UCLA
Department of Statistics Papers

Title
Women mathematicians in data-centric occupations (with a context)

Permalink
https://escholarship.org/uc/item/86n169xf

Author
Sanchez, Juana

Publication Date
2020-03-19

Supplemental Material
https://escholarship.org/uc/item/86n169xf#supplemental

Women mathematicians in data-centric occupations (with a context)

Juana Sanchez

UCLA, Department of Statistics

jsanchez@stat.ucla.edu

http://directory.stat.ucla.edu/faculty/single-page/?smid=159

Texas State University, San Marcos, Texas   2/28/2020
Joint Seminar of Statistics, Women Doing Math and Talk Math 2 Me groups

Thank you to

Talk Math 2 Me,
Statistics,
and
Women Doing Math
organizers for inviting me

Special thanks tor Dr. Vera Ioudina for making my whole stay an unforgettable experience and for introducing me to this amazing group of scholars and this university.

OUTLINE

- The context:
    - A little about you, me, Statistics at UCLA, and Statistics Education Research in the U.S. and around the globe
    - The data deluge and the start of another, yet more rushed, cycle of research?
- **Women mathematicians** in data-centric jobs
- Probability
    - When probability is the first stats course a student takes
    - Why I wrote a book on probability in the context given
- Conclusions

1. The context.
A little about you. Use color cards given to you to answer

Are you

(A) An undergrad, future school teacher?

(B) A grad, future school teacher?

(C) Grad or undergrad, future "some occupation other than teaching"?

(D) Other ?

A	B
C	D

A little about you

You

(A) Enjoy analyzing data with statistical methods and computers.

(B) Have never analyzed data with statistical methods or other methods.
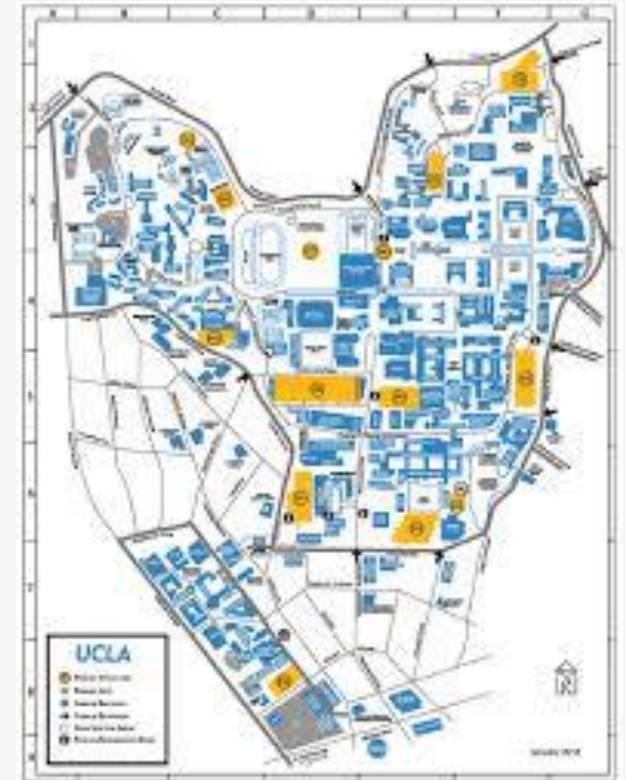
(C) Have modeled mathematically without using data

A little about me, and the Stats Dept at UCLA

Before 1998

Math dept

The division of stats housed by math, representing interests of the whole north campus departments
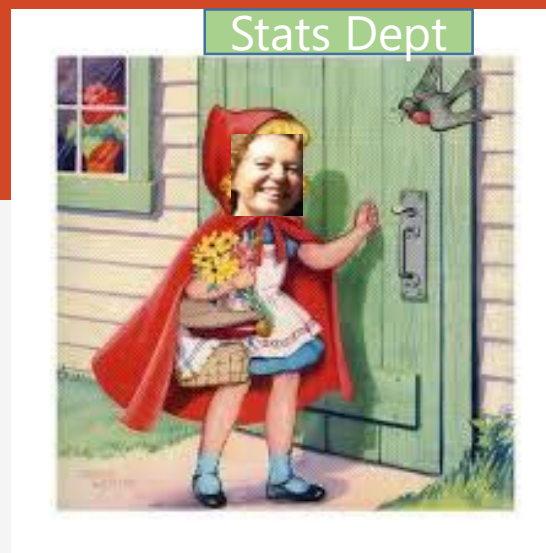
A little about me, and the Stats Dept at UCLA

Stats Dept

Before 1998

Math dept

The division of stats housed by math

1998 split

New Stats dept is born, offering Ph.D and, major and minor (later)

1998-2018 Stats dept

- **UCLA Stats majors and minors**
- **10 years of Datafest**
- **Preparing for math stats, machine learning and data science**

Motto

Motto →

> "**Statistics** is a **science**, not a branch of mathematics, but uses mathematical models as essential tools."
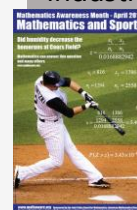>
> –John Tukey

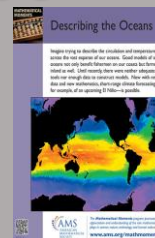**Stats as the "Science of data," not a subfield of mathematics.**

1998-2018 Math Dept

- **Pre-requisites for Stats**
- **Losing majors to Stats**
- **Math majors minoring in in Stats**

Posters by: Joint Policy Board for Mathematics (JPBM) – effort of the American Mathematical Society, Mathematical Association of America and Society for Industrial...

Describing the Oceans

WHAT WOULD YOU DO WITH ALL THIS DATA?

Mathematics Awareness Month - April 2010
Mathematics and Sports

When will I use math?

