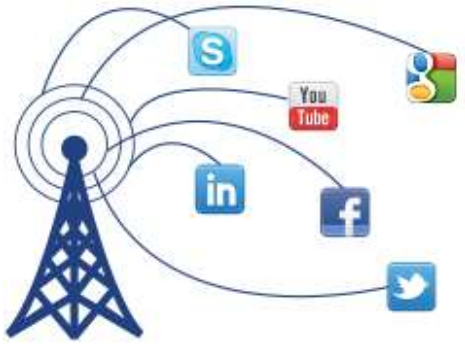
Impact from Cyber Incidents

TANGIBLE

- Destruction of infrastructure
- Incident handling costs
- Lost customers / clientele
- Legal judgements
- Regulatory fines
- Rectification of vulnerabilities

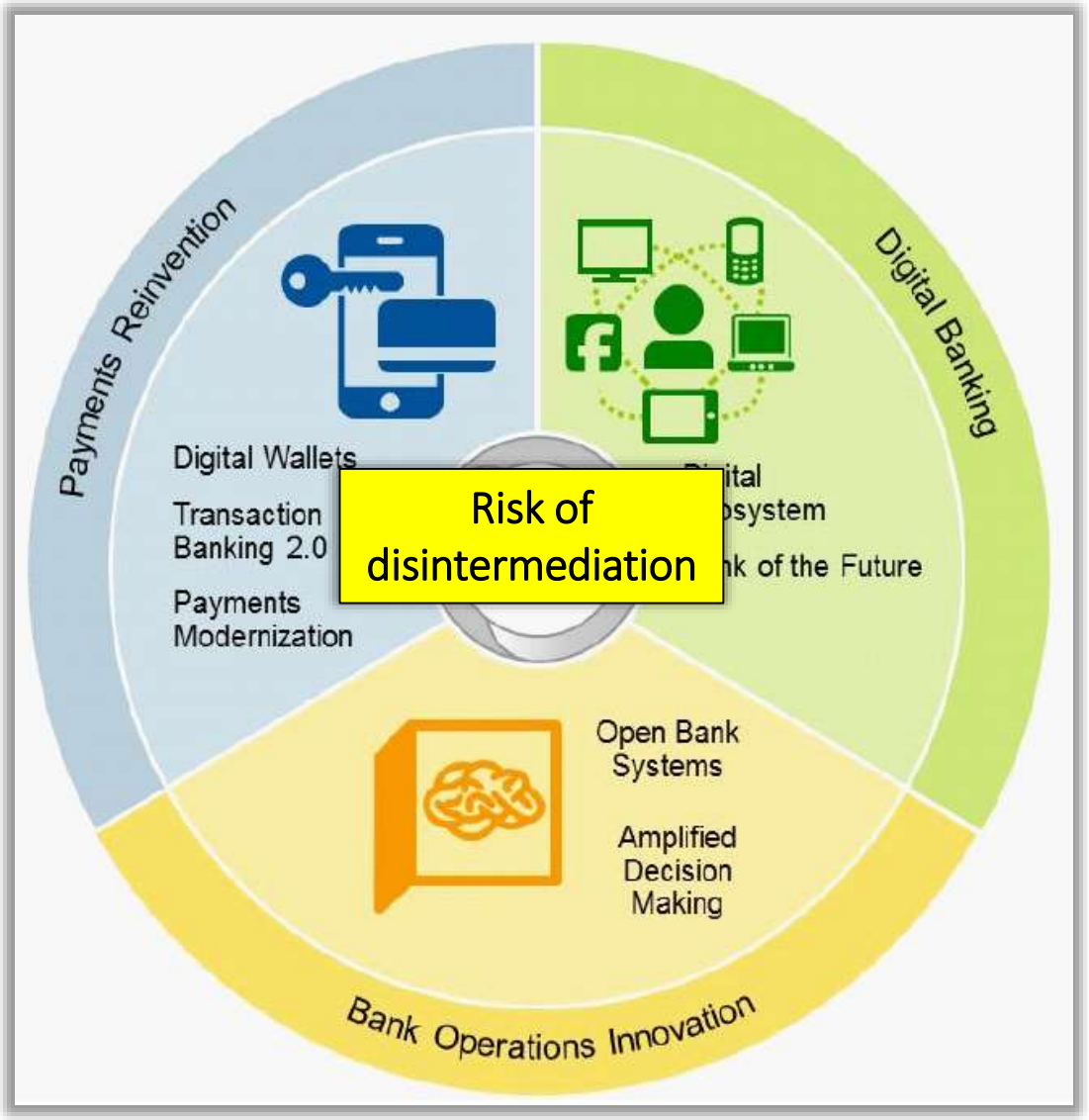
INTANGIBLE

- Loss of trust (customers, partners)
- Impact on strategic market position and share price
- Damage to reputation & brand
- Impact to credit rating



Reactive militarization…

Competitive ‘digital innovation’ pressures



Source:
Gartner. 2015. Agenda Overview for Banking and Investment Services.

Expanding digital offerings:

- Increasing access
- Complexity of systems
- Greater volumes of data

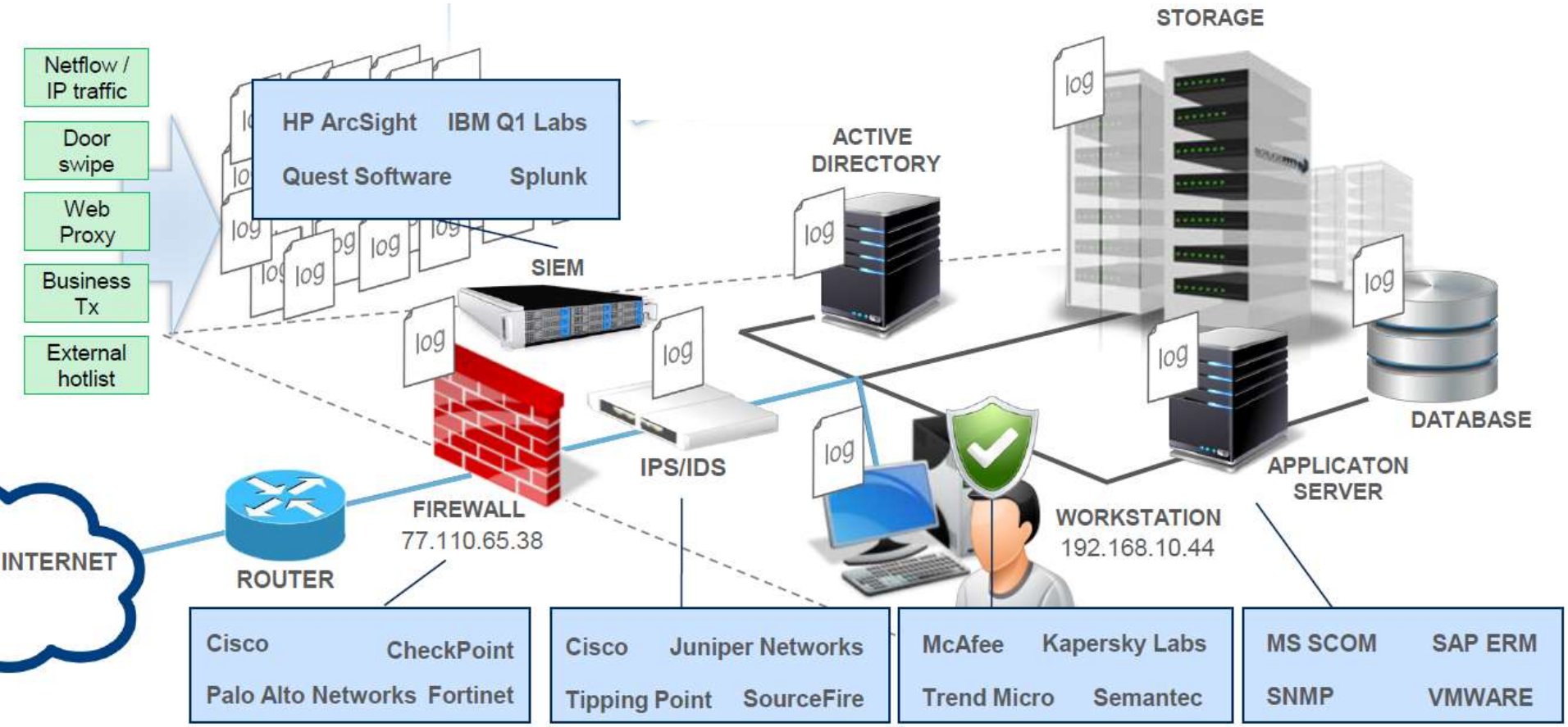
Proliferation of devices:

- BYOD
- VMs / containers
- IoT / smart devices
- ICS SCADA

In Search of: Targeted, Relevant, Actionable Alerts…

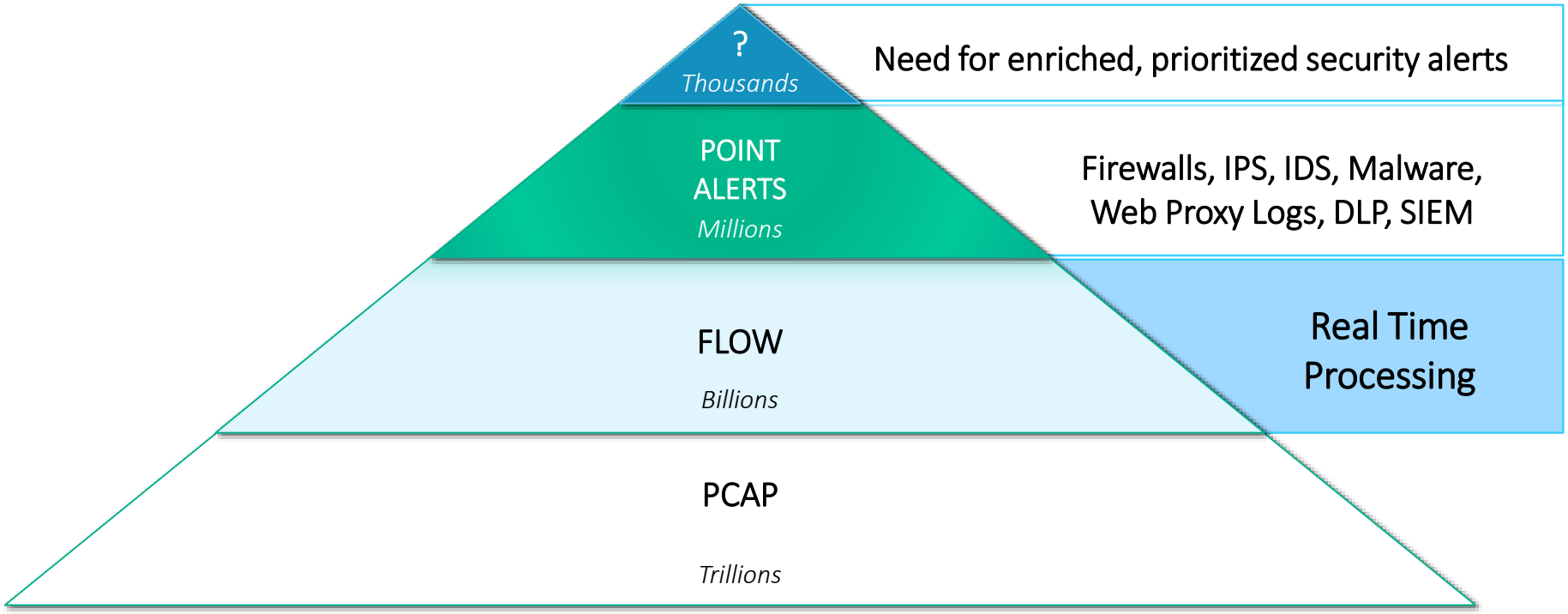
Data-Driven Cyber Risk Mangement

Many data sources… increasing data volume



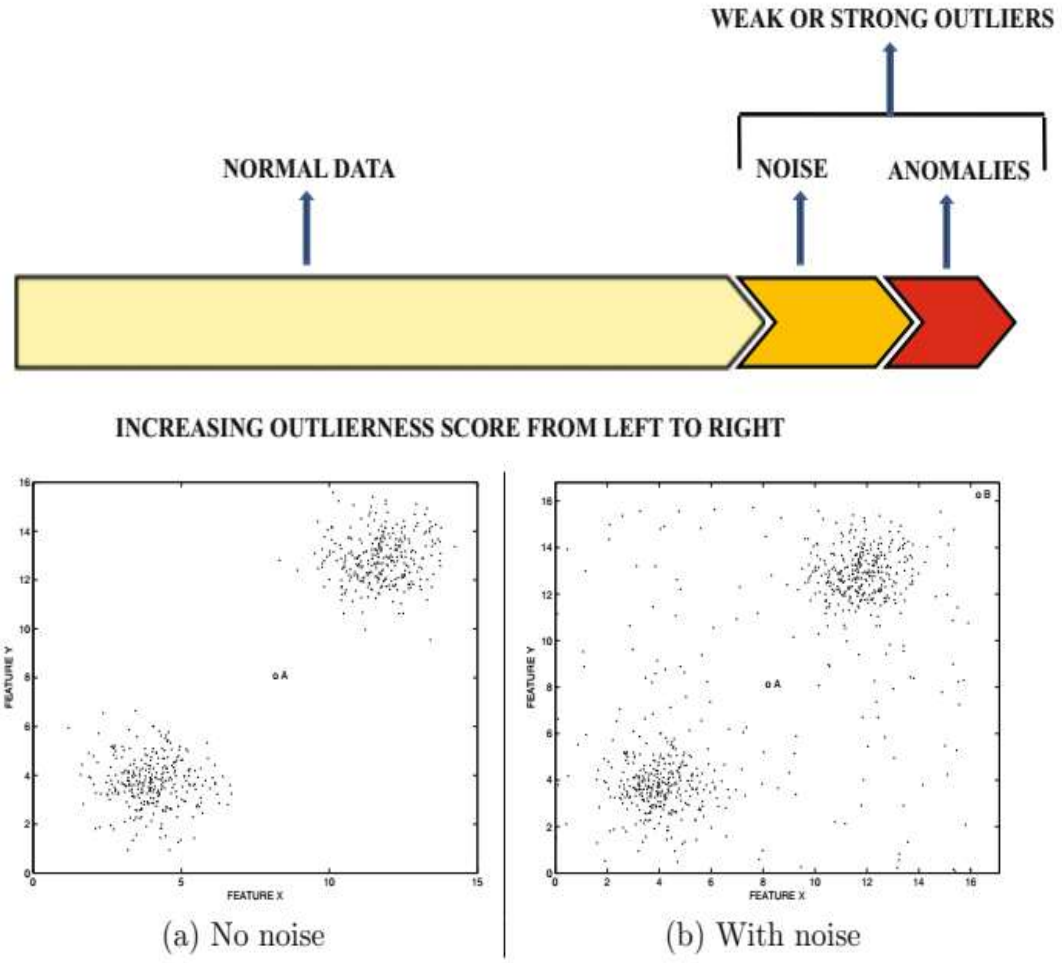
Source
Cyber Security Solutions, 2014.

Cyber Data Types and Monthly Volumes



Simply Complex

Identifying targeted anomalies amongst and ocean of noise…



SOURCE

Aggarwal, Charu C. (2017). "Outlier Analysis: Second Edition". Springer International Publishing AG.

