Predictive Modeling
From A to B

IA Actuaries Club
Des Moines

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

November 29, 2012

Predictive Analytics: Three Levels of Discussion

- ## Strategy
  - Profitable growth
  - More consistent, accurate and economical underwriting
  - Retaining, cross-selling to preferred policyholders
  - Marketing / prospecting
  - Promoting wellness
  - …

- ## Methodology
  - Model <u>design</u>   ("actuarial science")
  - Modeling <u>process</u>   ("data science")

- ## Technique
  - Supervised and unsupervised
  - GLM vs. decision trees vs. support vector machines…

At the Center of It All: Data Science

Or: "The Collision between Statistics and Computation"

- Today, it is arguably appropriate to call certain aspects of property-casualty actuarial practice "**data science**".

- This often involves more than just knowing how to run regression models.

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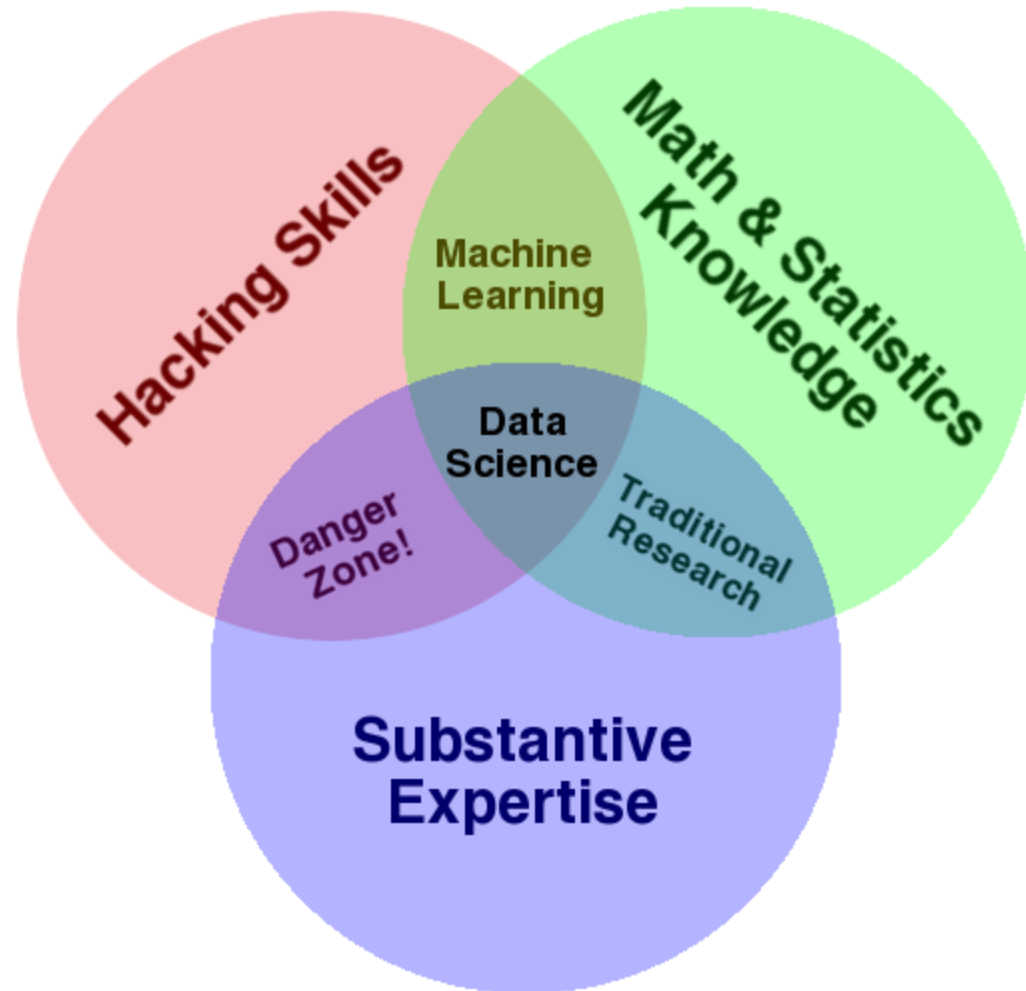 - Business intelligence [BI]
  - Information technology

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

Glamorous Models

(I'm not making this stuff up)

Analyzing Analytics

- Three thoughts capture a major reason why predictive analytics is ubiquitous...

