http://bizbench.com/wisdom-pyramid/

https://www.expertsystem.com/government-data-mining/

Bringing Data to Action

Insights Discovery at the Intersection of Multiple Data Sets

Kirk Borne

@KirkDBorne

Principal Data Scientist, Booz Allen Hamilton

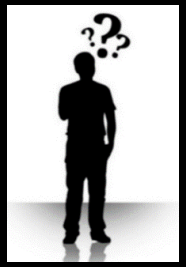
http://www.boozallen.com/datascience

What is Big Data's Biggest Challenge?

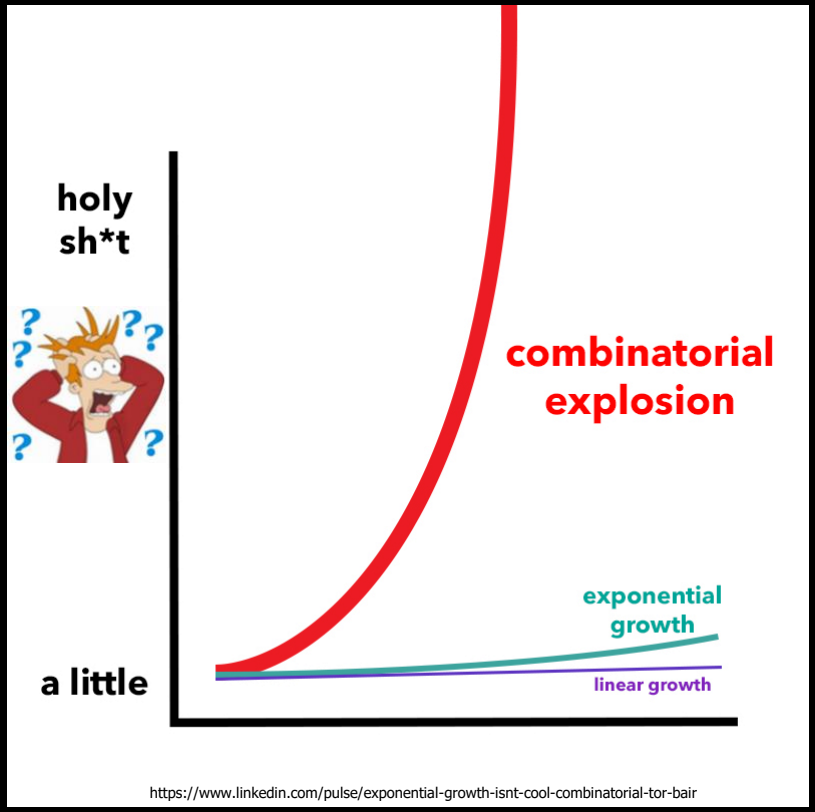
Hint: it's not Volume. Answer: it's Variety!

1. Every organization collects many different sources of data.
2. These multiple diverse data sets are often stored in separate silos (**).
3. Silos inhibit data science teams from integrating multiple data sets that (when combined) can yield deep, actionable insights to create value.
4. Diverse teams that can share and explore diverse data sets across an agile data lake have the power to change that entire story!

(**) Consider the Blind Men and the Elephant …



Ever since the beginning of humans, we have been curious…
So, we have collected evidence (data) to answer our questions,
which leads to more questions, which leads to more data collection,
which leads to more questions, …, which leads to **BIG DATA!**



https://www.linkedin.com/pulse/exponential-growth-isnt-cool-combinatorial-tor-bair

