

RSA[®]Conference2018

San Francisco | April 16 – 20 | Moscone Center

SESSION ID: STR-T10

FROM NO DATA TO DROWNING IN DATA - IT'S TIME FOR A REALITY CHECK

Jack Jones

EVP R&D

RiskLens

@jonesFAIRiq

#RSAC



Objectives...



- Understand why we want/need data
- Understand the nature of cyber data
- Understand the relationship between data and models
- Recognize where you can find different types of data
- Understand the “so what” of data
- Know how to gauge (and improve) the quality of your data
- Recognize some things to watch out for

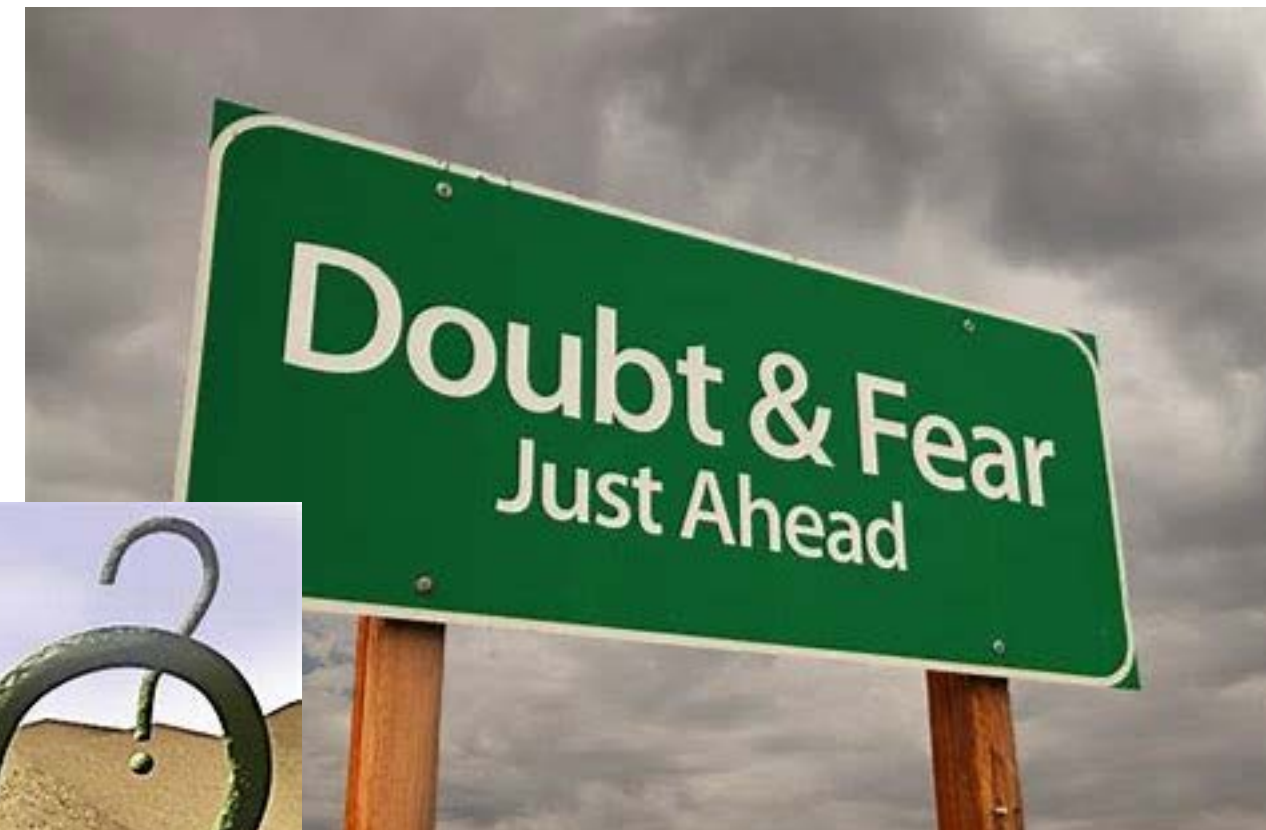
RSA®Conference2018



#RSAC

WHY DO WE WANT/NEED DATA?

The bottom line...

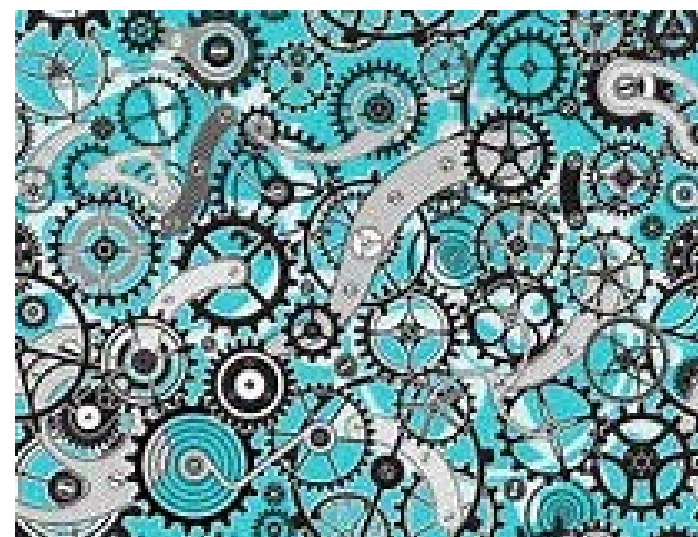


On important decisions,
we fear uncertainty

The risk landscape...



Complex



+

Dynamic



+

Limited resources



=

The need to
prioritize



Which means...



Prioritization



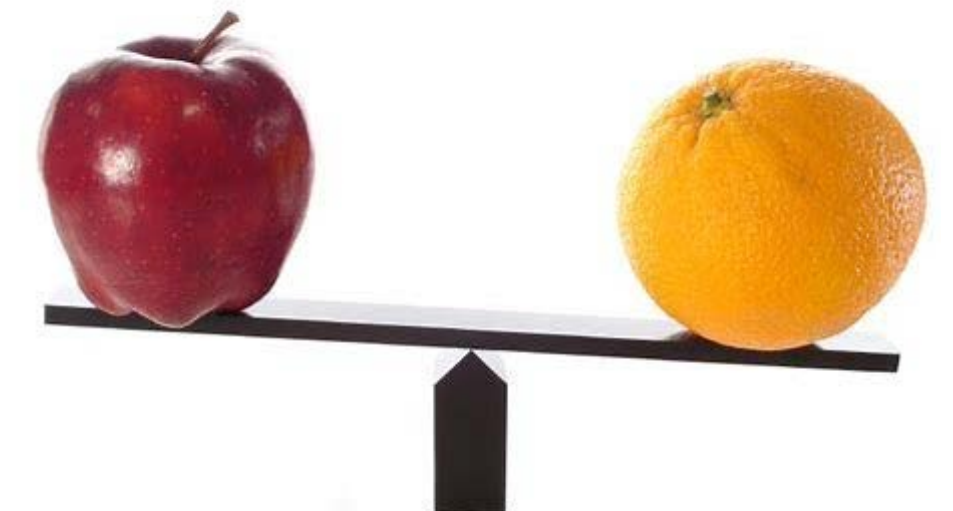
=

Decisions/Choices



Based on...

Comparisons



Comparisons are always based on measurement

Measurements always involve data

Which means...



Better

~~More~~ data = less uncertainty

RSA®Conference2018



#RSAC

THE TRUTH ABOUT DATA

How much data do we need?



“We don’t have enough data.”

How much is enough?

How much uncertainty is too much?

How much data do we need?



In science, “half-life” refers to when 50% of a substance has materially changed.

