Things every actuary should know about data science

A guide for the perplexed

CAS Annual Meeting
Philadelphia
November 17, 2015

James Guszcza, PhD, FCAS
Deloitte Consulting
jguszcza@deloitte.com

What is data science?

At the Center of It All: Data Science

Or: “The Collision between Statistics and Computation”

- The skill set underlying business analytics is increasingly called **data science**.

- Data science goes beyond:
  - Traditional statistics
  - Business intelligence [BI]
  - Information technology

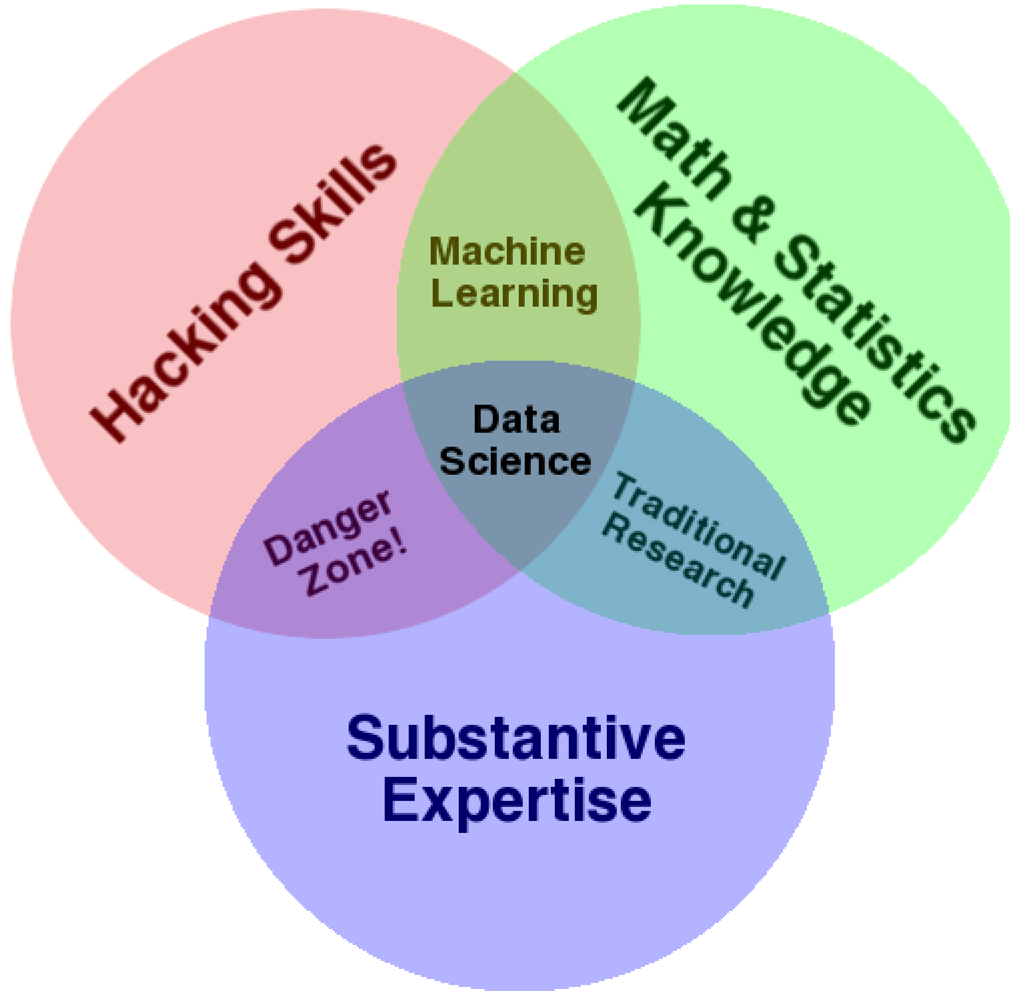


Image borrowed from Drew Conway’s blog
http://www.dataists.com/2010/09/the-data-science-venn-diagram

At the Center of It All: Data Science

Or: “The Collision between Statistics and Computation”

Is the actuarial profession here?
Should it be?