Applied math infiltrating data, posters on data problems

8

1998-2019, meanwhile,outside stats and math dept

Big prediction error by BLS. Make this three times more

9

1998-2018 Meanwhile, Statistics education research leading new ways to teach the Intro Stats course
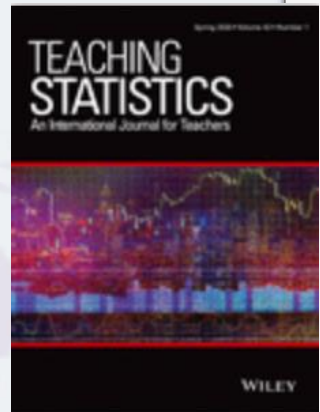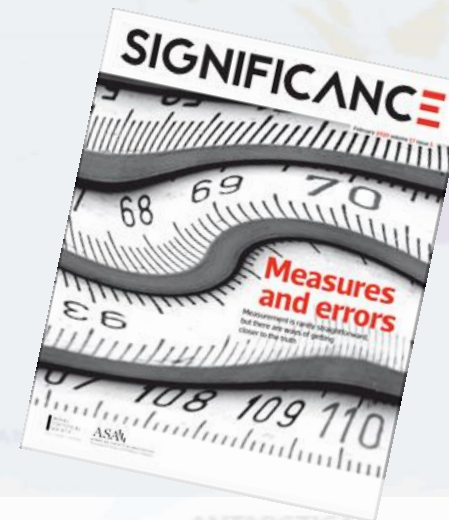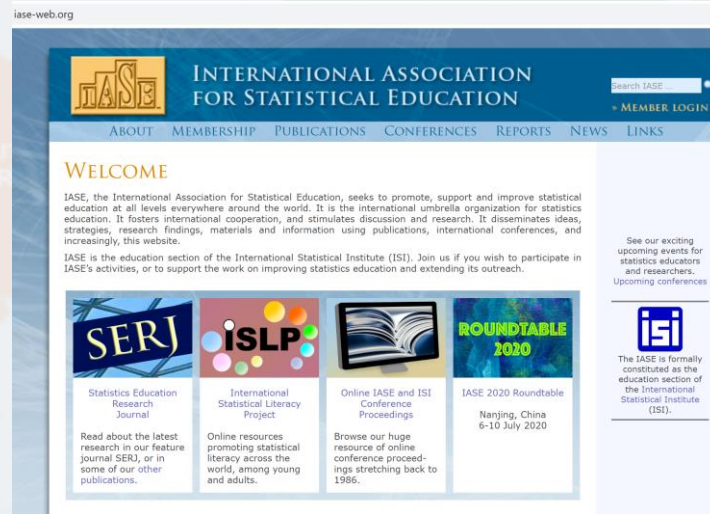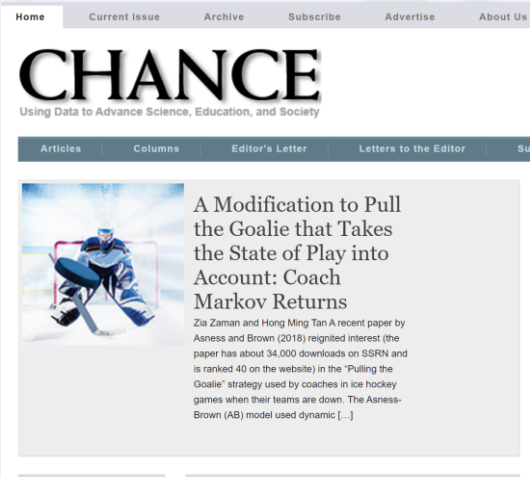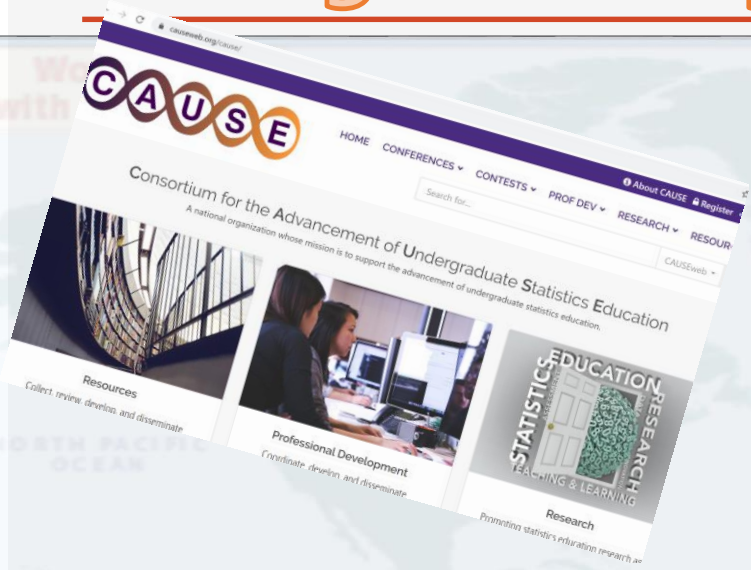
1998-2018 Implementing Stats education research in schools, professional development for teachers

https://www.amstat.org/ASA/Education/K-12-Educators.aspx

https://www.nctm.org/Standards-and-Positions/Principles-and-Standards/Data-Analysis-and-Probability/

3 strands running through all Levels K-12

Statistical Investigation
Skills & mental habits for "doing"

Probability
includes math. connections

Statistical Literacy
Skills & mental habits for interp./critiquing reports of investigations done by others
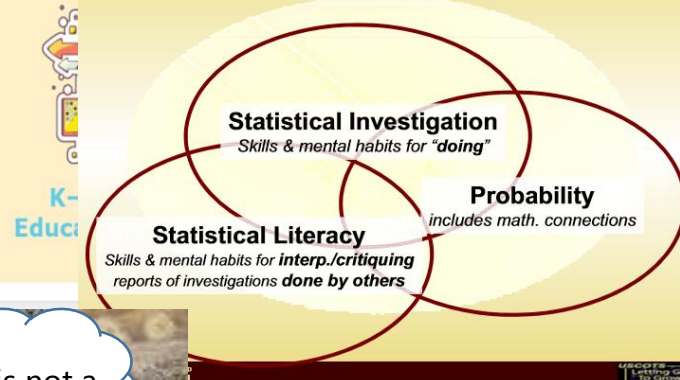
Statistics is not a branch of Math

https://www.amstat.org/asa/education/Guidelines-for-Assessment-and-Instruction-in-Statistics-Education-Reports.aspx

According to the new intro stats course and stats education in school model (reflected in poster), statistics consists of

(A) Making sense of figures
(B) Systems guides for, tough, fuzzy issues wherever the collection and interpretation of data involved
(C) The kernel of the process of inquiry-the common core that is left when one stripes away the contexts of particular investigations.
(D) The study of the process of scientific enquiry