What would the half-life be, of a cyber-related actuarial table?

Threat
capabilities

Assets

Attack
vectors

Control
conditions

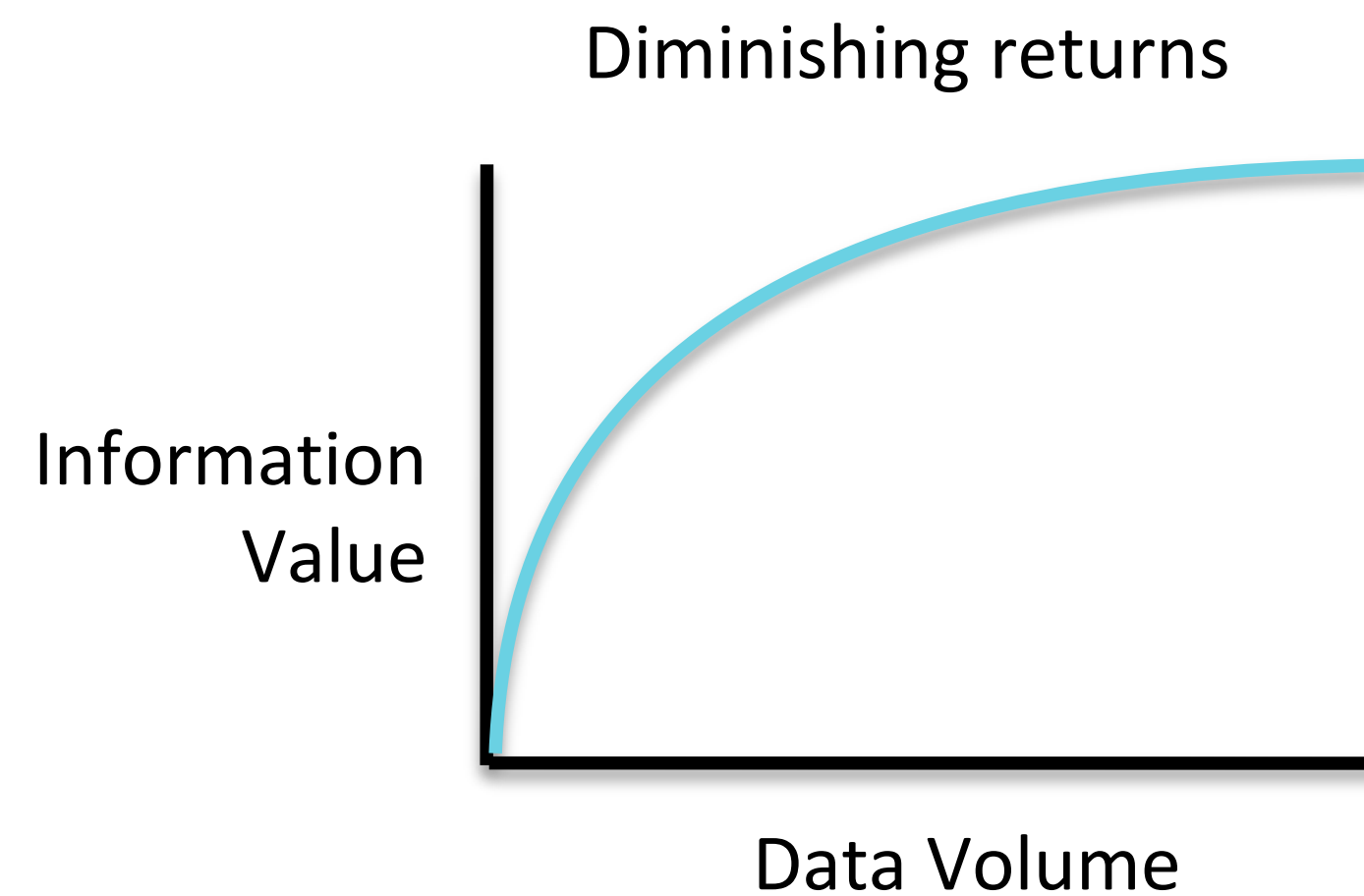
Loss
implications?

Not all data has equal value



The first few data points are vastly more valuable in reducing uncertainty than the 100th.

You have more data than you think you do.
(Douglas Hubbard)



You need less data than you think you do. (Douglas Hubbard)

Not all data has equal value



Data regarding your most important assets is more valuable than data regarding less valuable assets.

The trick is knowing which is which.

How much data do we need?

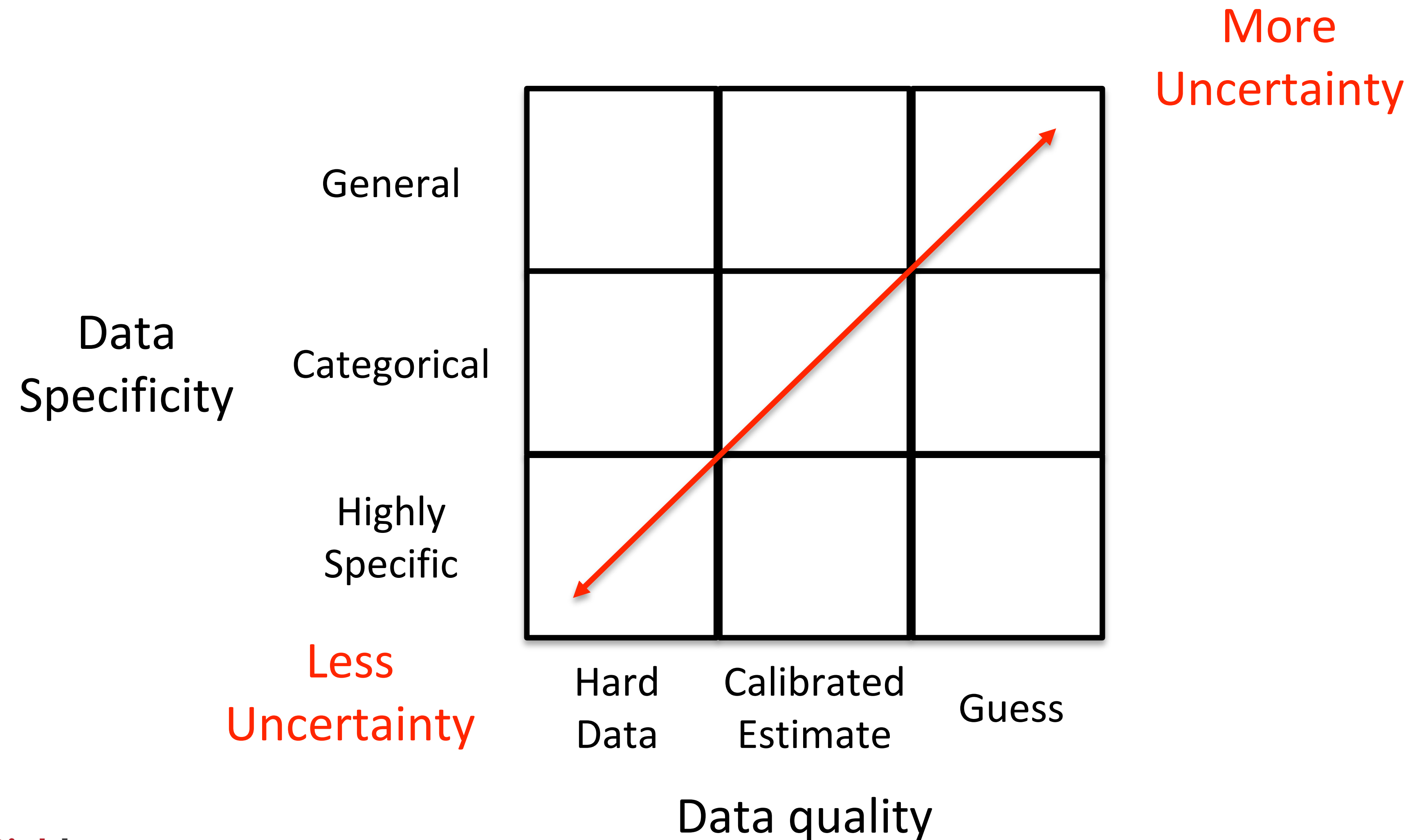


We have vast amounts of hard data for some areas of cyber risk, and almost none for others.

