

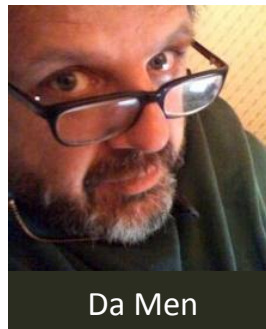


Comments, panel session: ICSE'14: “After the gold rush”

tim.menzies@gmail.com
Dagstuhl workshop on SA
June 23, 2014



Da bird



Da Men



Da Mann

Organized by....

Welcome to the Wild West

- Gold rush:
 - To extract nuggets of insight.
- Too many inexperienced “cowboys”:
 - No use of best or safe practices.
- “Goldfish bowl” panel:
 - Best practices, what not to do



Street light effect



"I'm looking for my keys."

- Moral #1:
 - look at the real data, not just the conveniently available data
- “Before we collect the data, need to redefine the right data to collect.”
- “Garbage in, garbage out”
- “Before analyzing terabytes of data, reflect some on user goals.”

Can we reason about data, without detailed background knowledge?

- Traditional view:
 - GQM
 - Traditional science:
 - Define data to collected
 - Collect
 - Infer
- “Newer” view:
 - Operational science (Mockus, keynote, MSR, 2014)
 - Find data
 - Reason about it
 - Prone to “streetlight effect”



Empiricists : Rationalists

- Observation : use of background knowledge
 - Mockus : Basili
 - Norvig : Chomsky
 - Locke : Leibnitz
 - Aristotle : Plato
 - Newton : Descartes
 - Ike said:
 - “I make no hypotheses.”
 - “I feign no hypotheses.”

Working in the light



*"I'm exploring the data
without pre-conceived biases."*

- Moral #2:
 - “before we rush to the new, lets reflect on what we can learn from what we can see right now. “
- “Too fast, too early to expect actionable data for some specific area. “
- “Building models on what we have before pushing ahead.”
- “Need to invest more in data science infrastructure”

Full disclosure: Menzies is not rational (he's an empiricist)

Wikipedia lists of human cognitive biases

- 100+ entries
 - The way we routinely get it wrong, every day.

Accidental discoveries

- America
- Penicillin
- Anesthesia
- Big bang radiation
- Internal pacemakers
- Microwave ovens
- X-rays
- Safety glass
- Plastics (non-conductive heat resistant polymer)
- Vulcanised rubber
- Viagra