- **"The whole of science is a refinement of everyday thinking."**
  - Albert Einstein

- **"Decision-making is at the heart of administration."**
  - Herbert Simon

- **"Human judges are not merely worse than optimal regression equations; they are worse than almost any regression equation."**
  - Richard Nisbett & Lee Ross
  - *Human Inference: Strategies and Shortcomings of Social Judgment*
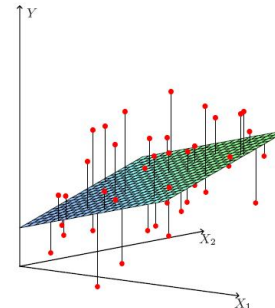
Agenda

Introduction

Why Business Analytics is Now Ubiquitous

Predictive Modeling: Levels of Discussion

Strategy

Methodology

Technique

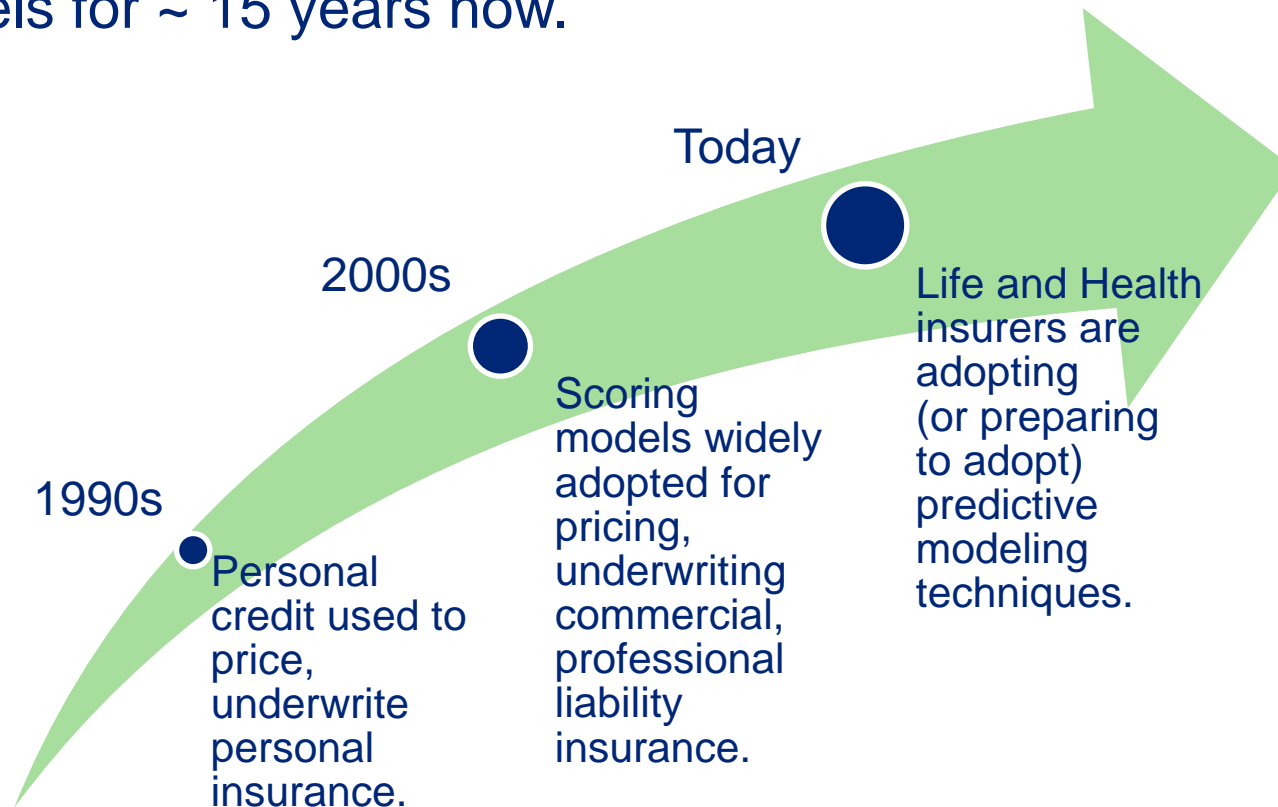# Why Business Analytics is Ubiquitous

Is Business Analytics New?

- Not really.

- Easy example: credit scoring is an early example of business analytics.

- It has been around for many decades.

Is Business Analytics New?

- Predictive Modeling in insurance isn't new either.
  … But it's rate of adoption continues to increase.

- Property-Casualty insurers have been extensively using predictive models for ~ 15 years now.