https://new.censusatschool.org.nz/resource/data-detective-poster/
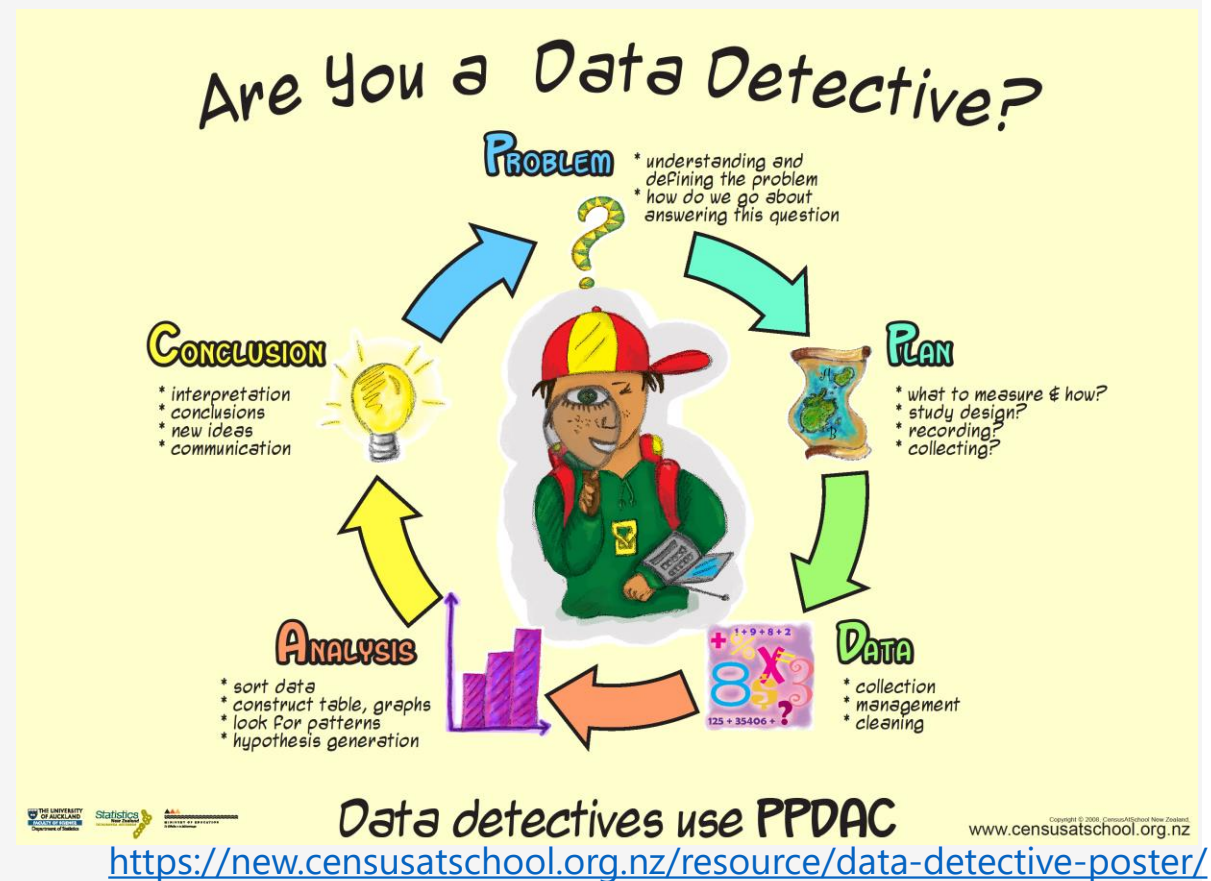
The Data Deluge

It's here before we are ready

Source of slide: Chris Wild, USCOTS 2013

13
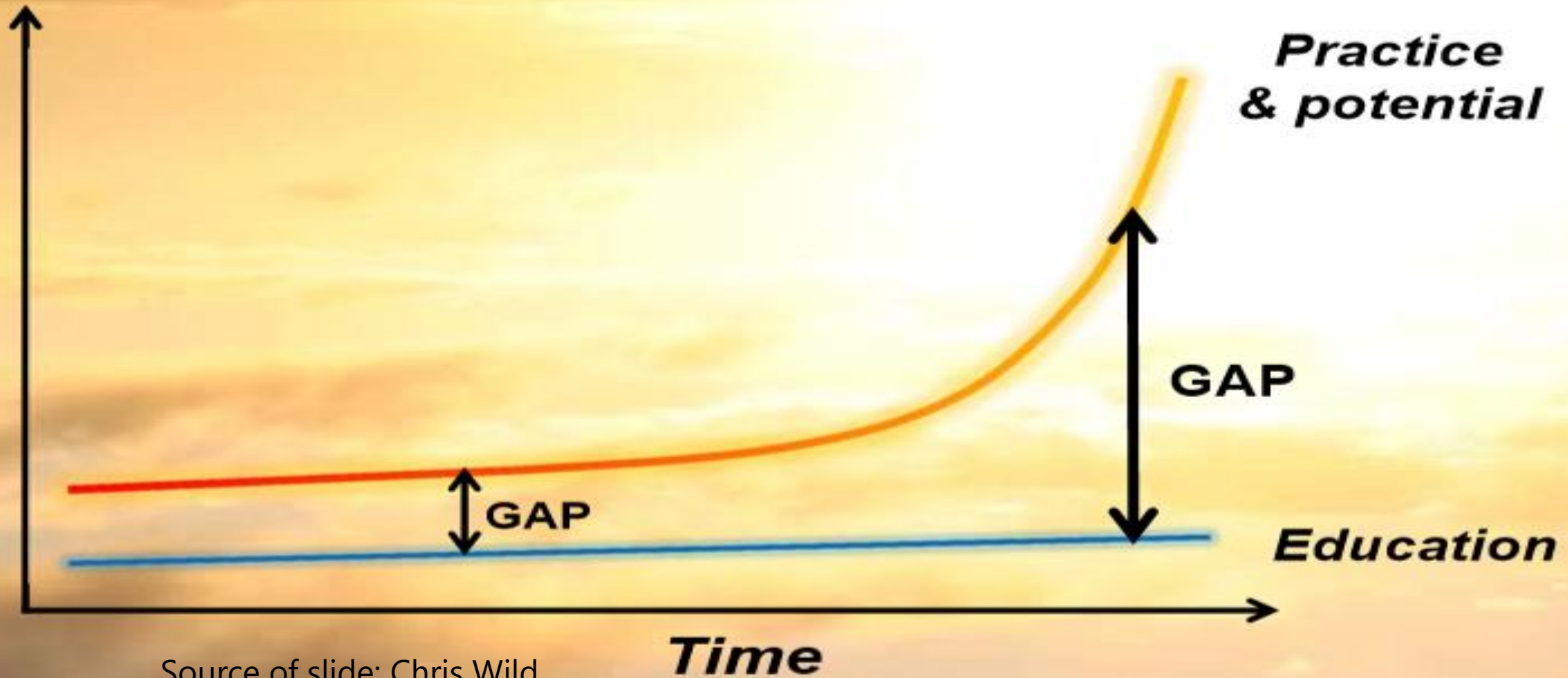
Source of slide: Chris Wild,
USCOTS 2013

2016 Mathematics Awareness Month
The Future of Prediction

TELEPHONES
WATCHES
COMPUTERS
INTERNET OF THINGS

Yogi Berra, paraphrasing Niels Bohr, said "It's tough to make predictions, especially about the future." Throughout Mathematics Awareness Month 2016, we will explore how mathematics and statistics are the future of prediction, providing insights and driving innovation. During the month, we will be asking the question, "What's next?" and exploring how mathematicians and statisticians contribute to the future. So, visit the website and join the quest to help us predict future timeline entries.

"Want to make accurate predictions? Get lucky, get psychic, or get data! Good data + great statistical methods = great future for prediction!"
JESSICA UTTS, 2016 American Statistical Association President

JOIN THE DISCUSSION! What does the "Future of Prediction" mean to you? Share your thoughts by tweeting @mathaware or sharing on www.mathaware.org

2011
Nest Labs, Inc., architect of the thoughtful home, introduces the Nest Learning Thermostat.

1971
A team of IBM engineers invent the floppy disk.

1994
Reza Raji described the concept in *IEEE Spectrum* as "[moving] small packets of data to a large set of nodes, so as to integrate and automate everything from home appliances to entire factories."

1924
John Harwood invents the first successful self-winding watch system.

1917
Finnish inventor Eric Tigerstedt files a patent for a "pocket-size folding telephone with a very thin carbon microphone."

1948
Claude Shannon publishes "A Mathematical Theory of Communication" in the *Bell System Technical Journal*.

1969: The quartz watch is introduced. In place of a balance wheel (oscillates at 5 beats per second), it uses a quartz crystal resonator that vibrates at 8,192 Hz, powered by an oscillator circuit.

