BIG DATA CALLS FOR BIG SECURITY!

Branden R. Williams
RSA Security

Jason Rader
RSA Security

Session ID: DSP-F41

Session Classification: Intermediate

WHAT IS THIS SESSION ABOUT?

- Big Data Analytics
  - They are the future!
  - How do we deal with them in the security world?
- Unique Security Concerns With Big Data
  - How can we learn from other security issues?
  - What snakes may be lurking in the grass?
- Applied Knowledge
  - A scenario!
  - What to do?
  - DO IT!

PREDICTIVE ANALYTICS – THE FUTURE OF BI

- Big Data
  - Not LOTS of data
  - But many disparate sources with mixed structure
- Industry trends driving Big Data & analytics
  - Storage is cheap
  - “Anywhere” computing
  - Hardware over-engineering
  - (or distributed computing)
- Math is pretty awesome

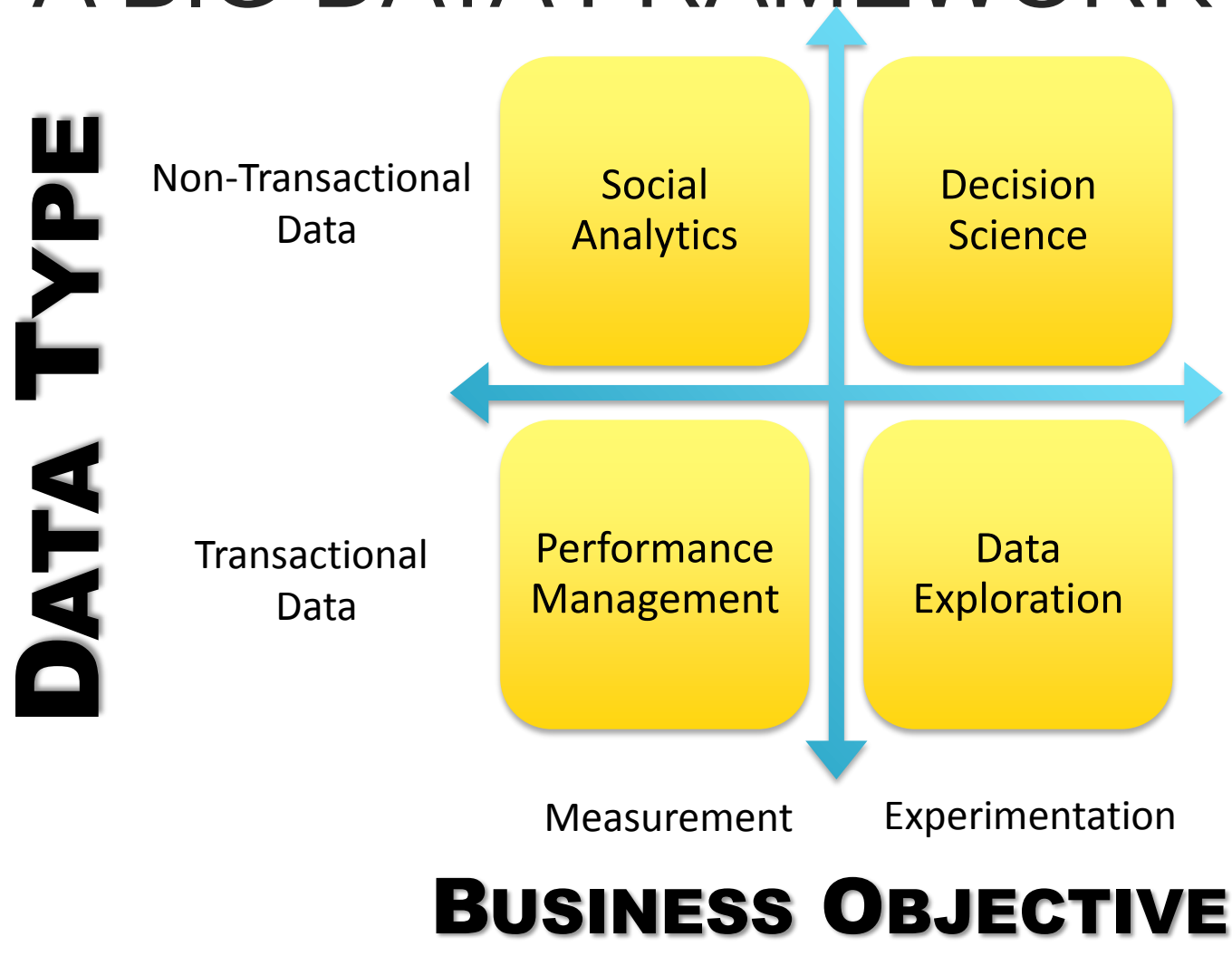
WHAT ARE PREDICTIVE ANALYTICS?

