ASA, Statistics, and "Data Science": Making an Impact

Karen Kafadar

Virginia Chapter of the ASA

25 October 2019

University of Virginia

From the American Statistical Association:

Welcome to the Virginia Chapter!

- ASA Mission:

  Promoting the Practice and Profession of Statistics

- Vision:

  A world that relies on data and statistical thinking to drive discovery and inform decisions

ASA

- Founded 1839
- ~18,000 members from 90 countries
- 50% academe, 40% business & Industry, 10% government
- 29 Sections (Surveys, Computing, Biometrics, Data Science, …)
- 10 Interest Groups (Astrophysics, Text Analysis, Auditing, …)
- 70 Committees (Leadership, Advisory, Awards, Education, …)
- 76 local ASA chapters in U.S. + Puerto Rico (now including VA!)
- Activities in Publications, Education, International Partnerships

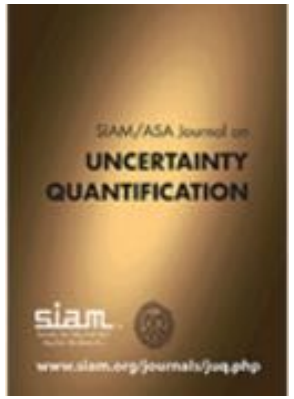
Causal Actuarial Society
Int'l Biometrics Society
Int'l Chinese Statistical Assoc
Int'l Indian Statistical Assoc
Int'l Society for Bayesian Analysis

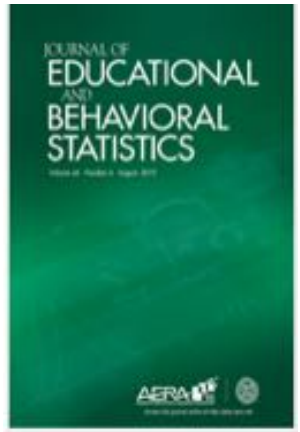
Int'l Statistical Institute
Institute for Math Statistics
Royal Statistical Society
Statistical Society of Canada
Korean Int'l Statistical Society

ASA

Repeat Performance:
UVA Nov 21, 12.30pm
Clark 109


COMPUTATIONAL AND
GRAPHICAL STATISTICS


SIAM/ASA Journal on
UNCERTAINTY
QUANTIFICATION
siam
www.siam.org/journals/juq.php