1982
The concept of a network of smart devices was demonstrated with a modified Coke machine at Carnegie Mellon University, becoming the first Internet-connected appliance.

2004
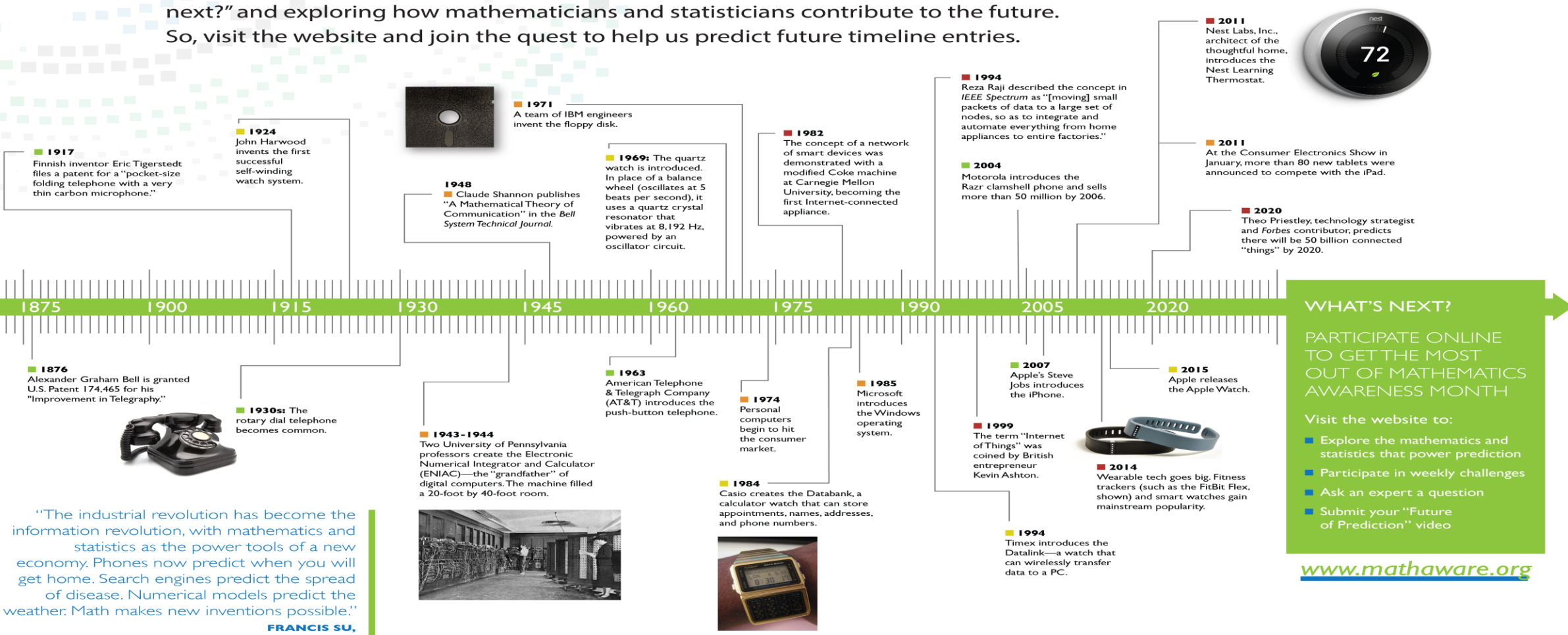Motorola introduces the Razr clamshell phone and sells more than 50 million by 2006.

2011
At the Consumer Electronics Show in January, more than 80 new tablets were announced to compete with the iPad.

2020
Theo Priestley, technology strategist and *Forbes* contributor, predicts there will be 50 billion connected "things" by 2020.

1875 · 1900 · 1915 · 1930 · 1945 · 1960 · 1975 · 1990 · 2005 · 2020

1876
Alexander Graham Bell is granted U.S. Patent 174,465 for his "Improvement in Telegraphy."

1930s: The rotary dial telephone becomes common.

1943–1944
Two University of Pennsylvania professors create the Electronic Numerical Integrator and Calculator (ENIAC)—the "grandfather" of digital computers. The machine filled a 20-foot by 40-foot room.

1963
American Telephone & Telegraph Company (AT&T) introduces the push-button telephone.

1974
Personal computers begin to hit the consumer market.

1985
Microsoft introduces the Windows operating system.

1984
Casio creates the Databank, a calculator watch that can store appointments, names, addresses, and phone numbers.

2007
Apple's Steve Jobs introduces the iPhone.

1999
The term "Internet of Things" was coined by British entrepreneur Kevin Ashton.

2015
Apple releases the Apple Watch.

2014
Wearable tech goes big. Fitness trackers (such as the FitBit Flex, shown) and smart watches gain mainstream popularity.

1994
Timex introduces the Datalink—a watch that can wirelessly transfer data to a PC.

"The industrial revolution has become the information revolution, with mathematics and statistics as the power tools of a new economy. Phones now predict when you will get home. Search engines predict the spread of disease. Numerical models predict the weather. Math makes new inventions possible."
**FRANCIS SU,
2016 Mathematical Association of America President**

WHAT'S NEXT?

PARTICIPATE ONLINE TO GET THE MOST OUT OF MATHEMATICS AWARENESS MONTH

Visit the website to:
- Explore the mathematics and statistics that power prediction
- Participate in weekly challenges
- Ask an expert a question
- Submit your "Future of Prediction" video

www.mathaware.org

15

2019-- Stats dept

Data theory major
jointly administered with math

Stats + math

Students are waiting, observing..

2019—Math Dept

Data theory major (first in the world) jointly administered with stats

https://www.math.ucla.edu/sites/default/files/ugrad/Majors,%20Minors%20&%20Spec/pdf/Data%20theory%20page.pdf