Data Science for Cybersecurity

CHALLENGES

Data overload

Disconnected &
low quality data

High false positive alerts

Unknown unknowns –
no baseline

Slow and manual investigation
processes

CYBER DATA SCIENCE

Focused insights from Big Data

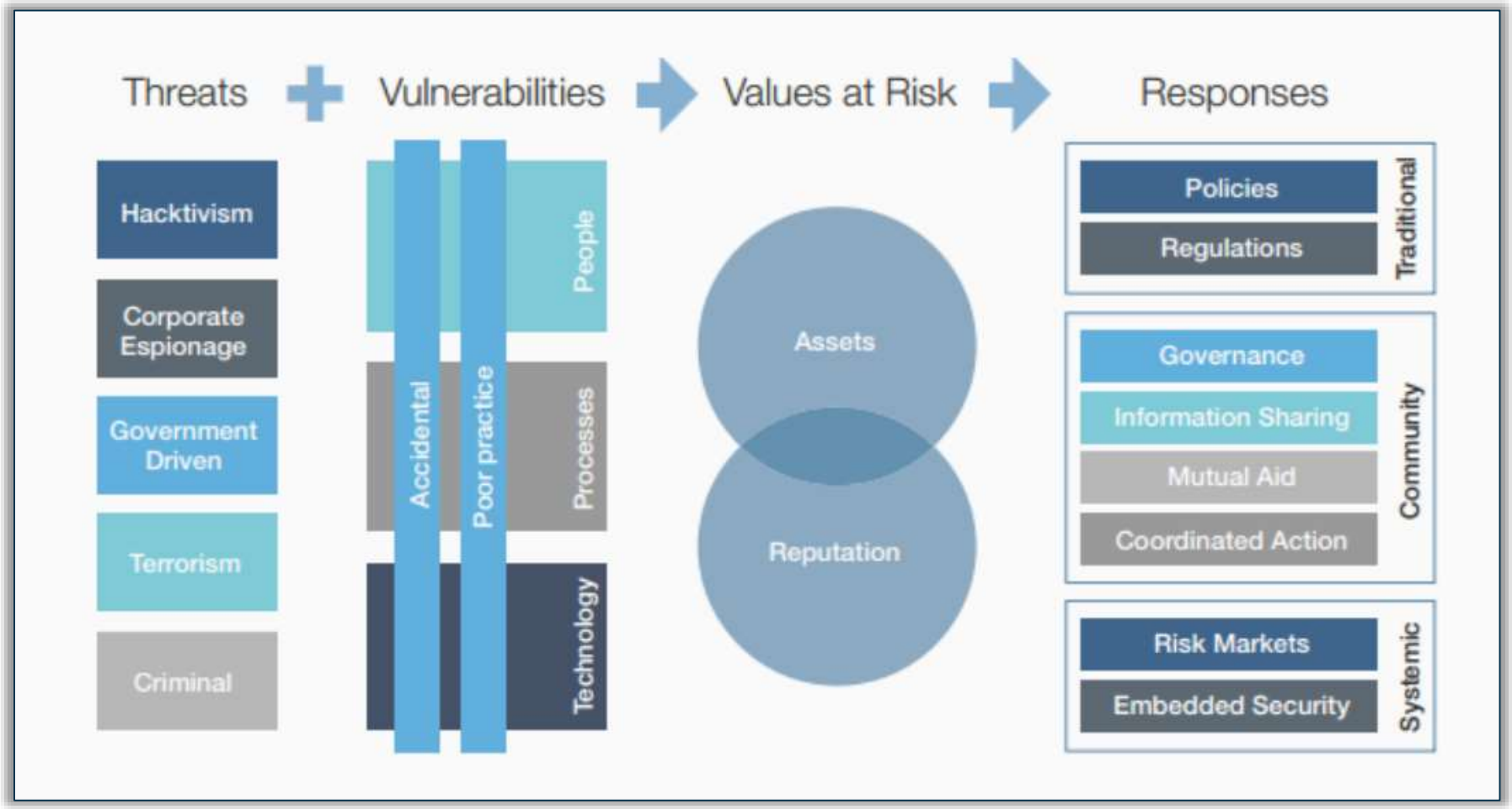
Managing and rationalizing data

Machine learning identifies
hidden patterns

Diagnostics for understanding
'normal'

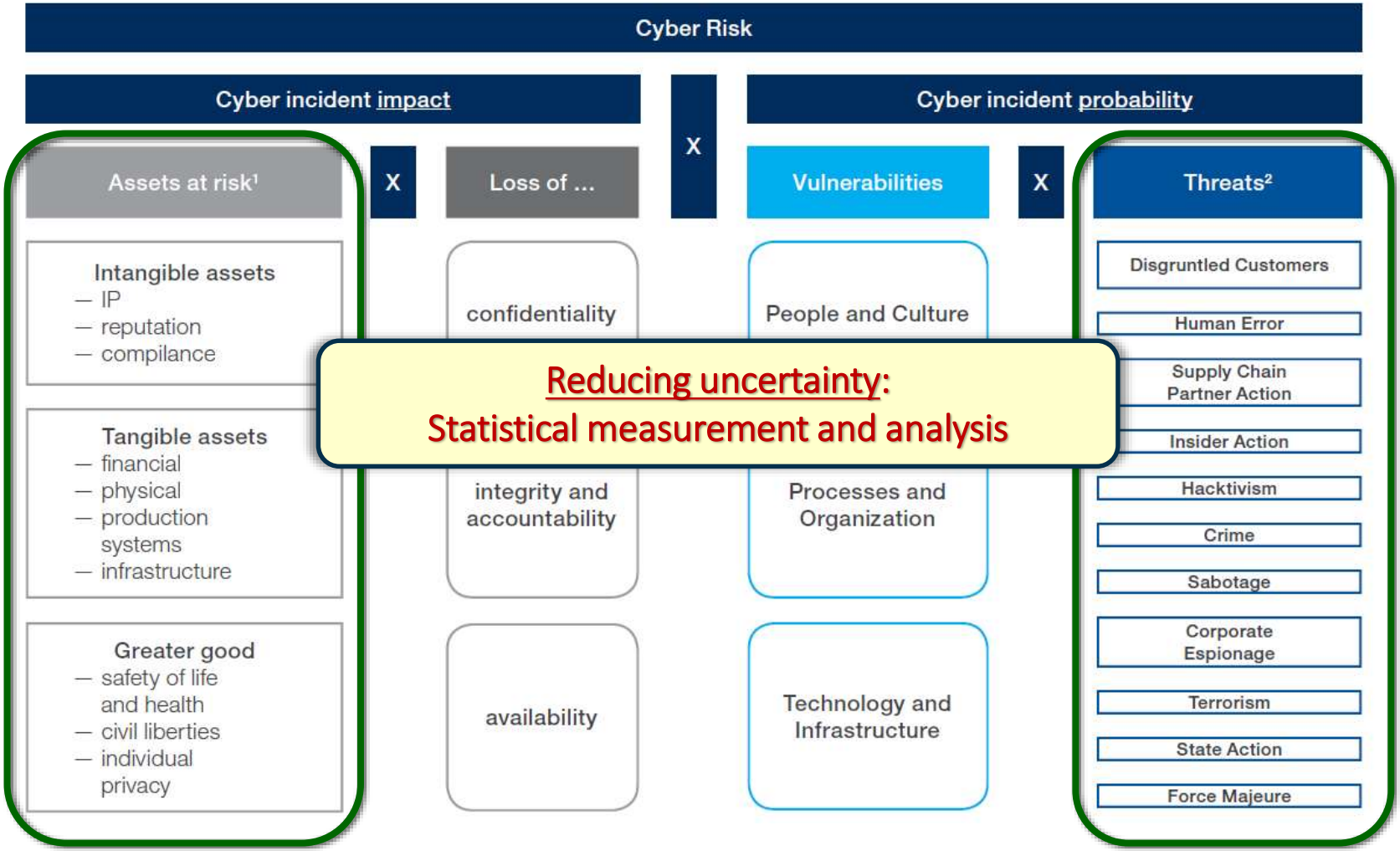
Targeted alerts based on
anomalies

Data Science => Uncertainty Reduction



SOURCE
Partnering for Cyber Resilience: Towards the Quantification of Cyber Threats
WEF report in collaboration with Deloitte:
http://www3.weforum.org/docs/WEFUSA_QuantificationofCyberThreats_Report2015.pdf

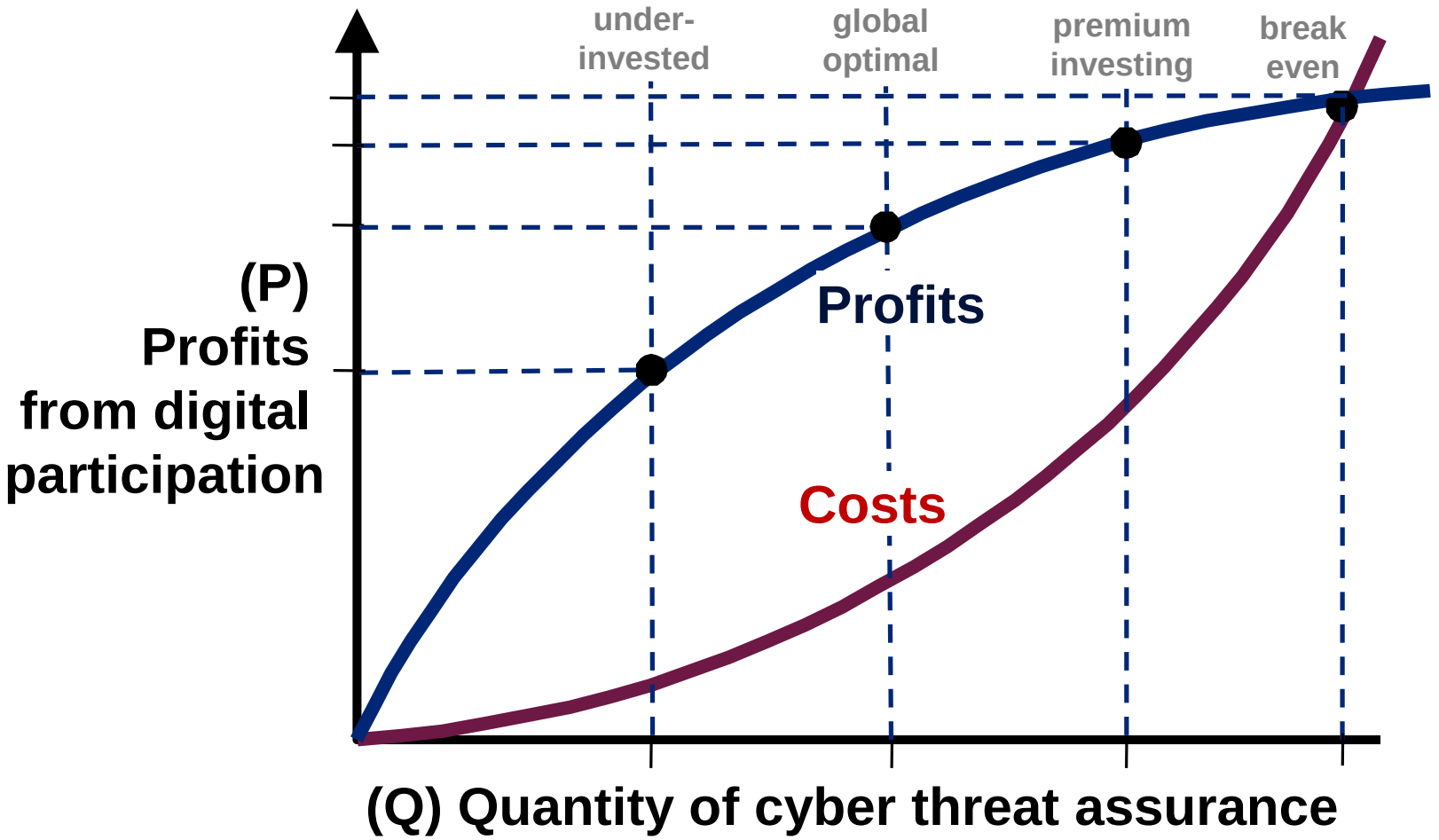
Data Science => Measurement



Advancing Cyber Resilience Principles and Tools for Boards
http://www3.weforum.org/docs/IP/2017/Adv_Cyber_Resilience_Principles-Tools.pdf

Optimizing Accessibility Versus Exposure

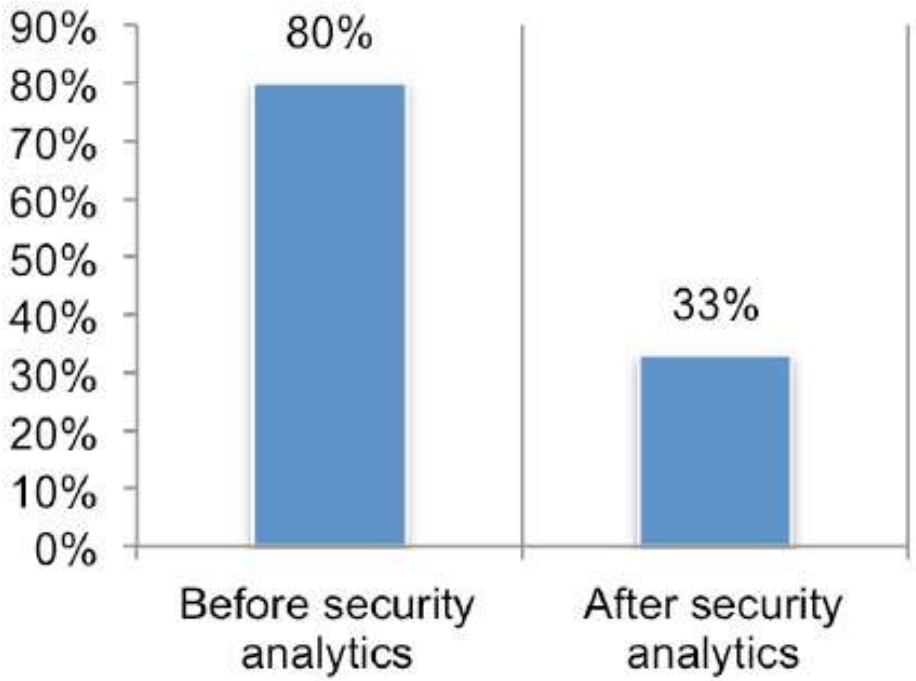
Invest to point of optimality



SOURCE

Partnering for Cyber Resilience: Towards the Quantification of Cyber Threats
WEF report in collaboration with Deloitte:
http://www3.weforum.org/docs/WEFUSA_QuantificationofCyberThreats_Report2015.pdf

Level of difficulty in reducing false alerts*



** Survey of 621 global IT security practitioners*

https://www.sas.com/en_us/whitepapers/ponemon-how-security-analytics-improves-cybersecurity-defenses-108679.html

Poll Question 1

1. At what stage are you in deploying a cybersecurity analytics program?

a) Not planning
b) Planning in next 3 to 12 months
c) Planning in next 12 to 24 months
d) Implementation underway
e) Have completed

Sponsored by sas

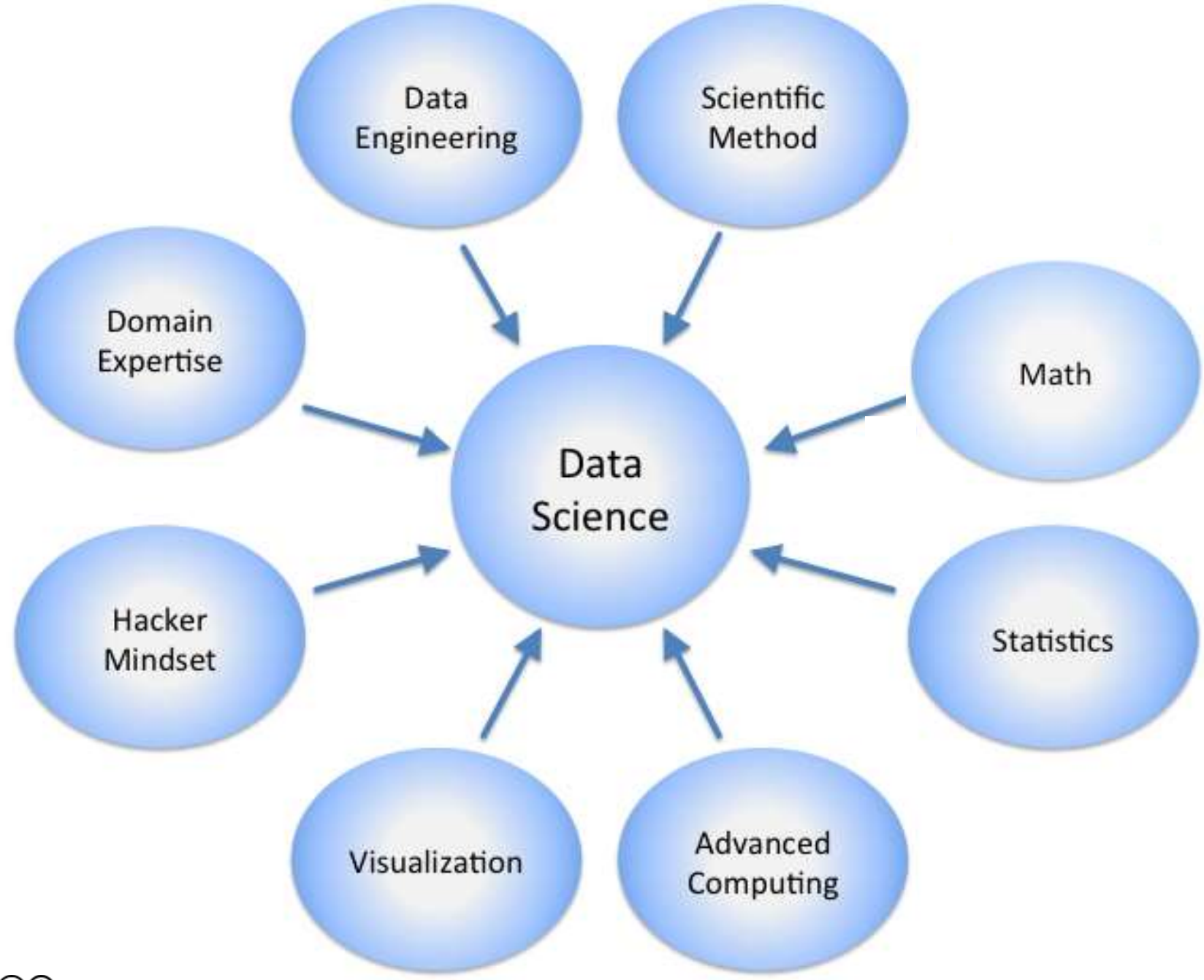
At what stage are you in your security analytics program?