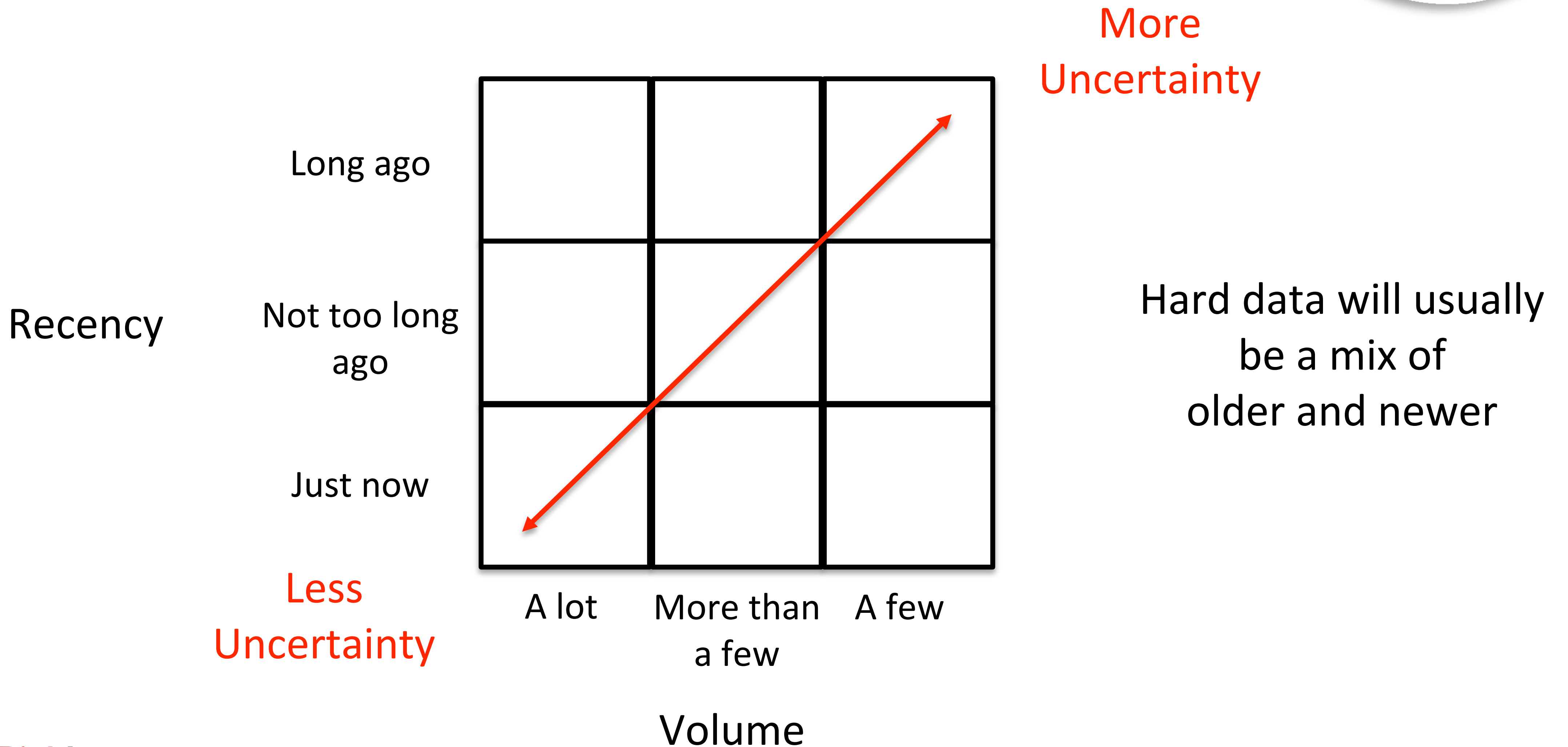
What are some examples of data we have a lot of?

What are some examples of where we have little data?

Gauging data-related uncertainty



Gauging hard data-related uncertainty



RSA[®]Conference2018



#RSAC

OKAY, SO WE HAVE DATA. SO WHAT?

What does the data mean?




Your organization has data regarding umpteen thousand unpatched vulnerabilities...

So what?

What decisions need to be made?

Two categories of data



- Raw data, for example:
 - The number of SQL injection vulnerabilities
 - The number of malware stopped at the perimeter
 - The number of loss events
 - \$\$\$ impact that materialized from loss events
- Interpreted data  Requires models
 - CVSS severity scores
 - Any risk rating or score

What does the data mean — i.e., so what?



If the model is badly broken, it doesn't matter how much data you have.

“All models are broken; some models are useful.”
(George Cox)

What does the data mean?



What's the most commonly used cyber security model?

Assumptions

Scope?

Formula?



Mental models

Measuring risk



- Every risk measurement involves two models:
 - The scope of what's being measured
 - What asset
 - What threat
 - Which vector
 - Which controls are relevant
 - What type of event (e.g., C, I, A)
 - An analytic model used to evaluate data (E.g., FAIR)

Formal models provide...



- A guide to what data you need for an analysis
- A means of interpreting data (deriving the “so what”)
- An expression of your analytic assumptions
 - Which can then be challenged and/or validated