Today

2000s

1990s

Personal credit used to price, underwrite personal insurance.

Scoring models widely adopted for pricing, underwriting commercial, professional liability insurance.

Life and Health insurers are adopting (or preparing to adopt) predictive modeling techniques.

So Why Now?   (Answer #1)

- Technology (Moore's Law)
  - Cost of storage and computing power has decreased exponentially

- Data
  - Third-party data is becoming increasingly available
  - Companies are learning to do more with their internal data

- Software and algorithms
  - Great analytic ideas keep coming from statistics, economics, machine learning, marketing, …
  - Free tools like R
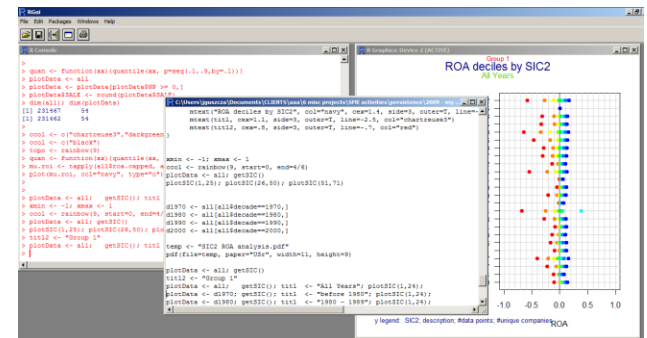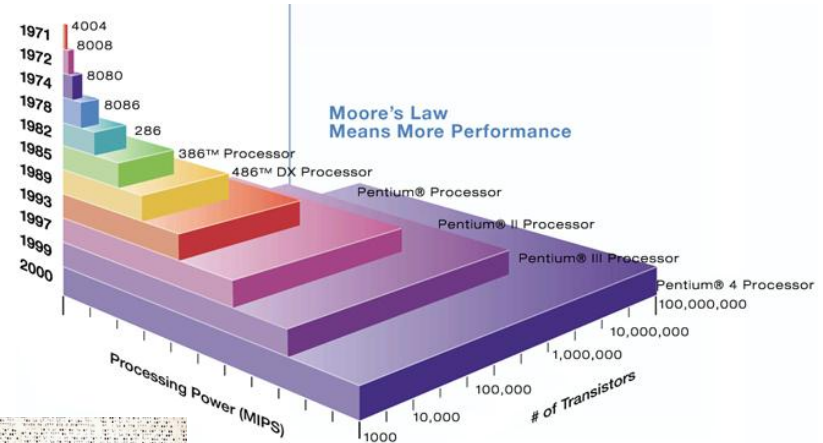
Exponential Growth of Data Availability

• More and more data is gathered, stored, and available to be
analyzed
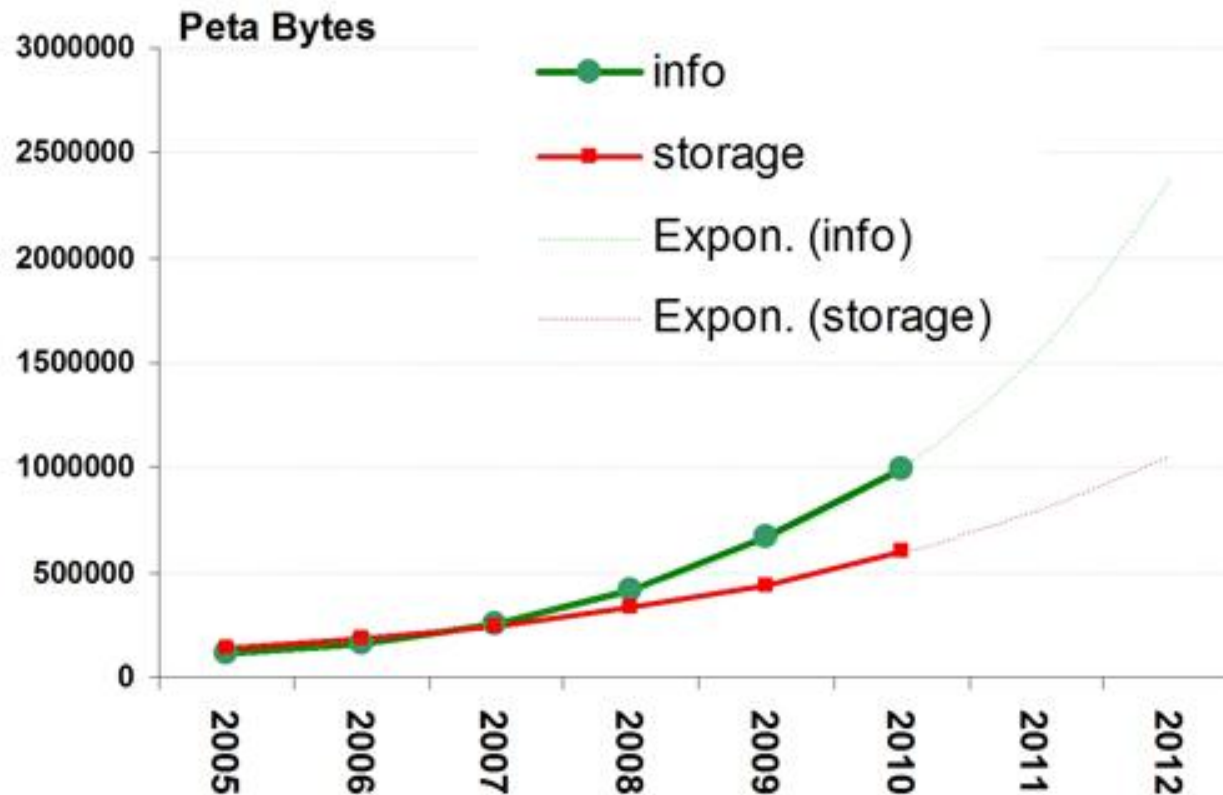
"Time Challenges – Challenging Times for Future Innovation Search"
http://www.dlib.org/dlib/may09/mestl/05mestl.html