We are not planning an analytics program	36%
We are planning to implement in the next 6 months	12%
We are planning to implement in the next 6-12 months	15%
We are planning to implement in the next 12-24 months	13%
We are in the middle of implementing our program	11%
We have completed our implementation	14%

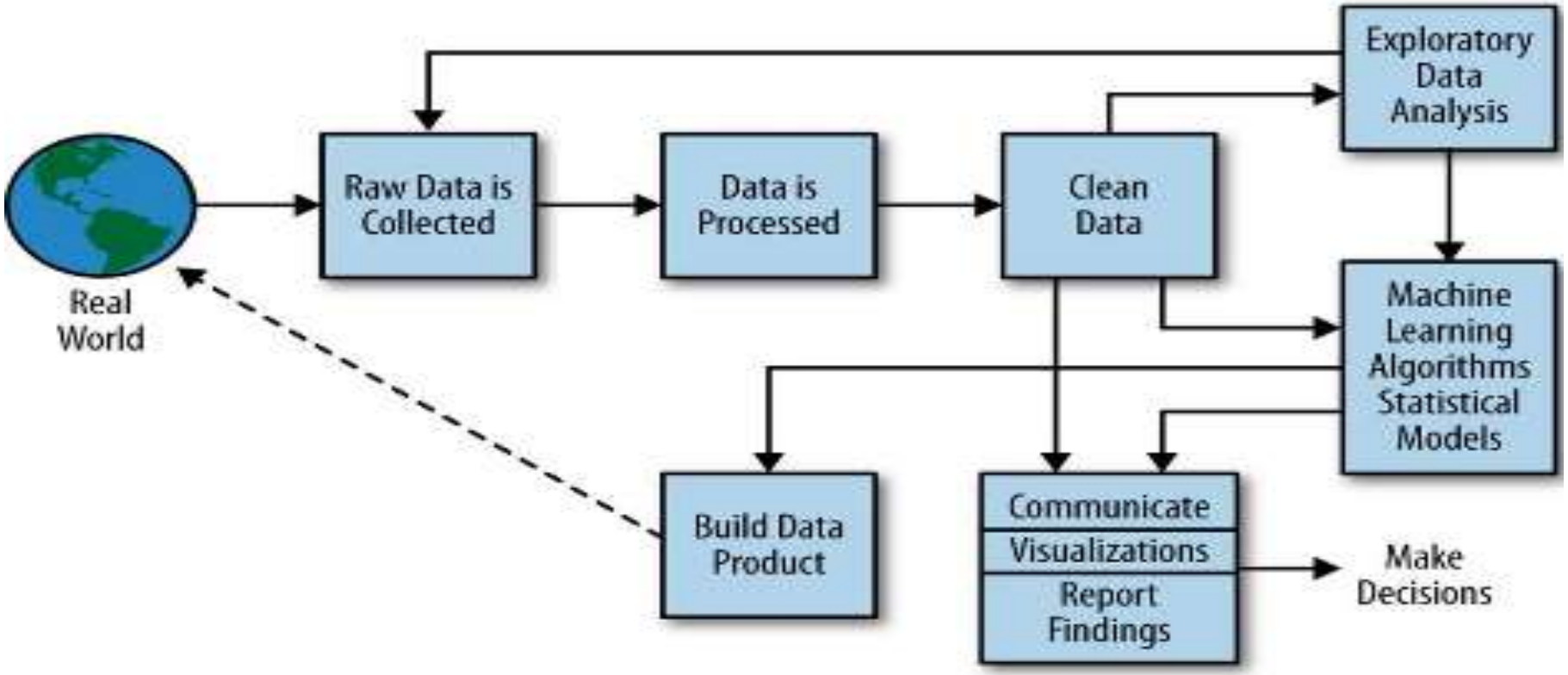
SOURCE

Security Brief Magazine. (2016). "Analyze This! Who's Implementing Security Analytics Now?" Available at https://www.sas.com/en_th/whitepapers/analyze-this-108217.html

Overview
Data Science



 Calvin.Andrus (2012) http://en.wikipedia.org/wiki/File:DataScienceDisciplines.png

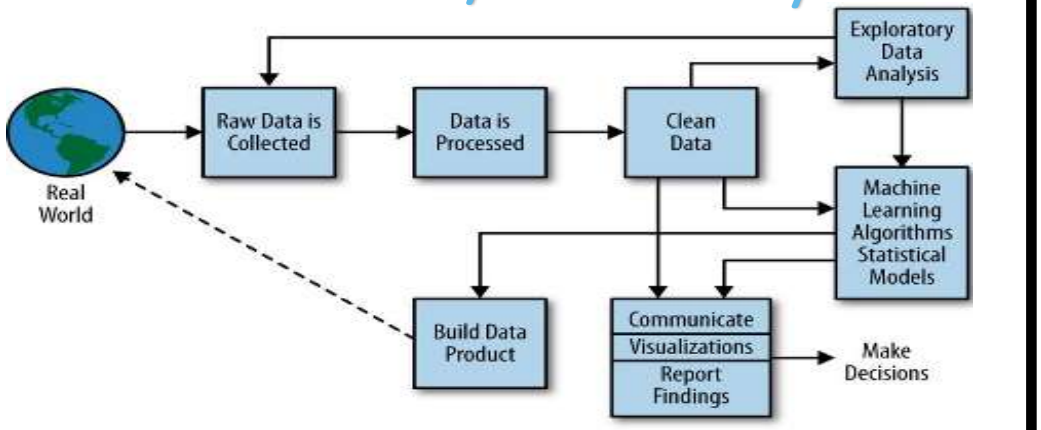


Schutt, Rachel; O'Neil, Cathy (2013).
Doing Data Science: Straight Talk from the Frontline. O'Reilly Media.

Business (Enterprise) Analytics

- Stakeholder alignment
- Goals
- Problem framing
- Uncertainties
- Risks

Data Science / Data Analytics



- Modelling
- Testing
- Validation
- Communicate
- Maintenance

organizational process

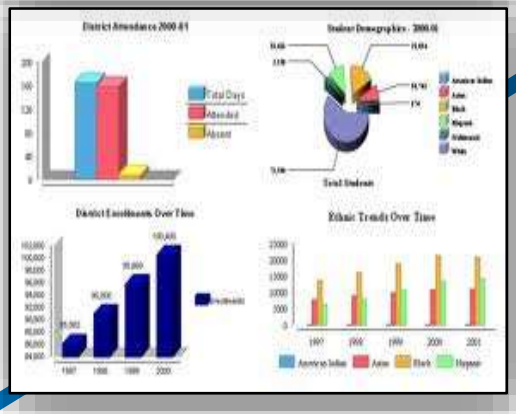
Historical View

SOPHISTICATION

VALUE

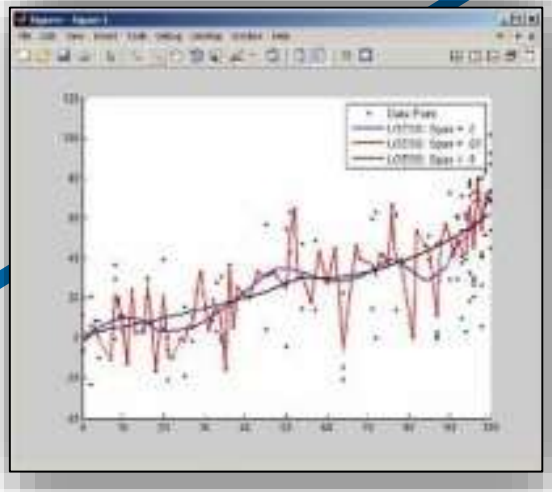
What happened?

DESCRIPTIVE



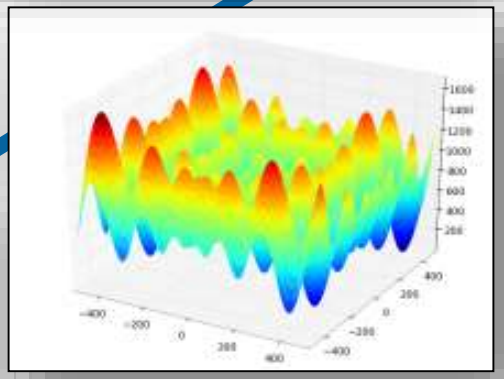
What are trends?

PREDICTIVE



What to do?

PRESCRIPTIVE



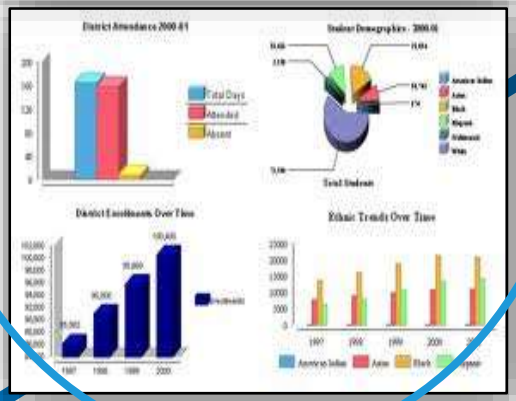
Historical View

SOPHISTICATION

VALUE

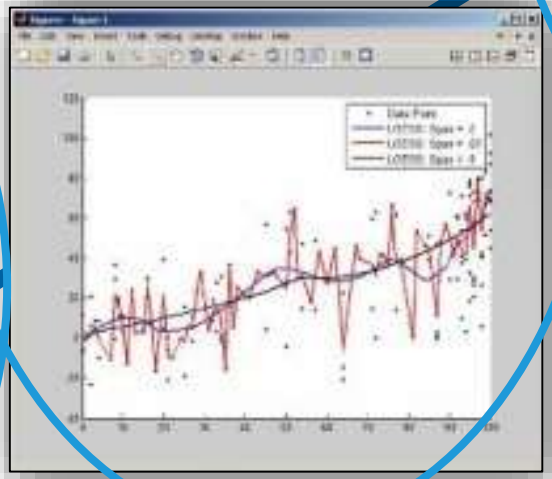
Business Intelligence (BI)

DESCRIPTIVE



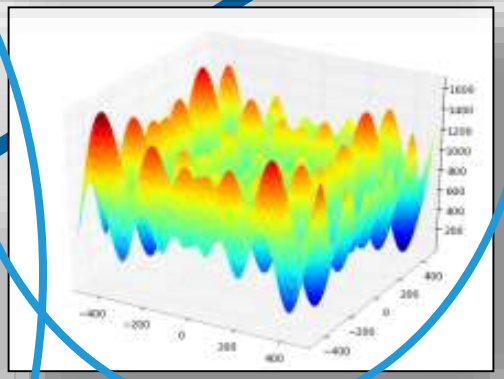
Econometrics Forecasting Machine Learning

PREDICTIVE



Operations Management

PRESCRIPTIVE

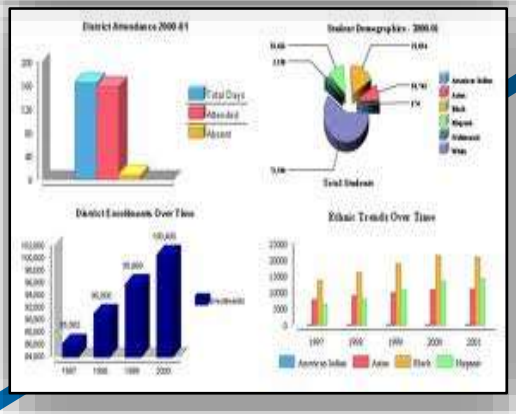


SOPHISTICATION

VALUE

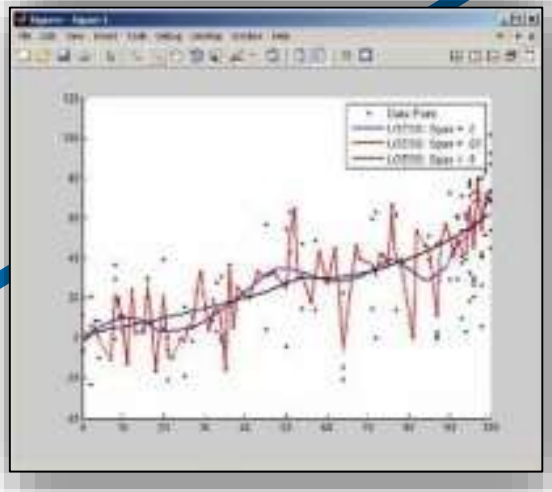
What happened?

DESCRIPTIVE



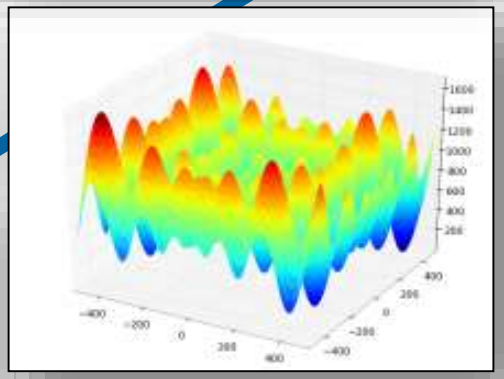
What are trends?

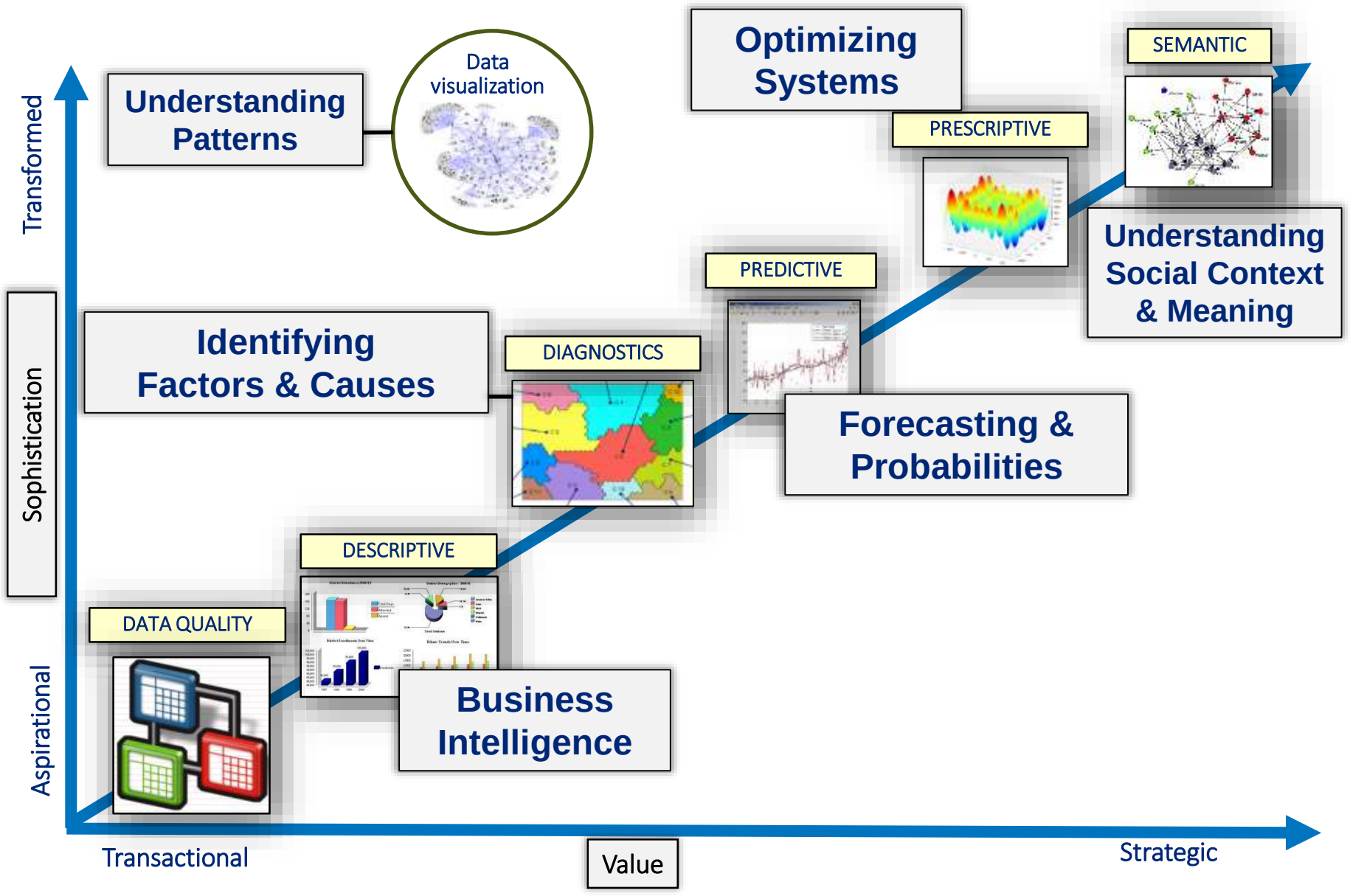
PREDICTIVE



What to do?