BTW — until AI takes over, all models are generated by humans

RSA[®]Conference2018



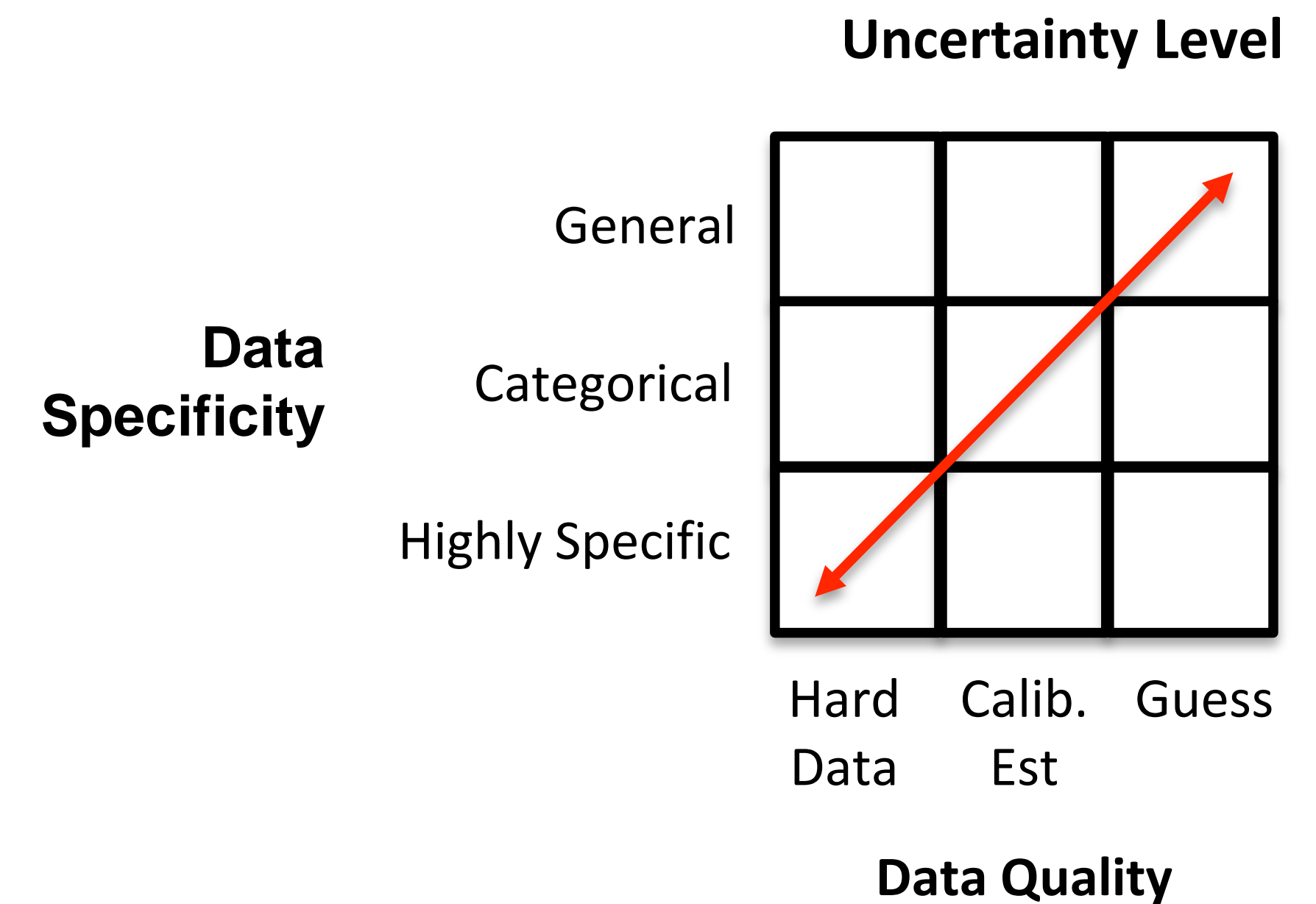
#RSAC

COMMON SOURCES OF CYBER DATA

Security Telemetry



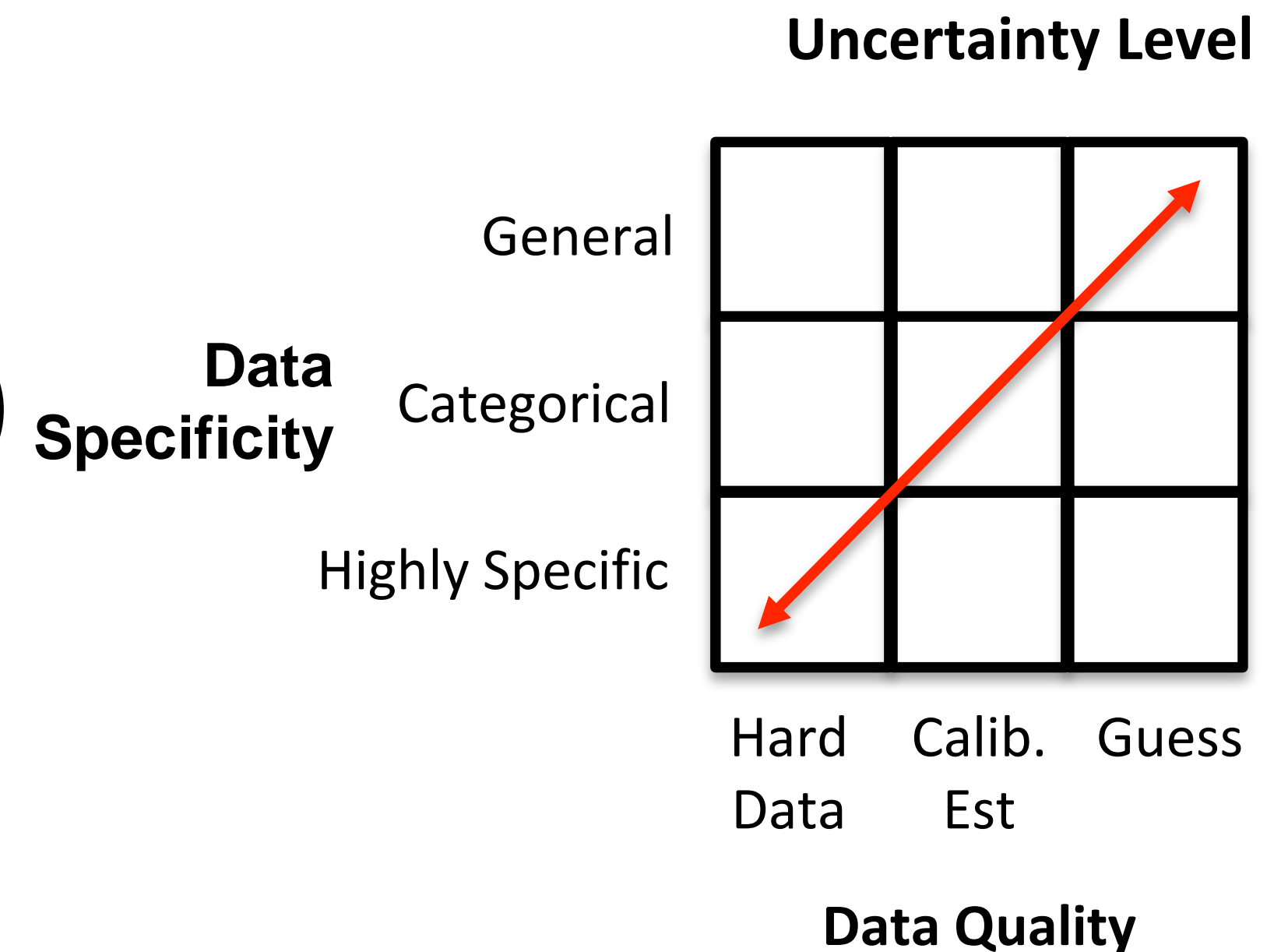
- Logs (typically raw data)
 - IDS
 - System
 - Application
 - Firewall
 - Anti-malware
- Scan results
 - Vulnerabilities (raw)
 - Vulnerability severity (interpreted)



