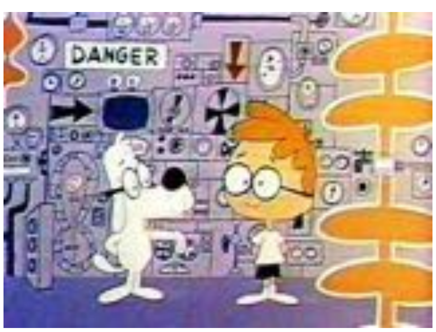
The Modern Statistics Student is....Younger

eCOTS 2022
Rob Gould
rgould@stat.ucla.edu

The WABAC Machine



1948

2011

Now

Hotelling's Questions

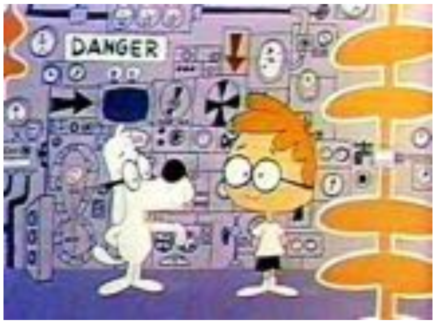
- Who are our prospective students of statistics?
- What should they be taught?

Hotelling (1948), The Teaching of Statistics, *Annals of Mathematical Statistics*

Hotelling's *Answers*

- All College Students
- "Consumers"
  Business executives, administrators, some researchers and teachers
- Users
  Sciences, social sciences, business, econ
- Producers of Methods
  Statisticians and mathematicians

The WABAC Machine



1948

2011

Now

2011

USCOTS

Citizen Statisticians

- We should teach with the goal of producing citizen statisticians: amateur statisticians who have been prepared to live in a world of complex, highly structured, readily available data and access to inexpensive and ubiquitous analysis tools.

Answers to the first two questions

- all students are both consumers and users of data
- sometimes knowingly, sometimes unwittingly

Statistical Literacy

- Familiarity with both basic tools and the language statisticians use (Garfield & Ben-Zvi, 2007)
- Having the skills necessary for understanding reports of medicine and science in media (Utts, 2003)
- Understands statistics as used in civic discourse (Rumsey, 2002)

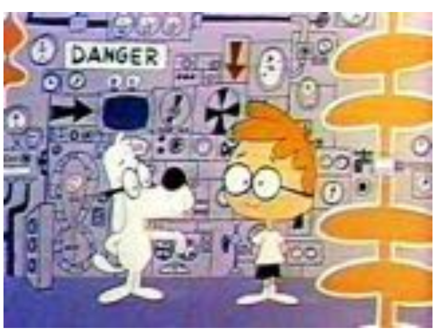
Statistical Thinking

- Thinking (problem-solving and problem-identifying) like a statistician (Wild & Pfannkuch, 1999)
- Using probabilistic descriptions of variability in inductive reasoning and analysis (Brown & Kass, 2009)

Questions for Educators

- What is "data literacy"?
- What is its relation to statistical literacy and statistical thinking?
- What do all citizens need to know about data?
- What do future researchers? Future statisticians?

The WABAC Machine



1948

2011

Now

The need

Perils

- Ignorance-based decision making
- Privacy weakened, loss of autonomy
- Security compromised
- Increased social inequity

Promises

- Quality decision-making in the presence of uncertainty
- Improved control over educational and professional career
- Insight into daily lives
- Improved social equity
- Increased autonomy

K-12 Data Science Education

- Data Science 4 All (Center for Risc)
- States that have adopted K-12 Data Science Standards:
  - Alabama, Georgia (in CS standards), Ohio,Massachusetts, Oregon, Virginia
- State that are Considering or Progressing
  - California, New York, Oklahoma

Why "Data Science" (and not "Statistics")?