- Simply, deriving value from big data!
- Complexly (is that a word?)
  - Organizing the unorganized
  - Heavy math use
  - The search for leading indicators
- How do we use it?
  - Business strategy
  - Public safety (CDC,NOAA)
  - Predict the future

WHO DOES THIS WELL TODAY? INVESTORS!

- Massive historical data
- Lots of math
  - Sometimes bad math
  - The search for trends
  - Execute trades!
- The next step?
  - Implications of social media
  - See through the corporate veil
  - Now we can focus on the individual!

A BIG DATA FRAMEWORK



Parise, S., Iyer, B., & Vesset, D. (2012). Four strategies to capture and create value from big data. *Ivey Business Journal*, *76*(4), 1-5.

PREDICTIVE ANALYTICS – THE WHY

- Businesses have data coming out of their ears
- SO DO THE BAD GUYS!
- Huge shift from the past
  - Forensics example
- Unused data
- Can we find value?
- Business leaders must innovate[1]
  - But focus on business model innovation
  - PA can drive this!

[1] Amit, R., & Zott, C. (2012). Creating value through business model innovation. *MIT Sloan Management Review, 53*(3), 41-49.

DERIVED DATA & ANALYTICS

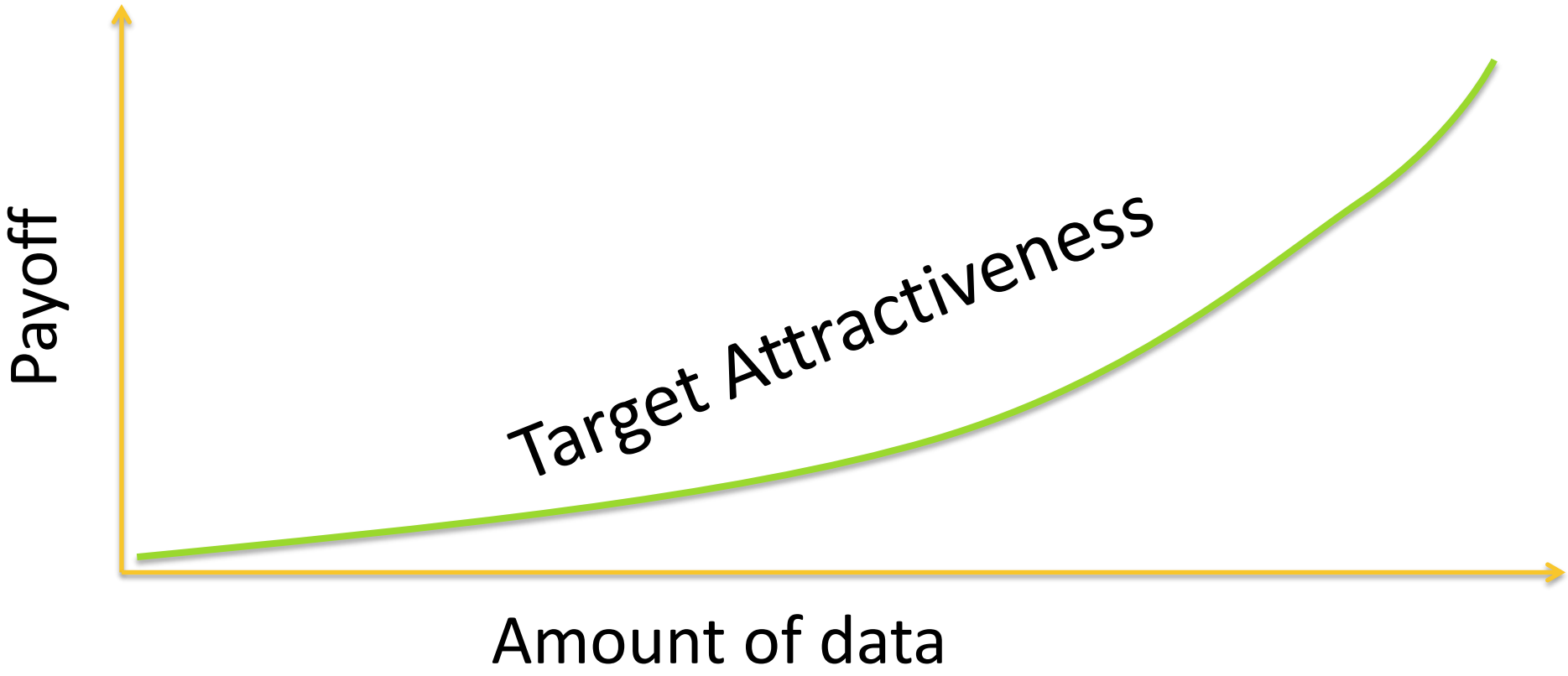
(MORE SENSITIVE THAN ORIGINAL DATA!)