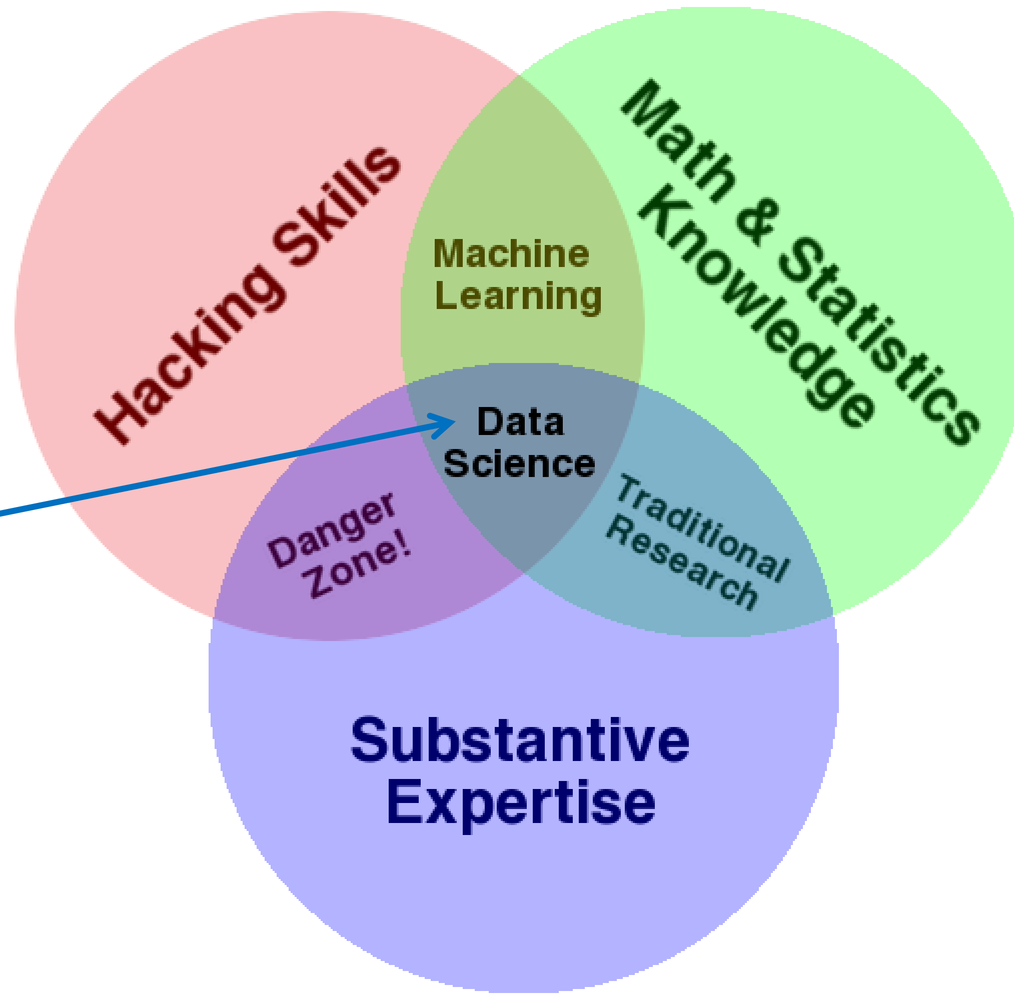


Image borrowed from Drew Conway’s blog
http://www.dataists.com/2010/09/the-data-science-venn-diagram

At the Center of It All: Data Science

Or: "The Collision between Statistics and Computation"

Or is it here?
Is that ok?

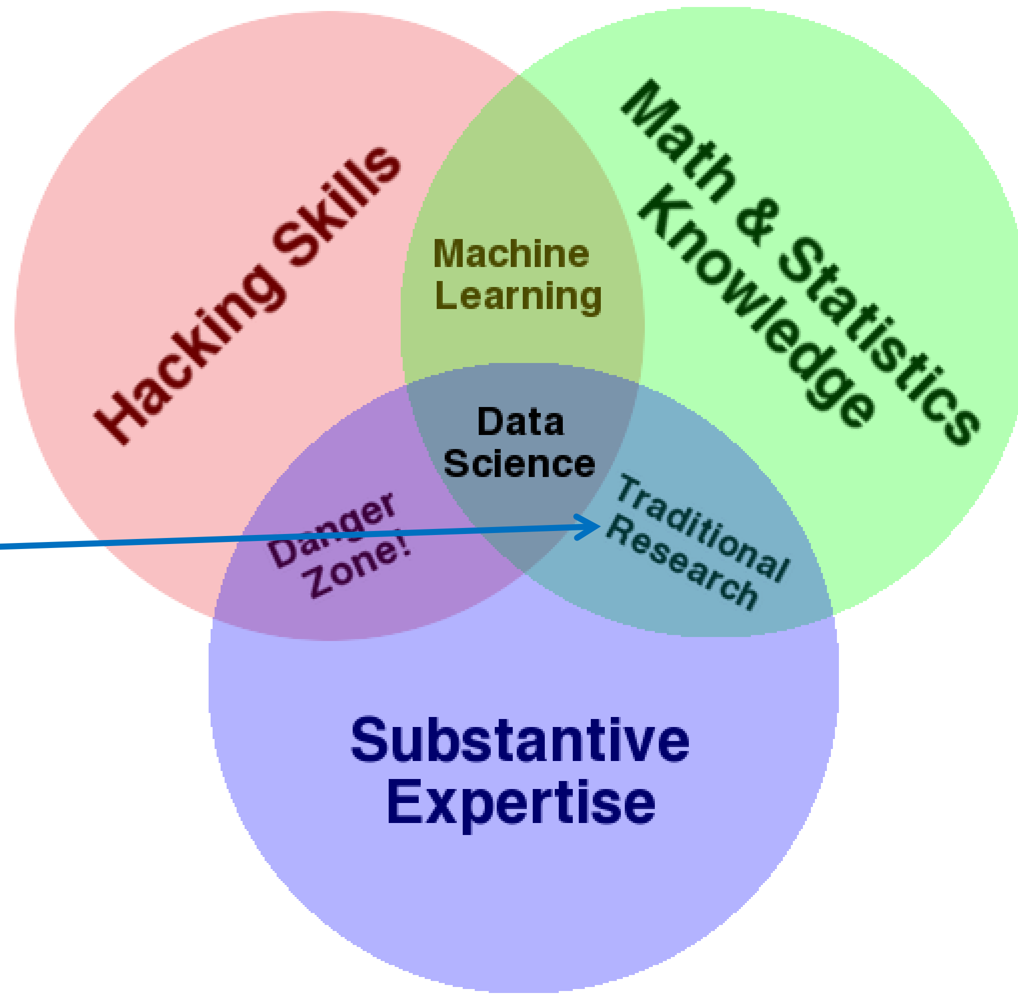
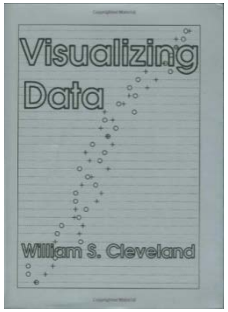