Testing



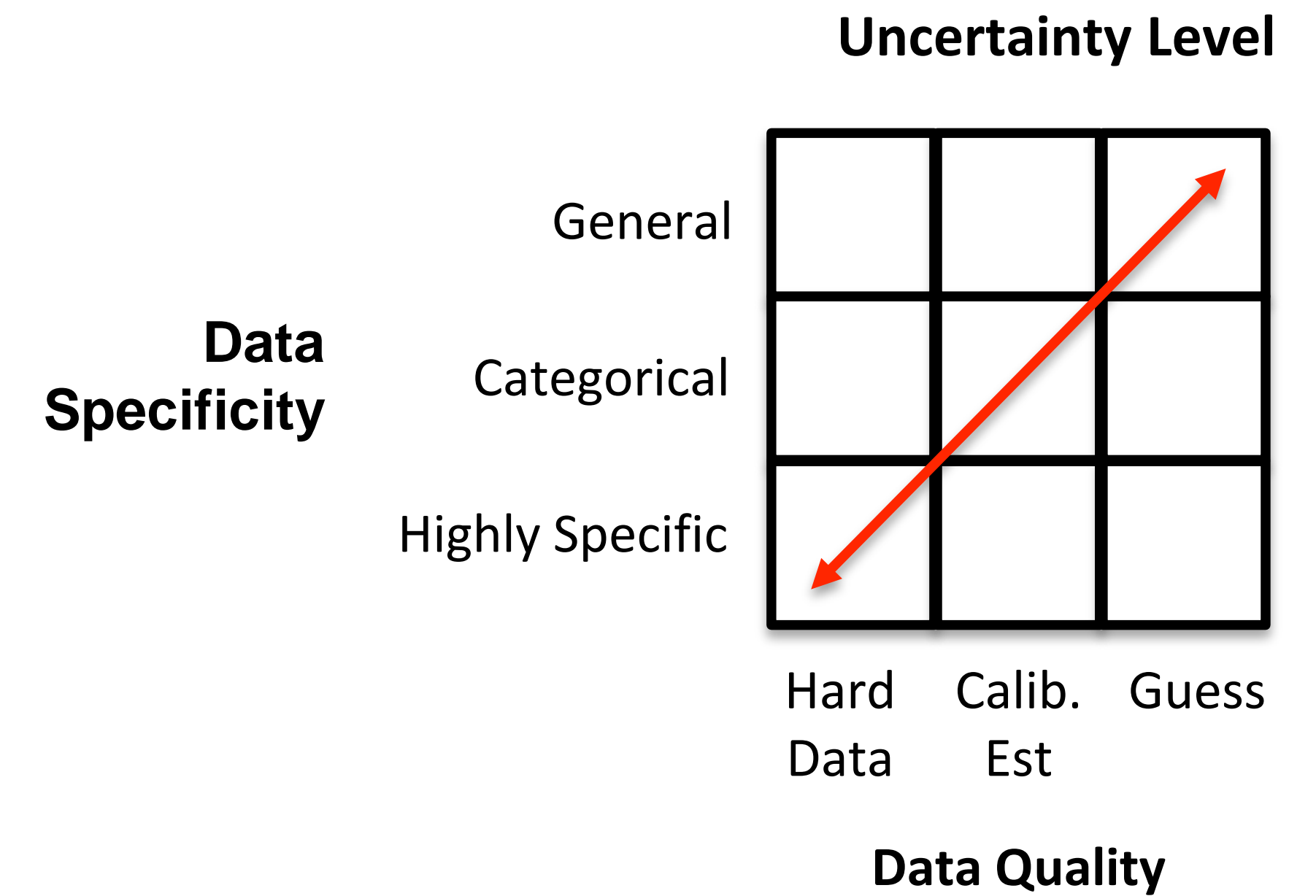
- Penetration tests
 - Weaknesses identified (raw)
 - Severity (interpreted)
- Audits
 - Control deficiencies (raw)
 - Severity (interpreted)
- Survey assessments (e.g., 3rd party questionnaires)
 - Control deficiencies (raw)
 - Severity (interpreted)



Threat intelligence



- Activity levels (raw)
- TTPs (raw)



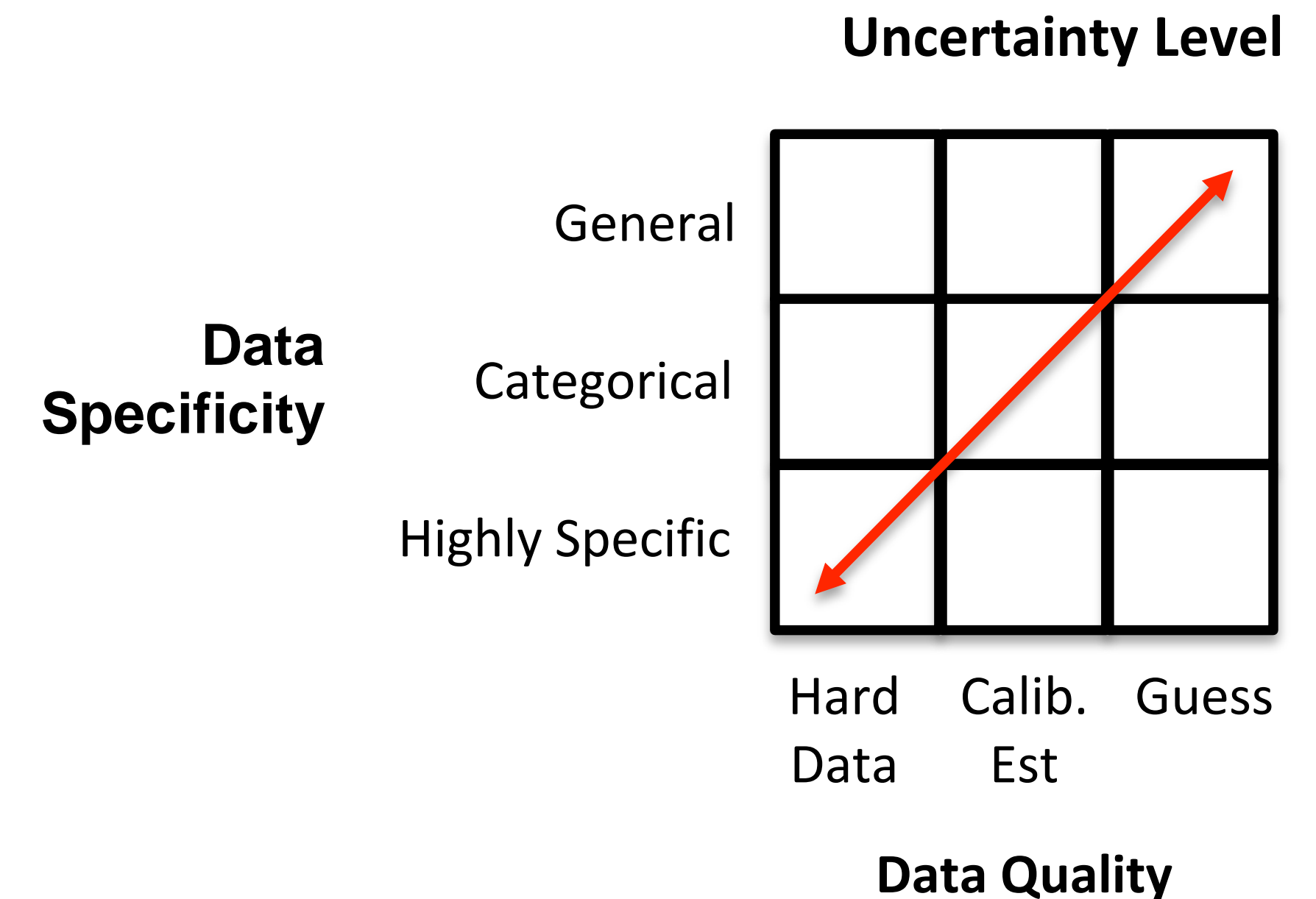
Industry resources/research



- Examples:
 - Verizon
 - Ponemon
 - Cyentia
 - Advisen
 - FAIR Institute
 - Universities
 - Eventually, the insurance industry

Usually a mix of raw data and their interpretation of it.