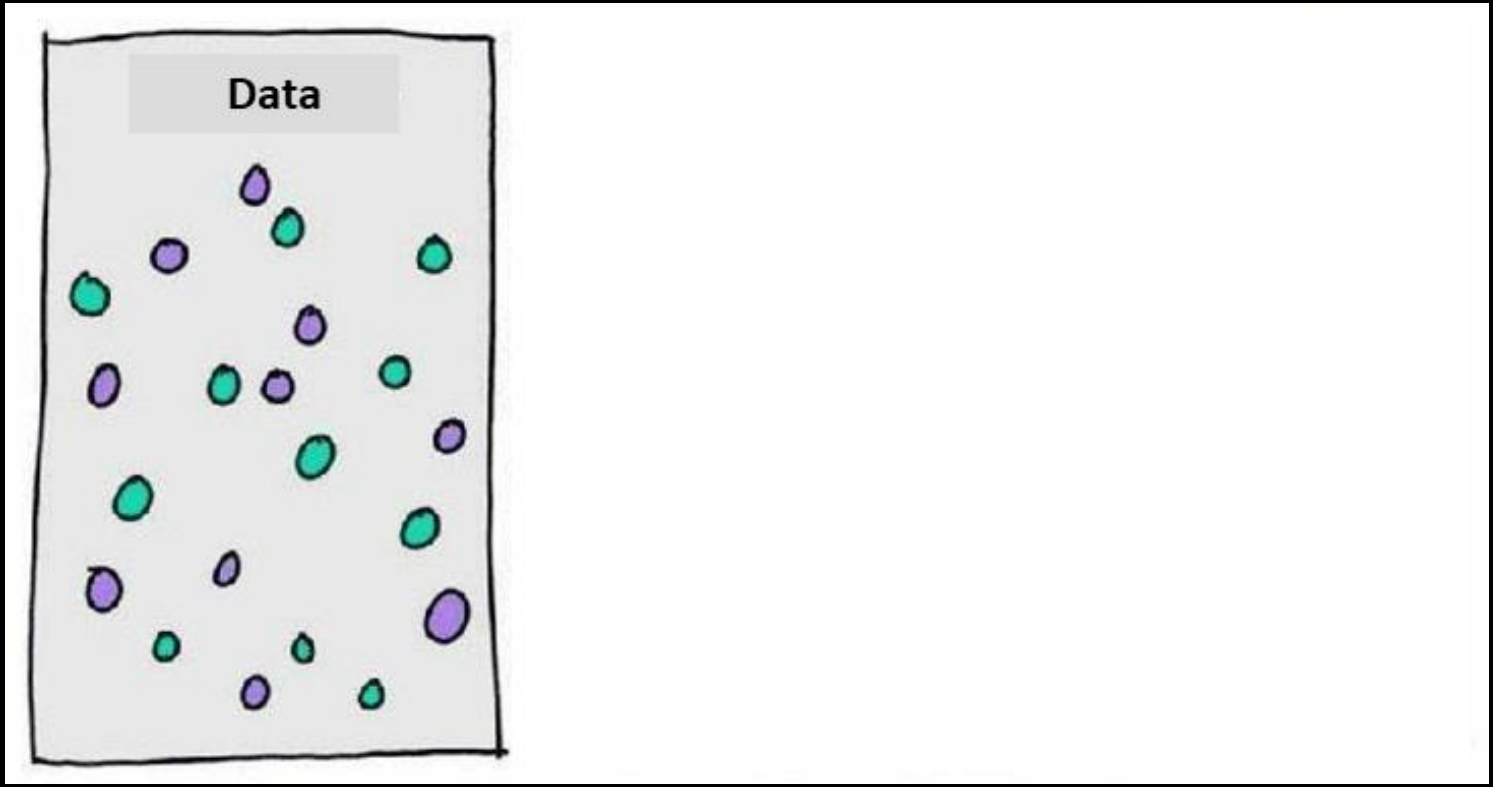
Knowledge is about connecting the dots.
@KirkDBorne

$y \sim x! \approx x \wedge x$
→ Combinatorial Growth!
(all possible interconnections, linkages, and interactions)