Image borrowed from Drew Conway's blog
http://www.dataists.com/2010/09/the-data-science-venn-diagram

The origin of "Data Science"



Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics

William S. Cleveland
Statistics Research, Bell Labs
wsc@bell-labs.com

Abstract

An action plan to enlarge the technical areas of statistics focuses on the data analyst. The plan sets out six technical areas of work for a university department and advocates a specific allocation of resources devoted to research in each area and to courses in each area. The value of technical work is judged by the extent to which it benefits the data analyst, either directly or indirectly. The plan is also applicable to government research labs and corporate research organizations.

Glamorous Models

(I'm not making this up)

Harvard Business Review

THE MAGAZINE | BLOGS | AUDIO & VIDEO | BOOKS | WEBINAR

Guest | limited access

Register today and save 20%* off your first order

THE MAGAZINE

October 2012

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

The culture of data science

"The best thing about being a statistician is that you get to play in everyone's back yard."

-- John Tukey
Princeton/Bell Labs

"The dominant trait among data scientists is an intense curiosity… This often entails the associative thinking that characterizes the most creative scientists in any field."

-- D.J. Patil
Linkedin

Points on a curve

Explaining variation in "data science"

- "Data science" is an umbrella term
- It spans multiples disciplines and statistical / computer science paradigms
- A continuum of paradigms

Modeling the world	Collective Intelligence	Bayesian Data Analysis	Clinical vs Actuarial Prediction	Statistical Learning	Machine Learning	Modeling the data

- Some fall more naturally within "actuarial science" than others
- Ok to embrace fuzzy concepts – as long as we remember they are fuzzy
- ("What happens in vagueness stays in vagueness")

The second machine age

Collective Intelligence	Bayesian Data Analysis	Clinical vs Actuarial Prediction	Statistical Learning	Machine Learning

The second machine age

Collective Intelligence	Bayesian Data Analysis	Clinical vs Actuarial Prediction	Statistical Learning	Machine Learning