PRESCRIPTIVE





Data Science for Cybersecurity Risk Analysis

Poll Question 2

2. What is the most important objective for applying cybersecurity analytics?

a) Detect events in progress

b) Determine root cause of past events (forensics)

c) Provide advanced warning of potential internal threats and attackers

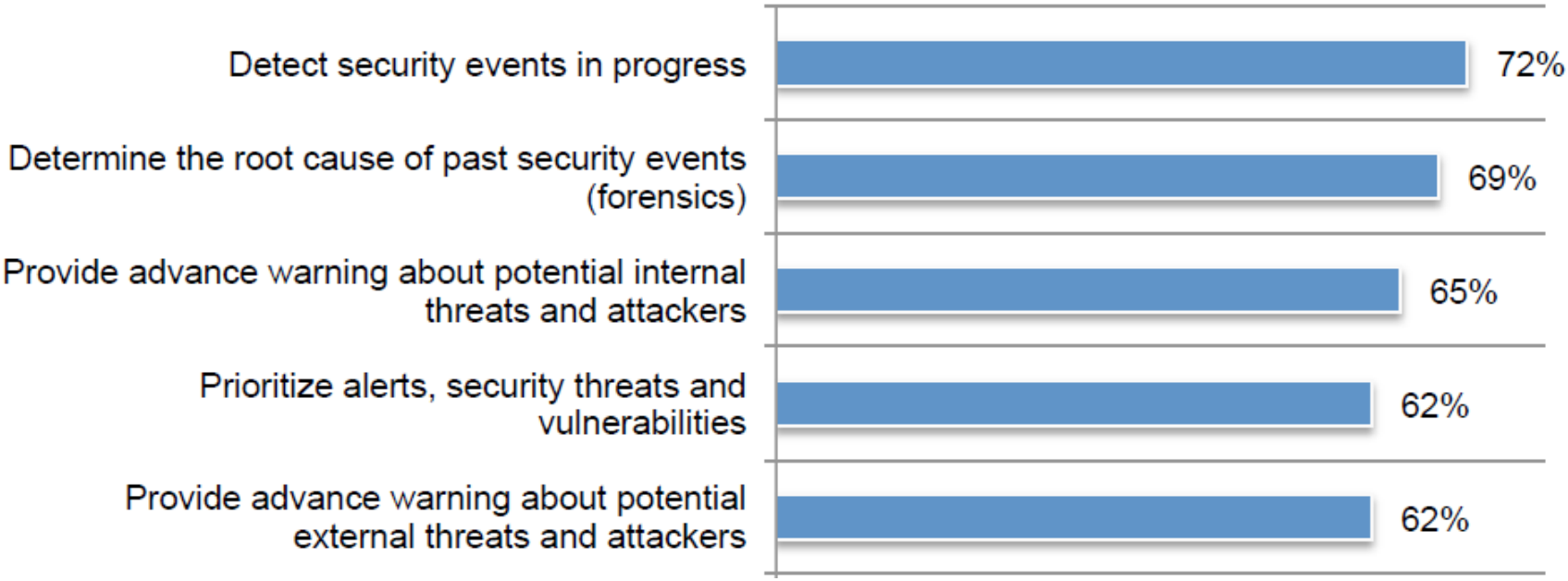
d) Prioritize alerts, security threats, and vulnerabilities

e) Provide advanced warning about potential external threats and attackers

Ponemon INSTITUTE

When Seconds Count: How Security Analytics Improves Cybersecurity Defenses

Most important objectives for a cybersecurity analytics solution*

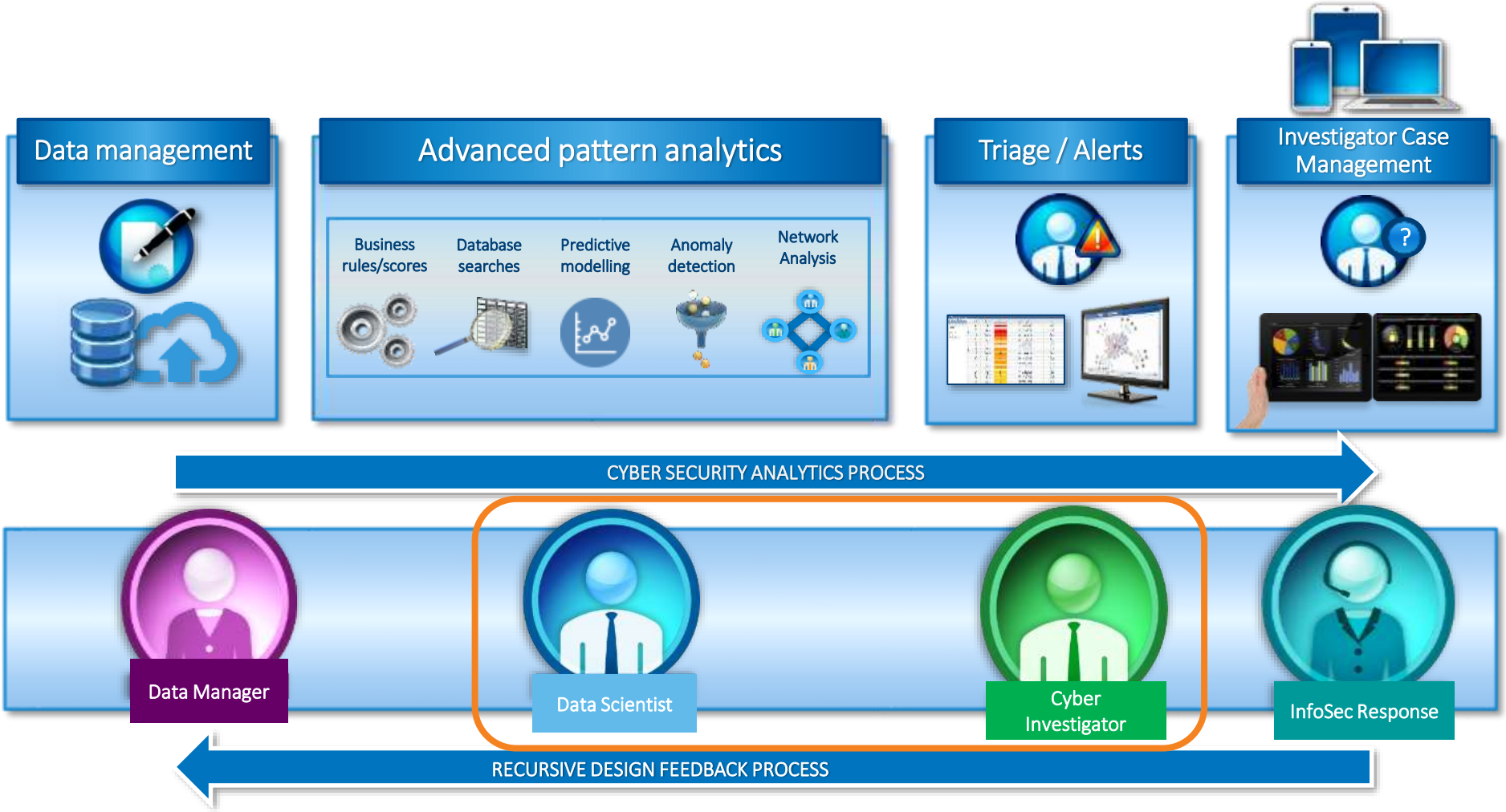


https://www.sas.com/en_us/whitepapers/ponemon-how-security-analytics-improves-cybersecurity-defenses-108679.html

** Survey of 621 global IT security practitioners*

sas

Enterprise Cyber Security Data Science

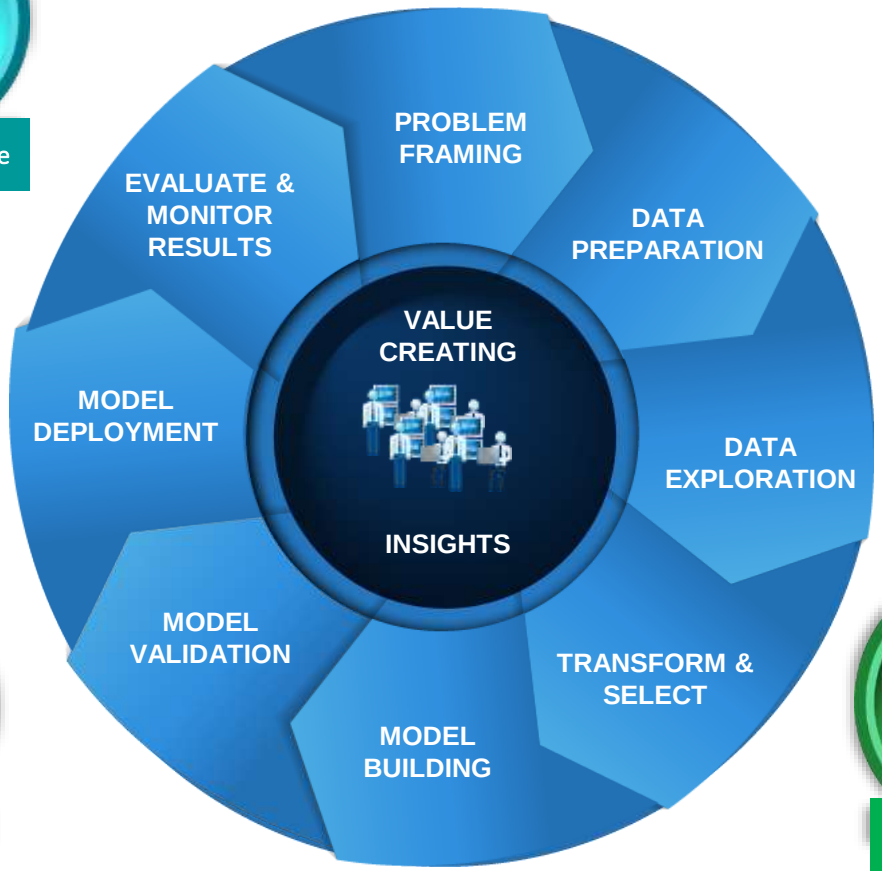


Cyber Data Science Lifecycle

InfoSec Response

Data Manager

Pattern Detection



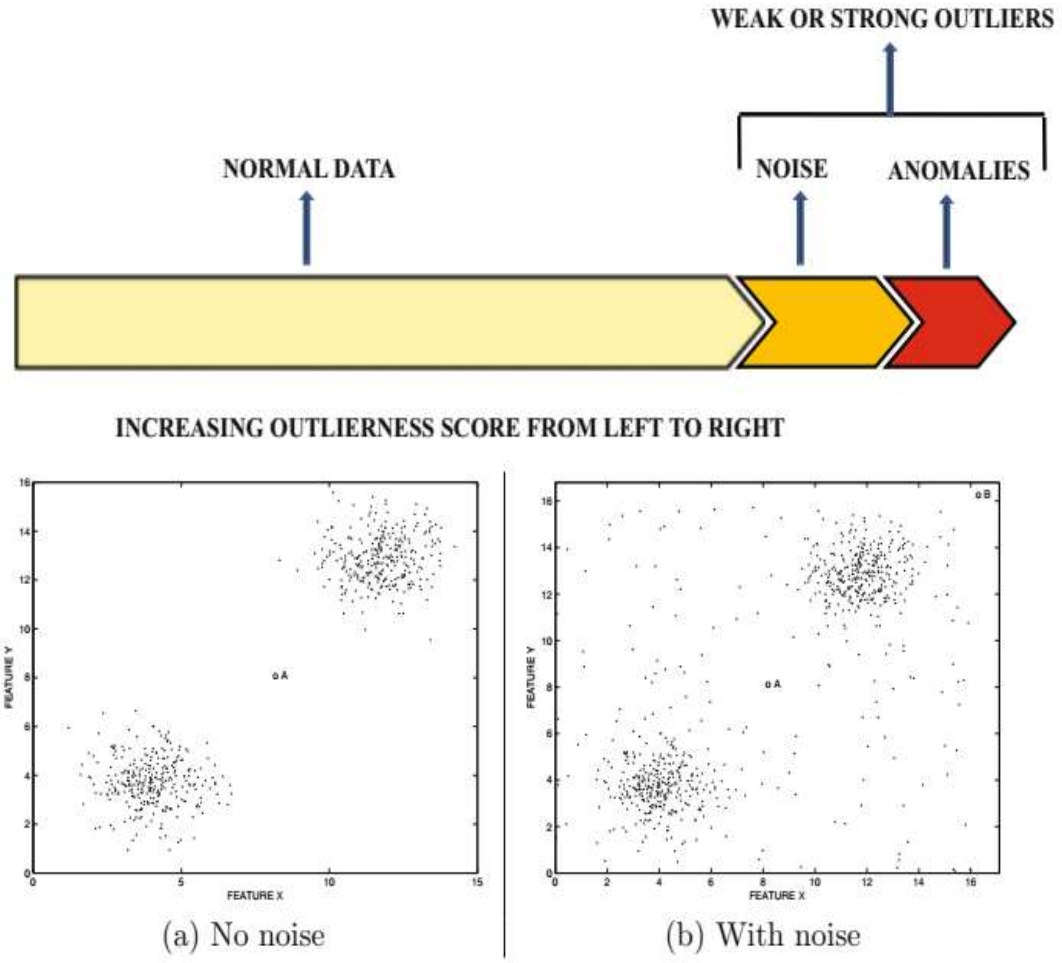
Exploration and Insights

Data Scientist

Cyber Investigator

Simply Complex

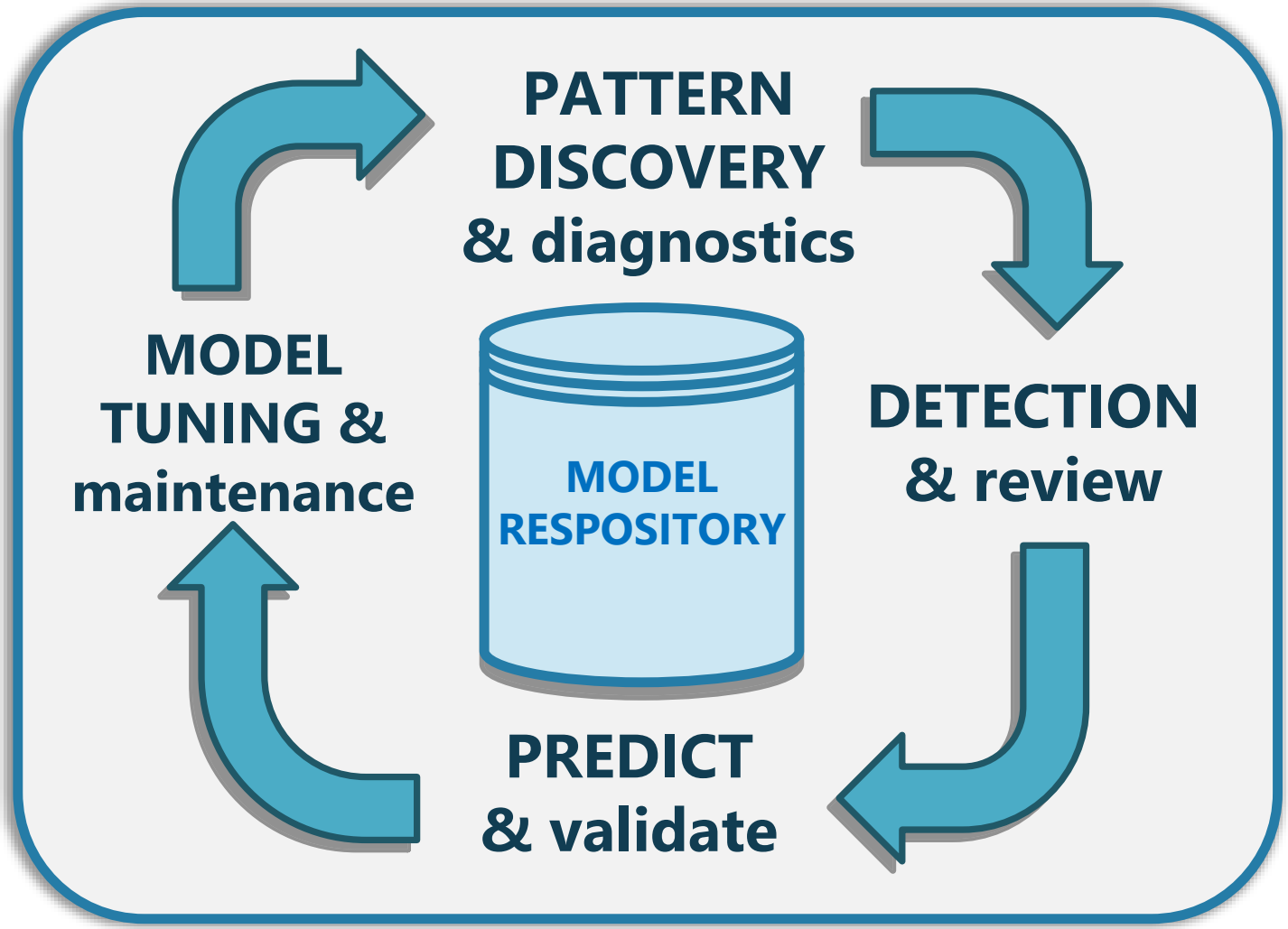
Identifying targeted anomalies amongst and ocean of noise…