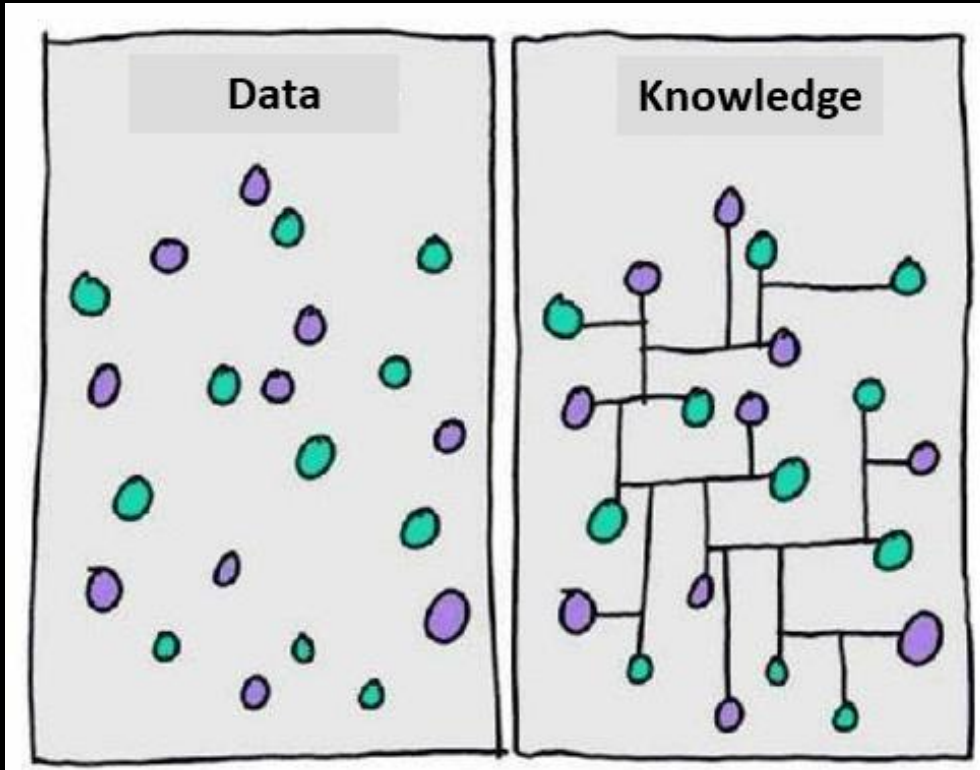
$y \sim 2 \wedge x$ (exponential growth)

$y \sim 2 * x$ (linear growth)

Data



Data **Data Science**



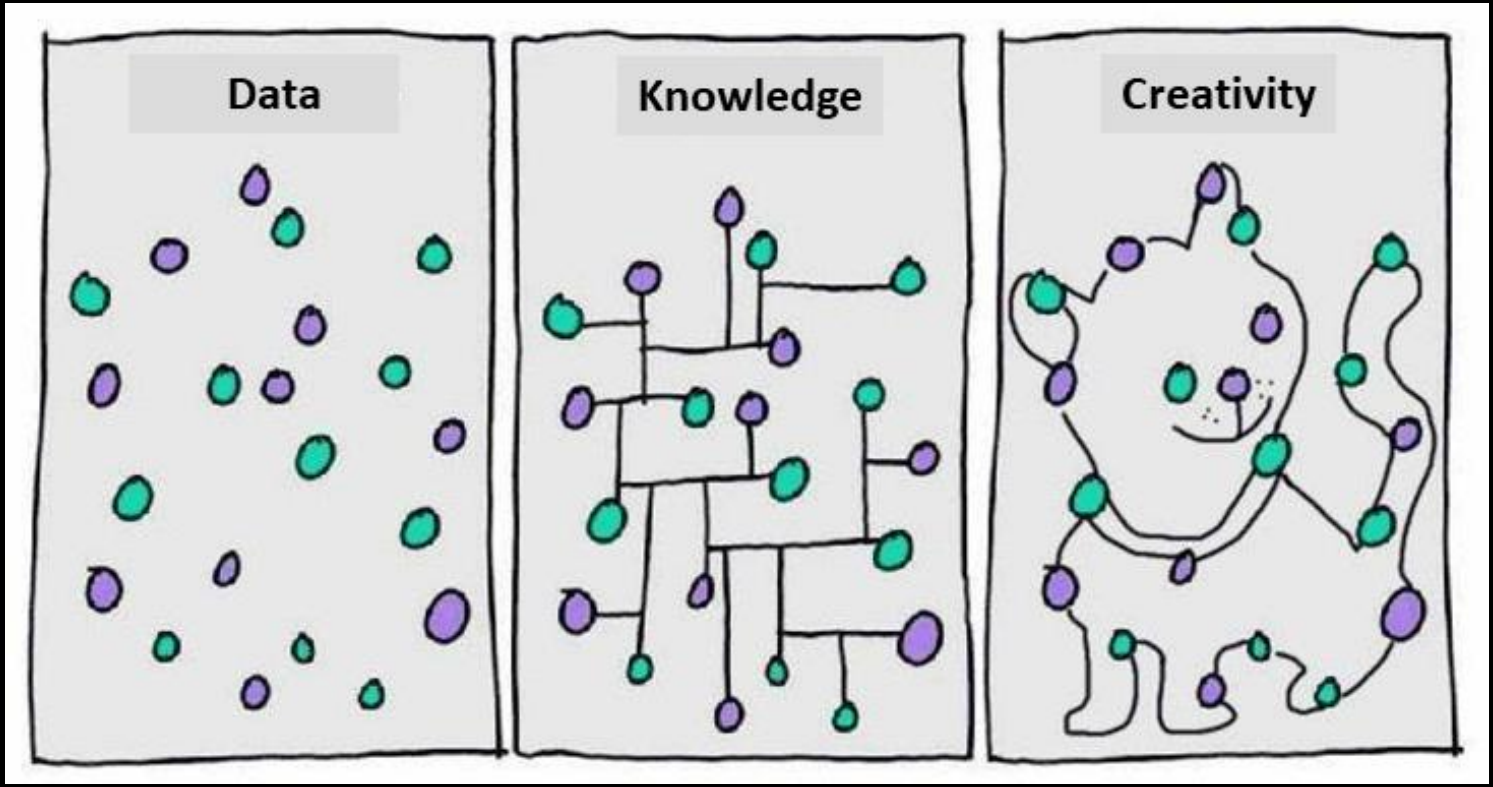
Knowledge is about connecting the dots.