- Analytics automate “data crunching”
- Could become the new IP!
- Businesses act on output
- What about deriving sensitive info?
  - PII
  - Privacy issues!
- If I accurately guess your info, do I need to protect it?

PROBLEM OF DERIVED SENSITIVE DATA

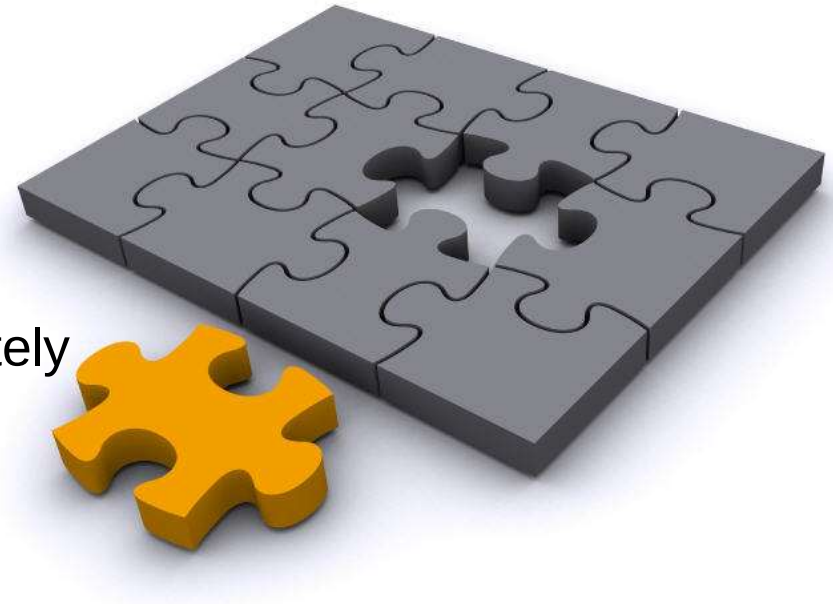
- If I derive your PII, must I protect it?
- Law is WAY behind
- Compliance might be farther
- Do I have to protect data from aggregation?
  - Flickr EXIF Reveals Client li
  - Geotags
- Huge impact to business
  - US = Case Law
  - EU Privacy Updates

THE RELEVANCE OF DATA MASS



REAL STORIES OF BIG DATA FUN

- How Target knew a teenager was pregnant before her own father did:
  - Pause for story time!
- Implications:
  - Target can predict due dates/sex
  - Aggregations of data can accurately predict the future
  - Better placement (upside?)
  - Price discrimination
- Not Singular pieces of data!
  - It's the CORRELATION that matters



THE COMPARTMENTALIZATION TECHNIQUE

- Big data requires big flexibility
  - But that impacts security
  - Think about SQL features:
    - Views
    - Complex Queries (with aggregate
    - Column-level RBAC
- Classic approaches (where have we solved this problem before)
  - DBAs still a risk (so is the Data Scientist)
  - Operators may not know what they are looking at