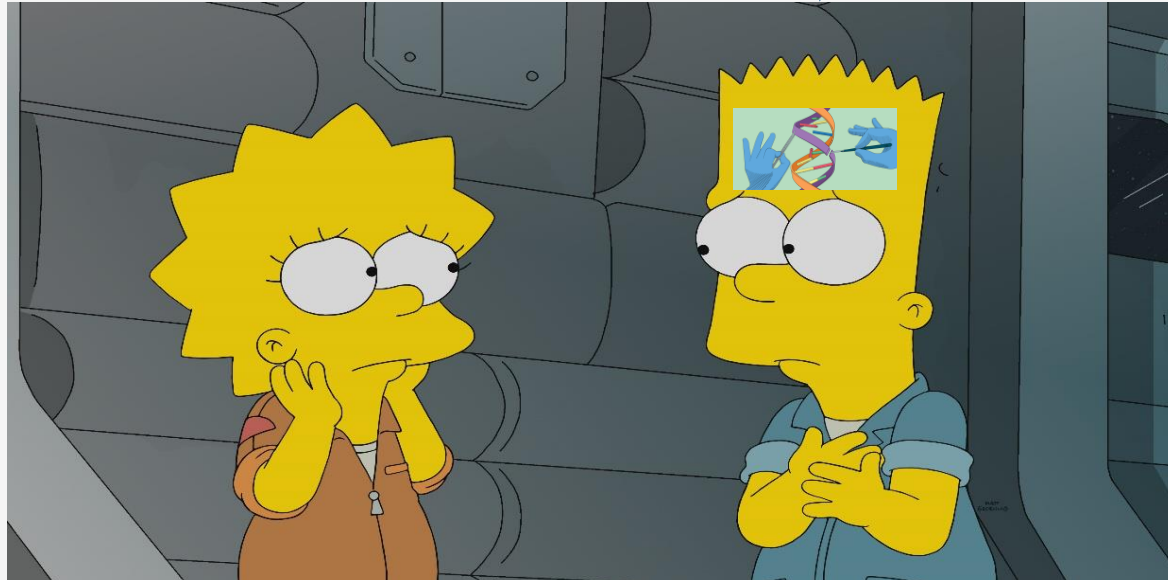**Stats +math**

While Computer Science creates a quick track

https://www.uclaextension.edu/digital-technology/data-analytics-management/certificate/data-science

All now flying together, while separate, and feel allright

The future for teachers--- The cycle begins again
Mobilize: partnership Center X, CENS at UCLA, LAUSD, CSTA

College Futures Foundation to Support Data Science Instruction in LA Schools

Data Science was introduced in the Los Angeles Unified School District (LAUSD) in the 2014-2015 school year and, in 2018, UCLA joined with the LAUSD to provide additional support for the course. Introductory Data Science (IDS), which fulfills the UC/CSU A-G subject matter requirement in mathematics, is currently available in 16 high schools in the school district.

https://www.mobilizingcs.org/

Introduction to Data Science
Our data. Our lives.

Curriculum | IDS Partnership | Technology | Professional Development | Publications, Presentations & News

Curriculum Overview

Unit #	Title	Description
Unit 1	Data and Visualizations	Introduces students to fundamental notions distribution and multivariate associations and interpreting visualizations of real-world proce
Unit 2	Distributions, Probability, and Simulations	Students use numerical summaries to descr probability through the lens of computer sim
Unit 3	Data Collection Methods: Traditional and Modern	Prepares students to learn about the various ways of collecting data, including Participatory Sensing, and the effect that data collection has on their interpretation of the patterns they discover
Unit 4	Predictions and Models	Students learn to make and how to use mathematical and statistical models to predict future observations and how data scientists measure the success of these predictions

And the former stats education research crusaders are running towards research and curriculum making again

International Data Science in Schools Project (IDSSP)

Purpose   About Us   Supporters   Contact   Resources   Framework   News

The last decade has seen unprecedented growth in the availability of data in most areas of human endeavor. Whole branches of science have been developed to allow corporations to transform the way marketing is conducted, to drive scientific progress in areas such as Bioinformatics, and to inform decision-making at all levels in governments and industry. Further, the scale and complexity of much of these data are beyond the capability of a single computer to manage or a single individual to analyze.

These realities generate a very significant imperative to ensure that there is an adequate supply of people entering the workforce who are equipped to handle the new challenges of learning from data. There is a compounding factor: on the evidence available, demand for data scientists is not only massively outstripping supply, but the situation is worsening, and this is a world-wide problem.

And beyond this, there is an equally pressing need for people in our societies to be more capable of understanding, interpreting and making decisions based on quantitative data as they cope with the vagaries of life.

The purpose of this international collaborative project is to transform the way education in Data Science is carried out in the last two years of school, with two objectives:

1. To ensure that school students acquire a sufficient understanding and appreciation of how data can be acquired and used to make decisions so that they can make informed judgments in their daily lives, as students and then as adults. In particular, we envisage future generations of lawyers, journalists, historians, and many others, leaving school with a basic understanding of how to work with data to make decisions in the presence of uncertainty, and how to interpret quantitative information presented to them in the course of ... es.

   ... ool students sufficient interest and enthusiasm for Data Science that ... s in Data Science with a view of making a career in the area.
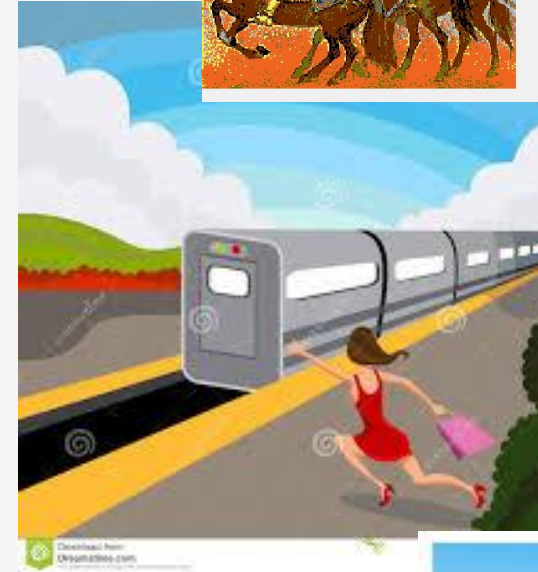
Will we see from now on thousands of articles on how to teach data science? On how to implement data science in the curriculum? Probably

21

A little about you

In your opinion,

(A) The NCTM will put the standards for data science as part of the math standards

(B) The NCTM will not add the standards for data science as part of the math standards

A B
C D

2. WOMEN MATHEMATICIANS IN DATA-CENTRIC JOBS

We always put women in academics or PhDs as role models. (on the left)

These are too far from students who these days jump into jobs after a B.S.

The context I presented has put young women, with B.S. in important data-centric jobs

Who was the first person to be what we would now call a coder (someone that writes computer programs to be executed by a computer?)

(A)Lady Ala Lovelace

(B)Charles Babbar

(C) Mary Allen Wilkes

(D) Arlene Gwendolyn Lee

Who was the first person to be what we would now call a coder (someone that writes computer programs to be executed by a computer?)
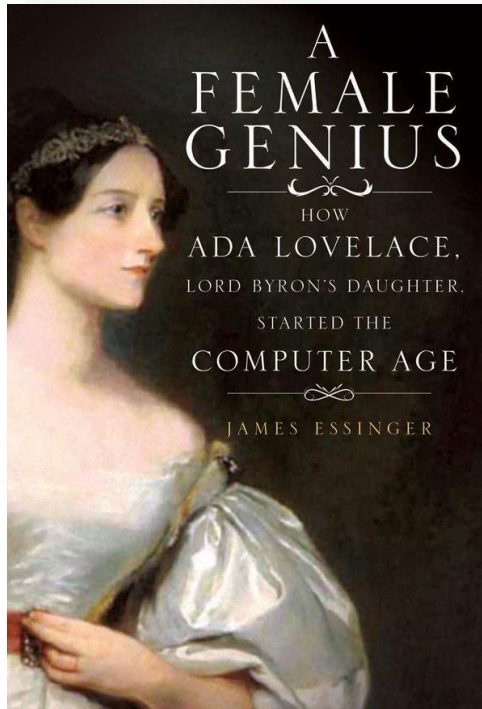
(A)Lady Ala Lovelace, as a young mathematician in England in 1833, wrote the first computer program in history, an algorithm with which an Analytical Engine designed by Charless Babbar would calculate the Bernoulli sequence of numbers.
Died at 36, never saw the code executed.