@KirkDBorne

Data

Data Science

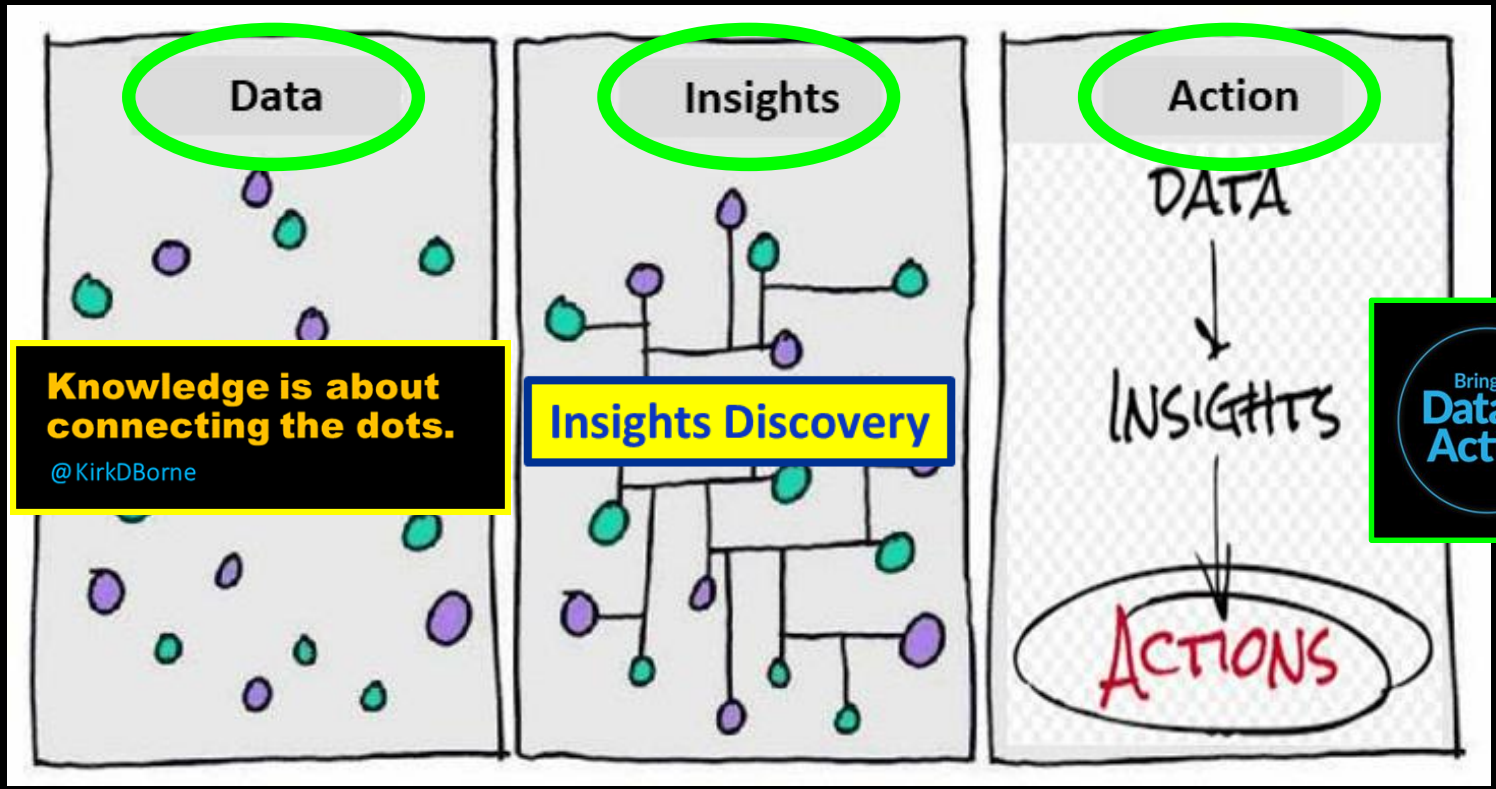
(Innovation and Value Creation)
Design Thinking