- Ubiquity: Data are everywhere and accessible to students. These provide opportunities for students to explore questions of direct interest to them. Analytical tools are accessible for many ages.
- Opportunity (Wise (2020)): Interesting and important questions can be addressed through acquiring disparate data sets, although we must be cautious of sometimes dubious or unknown provenance.
- Wise: a data scientist is "one who must build a bridge between ill- defined questions and unstructured messy data that may (or may not) be fit to address them, by assessing, cleaning, organizing, integrating and visualizing data, selecting suitable algorithms and code to be composed into a reproducible workflow, and communicating appropriate inferences in an ethical manner."

A.F. Wise, Educating Data Scientists and Data Literate Citizens for a New Generation of Data, *Journal of the Learning Sciences*, 29:1 (2020), 165-181,DOI:10.1080/10508406.2019.1705678

What should be the goal of K-12 Data Science education?

Oceans of Data report:

- "The data-literate individual understands, explains, and documents the utility and limitations of data by becoming a critical consumer of data, controlling his/her personal data trail, finding meaning in data, and taking action based on data. The data-literate individual can identify, collect, evaluate, analyze, interpret, present, and protect data."

Building Global Interest in Data Literacy: A Dialogue. Workshop Report, Education Development Corporation Oceans of Data, 2018.

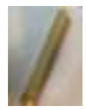
Data Acumen

- Data Acumen: "...the ability to give all students the ability to make good judgments, use tools responsibly and effectively, and ultimately make good decisions using data."
- Requires "mathematical, statistical and computational foundations", and "training in data acquisition, modeling, management and curation, data visuzliation, workflow and reproducibility, communication and teamwork, domain-specific considerations and ethical problem solving."

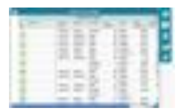
National Academies of Sciences, Engineering, and Medicine 2018. "Data Science for Undergraduates: Opportunities and Options." Washington, C: The National Academies Press. p. 3

Research Highlights

- Statistical and Inferential



- Computational

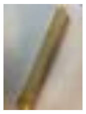


- Data
- Analysis

Lessons from Lit: Inference

- Inference is hard. Really hard.

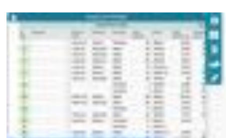
  - delMas (2004)
- Inferential thinking can be developed over time, and should be.
  - Informal inferential thinking (Rubin & Makar (2009))
- "What I see [through data] is not really the way it is." Inferential thinking is essential
  - Wild, Pfannkuch, Regan, Horton (2011)



Lessons from Literature: Computation

- Computer code can enhance and in some instances replace mathematical notation for helping students understand.
    - Kaplan (2007)
- Tidy data is necessary for analysis, but is the result of a process.Students should learn "data moves", or else their inferential thinking skills will go to waste.
    - Wickham (2014); Erickson, Wilkerson, Finzer, Reichsman(2019)
- Integrating statistical and computational thinking in a course creates new challenges for learners (and therefore educators)
    - Thoma, Deitrick, Wilkerson (2018)
- Coding integrated into mathematics and statistics can be "Purposeful", "Efficient" and "Empowering"
    - Heinzman (2020)

Lessons from Lit: Data

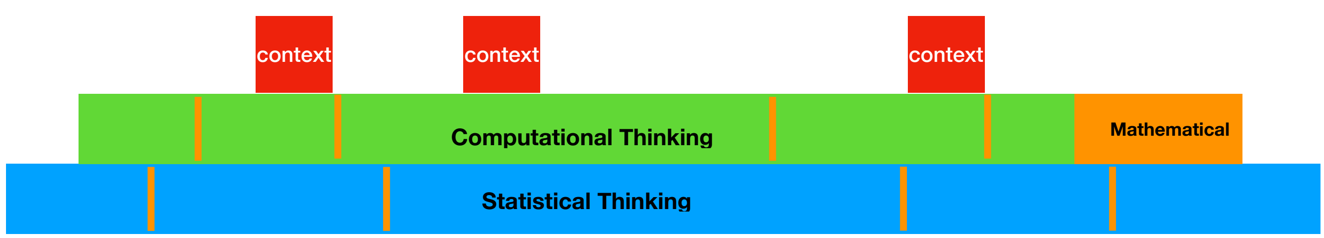


- Tidy data is what the computer wants, but is not necessarily what students have in mind. We need to build on students' understandings.
    - Konold, Finzer, Kreetong (2014)
    - Haldar, Wong, Heller, & Konold (2018)