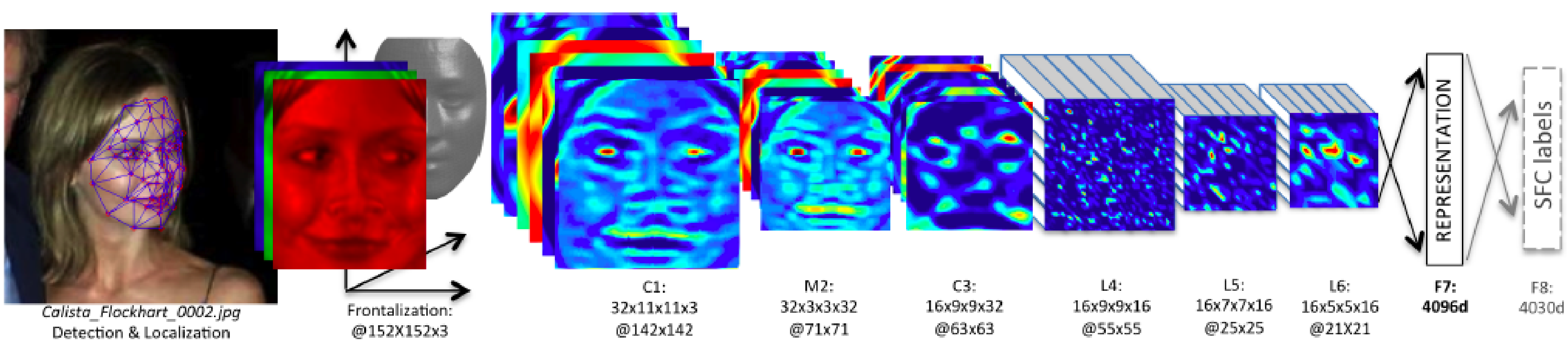
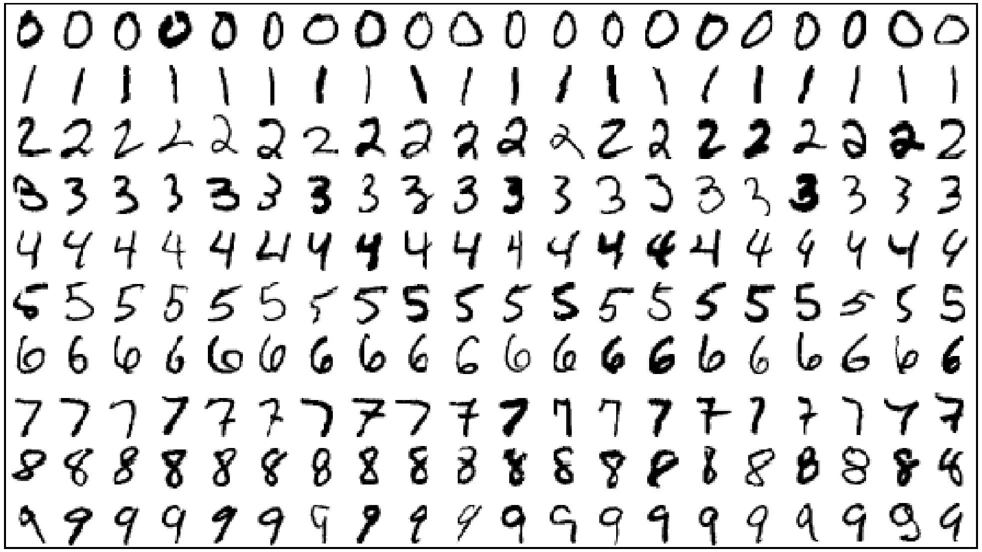
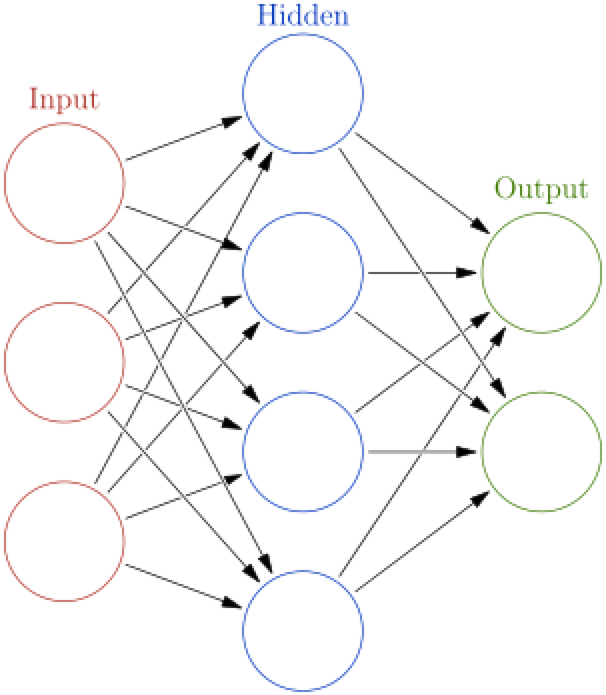


Figure 2. **Outline of the *DeepFace* architecture**. A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate outputs for each layer. The net includes more than 120 million parameters, where more than 95% come from the local and fully connected layers.

The second machine age

Collective Intelligence	*Bayesian Data Analysis*	*Clinical vs Actuarial Prediction*	*Statistical Learning*	*Machine Learning*





The second machine age

Collective Intelligence	Bayesian Data Analysis	Clinical vs Actuarial Prediction	Statistical Learning	Machine Learning

Examples:

- Natural language processing and "cognitive computing"
- Deep learning for pattern recognition
- Internet search
- Recommendation algorithms

Sample applications:

- Next-generation precision underwriting, fraud detection
- Image recognition – help adjust claims
- Speech recognition – customer service

Greater statistics – learning from data

Collective Intelligence	Bayesian Data Analysis	Clinical vs Actuarial Prediction	Statistical Learning	Machine Learning