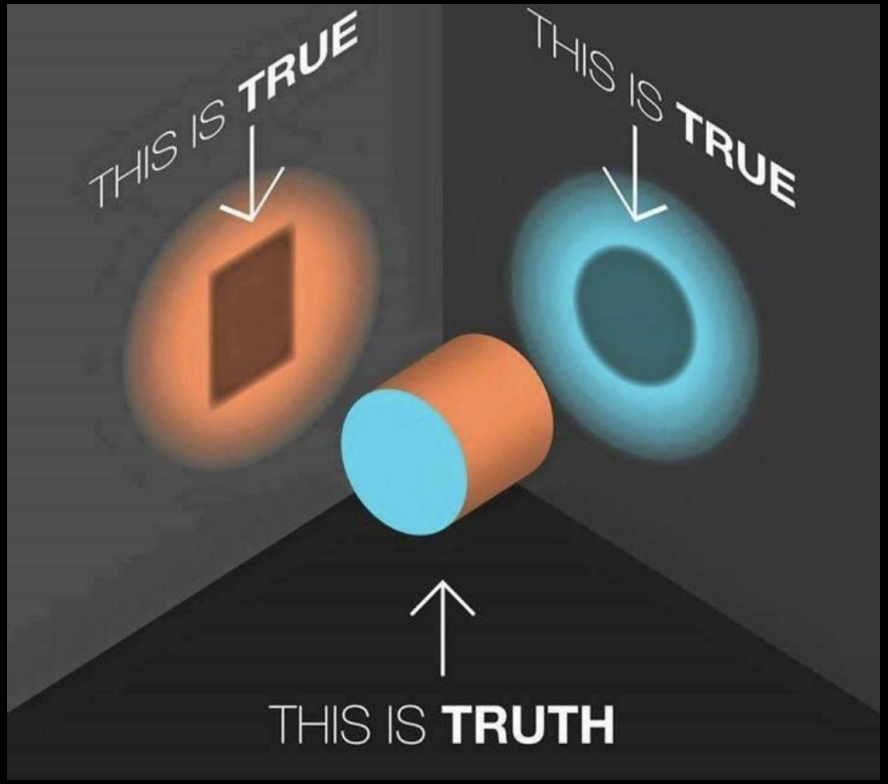
http://www.boozallen.com/datascience

@KirkDBorne

The Input (fuel)	The Engine	The Output
Data	Data Science	A. I.