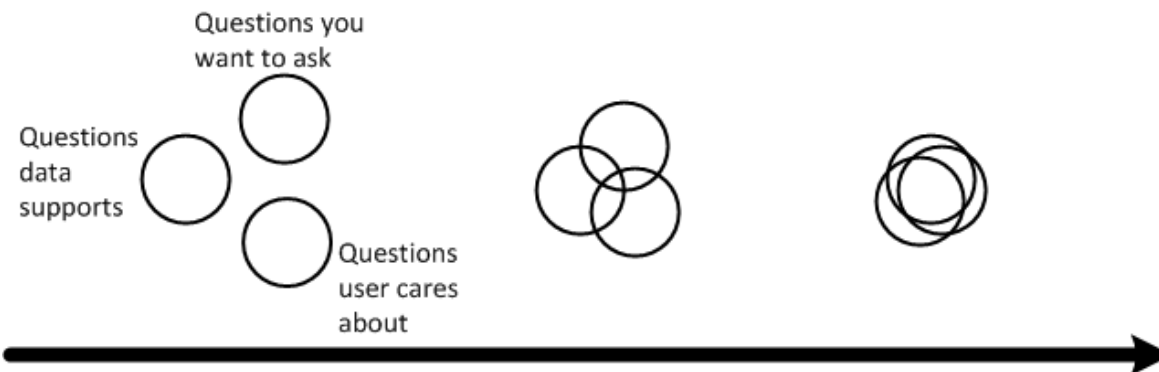
All (current) conclusions are wrong

Prone to revisions

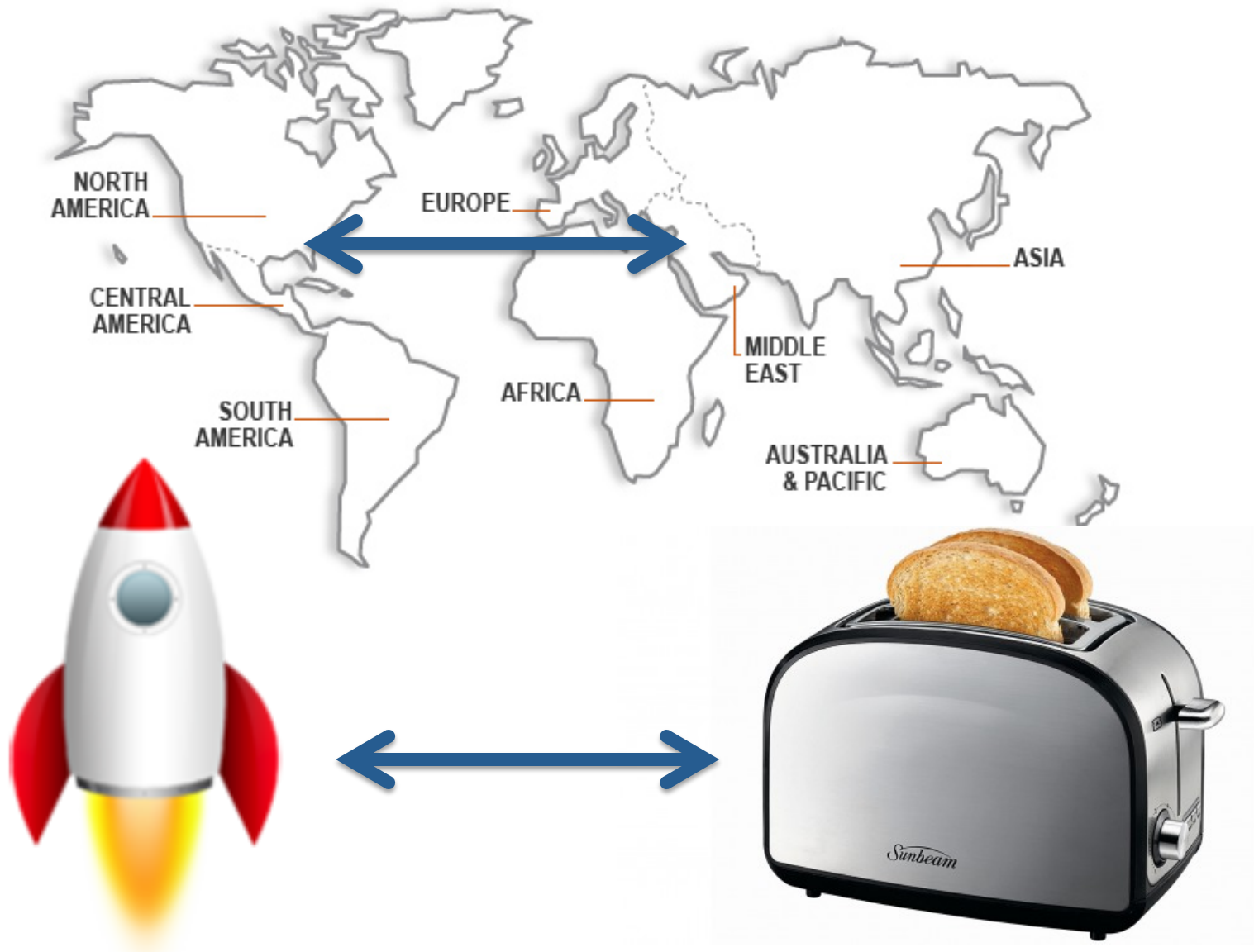
- By subsequent analysis
- Any current models is
 - Wrong, but useful
- Timm's Law:
 - Less conclusions, More conversations

To find better conclusions ...