Statistical Science
2001, Vol. 16, No. 3, 199–231

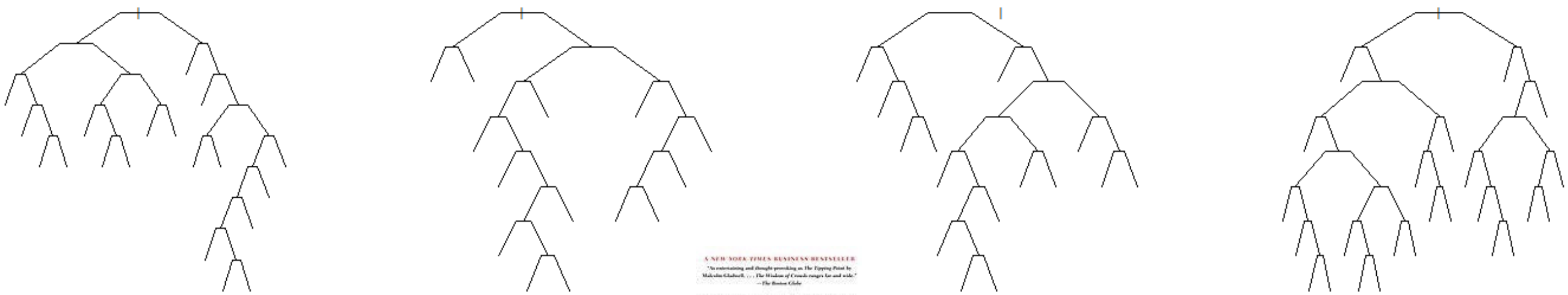
Statistical Modeling: The Two Cultures

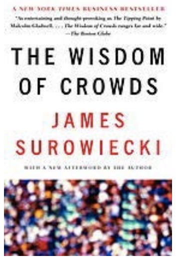
Leo Breiman

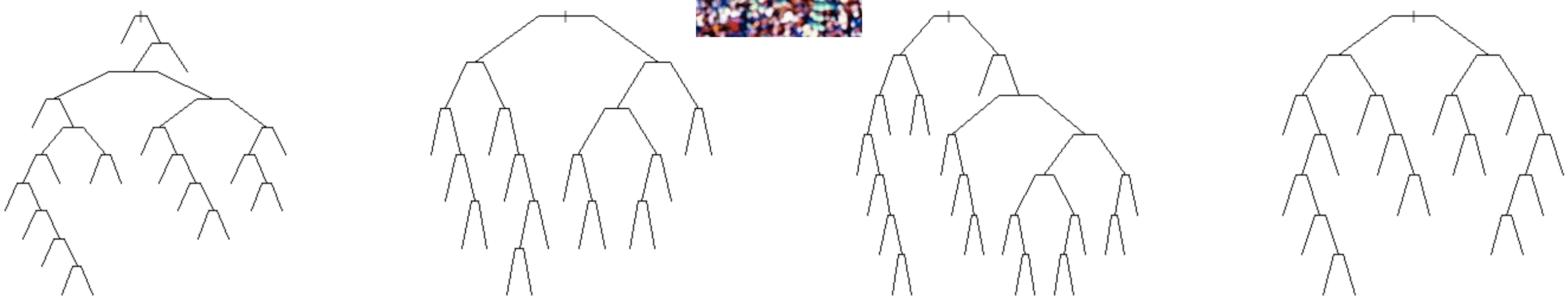
Abstract. There are two cultures in the use of statistical modeling to reach conclusions from data. One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown. The statistical community has been committed to the almost exclusive use of data models. This commitment has led to irrelevant theory, questionable conclusions, and has kept statisticians from working on a large range of interesting current problems. Algorithmic modeling, both in theory and practice, has developed rapidly in fields outside statistics. It can be used both on large complex data sets and as a more accurate and informative alternative to data modeling on smaller data sets. If our goal as a field is to use data to solve problems, then we need to move away from exclusive dependence on data models and adopt a more diverse set of tools.

Greater statistics – learning from data

- *Collective Intelligence*
- *Bayesian Data Analysis*
- *Clinical vs Actuarial Prediction*
- *Statistical Learning*
- *Machine Learning*







Greater statistics – learning from data

Collective Intelligence	Bayesian Data Analysis	Clinical vs Actuarial Prediction	Statistical Learning	Machine Learning

Examples:

- Tree-based modeling
- Ensembles
- Bagging, boosting
- GLM, regularized regression (lasso, ridge, …)
- Support vector machines
- Unsupervised learning
- …

Sample applications:

- Ratemaking, price optimization
- analysis of “sparse” digital breadcrumbs
- Credit scoring
- Analysis of telematics data
- Precision marketing, customer segmentation
- Setting case reserves

Playing Moneyball

Collective Intelligence | Bayesian Data Analysis | **Clinical vs Actuarial Prediction** | Statistical Learning | Machine Learning

Science 31 March 1989:
Vol. 243 no. 4899 pp. 1668-1674
DOI: 10.1126/science.2648573