JOURNAL OF
EDUCATIONAL
AND
BEHAVIORAL
STATISTICS

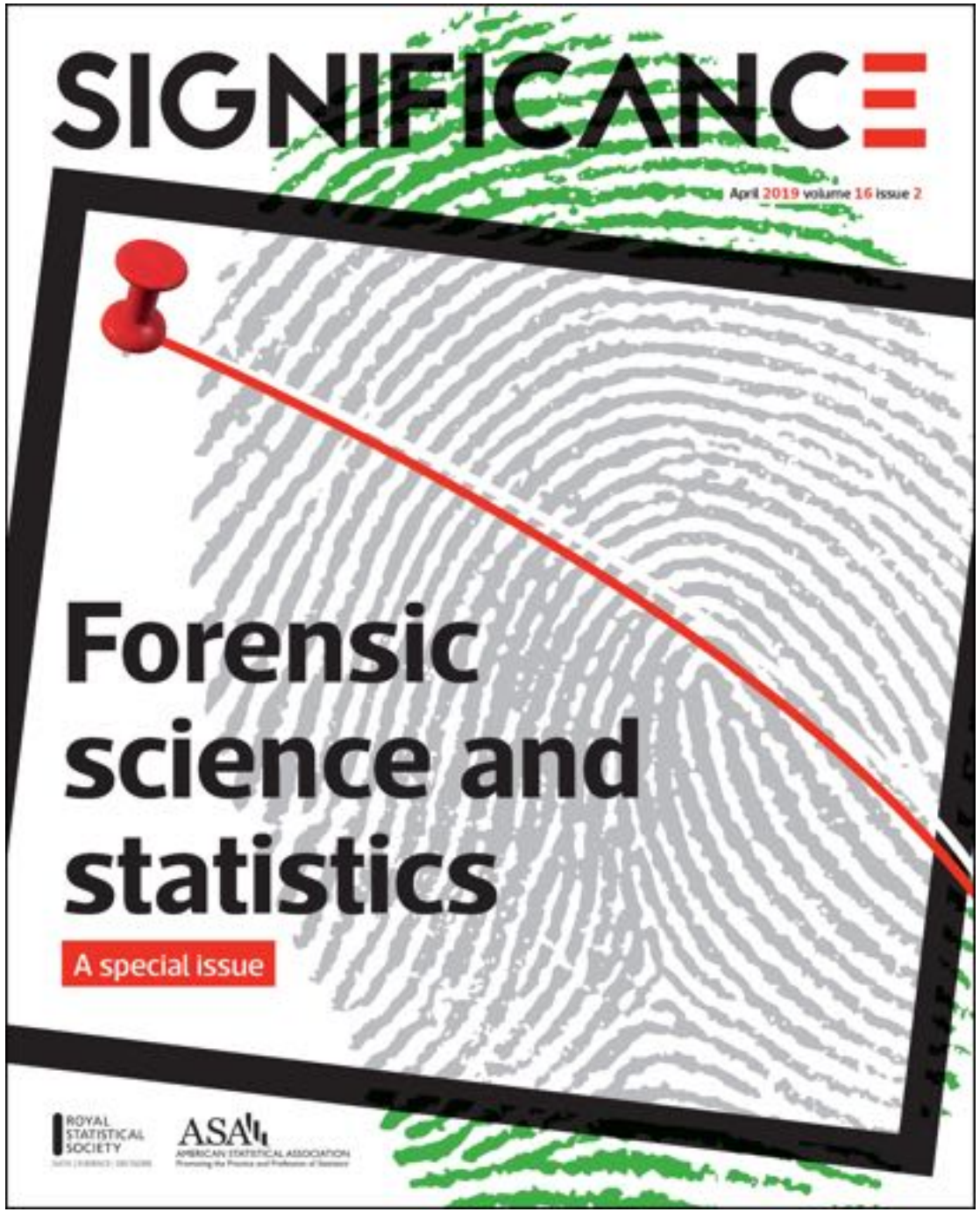

Journal of
Agricultural,
Biological and
Environmental
Statistics

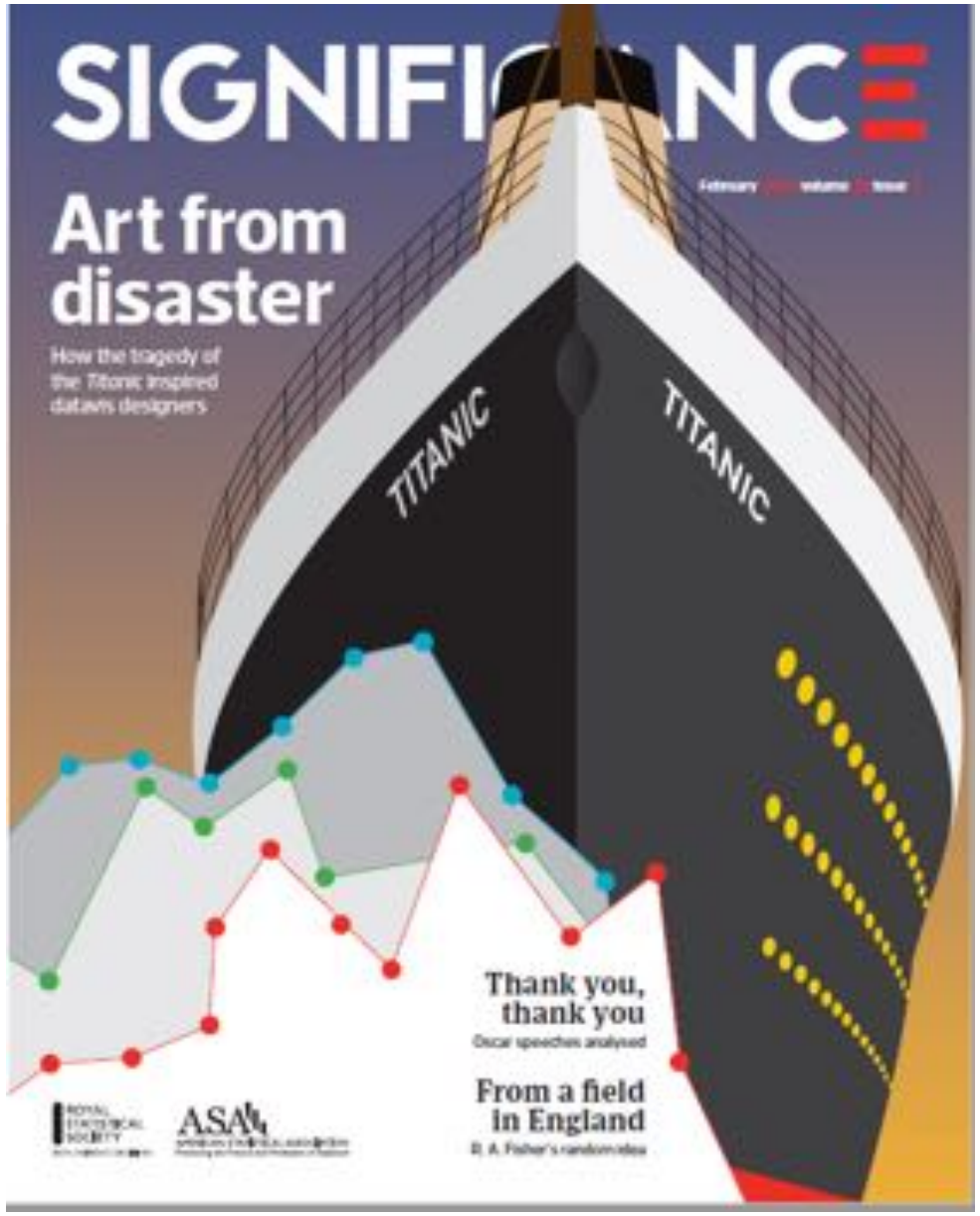

Technometrics


Journal of
Survey Statistics
and Methodology
OXFORD


STATISTICS
SURVEYS

ASA

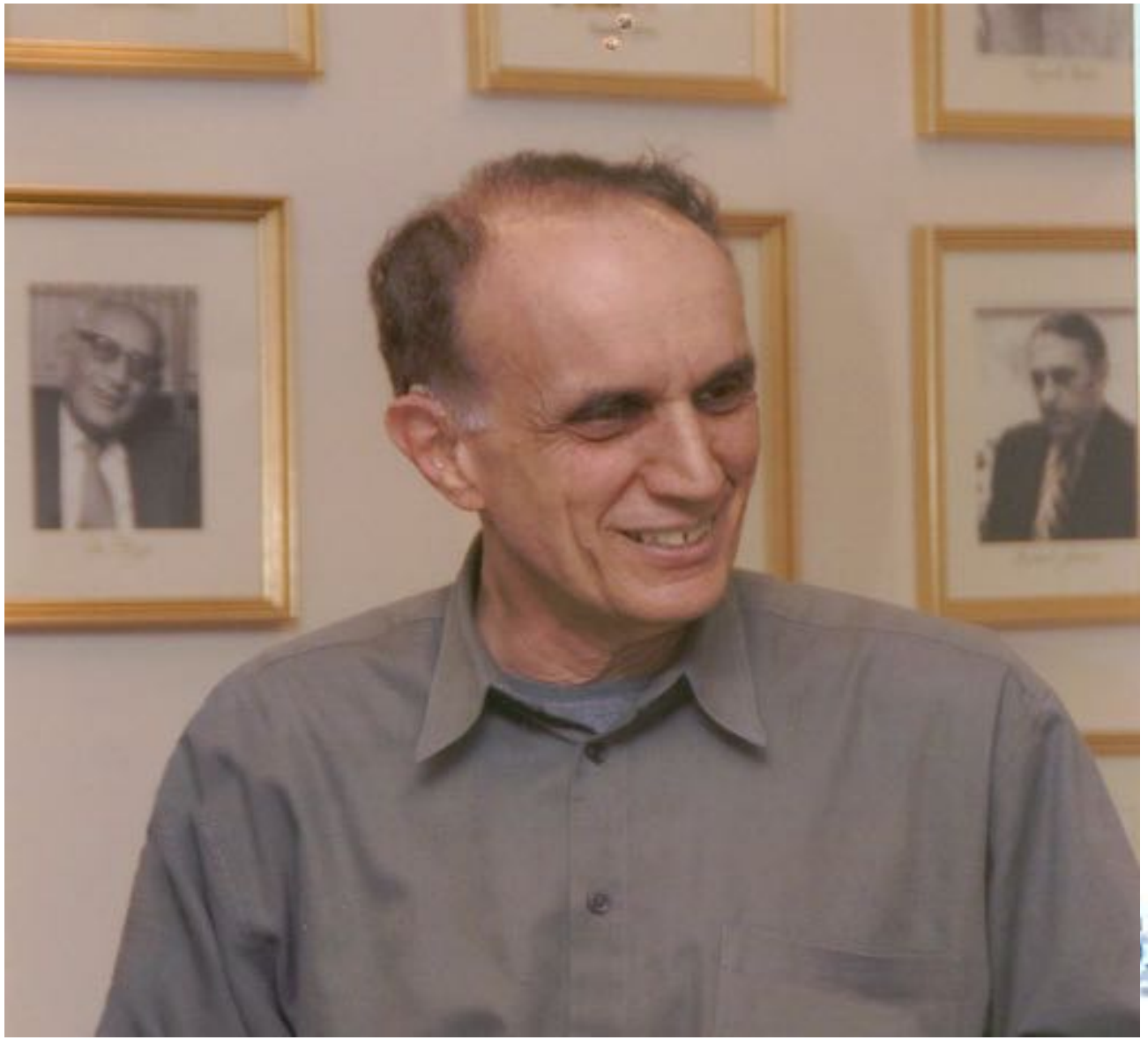

SIGNIFICANCE
April 2019 volume 16 issue 2
Forensic science and statistics
A special issue
ROYAL STATISTICAL SOCIETY
ASA
AMERICAN STATISTICAL ASSOCIATION


SIGNIFICANCE
Art from disaster
How the tragedy of the Titanic inspired dataviz designers
TITANIC
TITANIC
Thank you, thank you
Oscar speeches analysed
From a field in England
R. A. Fisher's random idea
ASA