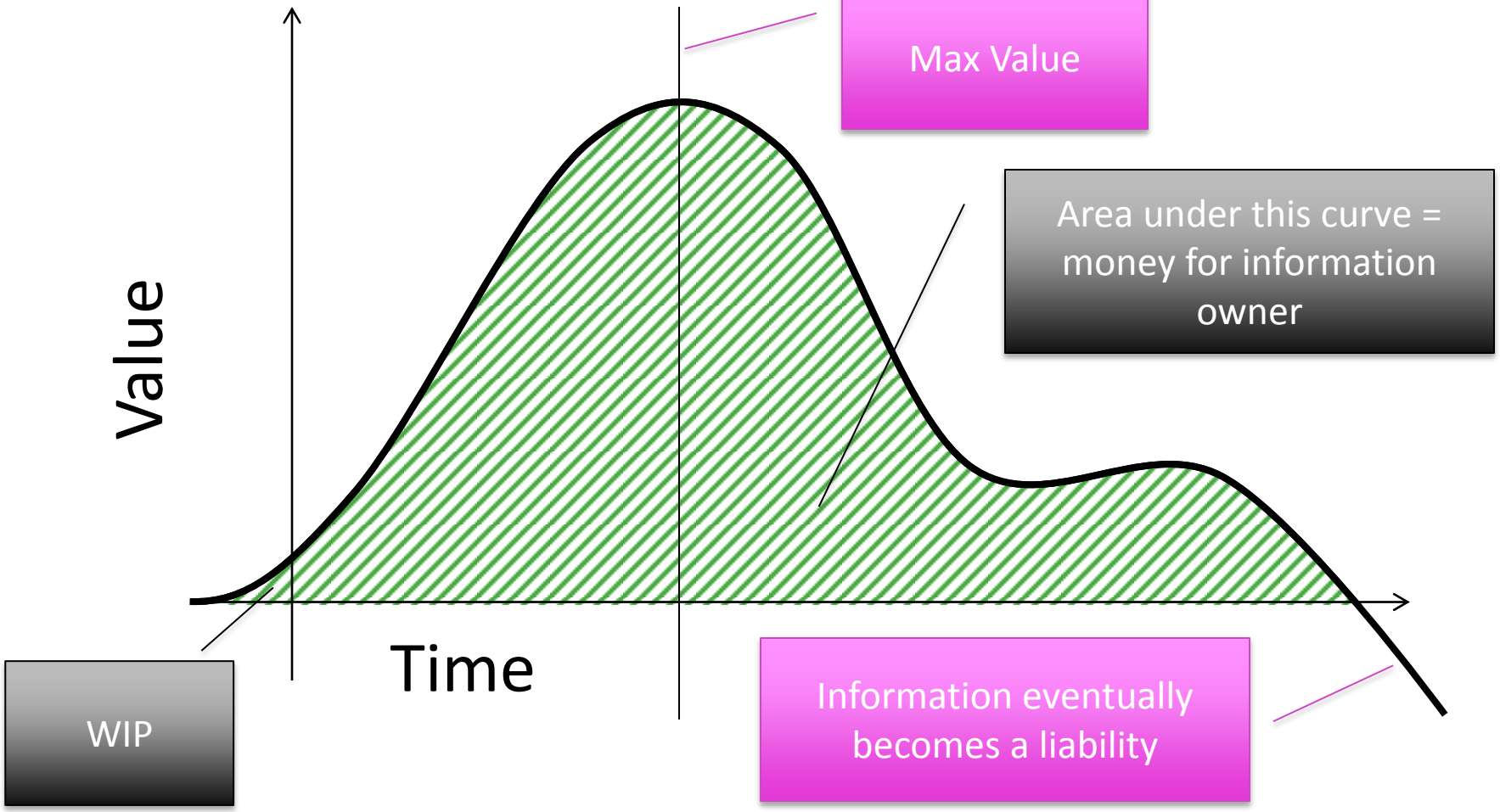
DATA SYNTHESIS TECHNIQUES

- Hacking the Coca Cola formula
  - Can you derive the formula?
  - YES YOU CAN!
  - Shipping/Receiving/Production data
    - Think Crypto plaintext attack
  - KFC would be harder, but possible!
- IP/Strategy Derivation
  - Attackers can figure out your IP by looking at data sets you don't necessarily protect
  - Macro trends repeat themselves over multiple sets of data, can I fill the gaps from missing data sets?