Exponential growth of Statistical Computing Availability

- The open-source R statistical computing environment has experienced exponential growth

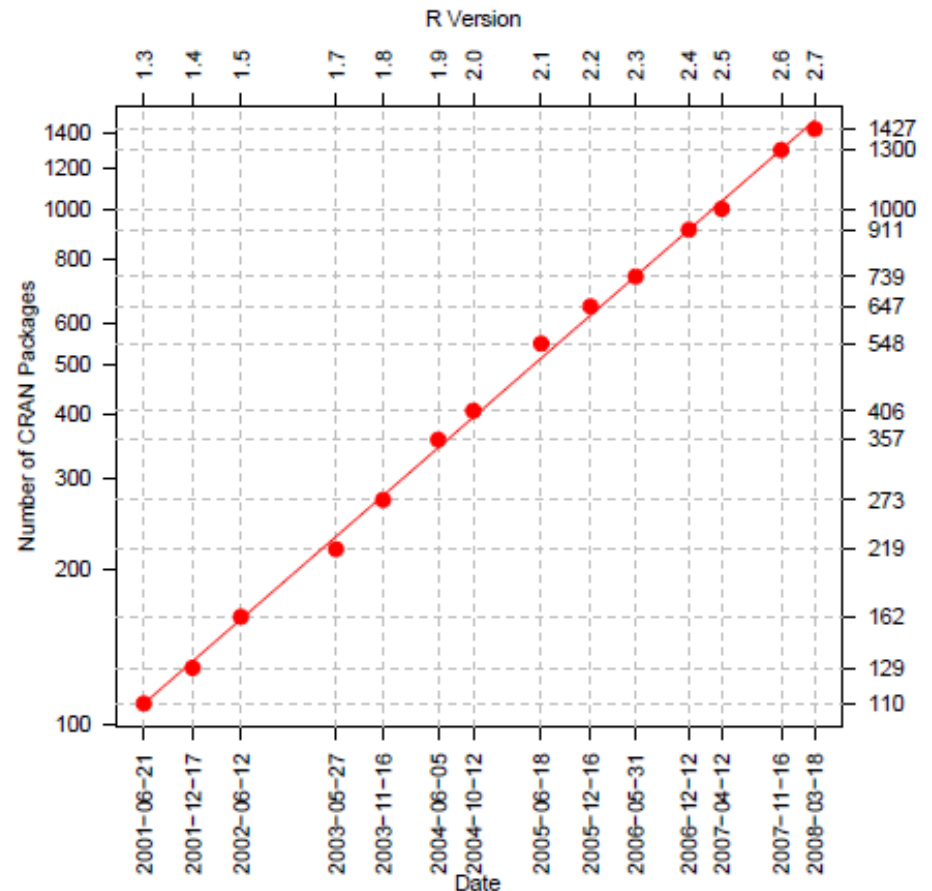- The number of contributed packages has grown exponentially

Figure 1: The number of R packages on CRAN has grown exponentially since R version 1.3 in 2001. *Source of Data*: https://svn.r-project.org/R/branches/.