ASA co-sponsors the International Prize in Statistics

2017: Sir David Cox

2019: Bradley Efron



Benefits of ASA Membership

- Complimentary Electronic Access to all wholly-owned ASA journals
- Complimentary print copies of *Amstat News* (Newsletter) & *Significance*
- “ASA Community” interaction (questions, comments, etc)
- Discounts to ASA-sponsored meetings & workshops
- Career resources! Job fair, Career Center at JSM, …
- Connections with statisticians worldwide

Education + Career Opportunities

- ASA DataFest
  - – Annual competition: teams of undergraduates work on complex dataset
- Data Expo
  - –At JSM: ASA Sections on Graphics, Statistical Computing, and Government sponsor DATA EXPO (special Poster Session): Participants report on analyses of common data set. UVA sends our local team to JSM: they won 2nd prize in Denver!
  - –Student Early Career Travel Fund
  - –ASA Student and Early Career Travel Fund to encourage students and early-career professionals to engage in the statistical community at ASA-sponsored meetings
- Internships and Fellowships
  - –Resource provides a list of internship and fellowship opportunities.
- ASA JobWeb
  - –Find the position that's right for you.
- Professional Development
  - –ASA offers technical & professional skills courses, in-person & virtual workshops.
  - –FREE virtual workshops 2019, 2020: Use of Blended Data & Combining Alternative Source Data: Methods, Challenges, Considerations in the Era of Big Data

2020-2021 Science and Technology Policy Fellow at USDA

gstat Graduate Statistician

pstat® Accredited Professional Statistician®

ASA DataFest

ASA Sections & Interest Groups: Common Research Interests

- 29 Sections: Sponsor events, JSM sessions, sponsor awards, etc.

Biometrics; Biopharmaceutical Statistics; Genomics & Genetics;
Statistical Learning & Data Science; Imaging; Consulting
Government, Defense & National Security; Environment;
SPES: Phys Sci & Engineering Q&P: Quality & Productivity
Statistical Computing; Statistical Graphics
Sports; Social Statistics; Risk Analysis; Lifetime Data Science;
Survey Research Methods; Marketing; Business & Economics

- 9 Special Interest Groups: Informal groups, may grow into Sections

Astrostatistics, Stat Auditing, Business Analytics/Stat Education;
History of Statistics; Pharmacometrics; Quantum Computing & ML;
Text Analysis, Transportation Statistics; Uncertainty Quantification

Participating in ASA

- Conferences & Workshops: Present research, chair/organize sessions
- Submit articles on interesting topics to *Amstat News* or ASA Connect
- Volunteer to serve on a Committee (Members appointed by Pres & VPs)
- Participate in Chapters & Sections & Special Interest Groups
- Apply for Fellowships
- Propose a Speaker for Speaker's Bureau (2020): Jack@amstat.org
- Make an Impact !

ASA Governance

- Council of Section Representatives
- Council of Chapter Representatives
- ASA Board of Directors: 14 Voting Members

  3 Presidents (Past, Current, Future): Elect 1 each year

  3 Vice Presidents: Elect 1 each year (3rd Year VP: Executive Comm)

  3 Reps from Council of Sections: Elect 1 each year

  3 Reps from Council of Chapters: Elect 1 each year

  1 Publications Representative (Elect 1 every 3 years: 2020, 2023, …)

  1 International Representative (Elect 1 every 3 years: 2020, 2023, …)

  (plus 2 appointed members: Treasurer & Secretary=ASA Exec Director)

What do Presidents do?

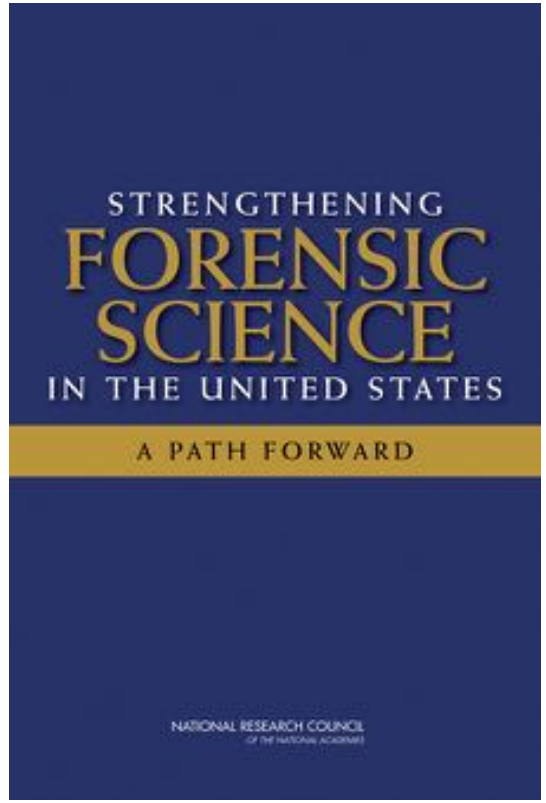
- Write monthly column in *Amstat News* (no one reads it)
- Respond to members who criticize column content
- Set JSM theme, give JSM address, invite JSM Speaker:

  see ww2.amstat.org/meetings/jsm/2019/webcasts/index.cfm
- Propose Initiatives for ASA to pursue
- 2004 (Efron): Electronic Pubs, *SBR;* 2006 (Keller): Science Policy
- 2018: Expert witness course ($995!), Leadership Institute
- 2019: Impact, “Fake News”, Diversity (Latina(o), Native American); “Significance”