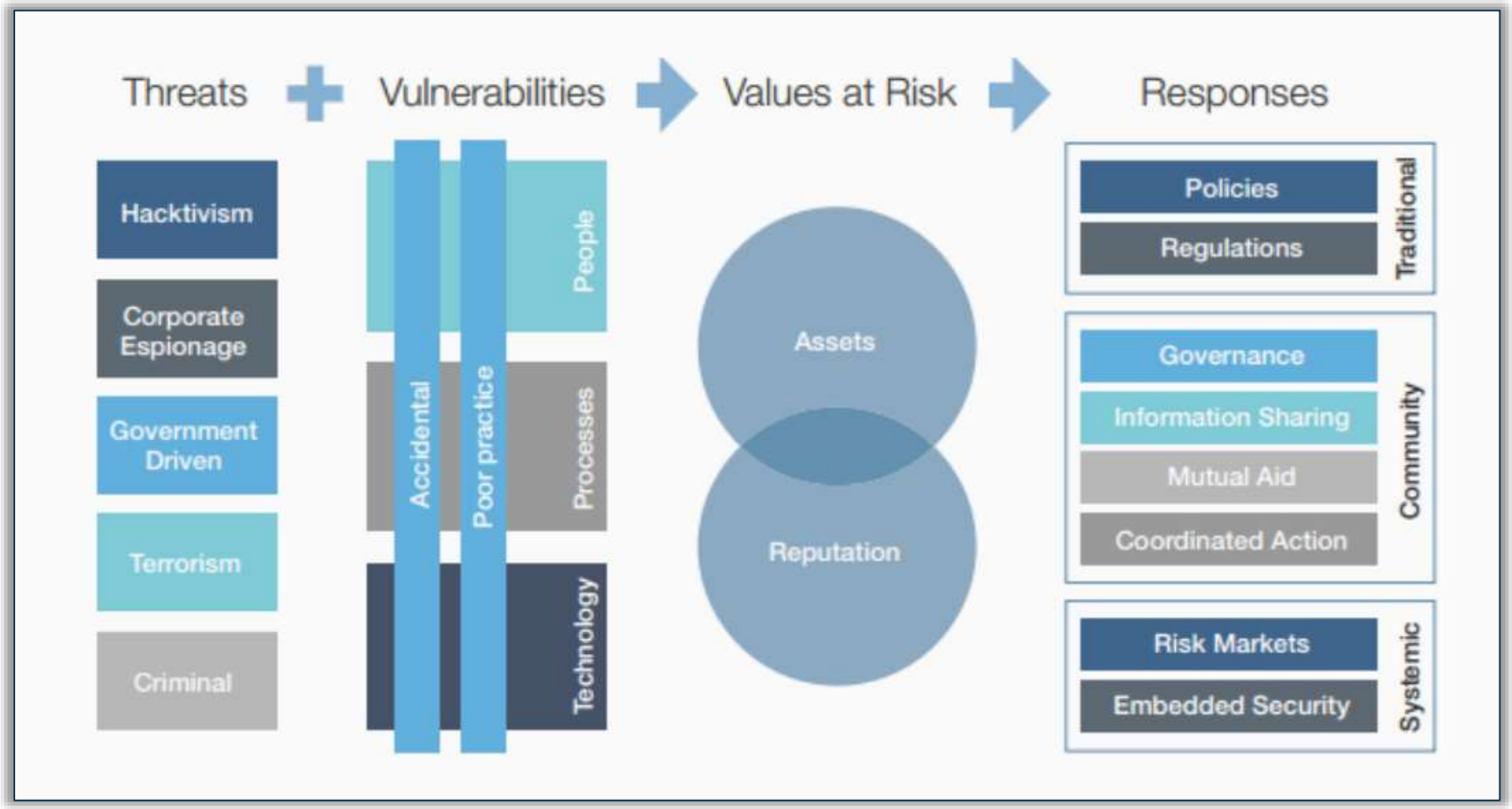
SOURCE

Aggarwal, Charu C. (2017). "Outlier Analysis: Second Edition". Springer International Publishing AG.

Data Science for Cybersecurity Supports Core Uncertainty Reduction

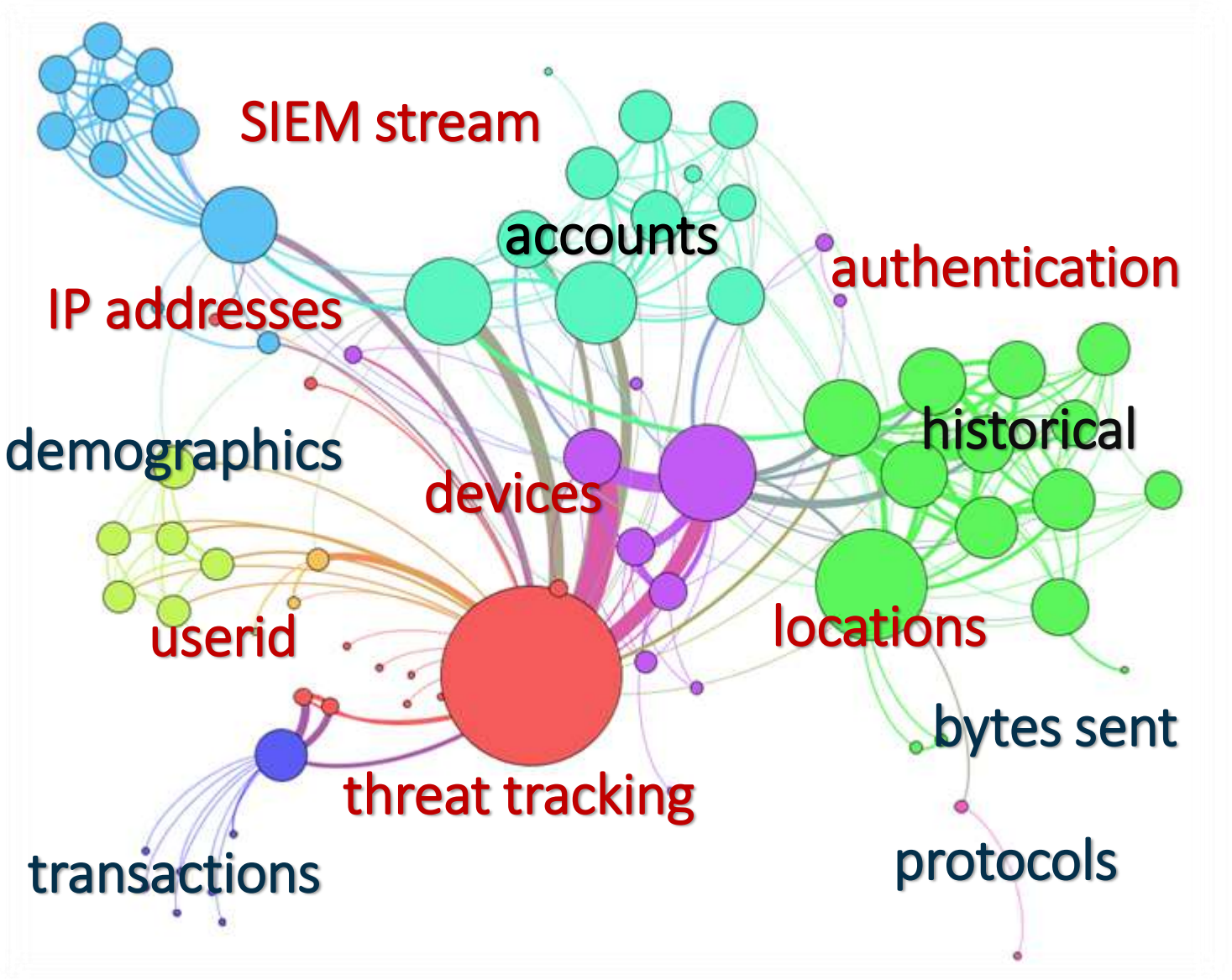


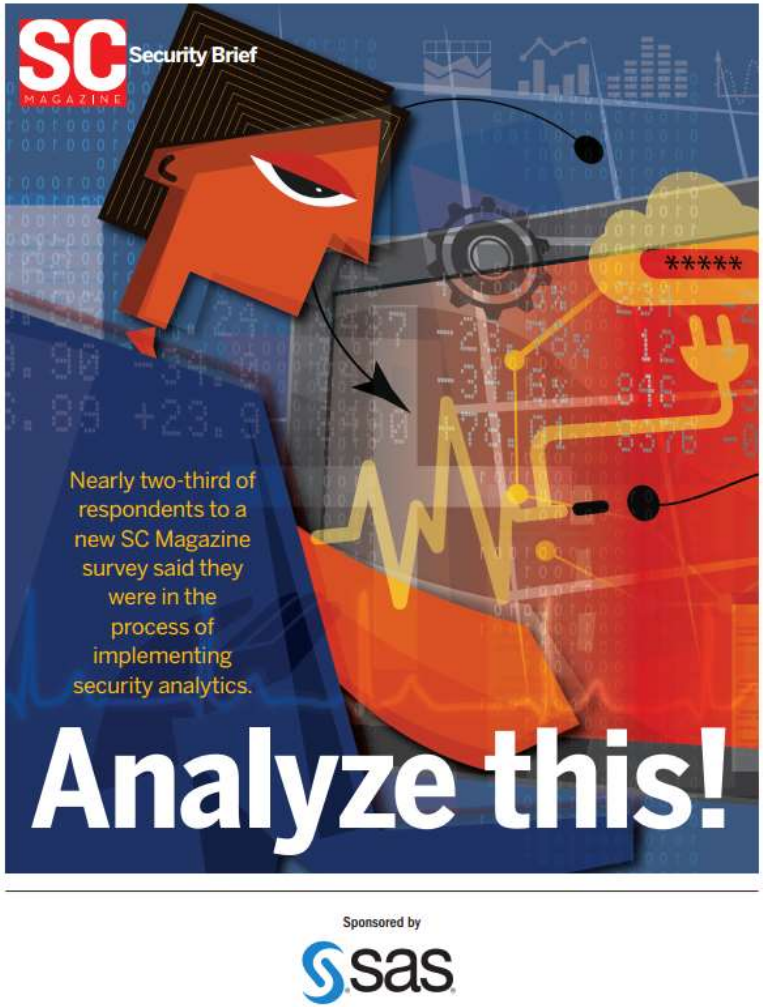
Data Science => Uncertainty Reduction

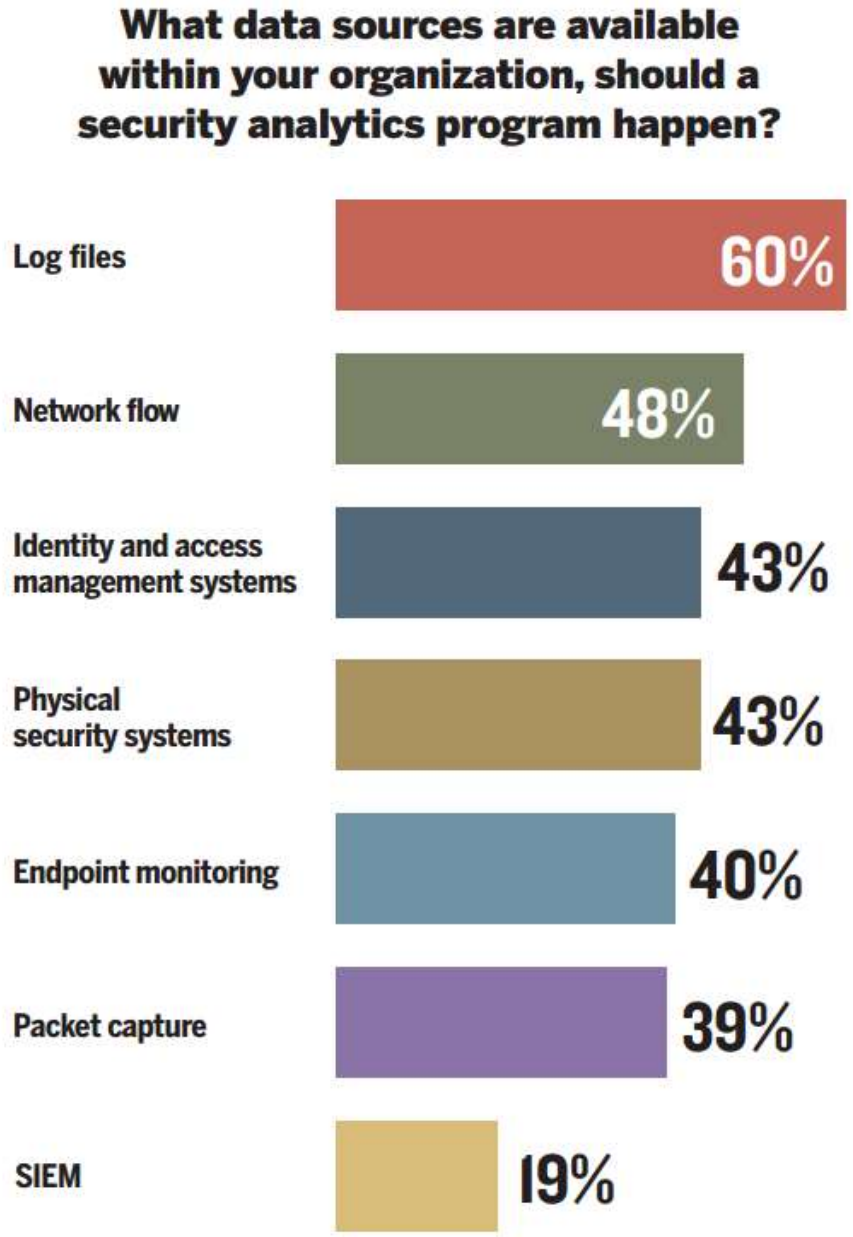


SOURCE
Partnering for Cyber Resilience: Towards the Quantification of Cyber Threats
WEF report in collaboration with Deloitte:
http://www3.weforum.org/docs/WEFUSA_QuantificationofCyberThreats_Report2015.pdf