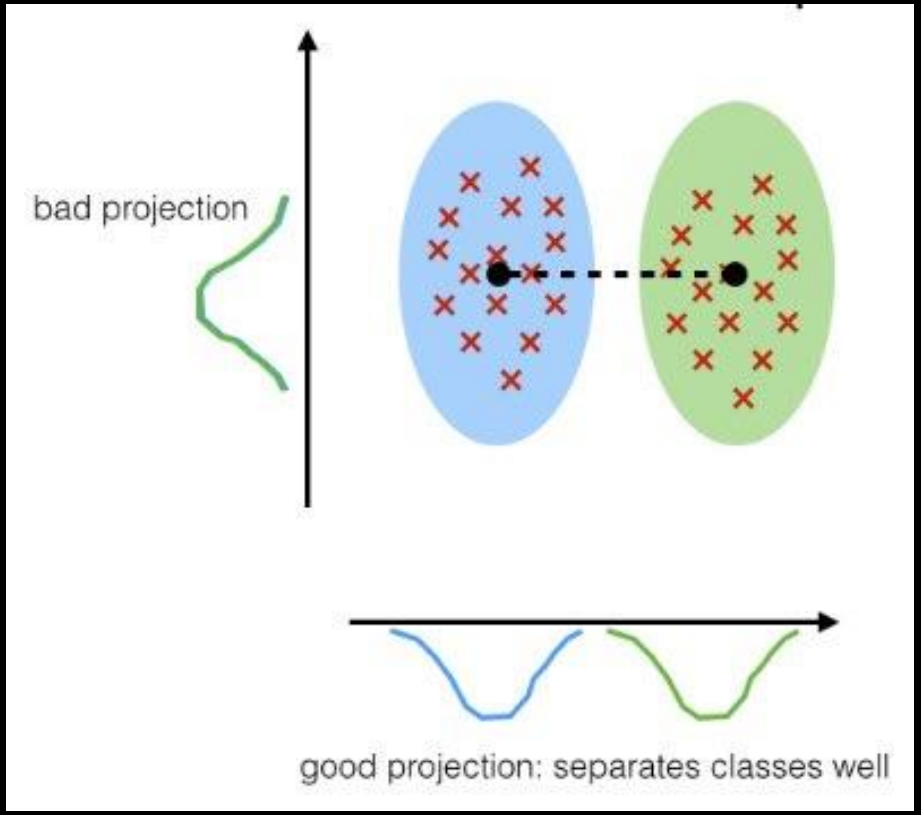
High-Variety Data can be a Bias-Buster. Projection Matters! ➡ https://bit.ly/2CGHZjN



Source: http://www.transformativeinsights.co.nz/blog/new-perspective-on-conflict

Your chosen data attributes represent a low-dimension projection of the full truth – the feature space (dimensions) in which you explore your data is a form of cognitive bias – … **it matters!**

Feature Selection and Projection

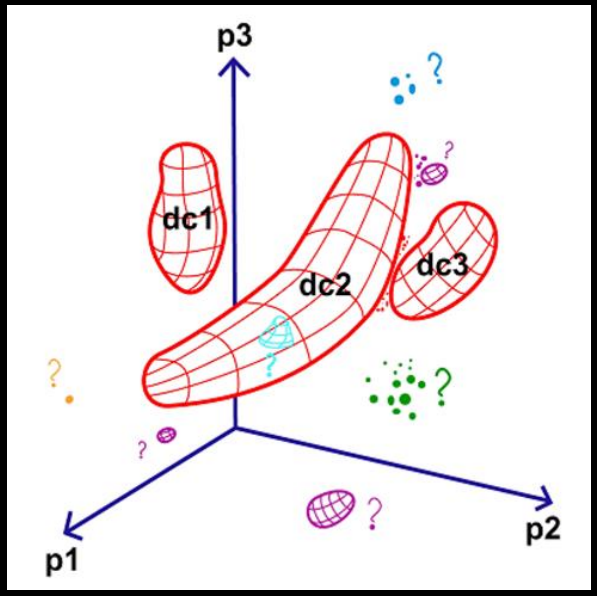


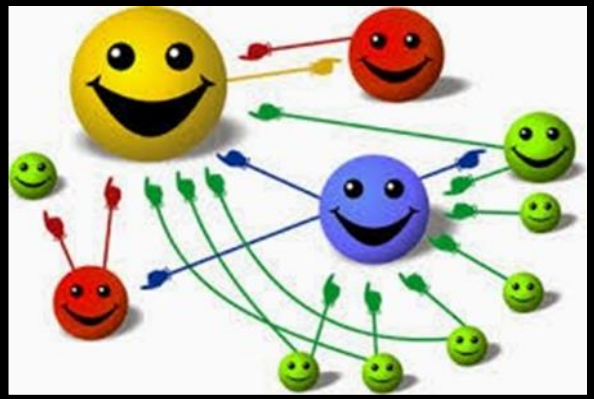
Source: https://www.quora.com/How-was-classification-as-a-learning-machine-developed

Feature Selection is important in order to disambiguate different classes.