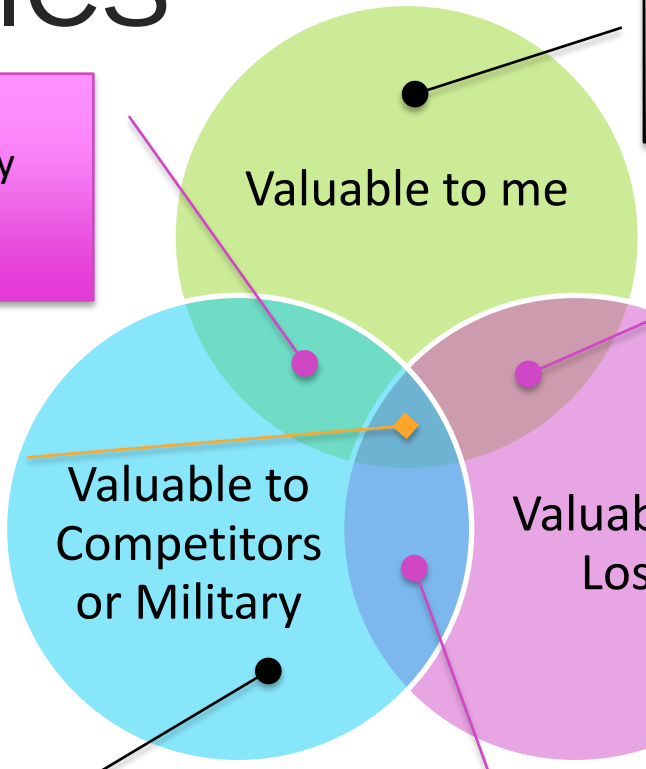
PROTECTION SCHEMES – WHERE TO START?

- Data mapping
  - Where does data live?
  - What is the full lifecycle of data
    - Raw
    - WIP
    - Data product
    - Disposal
  - How is it used by the business?
- Common Practices (with caveats)
  - Governance
  - Separation of Duties/RBAC
  - Dev/UA/Prod