(A)Charles Babbar

(A) Mary Allen Wilkes

(B) Arlene Gwendolyn Lee

A   B

C   D

A woman, Mary Allen Wilkes, wrote the software that would let a user control the first interactive personal computer in real time, in 1961

In the `1940s women were pioneers in writing software for the machines. Kathleen McNulty, Jean Jennings, Betty Snyder, Marlyn Wescoff, Frances Bilas, Ruth Lichterman, programmed to execute instructions. There was no google, no internet forums, just "their brain in a jar, solving hellish problems."

For men, the real glory lay in making the hardware.

In the '50s and '60s employers looked for candidates who were logical, good at math and meticulous and passed aptitude tests. Gender stereotypes worked in womens' favors.

"The computer did not care that I was a woman or that I was black. Most women had it much harder."
(Arlene Gwendolyn Lee, one of the early female programmers in Canada).

Katherine Johnson Dies at 101; Mathematician Broke Barriers at NASA

She was one of a group of black women mathematicians at NASA and its predecessor who were celebrated in the 2016 movie "Hidden Figures."

Katherine Johnson, part of a small group of African-American women mathematicians who did crucial work at NASA, in 1966. NASA/Donaldson Collection, via Getty Images

February 13th, 2019 by Clive Thompson, New York Times

The change after the 60s.

Managers began picking coders less on the basis of aptitude and more on how well they fit a personality type: the acerbic, aloof look, with incidence of rugged individualism or nonconformity

Times are changing again, women coders and data analysts are back in big ways

Women mathematicians in data-centric occupations

Kitu (B.S. in Stats 2018) works as a Machine Learning Engineer at O'Reilly Media. She is building a recommendation system to recommend personalized content to users in real-time, similar to how YouTube recommends personalized videos in real-time based on a user's profile. As an engineer in a data space, she gets to deploy her work to production. It is challenging because of how many new technologies she has to become comfortable with.

Kitu recommends:

Kitu says: "be capable of being a life-long learner. Also, as females, we often feel guilty and timid to admit when we have questions, but the only way to survive the data industry is to constantly be asking questions. Reach out for help often, and in turn guide others. Collaborating with others is immensely useful in data science since you learn and share much more in a group setting that comprises different skills and ideas than you would on your own. Data science is meant to be a collaborative art, so be comfortable working with others and challenging others' assumptions.

Women mathematicians in data-centric occupations

Janet (B.S. in Biochemistry,minor in stats)

She is Staff Research Associate I at UCLA's Brain Research Institute. She leverages information from electronic (confidential) hospital records and geospatial data sets to study the geospatial characteristics of mental illness in Colombia. She is looking at clusters or "hotspots" of increased incidence of mental illness to determine environmental factors affecting it. She plans to introduce genetic data in the future. She is crazy about the newly named "Computational Medicine dept." She says: "challenging myself to apply something new, coding or method, I learn faster."

Women mathematicians in data-centric occupations

Hena (B.S. in Stats 2018, minor in math) works as a Technology Analyst at Goldman Sachs. She has learned to effectively solve a wide range of business problems using quantitative techniques, working in teams, and communicating to technical and nontechnical colleagues. She also knows how to be okay with uncertainty – "our job as statisticians," she says, "is to help quantify some of it." It has shaped the way I view the world beyond just my career.

Women mathematicians in data-centric occupations

Connie (B.S. in Econ 2018, minor in stats)

works as database manager at an company that underwrites insurance policies for business. She says that the theories and IT learned in Stats help her complete tasks related to this new database which records insurance premiums and local taxes for firma with locations in various countries. She says that "models and realities clash and we have to do without the assumptions learned in the classroom."

Women mathematicians in data-centric occupations

Heidi (C. Phil alumnus 2001), senior project engineer and after that systems director at The Aerospace Corporation, was part of a team that won a President's Achievement Award in September 2015. The team demonstrated that on-orbit Global Positioning System IIR and IIR-M satellite batteries could be life-limiting for the satellites, and developed a charging protocol to extend the battery life, later proposing an orbital testing approach.

# 3. In this data-centric world context, with high demand for our majors, Probability, not Intro Stats, is

the first and only Statistics course many Engineers and Computer Scientists take at UCLA.

(Math 170A, satisfies the same pre-requisite as Stat 100A)

Probability has not triggered as much statistics education research and claims of need for reform as the intro stats class. Data is not seen as important in probability. Why?

Something must be done:

- Introduce probability as the goal of scientific inquiry: to model uncertainty in the real world.
- Captivate by relating probability to those who came up with many of the models, who were engineers, astronomers, surveyors, biochemists
- For example, Weibull and Gauss came up with their distributions (probability laws of the real world) through experimentation and data. Tell students how. Mathematicians then entertained themselves with the little f properties and the data was forgotten until students got to the stats class.
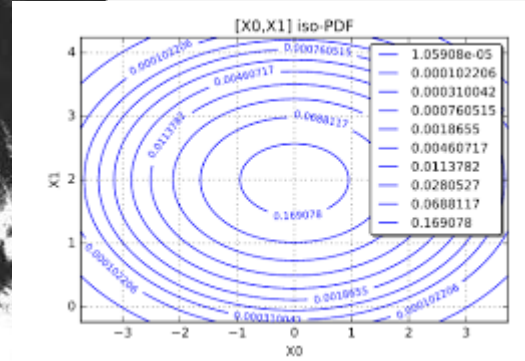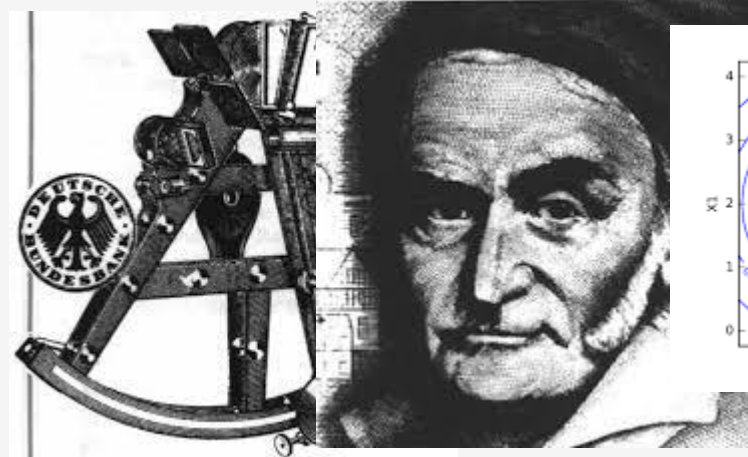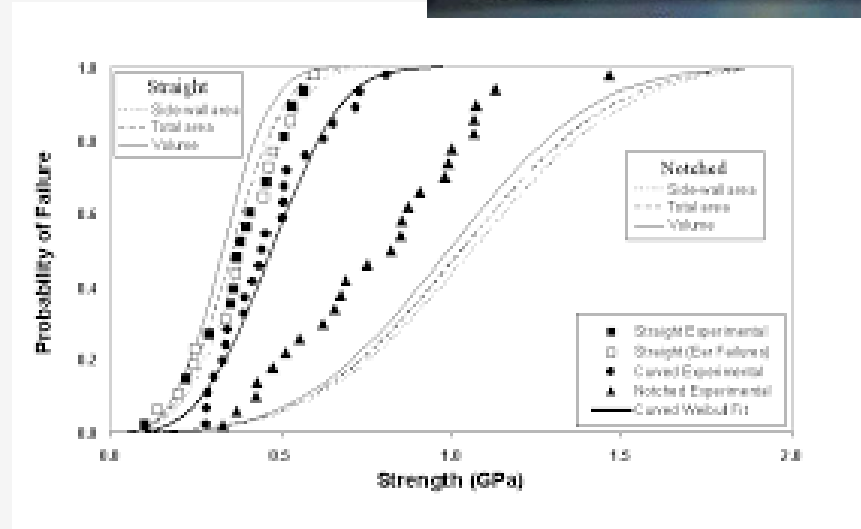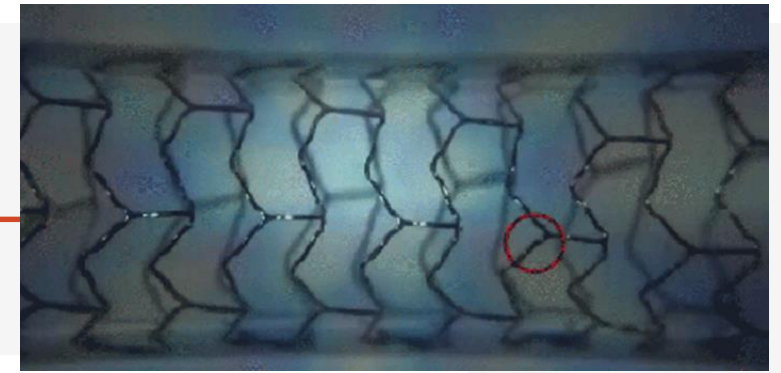
Something must be done:

- Challenge students prior beliefs, be honest about the role that data played in the discovery of the probability laws taught...

First 17th century gambler's problem

If you toss two fair six-sided dice and you have to bet on a sum of 8 or 7 which would you choose?

(A) 8
(B) 7
(C) Equally likely

Second 17 century gambler's problem

If you roll three fair six-sided dice and have to bet on a sum of 10 or 9, which one would you choose?

(A) 10
(B) 9
(C) Equally likely

Gamblers said in the 17th century that in both 1 and 2, C was the answer. But the data they collected every day said that this was not the case.

1

If you toss two fair six-sided dice and you have to bet on a sum of 8 or 7 which would you choose?

2

If you roll three fair six-sided dice and have to bet on a sum of 10 or 9, which one would you choose?

Math then explained why the data wins. It is not the number of partitions (gamblers model), count the permutations of each partition (the order)

1

(A) 8   Probability 5/36
(B) 7   Probability 6/36   (7 wins, data wins)
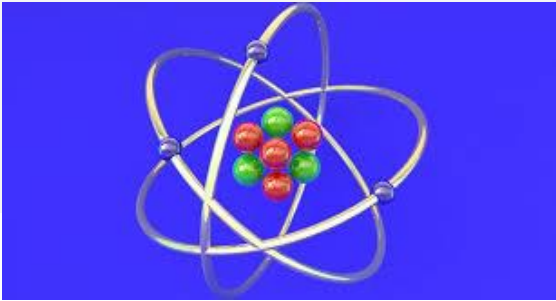(C) Equally likely   This is wrong

2

(A) 10   Probability 27/216   (10 wins, data wins)
(B) 9   Probability 25/216
(C) Equally likely   This is wrong.

Take it one step further, make students that got it wrong realize that they could have been thinking like some physicists

- *Maxwell-Boltzmann model of particles – Order matters (Microstates)*
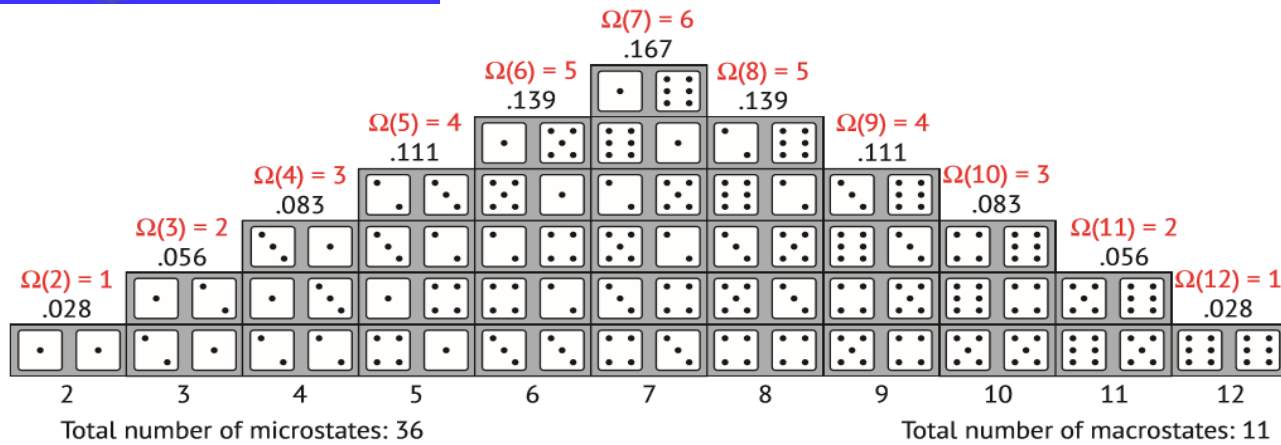- *Bose and Einstein model of particles – order not important*

Figure 1.3 A simple six-sided die model helps clarify a rather complicated physics concept.

Cheat sheet.

For 9, there are 6 partitions: 1/3/5, 1/2/6, 1/4/4, 2/2/5, 2/3/4, 3/3/3. But this is not what we should count, Galileo claims. Each of those partitions covers several possibilities, depending on which die exhibits the numbers. What we must count is the number of permutations of each partition. For three different numbers there are 6 permutations, for example. For the partitions given, we have the following 25 outcomes (out of 216): (1,3,5), (1,5,3), (3,1,5), (3,5,1), (5,1,3), (5,3,1), (1,2,6), (1,6,2), (2,1,6), (2,6,1), (6,1,2), (6,2,1), (1,4,4), (4,1,4), (4,4,1), (2,2,5), (2,5,2), (5,2,2), (2,3,4), (2,4,3), (3,2,4), (3,4,2), (4,2,3), (4,3,2), (3,3,3). Repeating the process for a sum of 10 points, we can show that there are 27 different dice-throws (out of 216).

It is not hard to do the same reasoning for the sum of two fair six sided dice.

Relate probability to today's concerns, like we have done with the intro stats course.