Linking and Managing ‘Big’ Cyber Data



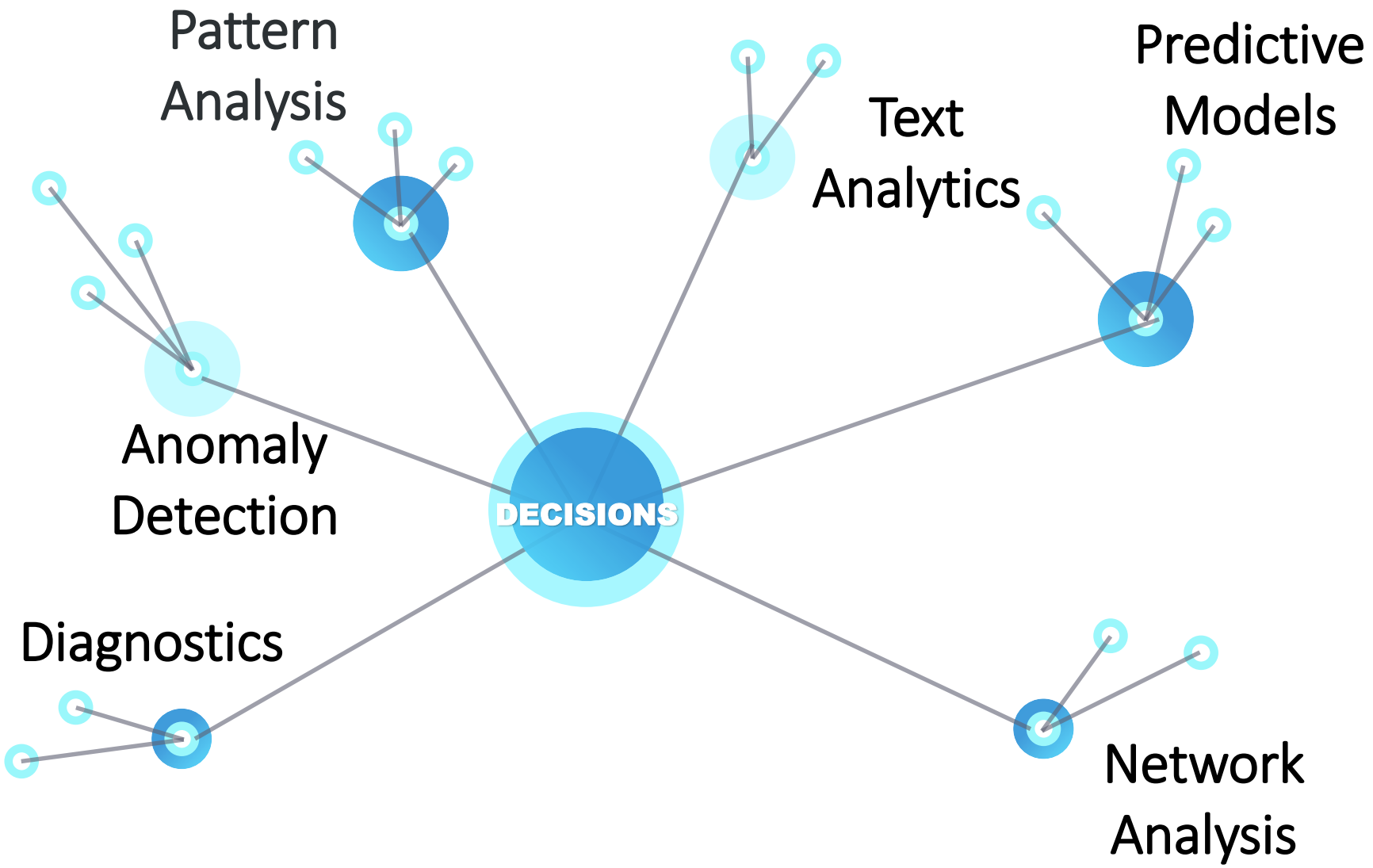




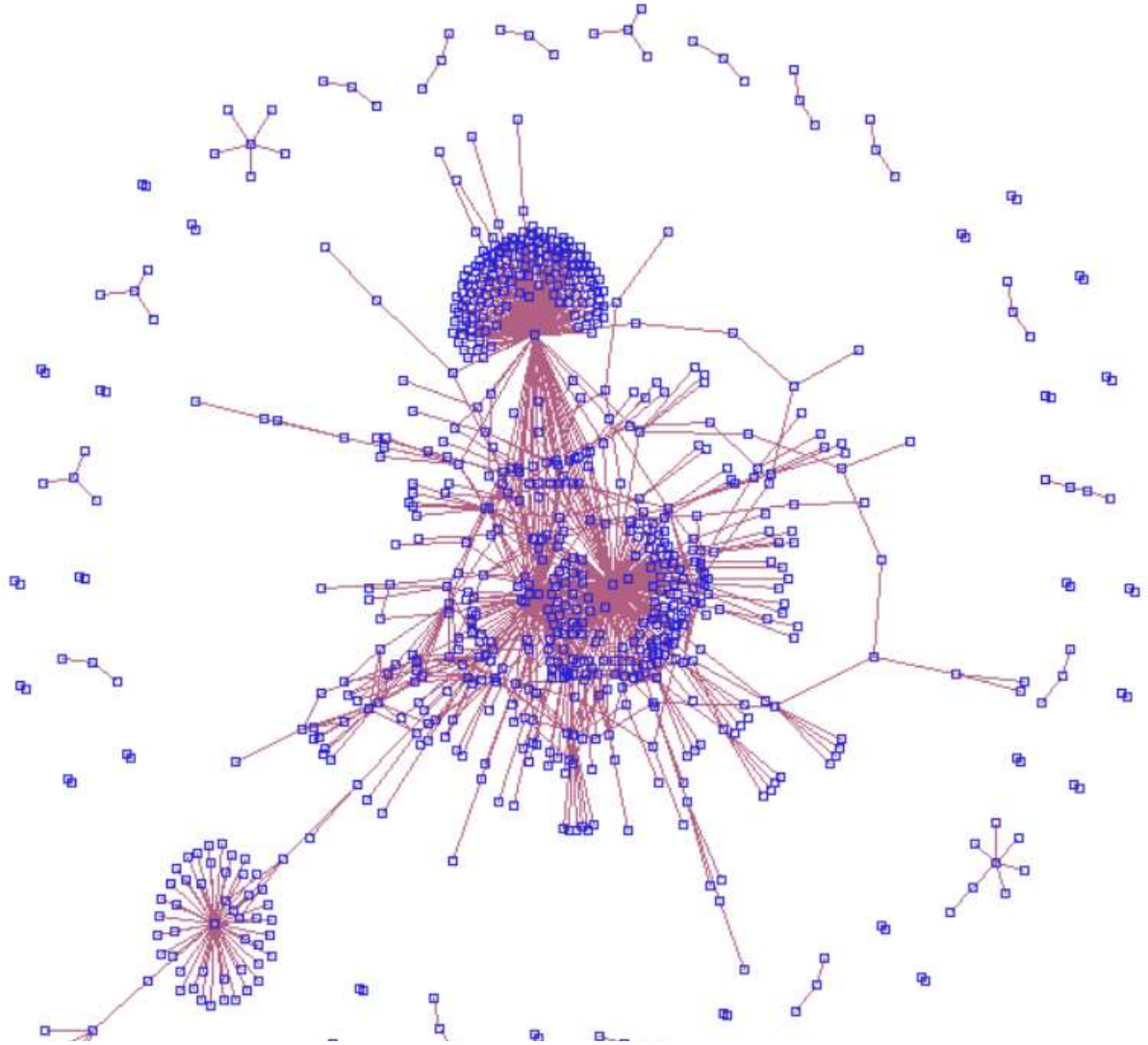
SOURCE

Security Brief Magazine. (2016). "Analyze This! Who's Implementing Security Analytics Now?" Available at https://www.sas.com/en_th/whitepapers/analyze-this-108217.html

Data Science: Multiple Analytics Methods



Statistical Methods: Network Discovery



MEASURES

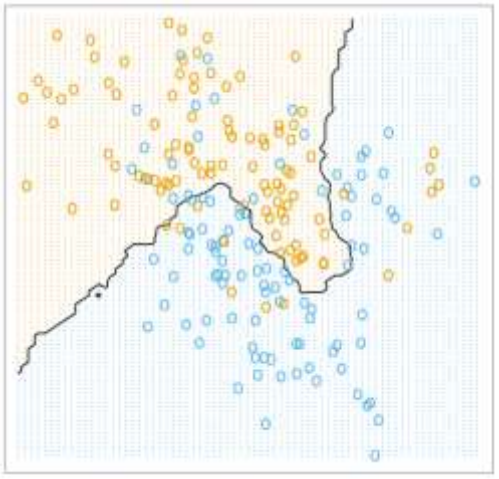
- Centrality
- Eigenvector
- Density
- Reach
- Strength
- Recopricity

Machine Learning

Discovering Patterns

Unsupervised machine learning

- You have a dataset, but little idea concerning the patterns and categories
- *Example*: your have a large set of Net Flow data, but do not know patterns



Cluster Analysis

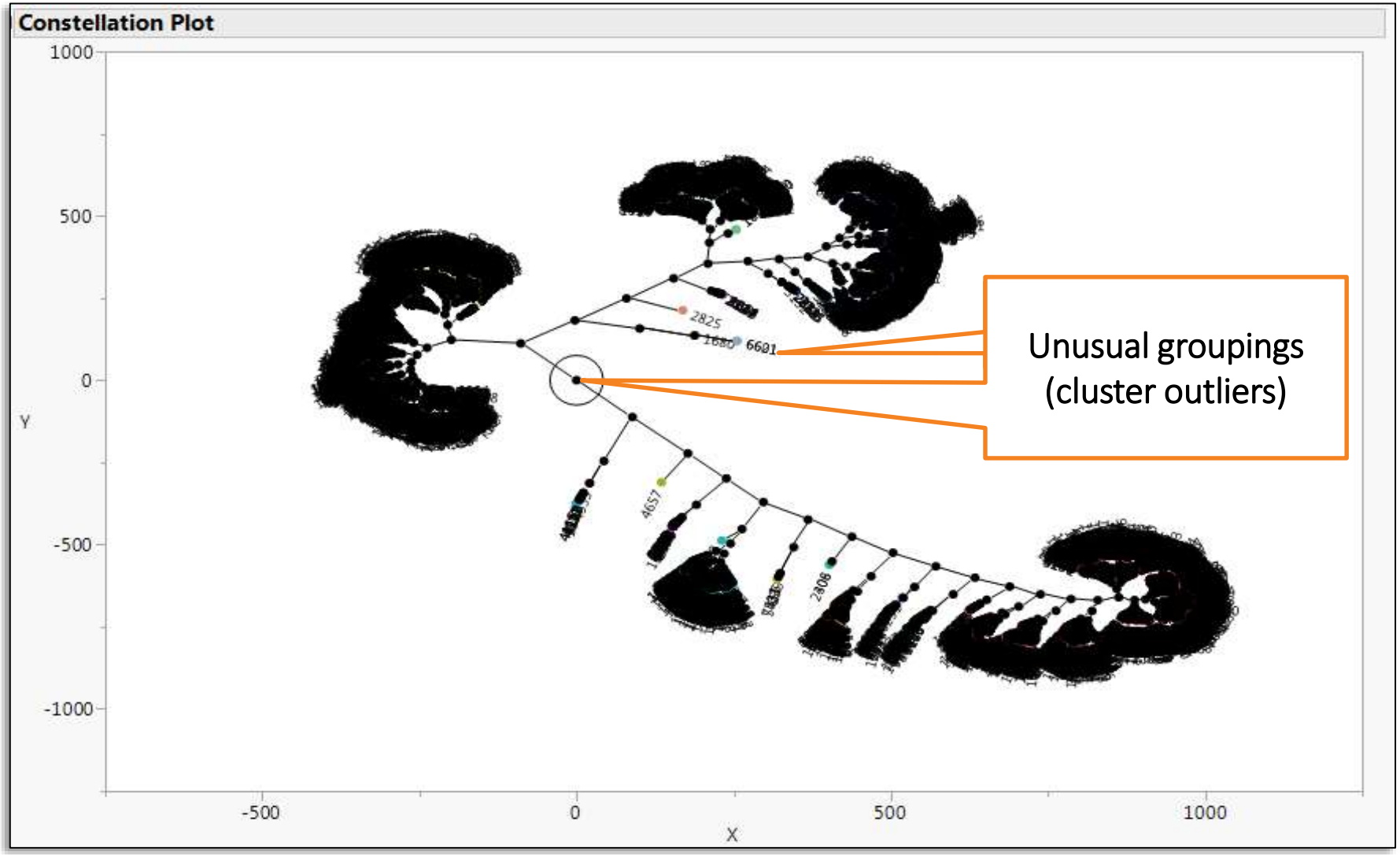
Detection / Prediction

Supervised machine learning

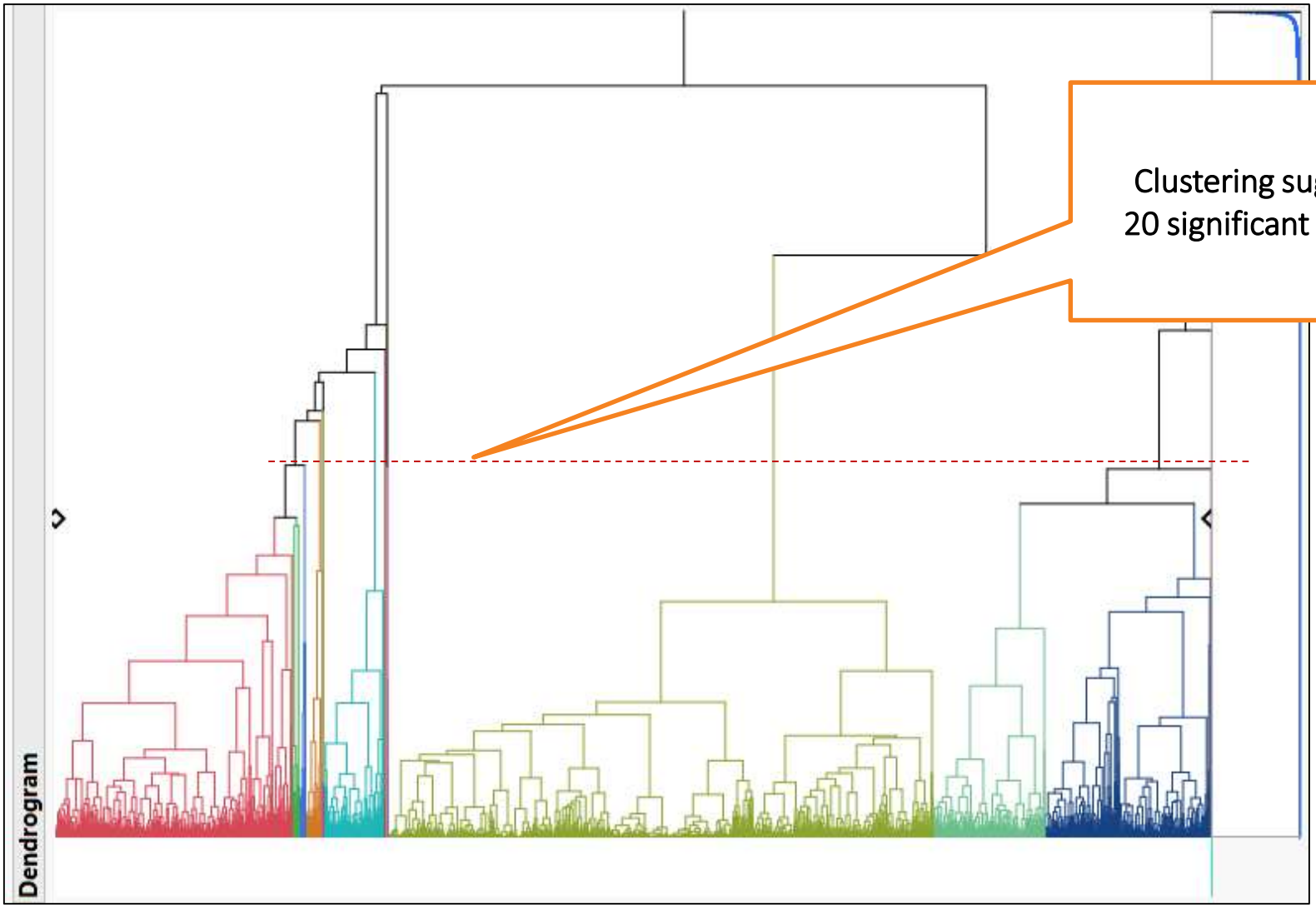
- You have a baseline: a dataset with examples of what you are attempting to predict or classify (random forests, boosted trees)
- *Example*: known examples of cyber attacks based on Net Flow data

Decision Trees

Unsupervised Machine Learning (identifying patterns)