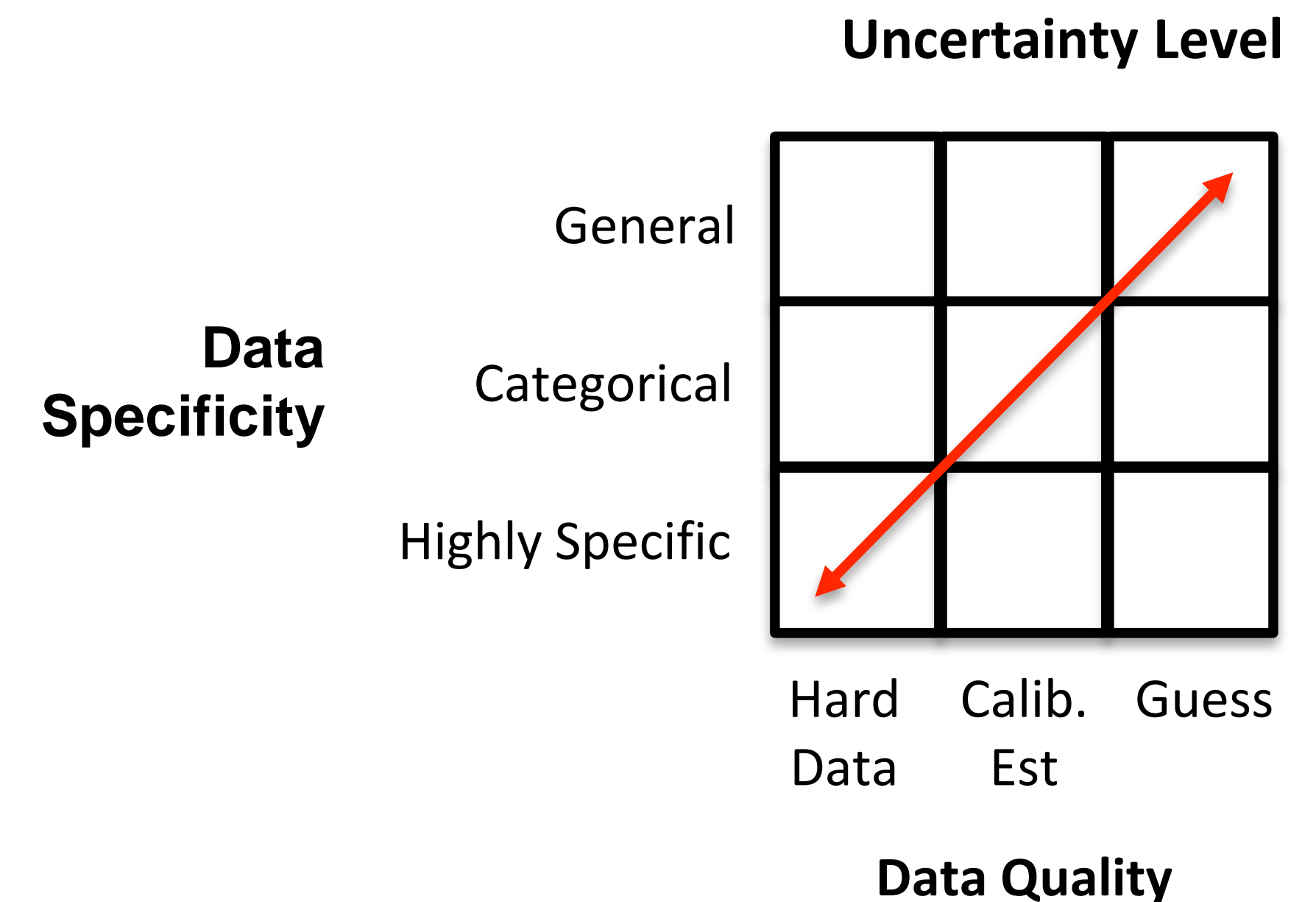
- Data quality varies
- Interpretation quality varies



Subject matter expert estimates



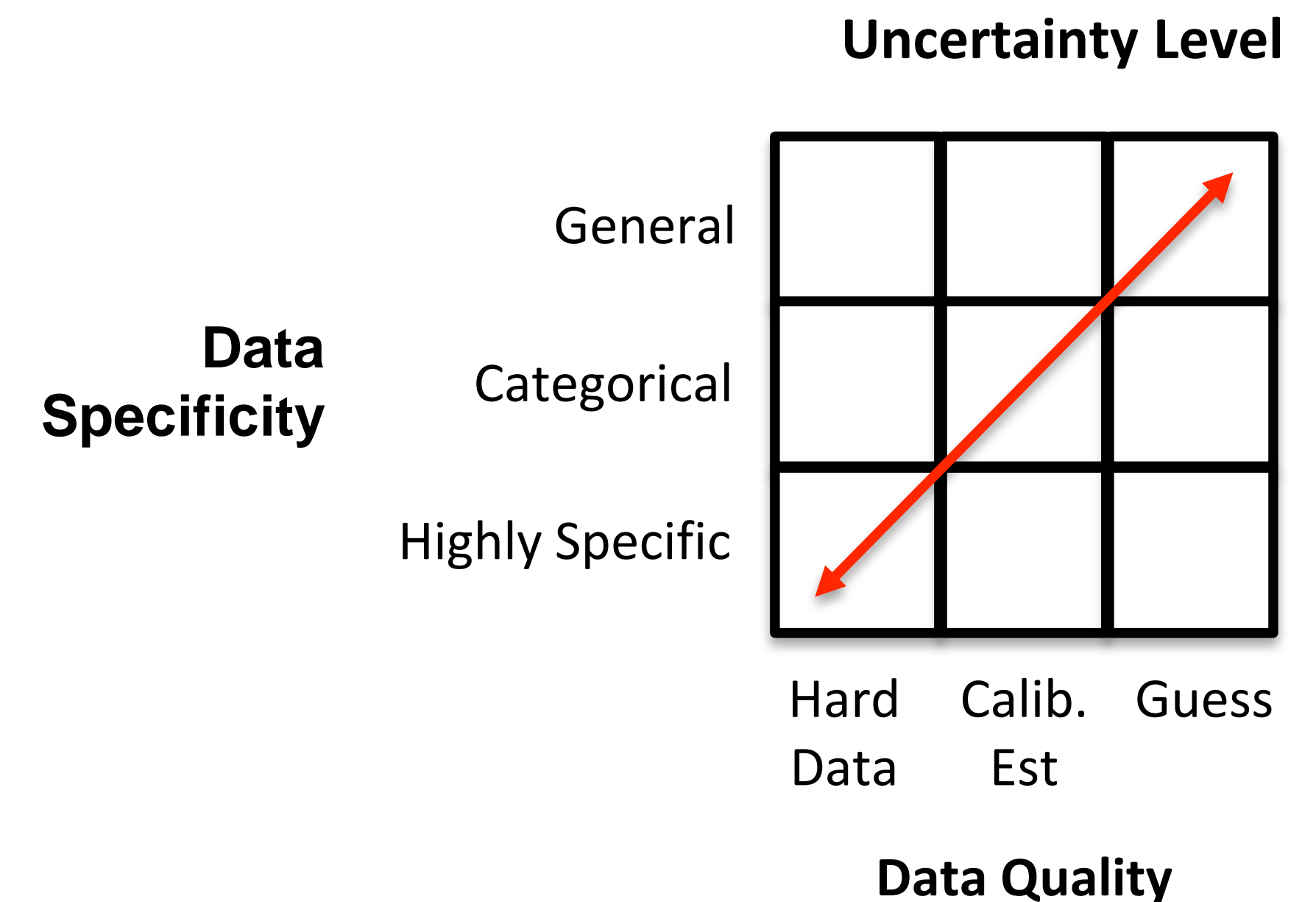
- Maybe all you have available to you are very low frequency data points, or your visibility is limited
- Estimates can be very high quality if:
 - They really are subject matter experts
 - The estimates are calibrated
- Should always be represented as ranges/distributions to reflect uncertainty



Incidents



- Incident data is rarely gathered or used effectively
- Can be an amazing data resource, particularly over time:
 - Actual losses
 - Response costs
 - Recovery costs
 - Lost revenue
 - Notification costs
 - Fines/judgments
 - Threat vectors
 - Control efficacy/deficiencies
 - Root causes