Why Now?   (Answer #2)

OR: Clinical vs Actuarial Judgment – the Motion Picture

Clinical versus actuarial judgment

RM Dawes, D Faust and PE Meehl

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.
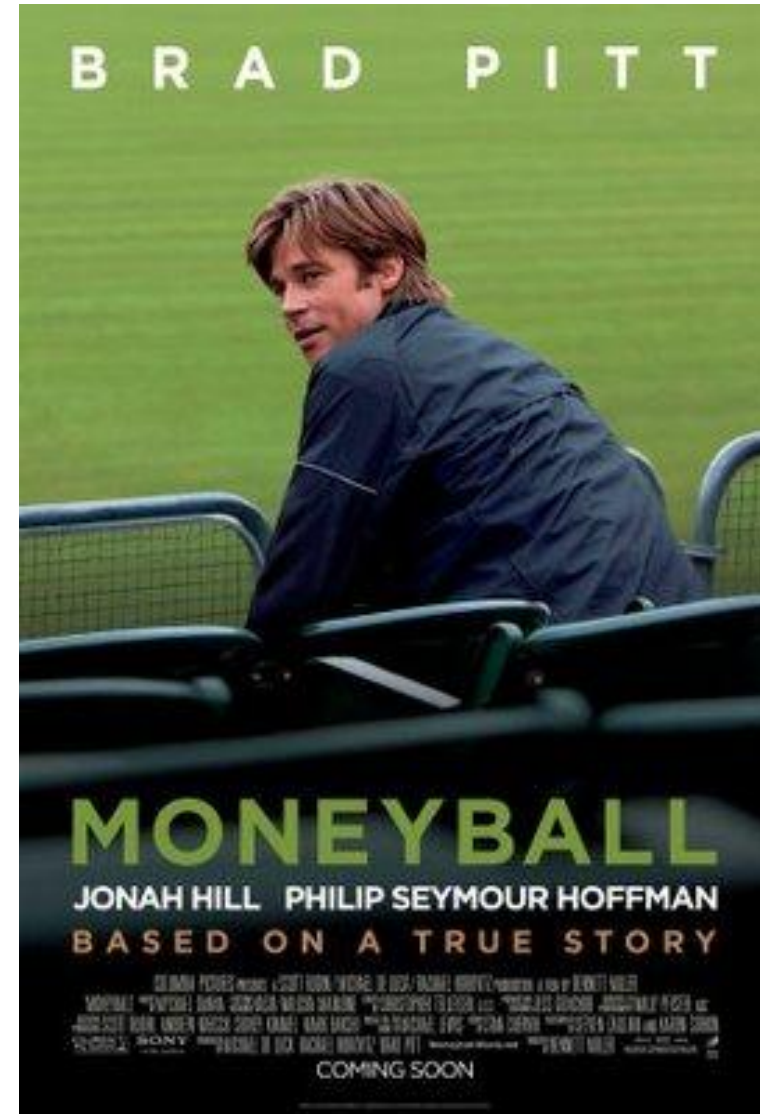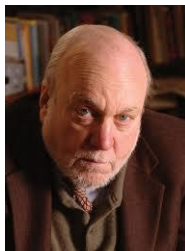
There's Something About Linda

- To illustrate, think about this person:

- Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations

There's Something About Linda

• Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations

<u>Now rank these possible scenarios in order of probability</u>:

• Linda is active in the feminist movement

• Linda is a bank teller

• Linda is a bank teller and is active in the feminist movement

There's Something About Linda

- Daniel Kahneman and Amos Tversky posed precisely this question to a group of people

- They found that 87% of the people thought that "feminist bank teller" was more probable than "bank teller"

There's Something About Linda

- Daniel Kahneman and Amos Tversky posed precisely this question to a group of people

- They found that 87% of the people thought that "feminist bank teller" was more probable than "bank teller"

F⊕RA.tv

Daniel Kahneman
Nobel Prize Winner in Economics

- But this is logically impossible.
  - Mathematical fact: Prob(A&B) < Prob(B)
  - So in particular: Prob(feminist & bank teller) < Prob(bank teller)