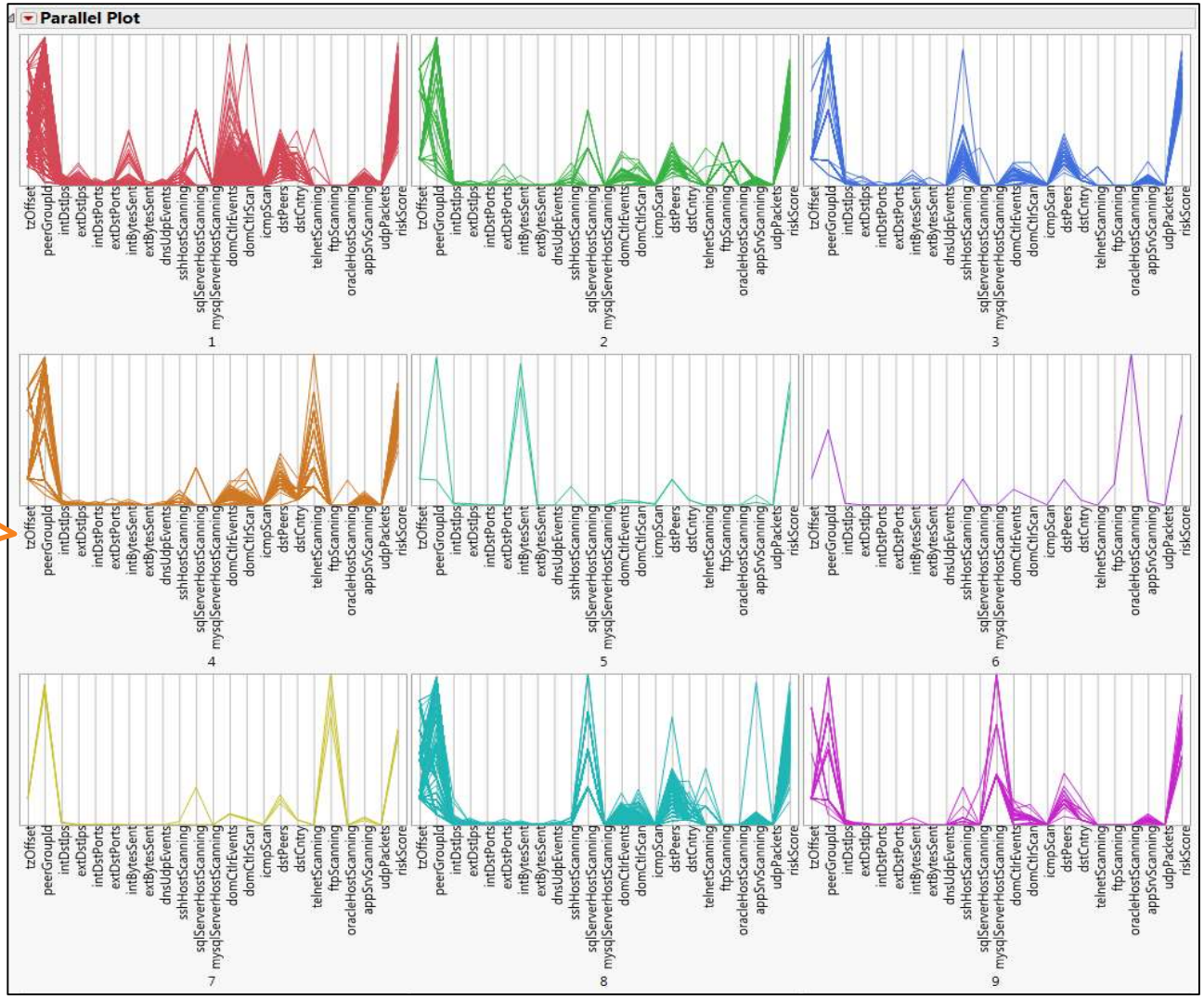


Not All Users are Alike…

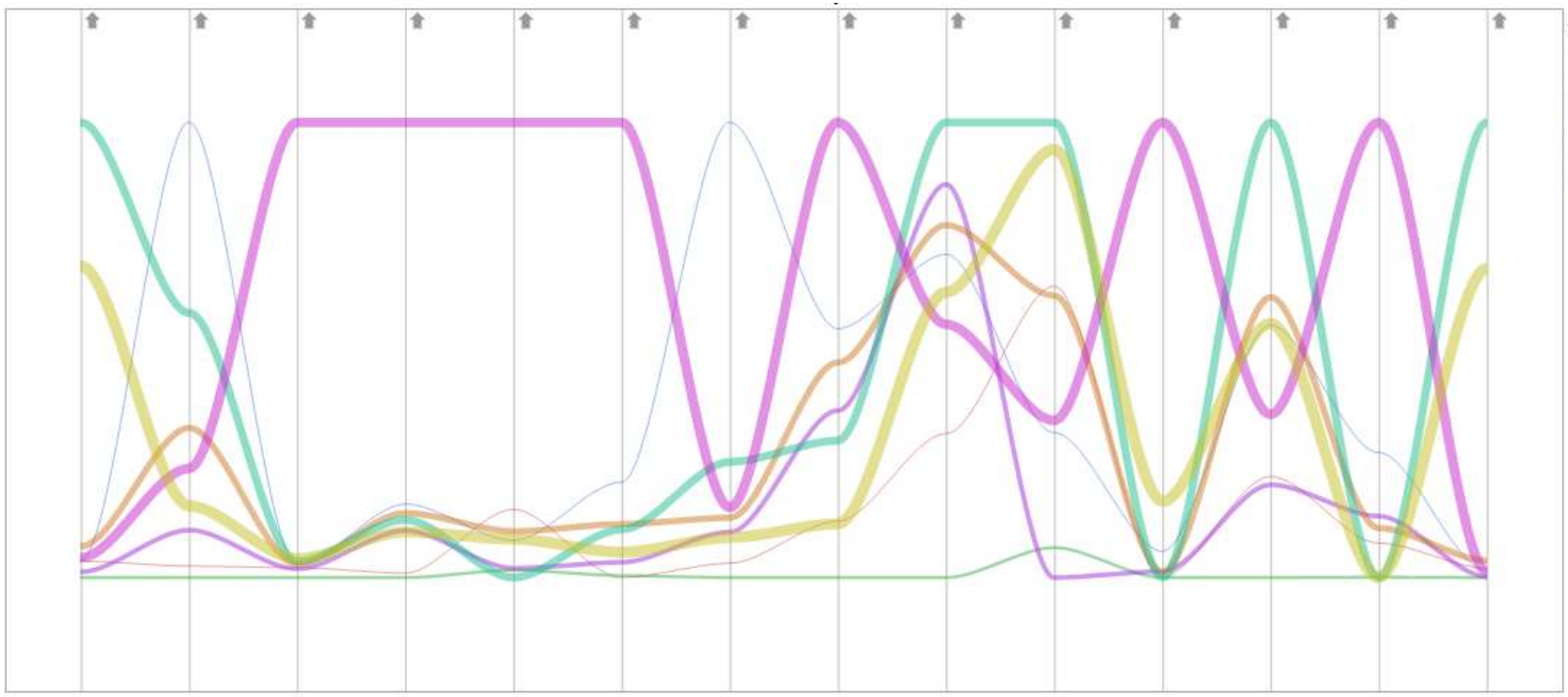


Patterns in Complexity: Cluster Analysis

Each cluster has a signature pattern of 22 measures (high and low)



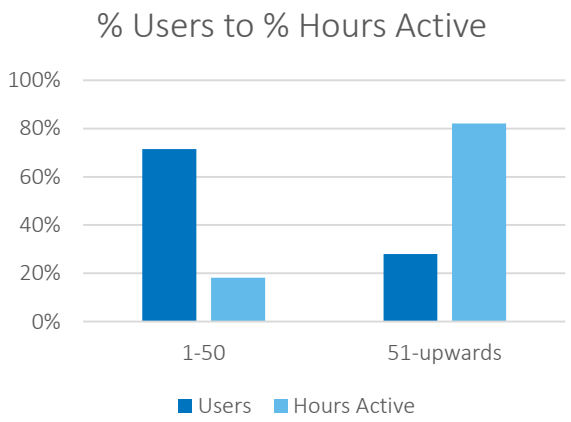
Patterns in Complexity: Cluster Analysis

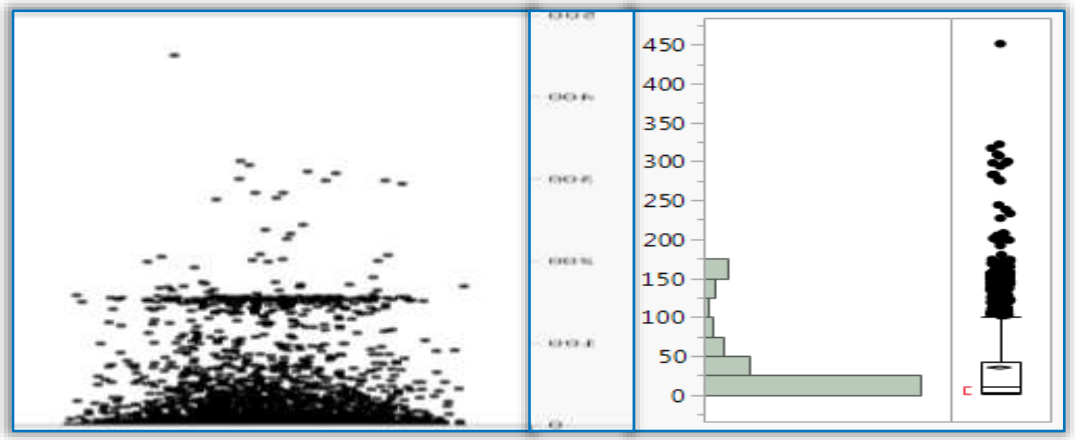


Learnings from the Field: User Patterns

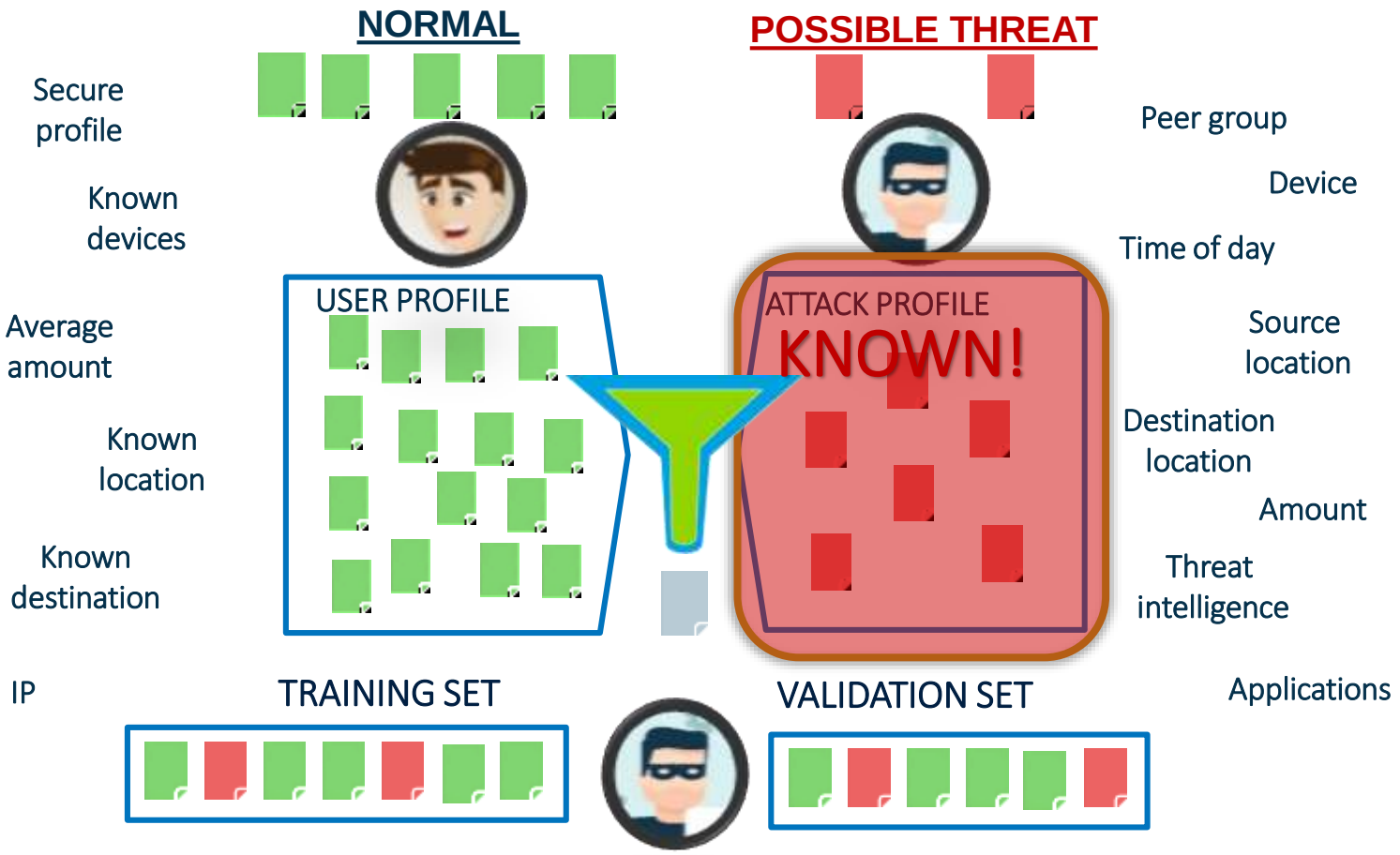
Pareto Principle

- **80/20%** pattern in network-usage (user hours online)
  - *Outliers*: multiple devices 24 hours online
  - High correlation (80-90%) between hours online and propensity to align with multiple usage patterns…
- Pattern has been observed across multiple samples



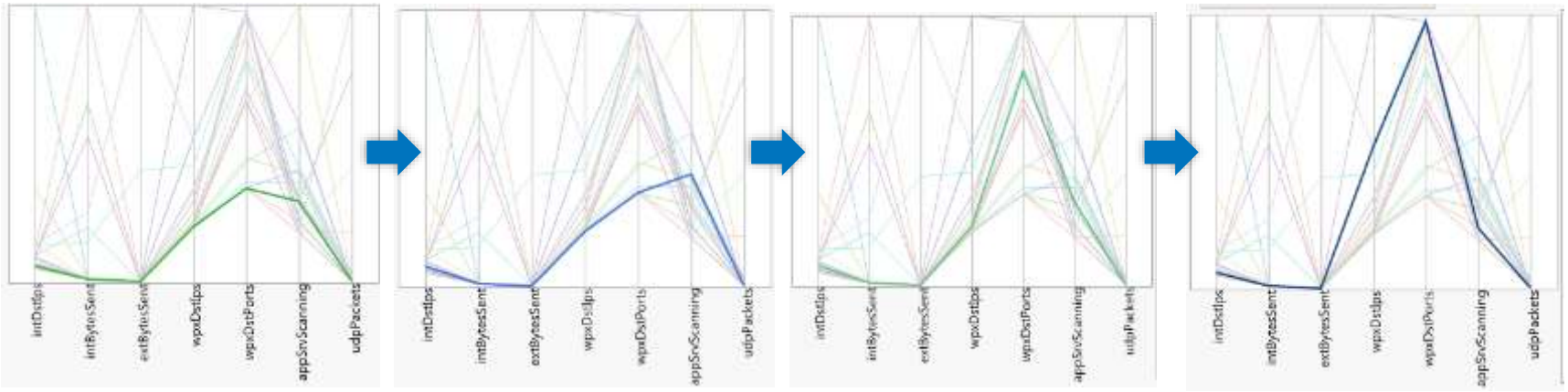


Supervised Machine Learning (Predictive)



Attack Pattern for Prediction

Signature pattern for identified INFECTED DEVICE



Web Proxy Host Scanning	Devices on the network that are anomalously scanning for external devices via the Web Proxy server
Web Proxy Destination Port Scanning	Devices on the network that are anomalously scanning for external devices via the Web Proxy server
Application Server Host Scanning	Devices on the network that are anomalously scanning for devices hosting an http or application server