THE VALUE OF INFORMATION OVER TIME



HOW TO VIEW DATA AND ANALYTICS



Customer Analytics
IT Configs
Biz Processes

Secret Sauce
Intellectual Property
Software Vuln DB
Corp Strategy

Derivative Data
Analytics for Sale
Medical Records

Crown Jewels
Easily Transferrable IP
Actionable IP
Encryption Keys

CC Data
PII/PHI Data
Unused Biz Data
Disinformation

COMPINT
Defense Information

Old Source Code
Old IP
Old/Retired Encryption Keys

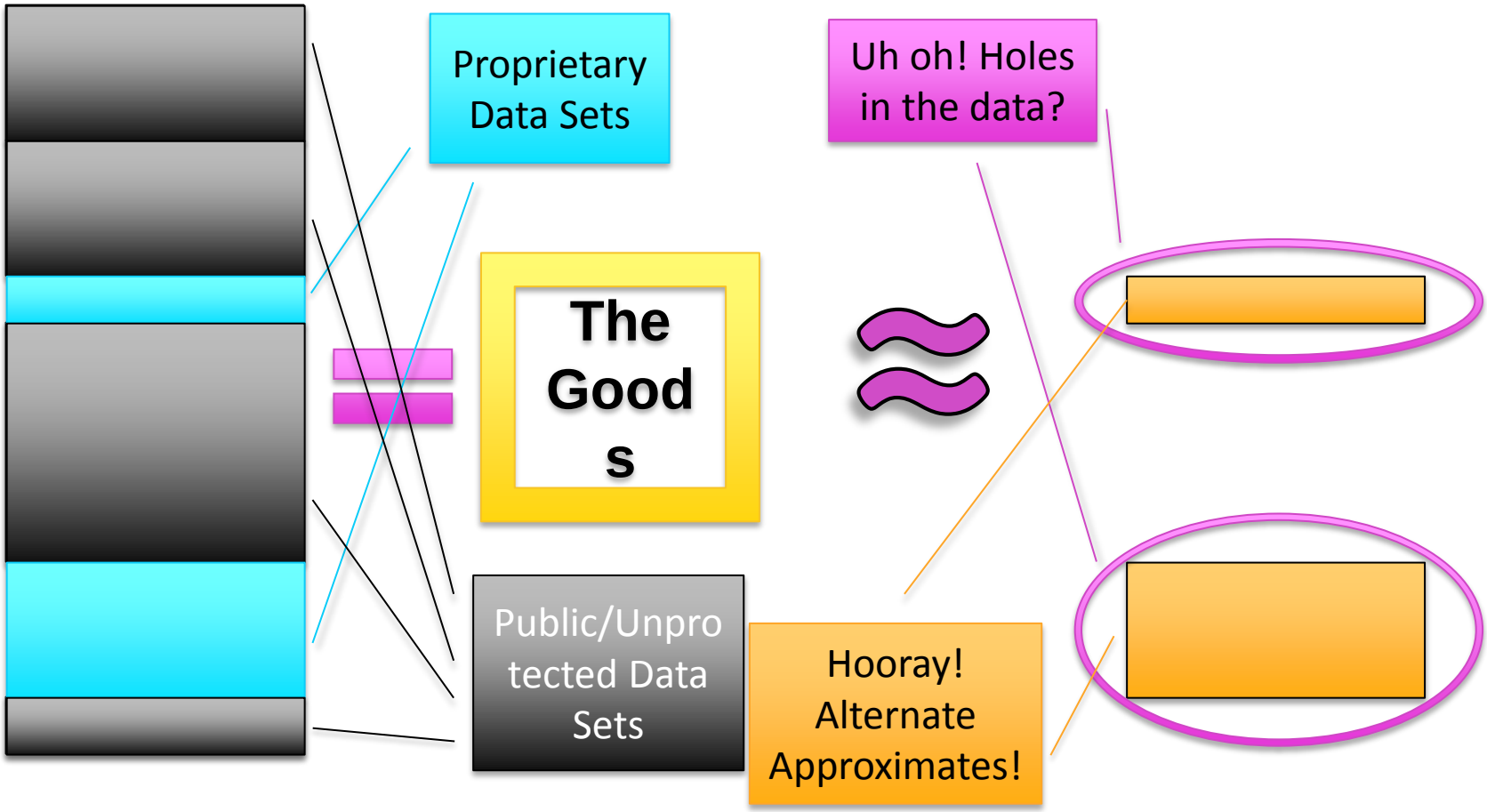
SCENARIO

- Taco Corp is evaluating land for leasing of mineral rights
  - They have proprietary data consisting of their own costs/profits related to past extractions
  - This data is critical in understanding when a particular lease becomes profitable, trigging the process of extraction
- Two critical assets:
  - Expected mineral output from plots of land
  - Profitability triggers
- Potential public info:
  - Existing leases with no extraction
  - Leases with active extraction
  - Commodity market for closing price of extracted minerals

SCENARIO (cont)

- Competitor (or real estate agent!) wants to get a jump on key land leases
- While they don't know the expected mineral outputs or the profitability triggers, can they synthesize?
  - May know approximate timing of extraction through permits, media
  - Can look up historical commodity valu
  - Search for patterns or thresholds
  - Derive profitability approximations
- Constantly refine over time
- Get the jump on Taco Corp!

MOCKUP OF DATA SYNTHESIS



WHAT DID WE LEARN?

- Large analytical computations with diverse data sets can be reconstructed (think RAID)
- What needs to be protected?
  - Inputs & data sources
  - Analytics process (“pre-IP”)
  - Outputs (duh)
- Analytics process bonus:
  - Learn where juicy data lives!
  - Understand what can be substituted!
  - What output is important to mgt!
  - New third party data could mean M&A

Q/A

Branden Williams
@BrandenWilliams

Jason Rader
@JasonRaderRSA

THANK YOU!