Feminists   Bank tellers

Feminist bank tellers
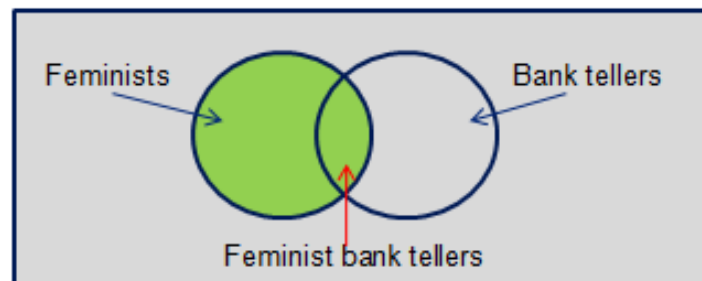
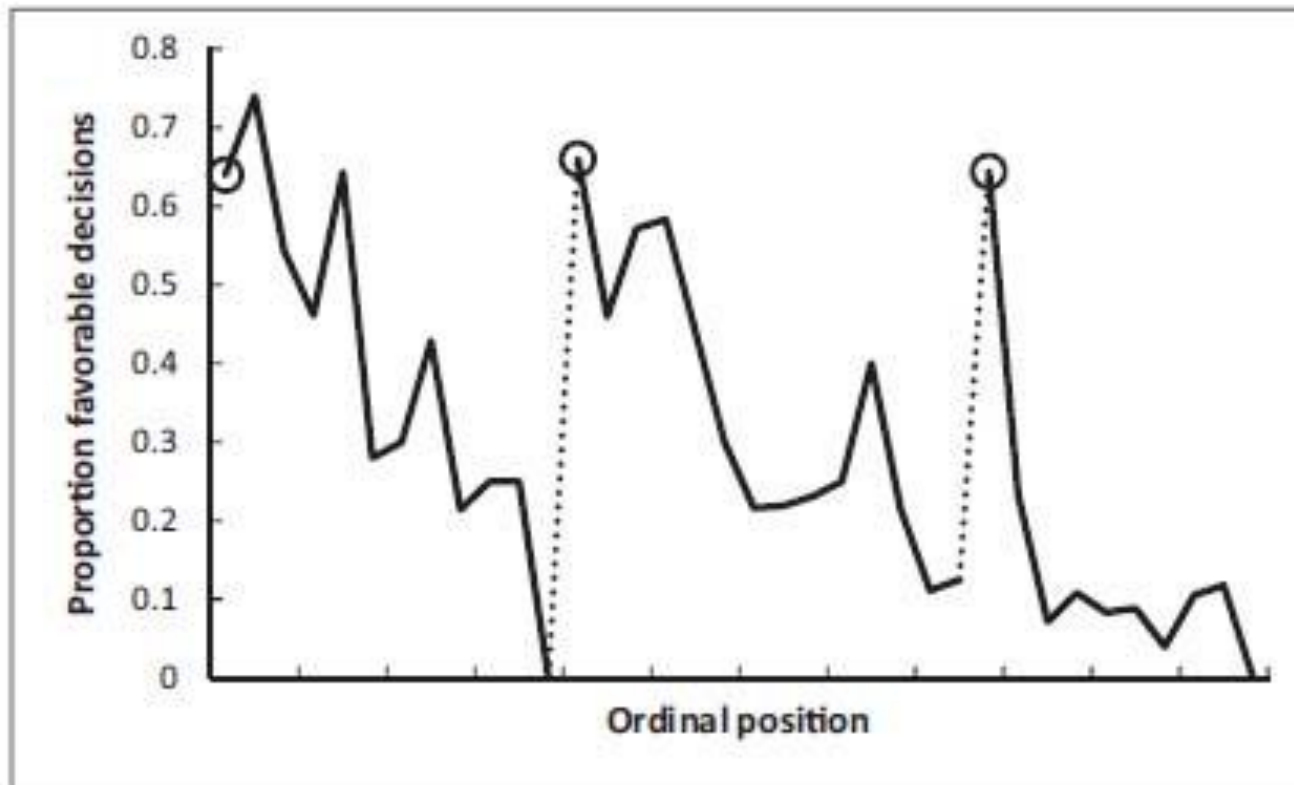# Decision Fatigue and Parole Decisions

Fig. 1. Proportion of rulings in favor of the prisoners by ordinal position. Circled points indicate the first decision in each of the three decision sessions; tick marks on x axis denote every third case; dotted line denotes food break. Because unequal session lengths resulted in a low number of cases for some of the later ordinal positions, the graph is based on the first 95% of the data from each session.