ASA & Public Policy: Steve Pierson, Daniel Elchert

Advocate for scientific integrity and evidence-based policymaking:

- 11th hour addition of citizenship question to 2020 census
- Move Economic Research Service from DC to Kansas City
- Use of reliable data and evidence in policymaking
- Strong Federal Statistical System (Census, BJS, BTS, BLS, …):
  (Federal Government employs many statisticians!)
- Support Forensic Science Reform →



Support Fellowship Opportunities:

- ASA Science Policy
- ASA-AAAS Fellowships: Data Science, Science & Technology

Is this person a citizen of the United States?

- ☐ Yes, born in the United States
- ☐ Yes, born in Puerto Rico, Guam, the U.S. Virgin Islands, or Northern Marianas
- ☐ Yes, born abroad of U.S. citizen parent or parents
- ☐ Yes, U.S. citizen by naturalization – *Print year of naturalization*

- ☐ No, not a U.S. citizen

THE KANSAS CITY STAR.

GUEST COMMENTARY

Sonny Perdue's claims about USDA agencies' move to KC don't hold up to scrutiny

The USDA's decision to move hundreds of research jobs out of Washington to Kansas City has triggered a backlash among federal employees. Employees stood in silent protest during a meeting about the move to Kansas City, in

Under ordinary circumstances, I would celebrate the U.S. Department of Agriculture's relocation

State Rep. Blake Carpenter explains his legislation on governors' age limits

Seven things about Mizzou's declining enrollment

STATISTICS WITHOUT BORDERS

Statisticians in compassionate service.

Statistics Without Borders (SWB) is a volunteer Outreach Group of the *American Statistical Association* that provides pro bono services in statistics and data science. We work to improve decision making and knowledge in efforts that promote welfare through the proper application of statistical principles and best practices, where access to such resources is limited.

Statistics Without Borders: Projects
swb.wildapricot.org

- Training for the Asante Africa Foundation (Kenya and Tanzania)
- Save the Children in Ethiopia (design analysis plan for cohort study of development in 2,000 children in Tigray, focusing on educational and early childhood interventions)
- Tacugama Chimpanzee Sanctuary in Sierra Leone (evaluating effect of environmental awareness campaign of the illegality of chimpanzee trading; creation of electronic data base for veterinarians)
- UNICEF Sierra Leone (evaluate health interventions)
- 2010 Earthquake in Haiti (estimating extent of impact and damage).

More challenges & partnerships ahead for Statistics:

- Ensure Election Security
- Combat Disinformation
- Estimate Human Trafficking
- Forecast Natural Disasters & their effects on public health & environment
- Data Science: Will we be replaced by AI?

Gary Marcus & Max Little:
https://www.statnews.com/2019/10/23/advancing-ai-health-care-trust/
(Forwarded to KK by Jordan Rodu)

"A statistician working alone is a statistician making mistakes."
David P. Byar (1938-1991)

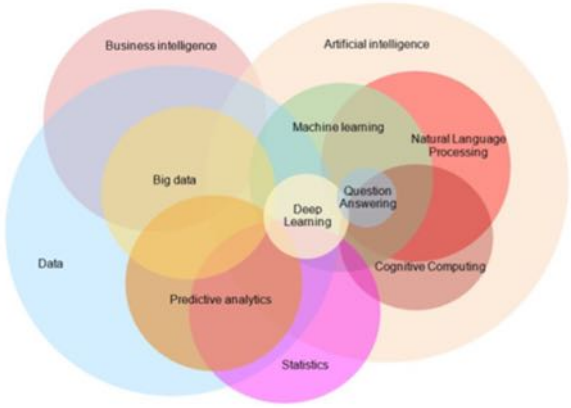
Advancing AI in Health Care: It's all about Trust

Three years ago, artificial intelligence pioneer Geoffrey Hinton said, "We should stop training radiologists now. It's just completely obvious that within five years, deep learning is going to do better than radiologists."

Today, hundreds of startup companies around the world are trying to apply deep learning to radiology. Yet the number of radiologists who have been replaced by AI is approximately zero. (In fact, there is a worldwide shortage of them.)

At least for the short term, that number is likely to remain unchanged. Radiology has proven harder to automate than Hinton — and many others — imagined. For medicine in general, this is no less true. There are many proofs of concept, such as automated diagnosis of pneumonia from chest X-rays, but surprisingly few cases in which deep learning ... has achieved the transformations and improvements so often promised.