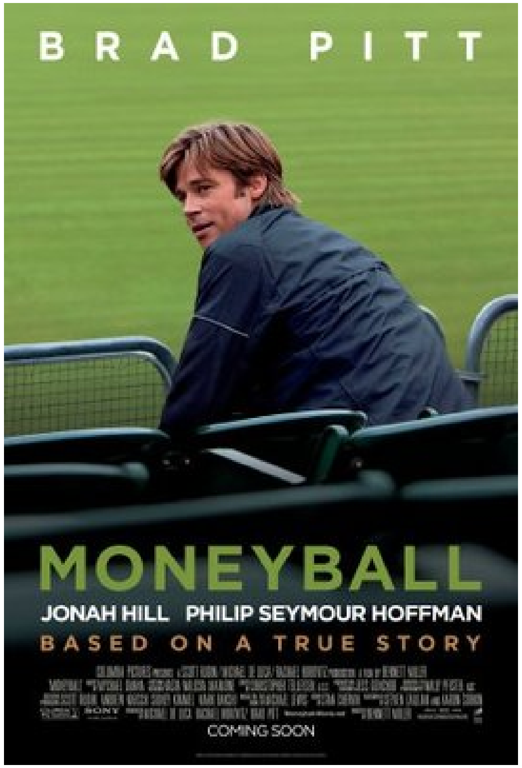
Clinical versus actuarial judgment

RM Dawes, D Faust and PE Meehl

+ Author Affiliations

ABSTRACT

Professionals are frequently consulted to diagnose and predict human behavior; optimal treatment and planning often hinge on the consultant's judgmental accuracy. The consultant may rely on one of two contrasting approaches to decision-making--the clinical and actuarial methods. Research comparing these two approaches shows the actuarial method to be superior. Factors underlying the greater accuracy of actuarial methods, sources of resistance to the scientific findings, and the benefits of increased reliance on actuarial approaches are discussed.

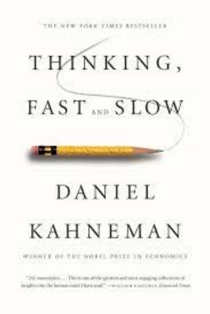


Playing Moneyball

Collective Intelligence	Bayesian Data Analysis	Clinical vs Actuarial Prediction	Statistical Learning	Machine Learning

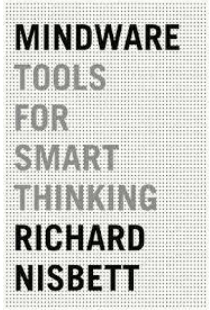
"Whatever else it produces, an organization is a factory that manufactures judgments and decisions."

– Daniel Kahneman, *Thinking, Fast and Slow*



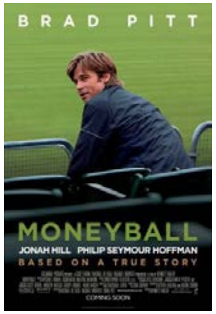
"Human judges are not merely worse than optimal regression equations; they are worse than almost any regression equation."

– Richard Nisbett and Lee Ross, *Human Inference*



"The market for baseball players was so inefficient… that superior management could still run circles around taller piles of cash."

– Michael Lewis, *Moneyball*



Playing Moneyball

Collective Intelligence	Bayesian Data Analysis	**Clinical vs Actuarial Prediction**	Statistical Learning	Machine Learning

- Kahneman: thinking fast (Type 1) vs thinking slow (Type 2)
- Type 1 is terrible at statistics
- Leads to inefficient markets a la Moneyball
- Research dating back to Paul Meehl in the 1950s
- ~~equations > experts~~
- (experts + equations) > experts

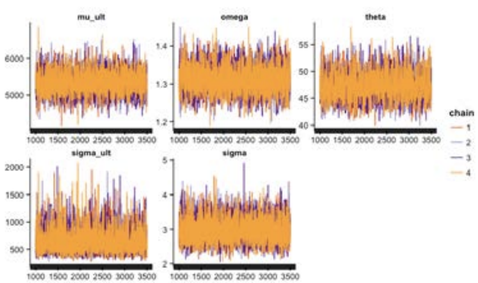
Sample applications:

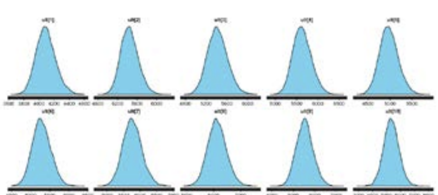
- Underwriting complex risks
- Claims triage
- Fraud investigation
- Premium audit
- Predictive hiring
- Risk management, safety analytics