http://economix.blogs.nytimes.com/2011/04/14/time-and-judgment/
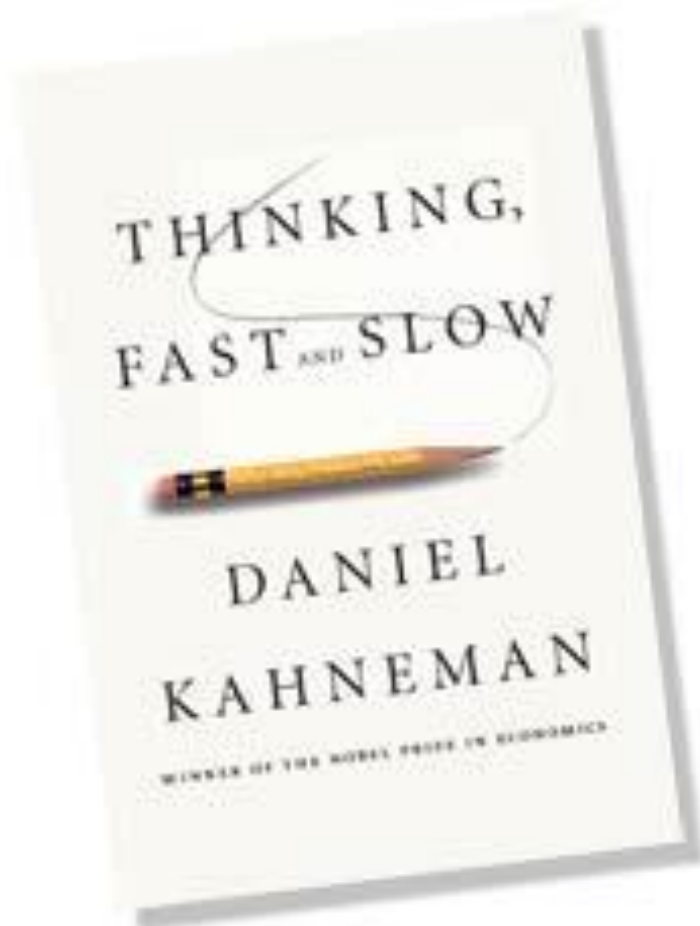
The Moral of These Stories

- Kahneman: there are two types of mental operations.

- Type 1: automatic, effortless, **associatively** coherent.

- Type 2: controlled, effortful, **logically** coherent.

- Most of our mental operations are "Type I" in nature.

- And "Type I" has a lot of trouble with statistics.

(2011 Book of the Year)

Analytics Everywhere

- The Obama 2012 reelection campaign used both marketing data as well as social science research to strategically get out the vote.

- Decision tree models are used to help ER doctors better triage patients complaining of chest pain.

- Predictive models are used to predict the price of different wine vintages based on variables about the growing season.

- Predictive models to help commercial insurance underwriters better select and price risks.

- Predict which non-custodial parents are at highest risk of falling into arrears on their child support.

- Predicting which job candidates will successfully make it through the interviewing / recruiting process… and which candidates will subsequently retain and perform well on the job.
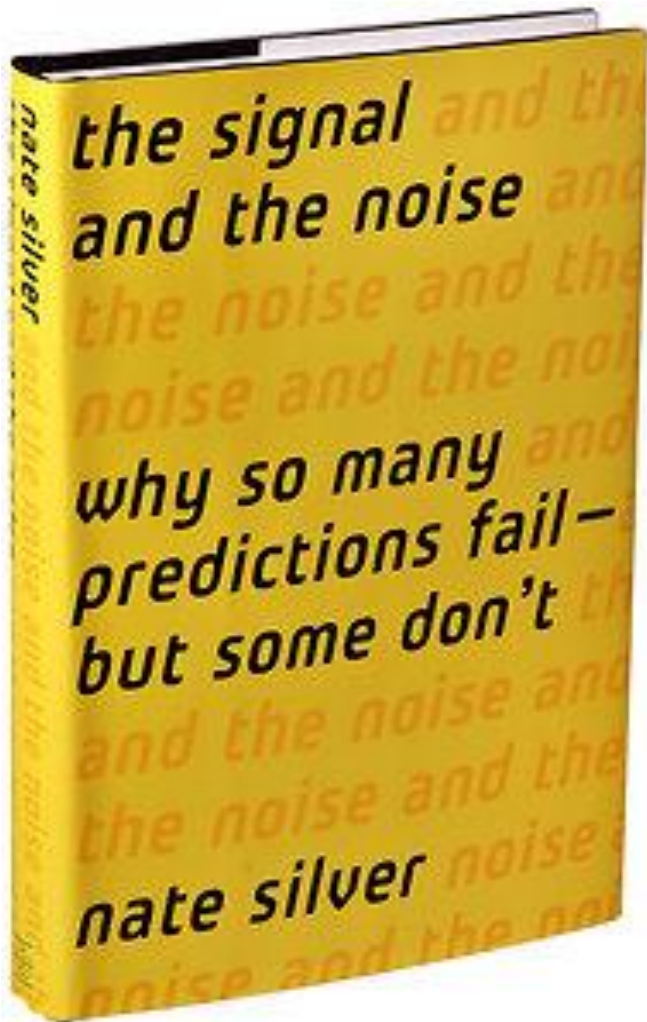
- Predicting which doctors are at highest risk of being sued for malpractice.