For example, help identify and eliminate corruption within the sports sector (Paulden 2016). Chris Gray (2015) (p=the probability of player A winning a point on serve).

$$P(A \; winning) = \frac{p^4(-8p^3 + 28p^2 - 34p + 15)}{p^2 + (1-p)^2}$$

Probability modeling in Artificial Intelligence, genetic counseling, spam email filtering, detection of hacking. ... There are many articles on probability in these contexts published in magazines written to popularized statistics to students (Significance, Chance, for example). Revisit good pedagogy of many years ago and multidisciplinary textbooks.

Students can be probability modelers before they engage in asymptotics

Data used to discover probability model

Probability law, model, F(x), f(x, e.g. Weibull)

New data on survival after surgery assumed to follow model

Fit model, determine if fit is good (stats inference)

Reach conclusion: Is model good for survival data? If not,
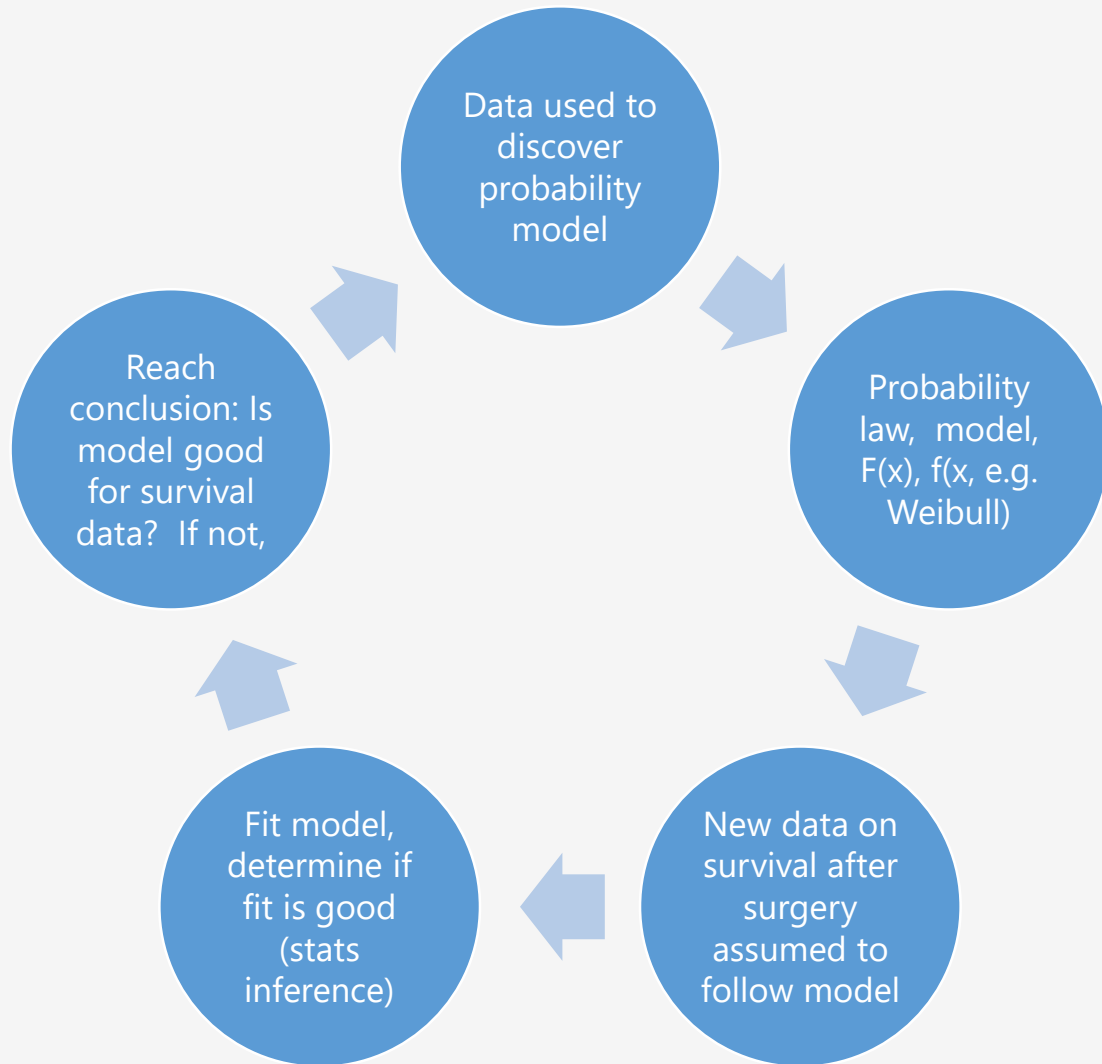
- By connecting data to probability we mean explaining to students how the density functions were discovered through experimentation and the resulting data, how laws of large numbers, central limit theorems and such were discovered through experimentation with data. Math came later to refine the formulas, to study properties.

- Experimentation is not simulation. Students need to experiment to discover new laws applicable in their discipline. With simulation they just prove empirically what was already discovered. The distinction between both is made in my book.

More education research on the intro probability course needed

- Why not take the same approach we have taken for the intro stats course these past 20 years, i.e., put probability in the context of this data-centric world we are in now? Make students modelers of uncertainty.
- My book is a summary of my modest research to bring together: (a) a lot of scattered contemporary and multidisciplinary applications of probability, (b) the role of data in the origins of the probability theory that we teach, and (c) creative ways of teaching probability by classical probabilists of years past together in one place.
- There are lots of material waiting to be harvested with the same enthusiasm that was dedicated to the intro stats by hundreds of stats education researchers.

Probability for Data Scientists

Juana Sánchez

FIRST EDITION
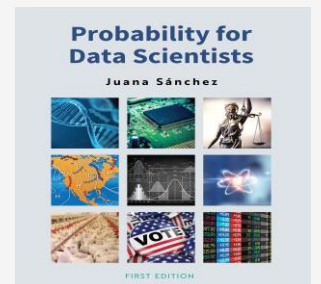
Conclusions

(A) Before the goal of creating a statistically literate society and having stats education fully implemented in schools has been reached, new curriculum on big data is expected to be introduced quickly in schools and universities. New teacher training is needed.

(B) In the middle of this demand, job opportunities have grown, other disciplines want us statisticians and mathematicians to teach our trade to students in a meaningful way to increase diversity and inclusiveness in their STEM disciplines, and to cater to the data that they collect, which varies by field and institution.

(C) Women with just a B.S or minor in statistics or the new applied mathematics are in high demand once again and are working in data-centric jobs after graduation. They play a crucial role in mentoring students and being role models and are an indication of what the job market for our majors is like.

(D) Research in probability education needs to catch up. We have enough of intro stats education research. My book is an attempt to make students aware of the role that data played in discovering the probability theory they study. It is a modest contribution to a much needed probability education research. (Note: the inventors of the density functions and probability laws we teach did not use big data, big data is not needed to make students experiment to discover models, but students faced with big data could discover new density functions with experimentation that results in big data). We need to make students feel that they are part of the scientific community.

Thank you to all of you for being here and to

Talk Math 2 Me,
Statistics,
and
Women Doing Math
organizers for inviting me

Special thanks tor Dr. Vera Ioudina for making my whole stay an unforgettable experience and for introducing me to this amazing group of scholars and this university.