The Role of Statistics in Modern Data Analysis
or
A Rose by Any Other Name



Hal Stern
Department of Statistics
University of California, Irvine
sternh@uci.edu
(KK edits in RED)

UCI Department of Statistics
Donald Bren School of Information & Computer Sciences

Quotation from a paper entitled “Data Analysis and Statistics: An Expository Overview”

Four major influences act on data analysis today:

1. The formal theories of statistics.
2. Accelerating developments in computers and display devices.
3. The challenge, in many fields, of more and ever larger bodies of data.
4. The emphasis on quantification in an ever wider variety of disciplines.

A quote from a paper entitled “Data Analysis and Statistics: An Expository Overview”

Four major influences act on data analysis today:

1. The formal theories of statistics.
2. Accelerating developments in computers and display devices.
3. The challenge, in many fields, of more and ever larger bodies of data.
4. The emphasis on quantification in an ever wider variety of disciplines.

DATA ANALYSIS AND STATISTICS: AN EXPOSITORY OVERVIEW *

1966, Proceedings Fall Joint Computer Conference

J. W. Tukey and M. B. Wilk

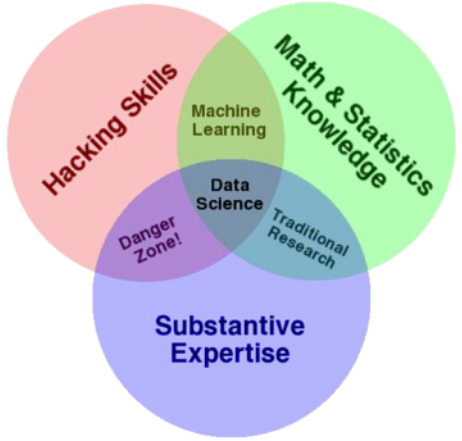
Princeton University and
Bell Telephone Laboratories, Inc.
Princeton and Murray Hill, New Jersey

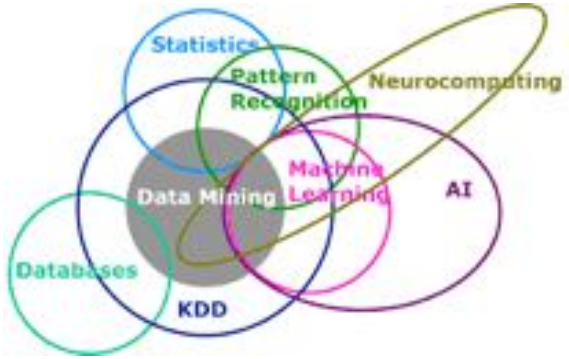
Terminology

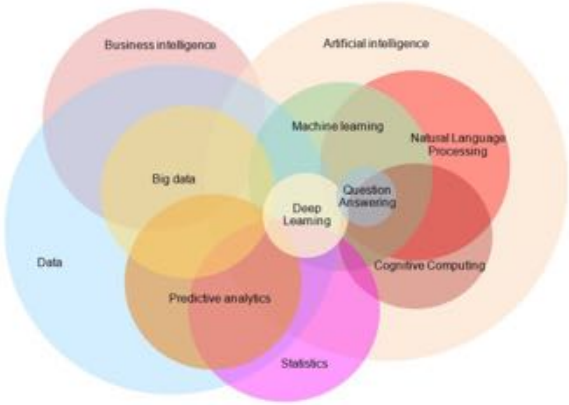
- Today, many terms associated with data analysis:
  - Statistics ☺
  - Data Science
  - Machine Learning
  - Big Data
  - Artificial Intelligence (AI)
  - Deep Learning (AKA Multi-Layer Neural Networks)
- Confusing? Can we clarify?