- Predicting the ultimate severity of injury claims.
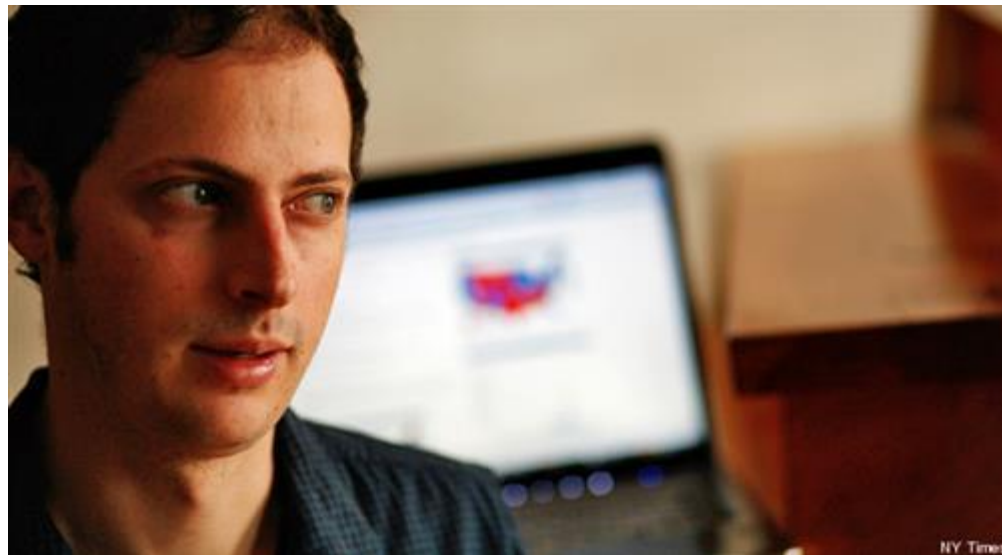
(Deloitte applications in green)

(2012 Book of the Year)

In the last days of the election, Peggy Noonan had a "feel" that things were moving Mitt Romney's way. George Will was more cerebral: his brain told him it would be Romney in a rout. And Michael Barone, who used to have a good divining rod to go along with an encyclopedic knowledge for all numbers political, also predicted a Romney landslide.

What they had in common, aside from putting up a brick Tuesday that completely missed the electoral net, was a last-hurrah push for the old-fashioned prediction by gut.
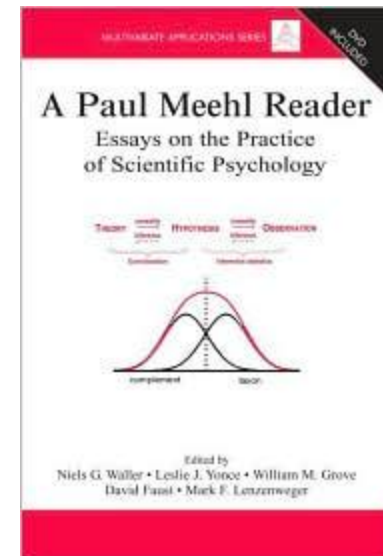
--- Timothy Egan, *The New York Times*
Nov 7, 2012

nate silver

the signal and the noise

why so many predictions fail— but some don't

nate silver

A Practical Conclusion

"There is no controversy in social science which shows such a large body of quantitatively diverse studies coming out so uniformly in the same direction as this one. When you are pushing over 100 investigations, predicting everything from the outcome of football games to the diagnosis of liver disease, and when you can hardly come up with half a dozen studies showing even a weak tendency in favor of the clinician, it is time to draw a practical conclusion."

-- Paul Meehl, "Causes and Effects of my Disturbing Little Book"

Deloitte.