Summary
Conclusion

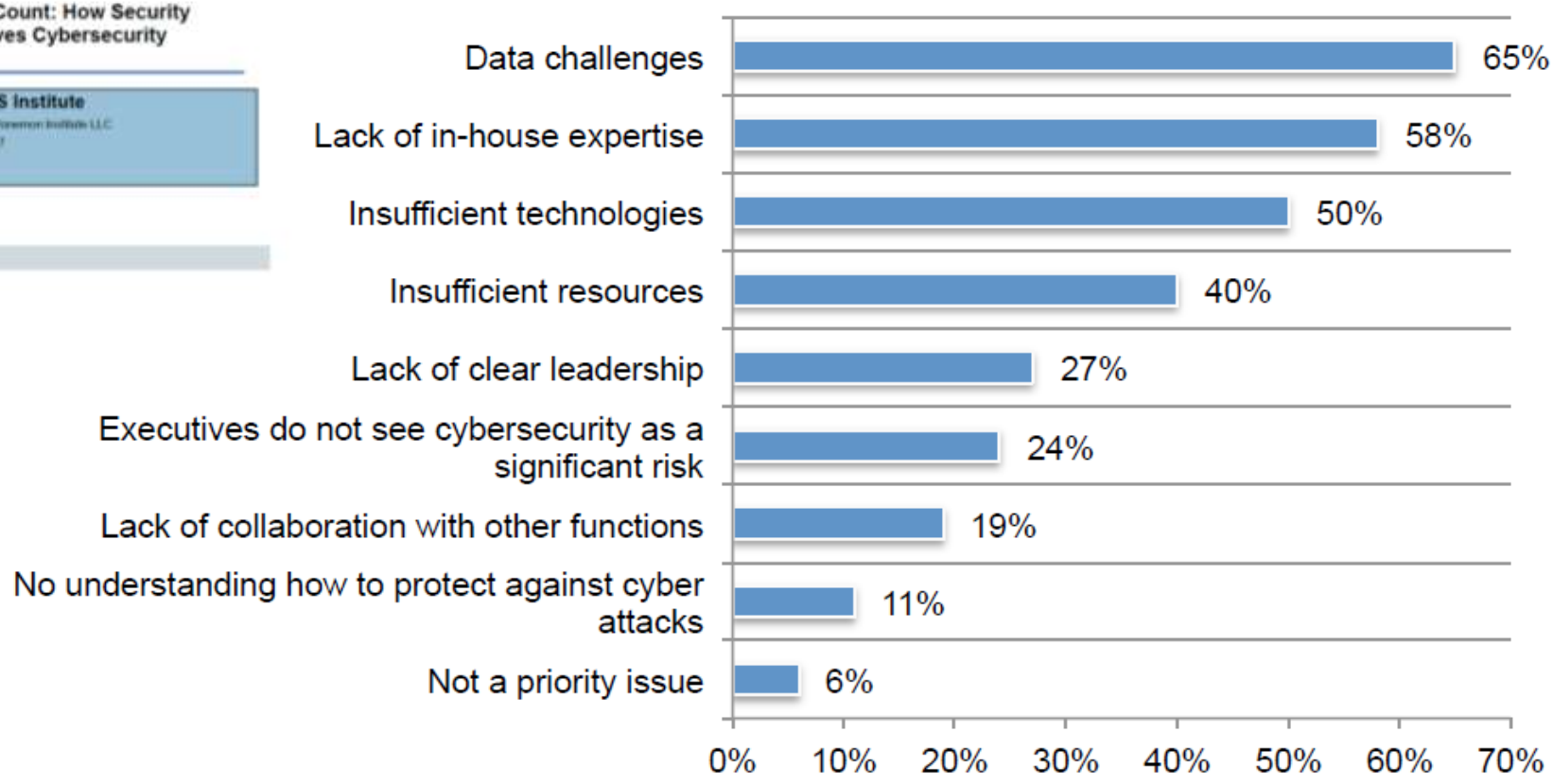
Ponemon INSTITUTE

When Seconds Count: How Security Analytics Improves Cybersecurity Defenses

Sponsored by SAS Institute

Independently conducted by Ponemon Institute LLC
Publication Date: January 2017

Challenges preventing successful use of cybersecurity analytics*



https://www.sas.com/en_us/whitepapers/ponemon-how-security-analytics-improves-cybersecurity-defenses-108679.html

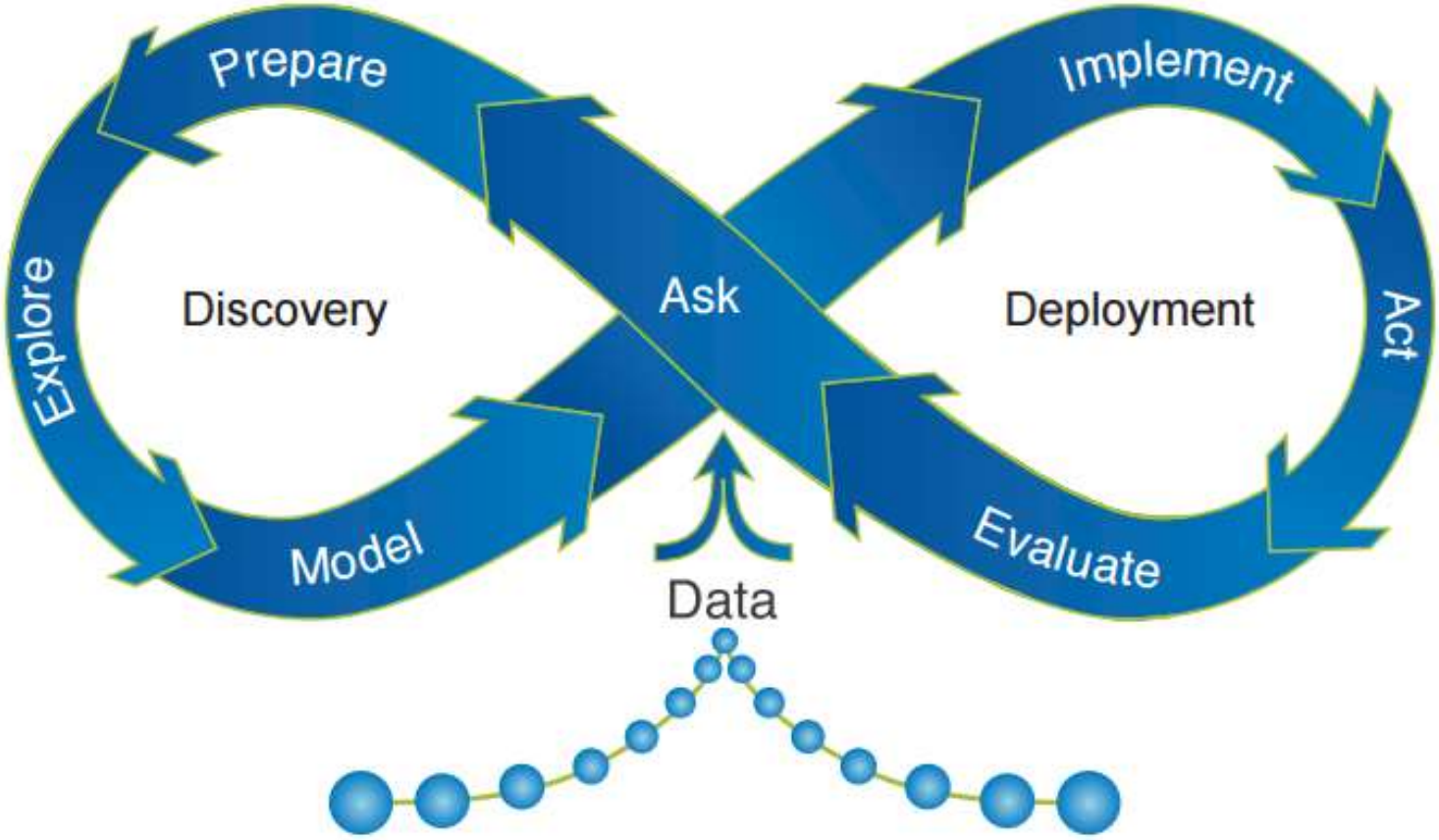
** Survey of 621 global IT security practitioners*

sas

Cyberanalytics context

CHALLENGES	APPROACH
Cyber detection is challenged by…	**Cybersecurity analytics…**
…high volumes of structured and **unstructured** data	…operation at **big data scale** at **high velocity**
…**disconnected** data sources of **variable quality**	…**assess, extract, transform,** and **aggregate** data
…high **false positive** alerts with rule-based approaches	…unsupervised machine learning identifies **hidden patterns**
…lack of **statistical baselines** to establish **validity**	…effective statistical **diagnostics** for model **validation**
…slow and **manual** investigation processes (needles in the haystack)	…supply hunters with **targeted alerts** based on demonstrable statistical anomalies

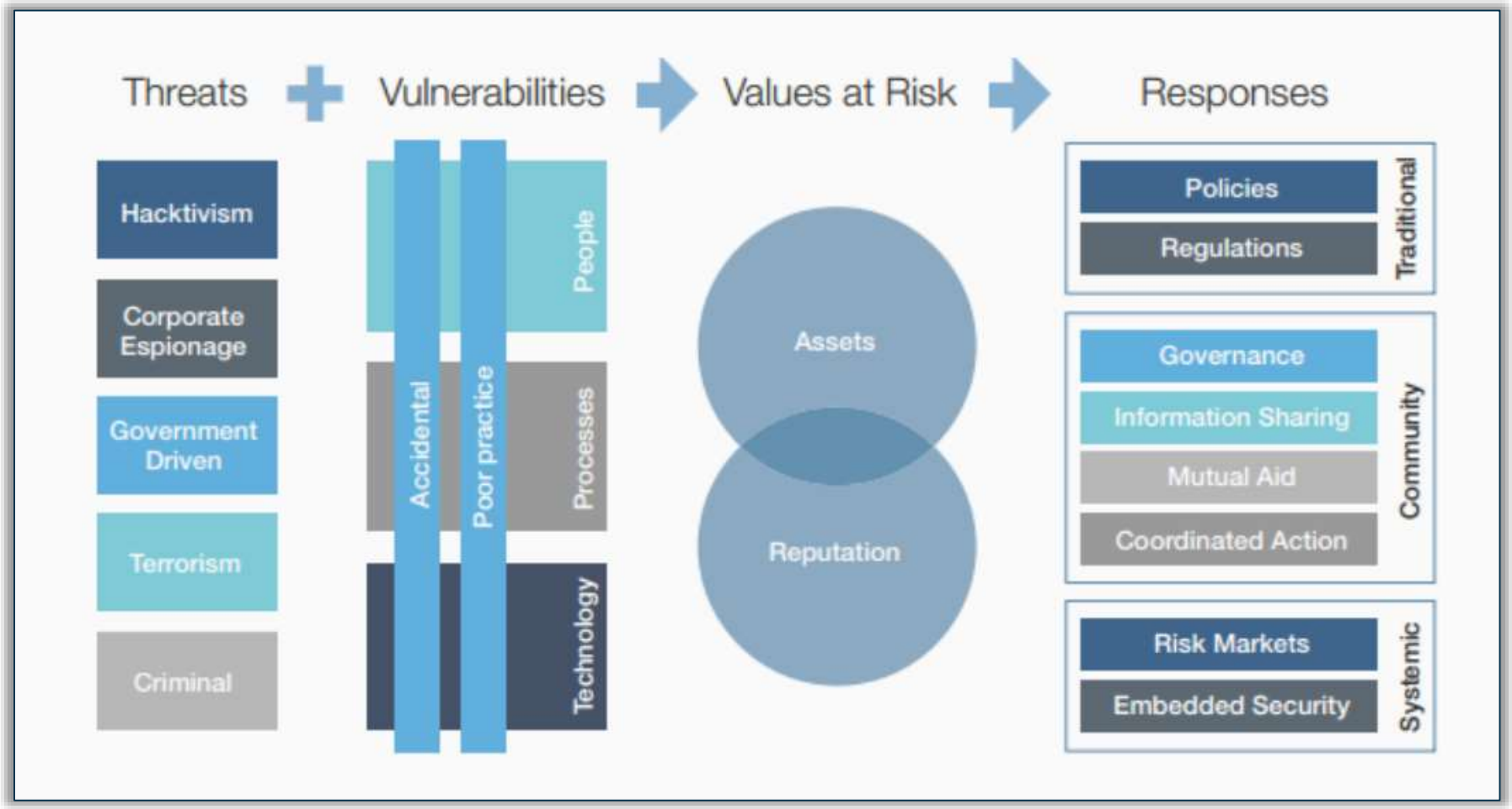
Analytics Lifecycle



SOURCE

SAS Institute. (2016). "Managing the Analytical Life Cycle for Decisions at Scale." Available at https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/manage-analytical-life-cycle-continuous-innovation-106179.pdf

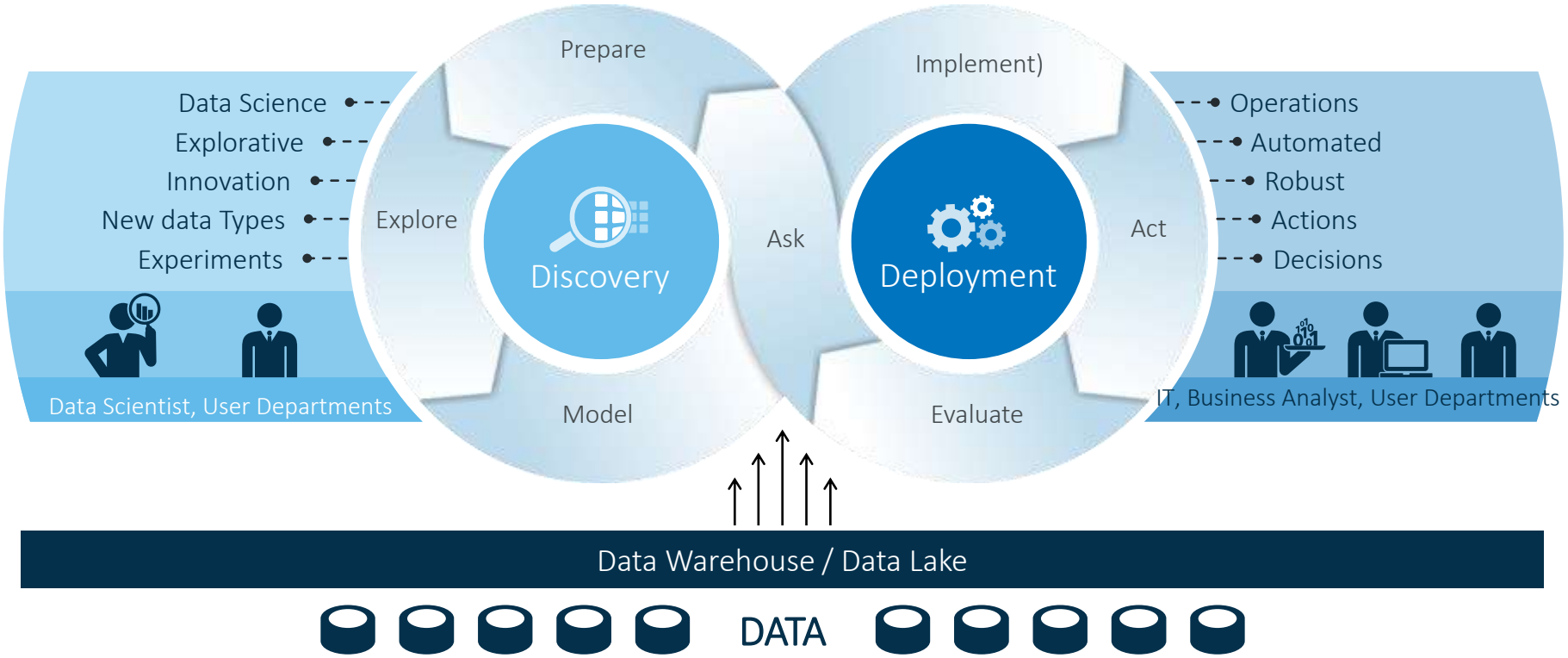
Data Science => Uncertainty Reduction



SOURCE
Partnering for Cyber Resilience: Towards the Quantification of Cyber Threats
WEF report in collaboration with Deloitte:
http://www3.weforum.org/docs/WEFUSA_QuantificationofCyberThreats_Report2015.pdf

Analytics Lifecycle

Building and Deploying Advanced Analytics



REFERENCES

- Aggarwal, C. (2013). "Outlier Analysis." Springer. http://www.springer.com/la/book/9781461463955
- Kirchhoff, C., Upton, D., and Winnefeld, Jr., Admiral J. A. (2015 October 7). "Defending Your Networks: Lessons from the Pentagon." Harvard Business Review. Available at https://www.sas.com/en_us/whitepapers/hbr-defending-your-networks-108030.html
- Longitude Research. (2014). "Cyberrisk in banking." Available at http://www.longitude.co.uk/wp-content/uploads/2015/02/cyberrisk-in-banking-106605.pdf
- Ponemon Institute. (2017). "When Seconds Count: How Security Analytics Improves Cybersecurity Defenses." Available at https://www.sas.com/en_us/whitepapers/ponemon-how-security-analytics-improves-cybersecurity-defenses-108679.html
- SANS Institute. (2015). "2015 Analytics and Intelligence Survey." Available at https://www.sas.com/en_us/whitepapers/sans-analytics-intelligence-survey-108031.html
- SANS Institute. (2016). "Using Analytics to Predict Future Attacks and Breaches." Available at https://www.sas.com/en_us/whitepapers/sans-using-analytics-to-predict-future-attacks-breaches-108130.html
- SAS Institute. (2016). "Managing the Analytical Life Cycle for Decisions at Scale." Available at https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/manage-analytical-life-cycle-continuous-innovation-106179.pdf
- SAS Institute. (2017). "SAS Cybersecurity: Counter cyberattacks with your information advantage." Available at https://www.sas.com/en_us/software/fraud-security-intelligence/cybersecurity-solutions.html
- Security Brief Magazine. (2016). "Analyze This! Who's Implementing Security Analytics Now?" Available at https://www.sas.com/en_th/whitepapers/analyze-this-108217.html
- UBM. (2016). "Dark Reading: Close the Detection Deficit with Security Analytics." Available at https://www.sas.com/en_us/whitepapers/close-detection-deficit-with-security-analytics-108280.html