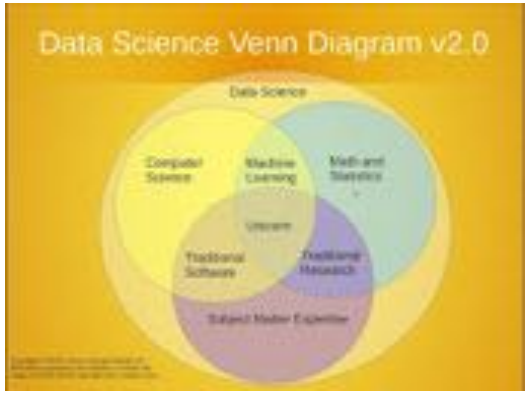
A Venn Diagram to Explain it All

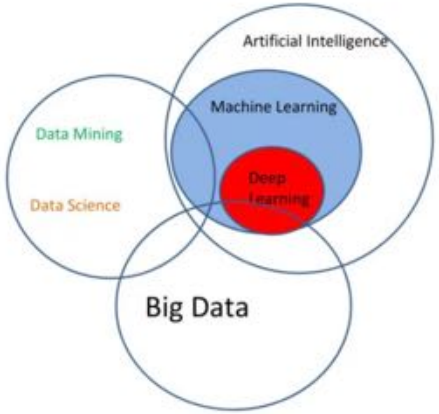
One of the first by Drew Conway

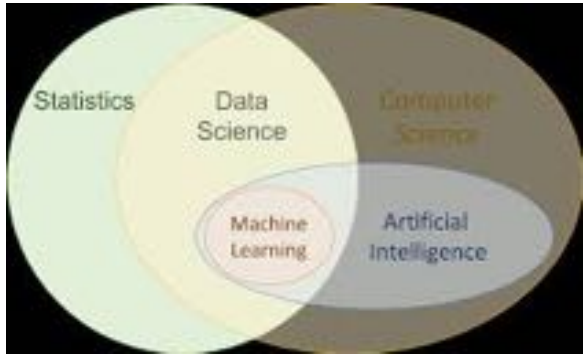












Terminology

A humorous take credited to Joel Grus:

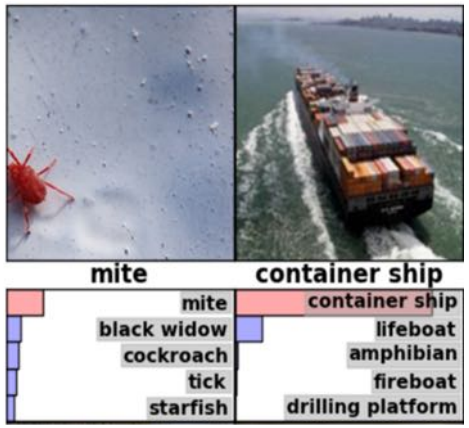
Terminology: AI, ML, “Deep Learning”

- AI contains ML contains Deep Learning
- AI goes back to at least the early 1950s
  - Has had several incarnations
  - Development of programmable computers
  - Neural networks
  - Expert systems
- Machine Learning terminology was coined early (1959) but emerged as a force in the 1990s
  - Some in CS viewed true AI as “too hard”
  - Led to a split in AI community
  - Result: ML develops algorithms to solve specific problems

Deep Learning

Deep Learning: a particular class of machine learning algorithms (multi-layer deep neural networks) that have proven **incredibly effective** at certain tasks (e.g., Siri: "What can I help you with?")

- Neural networks have a very long history
  (computational models of the neuron date back to the 1940s)
- Neural networks were extremely popular in the 1980s
- Started losing favor in the 1990s
  (they were not better than other predictive technologies)
- NNs returned in ML with a vengeance!



Terminology

- Relationship of AI / ML to Statistics?
- Statistics: focus on collection, analysis and interpretation of numerical data (Efron: "**Estimation & Attribution**")
- AI often refers to enabling computer systems to perform tasks normally assumed to require human intelligence
- Machine Learning

  Now often refers to the study of algorithms / models for performing tasks (e.g., classification)

  Efron: "Emphasis is on **Prediction**" (his talk in Aug 2019)

Terminology

- **Statistics' (stereotyped) view of ML**

"Computer science discovers the power of probability and statistical models to solve problems / analyze data"

"ML is not concerned about source of problems / data"

- **ML's (stereotyped) view of Statistics**

"Statisticians focus on mathematical theories for data analysis"

"Statisticians care more about interpretation / testing of models and less about making algorithms work"