RSA®Conference2018



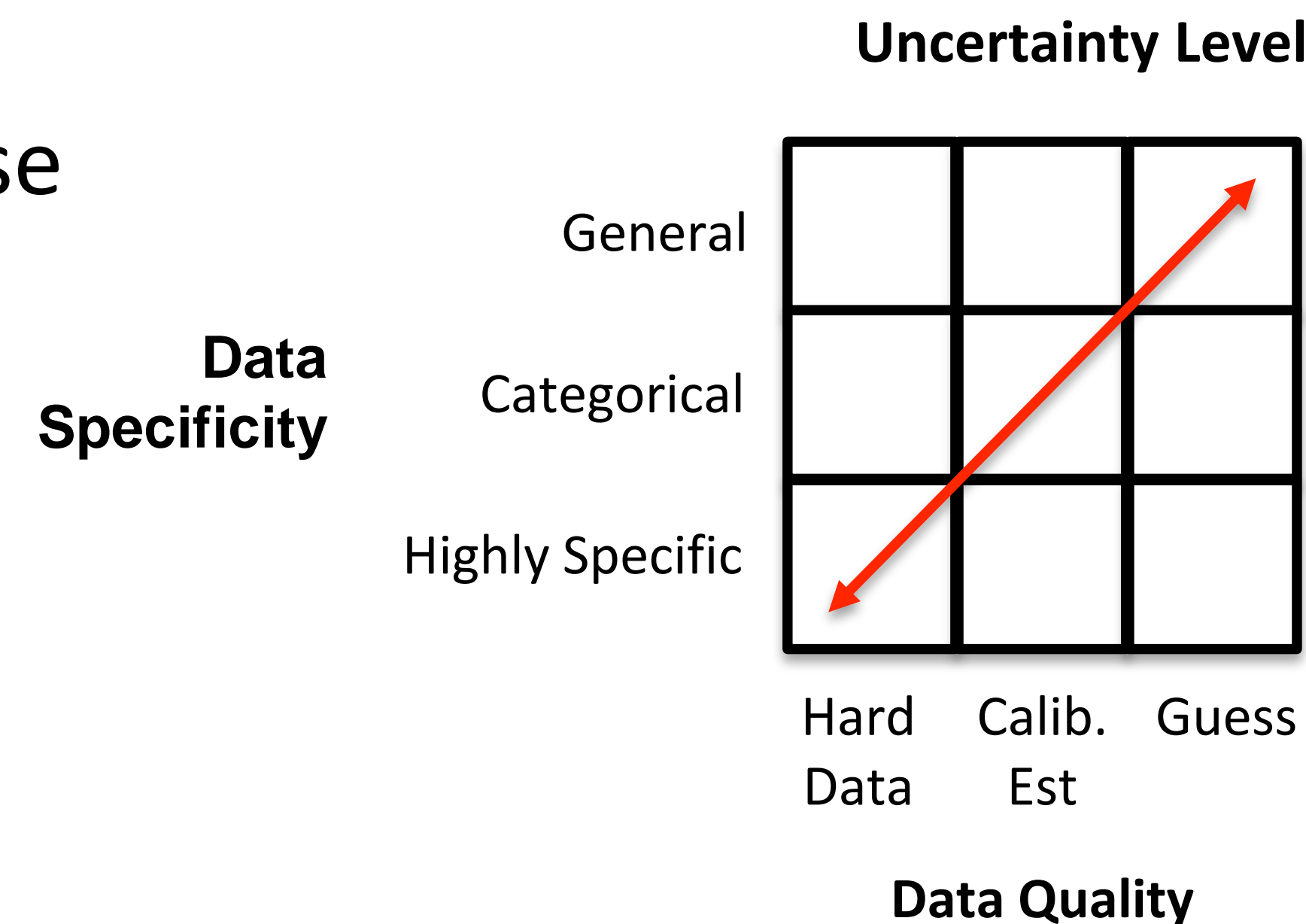
#RSAC

EXAMPLE

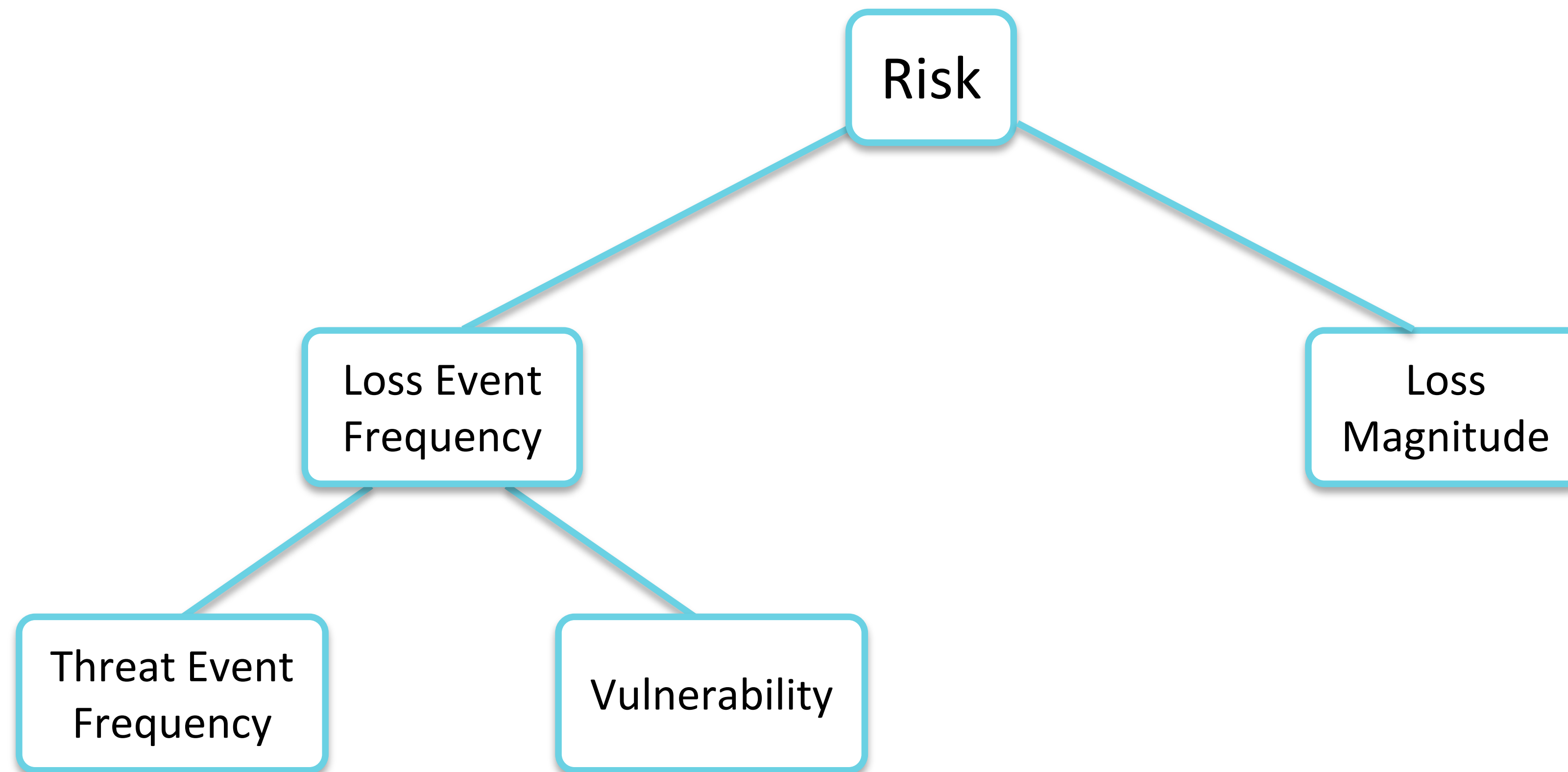
Phishing



- Your latest data shows
 - 15% of recipients are falling for simulated phishing e-mails
 - The organization is experiencing over 1000 phishing attacks per year, which is a 50% increase over the previous year
- So what?
- What decisions rest on this information?