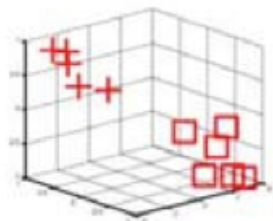
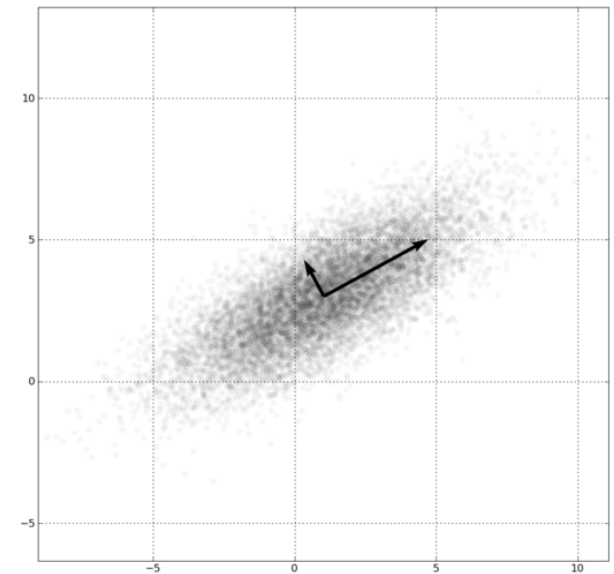
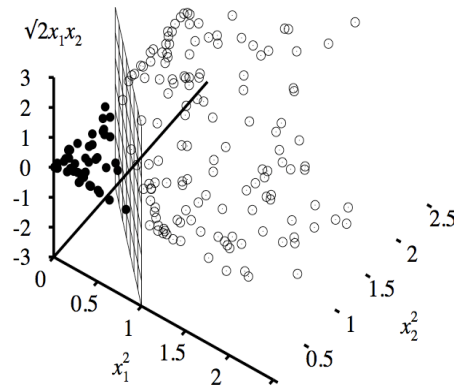
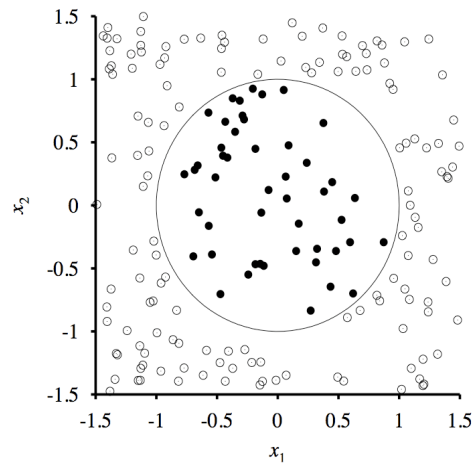
- Just keep looking
- For a community to find better conclusions
 - Discuss more, share more



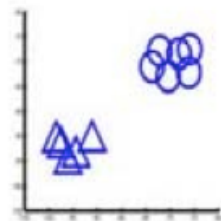
Between Turkish Toasters AND NASA Space Ships



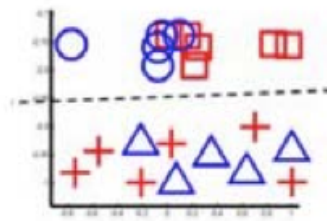
Raw dimensions less informative than underlying dimensions



(a) 3-D data



(b) 2-D data

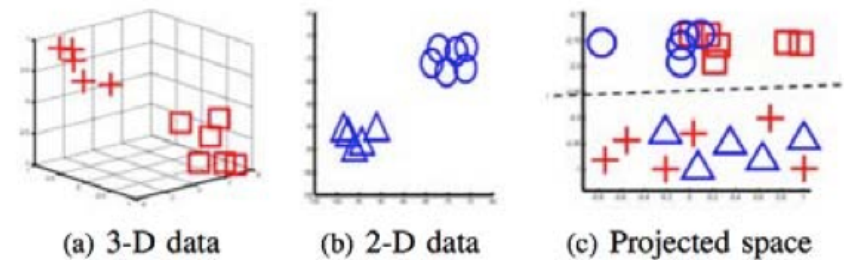


(c) Projected space

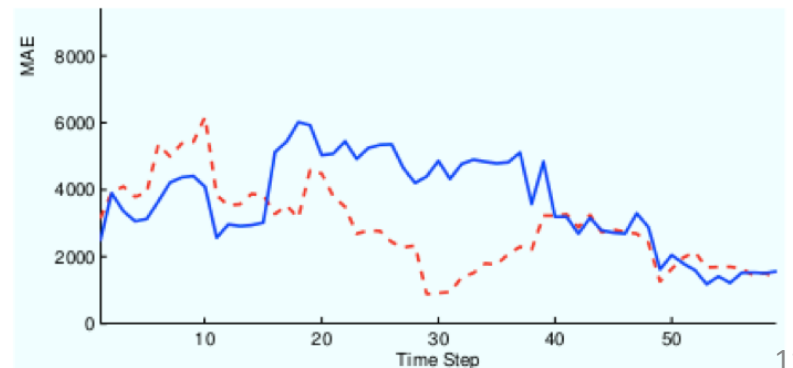
Q: How to TRANSFER Lessons Learned?

- Ignore most of the data
- relevancy filtering: *Turhan ESEj'09; Peters TSE'13*
- variance filtering: *Kocaguneli TSE'12, TSE'13*
- performance similarities: *He ESEM'13*

- Contort the data
- spectral learning (working in PC space or some other rotation)
Menzies, TSE'13; Nam, ICSE'13



- Build a bickering committee
- Ensembles *Minku, PROMISE'12*



A little bit of this, a little bit of that



"Aha! Bus tracks! I can follow these to the bus stop and not drive home drunk."

- Moral #3:
 - "My Father's house has many rooms."
- "False dichotomy between these two approaches. Really need to bridge these two modes."
- "Not the scientific method but scientific methods (plural)."