More importantly, **Class Discovery** depends on choosing the right projection and selecting the right features!

4 Types of Insights Discovery from Data:





1) **Class Discovery:** Find the categories of objects (population segments), events, and behaviors in your data. + Learn the rules that constrain the class boundaries (that uniquely distinguish them).
2) **Correlation (Predictive and Prescriptive Power) Discovery:** (insights discovery) – Find trends, patterns, dependencies in data that reveal the governing principles or behavioral patterns (the object's "DNA").
3) **Outlier / Anomaly / Novelty / Surprise Discovery:** Find the new, surprising, unexpected one-in-a-[million / billion / trillion] object, event, or behavior.
4) **Association (or Link) Discovery:** (Graph and Network Analytics) – Find both the usual and the unusual (interesting) data associations / links / connections across the entities in your domain.

Levels of Analytics Maturity in Data-Driven Applications

1) Descriptive Analytics

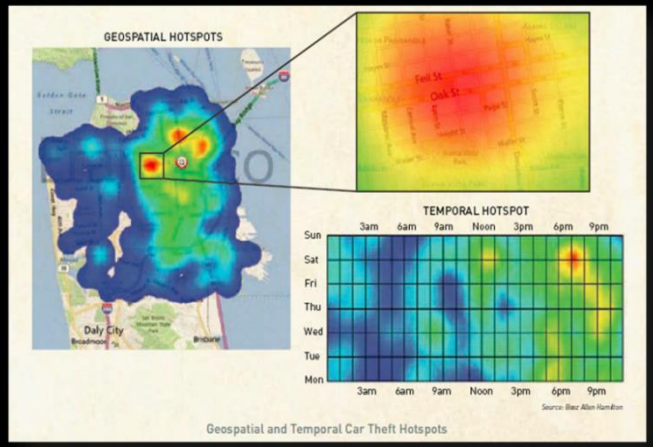
– **Hindsight** (What happened?)

2) Diagnostic Analytics

– **Oversight** (real-time / What is happening? Why did it happen?)

3) Predictive Analytics

– Foresight (What will happen?)



Geospatial and Temporal Car Theft Hotspots

5 Levels of Analytics Maturity in Data-Driven Applications

1) Descriptive Analytics

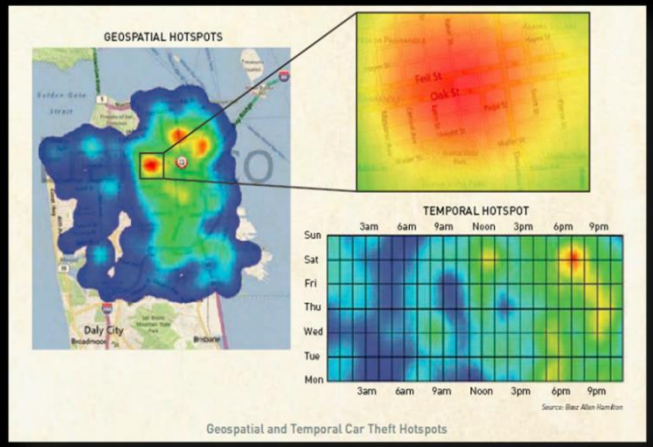
– **Hindsight** (What happened?)

2) Diagnostic Analytics

– **Oversight** (real-time / What is happening? Why did it happen?)

3) Predictive Analytics

– Foresight (What will happen?)



Geospatial and Temporal Car Theft Hotspots

4) Prescriptive Analytics

– **Insight** (How can we optimize what happens?) (Follow the dots / connections in the graph!) **Insights Discovery**

5) Cognitive Analytics

– **Right Sight** (the 360 view , **what is the right question to ask for this set of data in this context** = Game of Jeopardy)

– Finds the right insight, the right action, the right decision,… right now!

– Moves beyond simply providing answers, to generating new questions and hypotheses.