The FAIR model (partial)

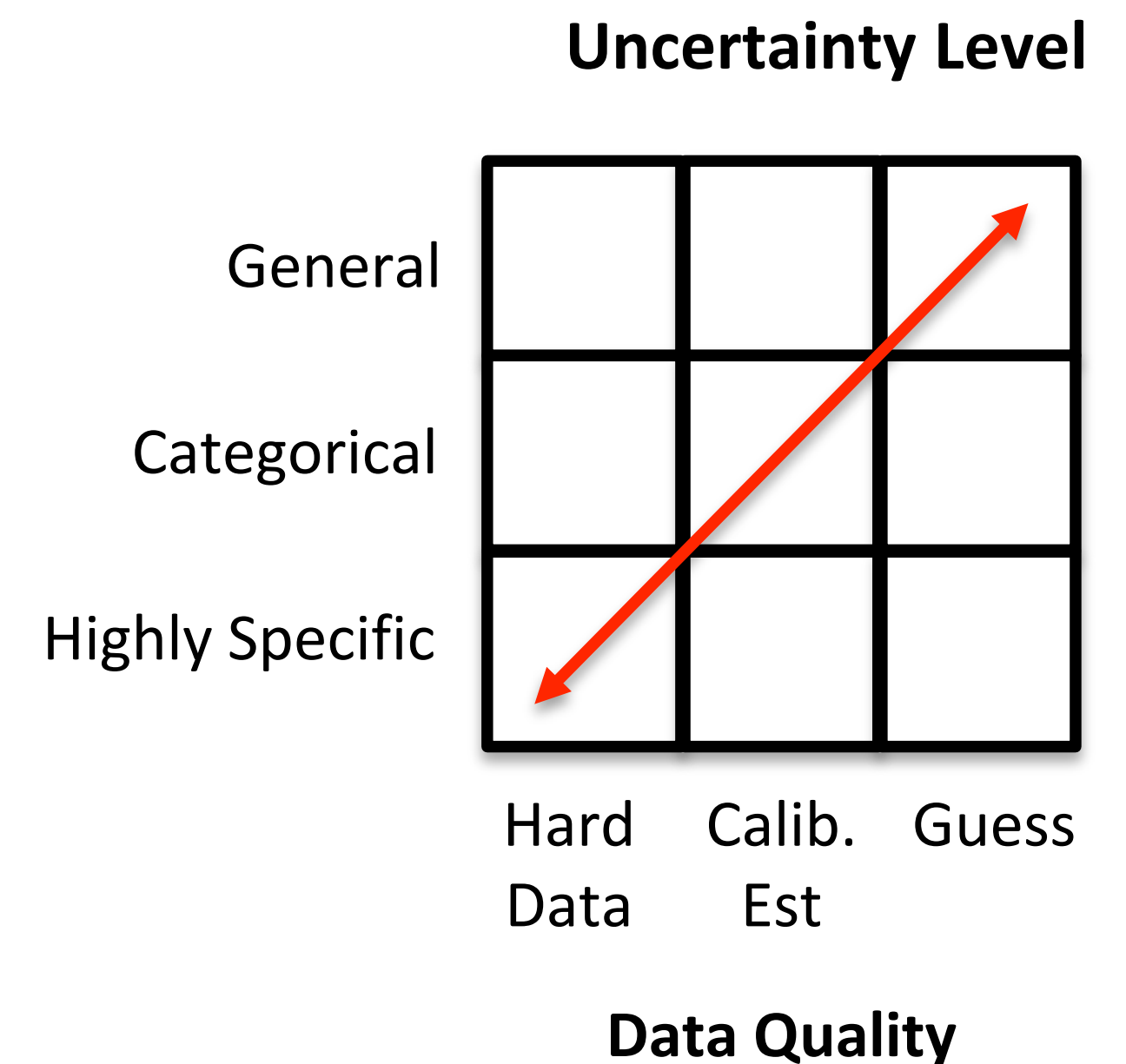


Phishing - the specifics matter



- Threat event frequency
 - **General:** E-mail phishing attacks
 - **Categorical:** E-mail spear phishing attacks
 - **Highly Specific:** E-mail spear phishing attacks where evidence suggests high sophistication and/or attempts to penetrate deeper (if compromise occurred)
- Vulnerability % (probability of user compromise)
 - **General:** All personnel
 - **Categorical:** Key personnel (or a department)
 - **Highly Specific:** A specific person
- Loss magnitude (\$\$\$)
 - **General:** Compromise of IP in general
 - **Categorical:** Compromise of specific IP type
 - **Highly specific:** Specific loss implications from specific IP compromise

**Data
Specificity**



RSA®Conference2018



#RSAC

THINGS TO WATCH OUT FOR

Vendor red flags...



- Risk quantification is becoming a bigger deal every day, which means vendors are climbing onboard in their marketing
 - “Take the human out of the equation”
 - “Use empirical data rather than random data (e.g., Monte Carlo)”
 - “Fully automated”
- Any tool that “scores risk” is using a model
 - What model(s)? Open? Proprietary?
 - What are their analytic assumptions?
 - What is their tool’s scope — i.e., what scenarios do/don’t they cover?

Models vs. frameworks



- NIST CSF, PCI DSS, FFIEC CAT, etc.,
 - Are NOT risk measurement models
 - ARE good practice frameworks
 - Provide data on control conditions, which can be used in risk measurement
- Results/gaps are interpreted by practitioners for the “so what”
 - Reliable interpretation requires the two models mentioned earlier
 - Analysis scope (assets, threats, vectors, event type, control conditions)
 - A risk analysis model (e.g., formula)

RSA®Conference2018



#RSAC

WRAPPING UP

The bottom line...



- We're drowning in some data, and have almost none of other data
- Regardless, we still have to prioritize as effectively as possible
- Uncertainty is unavoidable — no amount of data changes that
- Data, models and measurement is about reducing uncertainty
- Data quality is more important than data quantity
- Waiting for “enough” data is foolish
- Data without models is useless
- If you want to have a truly data-driven security/risk program, you need to understand the fundamentals regarding data and models



- Books
 - How to Measure Anything in Cyber Security Risk (Hubbard, Seiersen)
 - Measuring and Managing Information Risk: A FAIR Approach (Jones, Freund)
- White paper
 - Effectively leveraging data in FAIR analyses (FAIR Institute member resources)
- Web
 - The FAIR Institute blog (www.fairinstitute.org/blog)
 - The Open Group (www.opengroup.org)
 - Cyentia Institute (www.cyentia.com)

Applying What You Have Learned Today



- Next week you should:
 - Begin to leverage the resources mentioned previously
- In the first three months following this presentation you should:
 - Identify your organization's most important assets
 - Triage the data quality for those assets
 - Become trained in calibrated estimation
- Within six months you should:
 - Adopt or define a risk model for your organization (e.g., FAIR)
 - Evaluate the models your vendor products use to do risk/severity "scoring"
 - Evolve your security metrics to become risk metrics (add the "so what")

RSA®Conference2018



#RSAC

QUESTIONS?