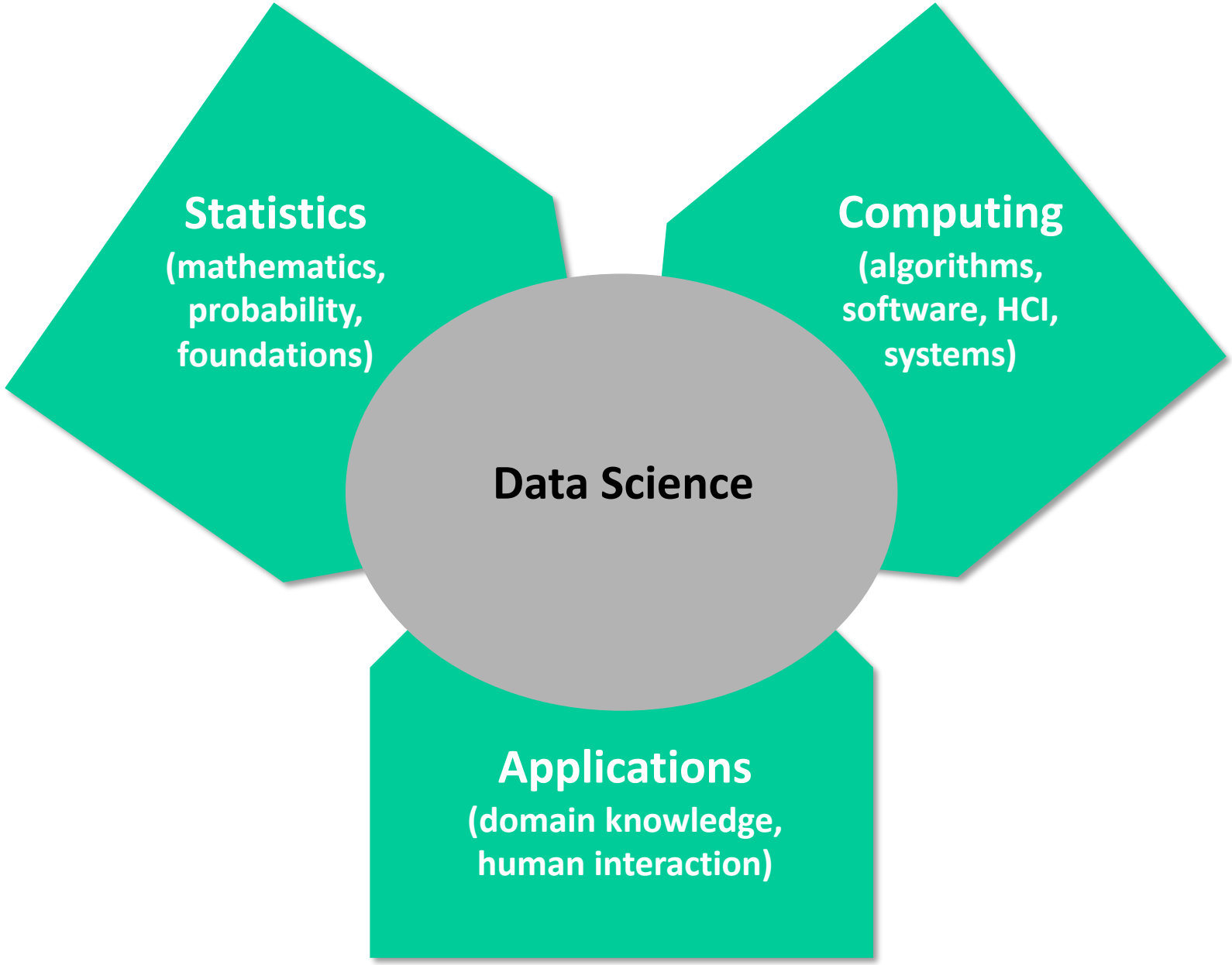
"Statistics cannot handle very large data sets"

Contributions of CS

- Key areas of expertise
  - Databases
  - Algorithms
  - Programming innovation
- These skills allowed ML researchers to
  - More easily obtain data (e.g., web crawling)
  - Manipulate large, heterogeneous data sources
  - Scale up algorithms to handle the large data sets

Statistics: Areas of Expertise

- Experimental design, statistical sampling for data collection
- Underlying mathematical theory that justifies procedures
- Estimation: Model, parameters, uncertainty
- Measures of uncertainty (e.g., confidence intervals)
- Emphasize distinction between correlation and causation


Statistics
(mathematics, probability, foundations)
Computing
(algorithms, software, HCI, systems)
Data Science
Applications
(domain knowledge, human interaction)

Modern Data Analysis

- Vast array of data analysis methods/strategies
- Includes techniques from statistics, machine learning, artificial intelligence
- Methods / models vary in many ways
  - Human input required to build the model
  - Size / dimension of the model (#parameters)
  - Amount of data required to fit the model
  - Strength of the assumptions required
  - Interpretability of results of applying the model

Modern Data Analysis

- Applications also vary
  - Objectives
    - Prediction / interpretation
    - Effects of causes / data exploration
  - Amount of data
    - e.g., logistic regression on millions of cases
    - e.g., presence / absence of training data
  - Heterogeneity of data types (images, text)
  - Frequency of analysis (once vs repeated model fitting)
  - Need for fairness / equity

Hal Stern's Conclusions: Stat & DS

- Statistics continues to play a large role in modern data analysis
  - Expertise in experimental design and data collection
  - Emphasize the importance of uncertainty and variability
  - Wide-range of techniques that have proven useful in science and policy
- The impact of statistics and statisticians will grow if we:
  - Embrace new methods/models, study their strengths & weaknesses
  - Include new developments in computer science in our programs
  - Continue to work well in collaborative teams

ASA Ad-hoc Committee on Statistics & Data Science:
Co-chairs Mark Glickman, Kathy Ensor

KK's conclusions

Statistics: A profession where you can work in almost any field

- Physical Science, metrology (NIST)
- Instrumentation, calibration (HP/Agilent)
- Cancer data: Geographic Epidemiology, Screening Trials
- Forensic Science: Stat foundations for Pattern-Matching Algorithms
- Law: Reliability of eyewitness evidence in the courtroom
- Genomics: Which genes affect protein production?
- Medicine: Clinical Trials
- Surveys: Census, Labor, Polling
- JW Tukey: "Play in everyone's sandbox"

It is hard for me to overstate the impact that my profession has had on my life

- Worked with incredibly smart people
- Many became good friends
- "*On apprend à tous les âges*" (ADK)

Thank you & Enjoy the conference!

Karen Kafadar

2019 President, American Statistical Association

Promoting the Practice and Profession of Statistics