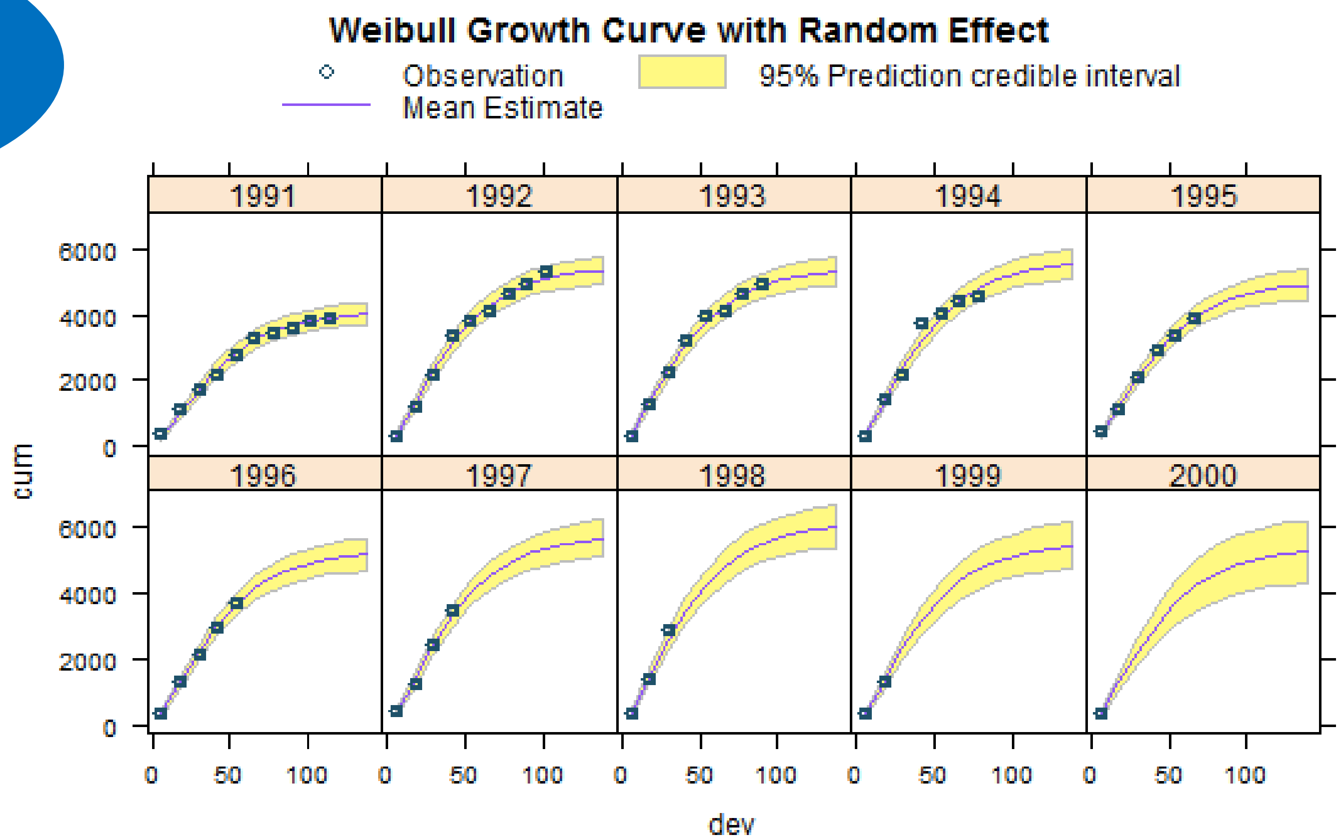
Statistics' "first culture"

Collective Intelligence | **Bayesian Data Analysis** | Clinical vs Actuarial Prediction | Statistical Learning | Machine Learning

Bayesian loss reserving example







Statistics' "first culture"

Collective Intelligence	Bayesian Data Analysis	Clinical vs Actuarial Prediction	Statistical Learning	Machine Learning

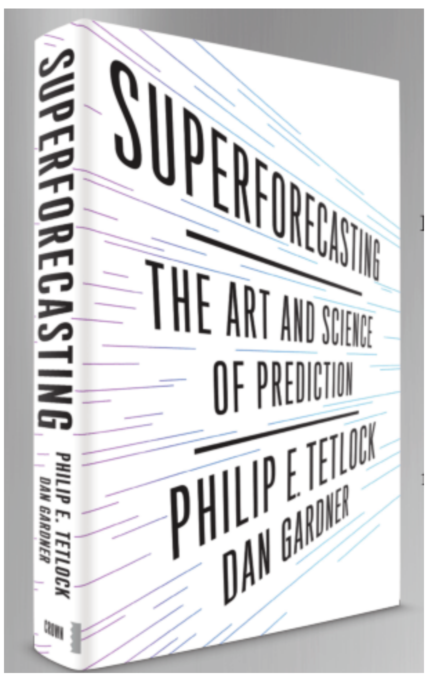
- Rather than just model the data we model the process that generates the data
- Appreciation for model risk
- Appreciation for parameter risk
- Necessary when you're in a situation where the data is useful but doesn't contain all of the information needed for predictions/inferences/forecasts

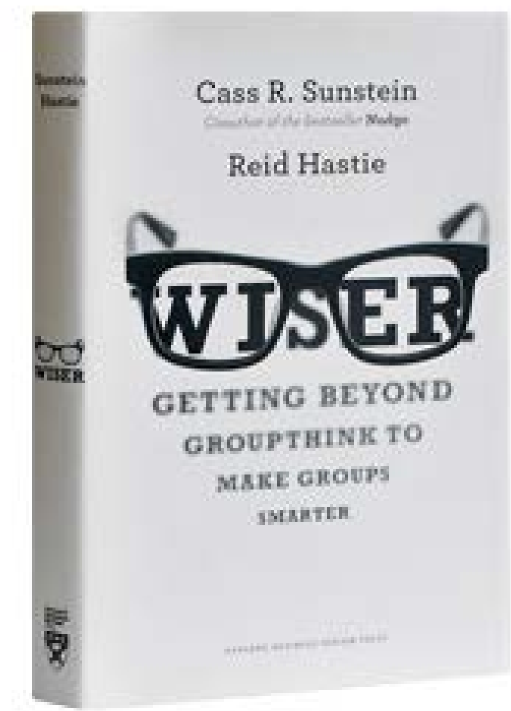
Sample applications:

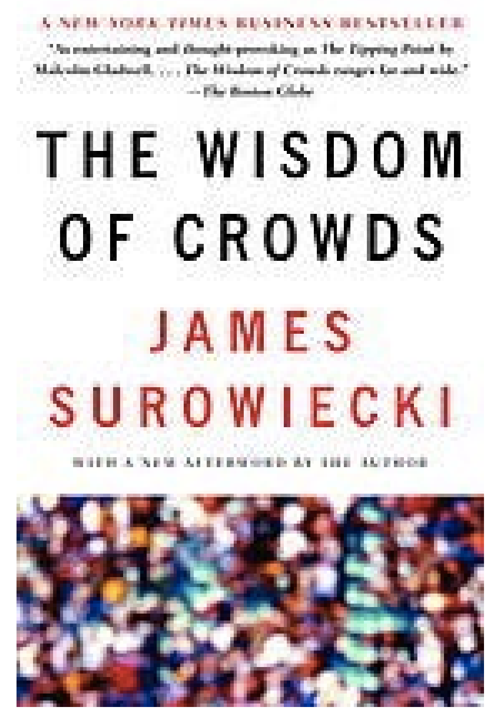
- Bayesian loss reserving
- Loss model analysis, VaR
- Precision medicine
- Social science research

I think we all agree about the opposite of groupthink

Collective Intelligence	*Bayesian Data Analysis*	*Clinical vs Actuarial Prediction*	*Statistical Learning*	*Machine Learning*







Collective intelligence

Collective Intelligence	Bayesian Data Analysis	Clinical vs Actuarial Prediction	Statistical Learning	Machine Learning

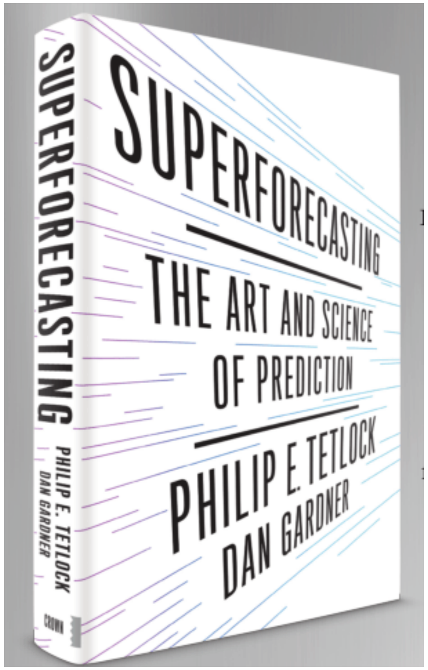
- Prediction markets
- Delphi method
- Combining forecasts
- Philip Tetlock's "Superforecasting"

Sample applications:

- Emerging risks (e.g. cyber security)
- Underwriting one-off risks
- Hiring decisions
- Strategic, investment decisions

"Superforecasting"

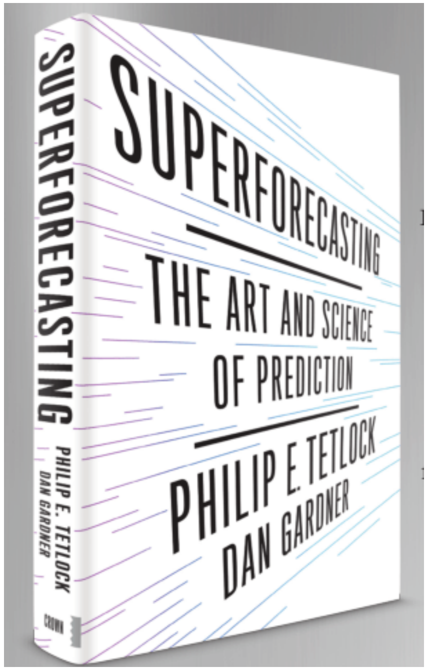
Collective Intelligence	Bayesian Data Analysis	Clinical vs Actuarial Prediction	Statistical Learning	Machine Learning



- "Foxes [not hedgehogs] choose their ideas from a variety of schools of thought."
- "Reality is infinitely complex"
- "[Be] probabilistic. Judge using many grades of maybe"
- [Be] intellectually curious
- "Beliefs are hypotheses to test, not treasured to be guarded"

“Superforecasting”

Collective Intelligence	Bayesian Data Analysis	Clinical vs Actuarial Prediction	Statistical Learning	Machine Learning



- “Check thinking for cognitive and emotional biases.”
- “[Be] reflective – introspective and self-critical”
- “Believe it’s possible to get better.”
- “Value diverse views”
- “Be determined to keep at it no matter how long it takes”