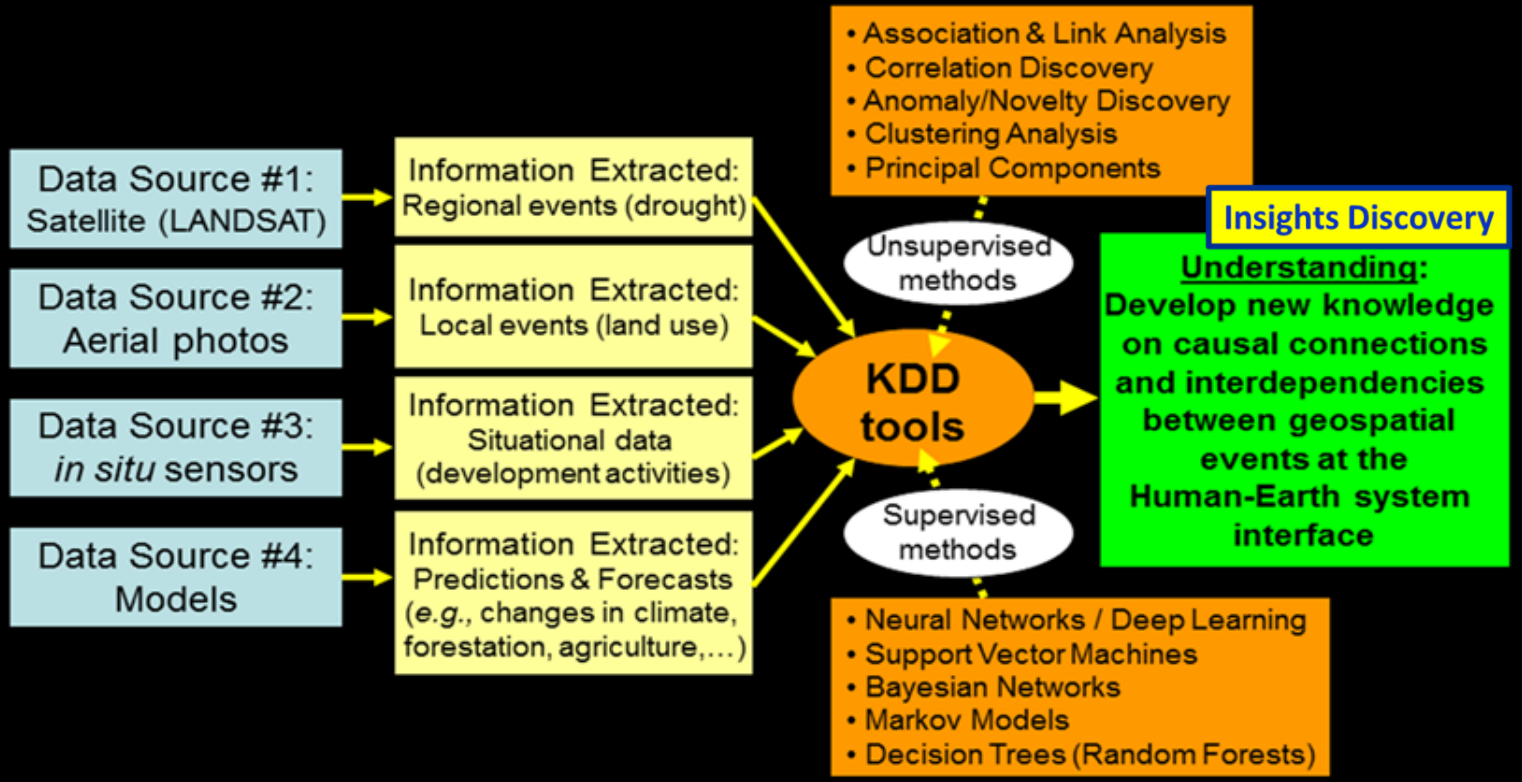
An "Easy Button" for Extracting Value from Data through Data Science and AI

- Pattern Discovery (Detection)
  - D2D: data-to-discovery
- Pattern Recognition (Learning)
  - D2D: data-to-decisions
- Pattern Exploration (DataOps)
  - D2D: data-to-dollars (innovation)
- Pattern Exploitation (Agile Deployments)
  - D2V: Data-to-Value (actionable insights)
  - D2A: Data-to-Action (value creation)

GEOSPATIAL DATA SCIENCE USE CASE: ENVIRONMENTAL SCIENCE

From Data to Information to Knowledge to Understanding

Early Warning and Monitoring Systems for Geospatial Event Discovery



The Real Power of A.I. comes through insights from many different sources (multiple data sets) – – There is nothing "artificial" about it!