Lessons from Lit: Data Analysis

- What exactly *is* data analysis, anyways?
    - Hicks & Peng (2019)
- Questions are key (particularly statistical investigative questions)
    - Arnold (2013)
- Questioning is hard
    - Frischemeier and Biehler (2018); Gould, Bargagliotti, Johnson (2017)

K-12 DS Schematic
aka
Data Acumen Development



GAISE preK-12 Revision

- See Harvard Data Science Review for overviews of Levels A, B, C
- Included (much) greater emphasis on computation and technology
- Encourages and provides examples for greater integration with sciences (specifically life sciences)
- Provides a general framework for supporting the acquisition of data acumen in K-12

What does High School Data Science Look like?

- International Data Science in Schools Project (www.idssp.org)
- Bootstrap World (Brown University, CS and Education)
- CourseKata (UCLA and Cal State LA, Psychology)
- YouCubed (Stanford, Math Education)
- ProDaBi (Paderborn University, Math & Statistics Education, CS Education)
- Mobilize Introduction to Data Science (IDS) (UCLA, Statistics, CS, Education)

IDS

Introduction to Data Science

IDS
Introduction to Data Science

- Lessons structured in pieces
  - Classroom activities, discussions, investigations develop definitions and concepts.
  - Rstudio interactive labs to apply concepts
  - Frequent "practicums" synthesize materials
  - End-of-unit milestones serve as assessments

IDS
Introduction to Data Science

- Full curriculum at: www.introdatascience.org
- 48 districts; 135 High schools; 28,500 students taught to date

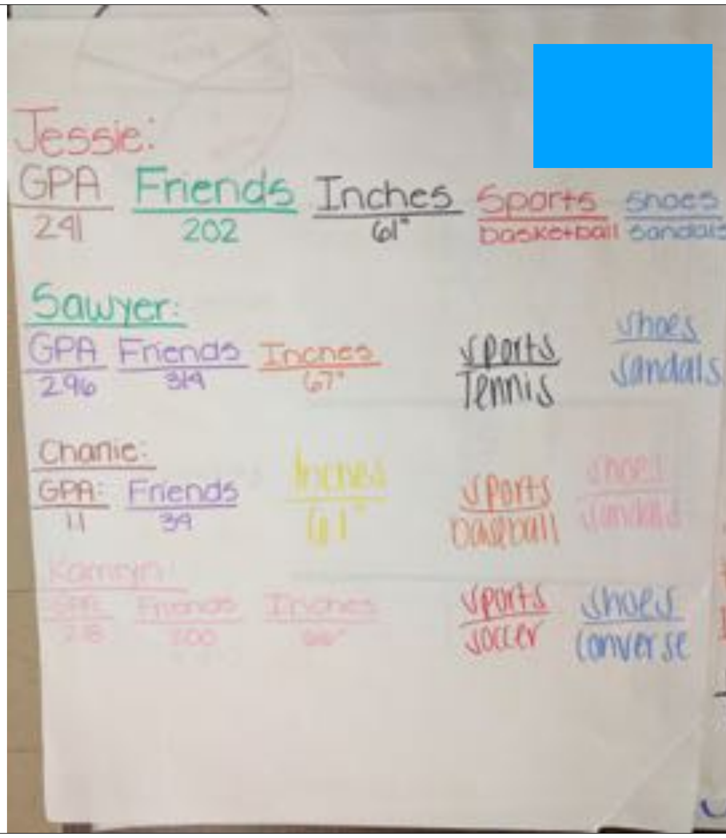
Unit 1, Lesson 2: Stick Figures

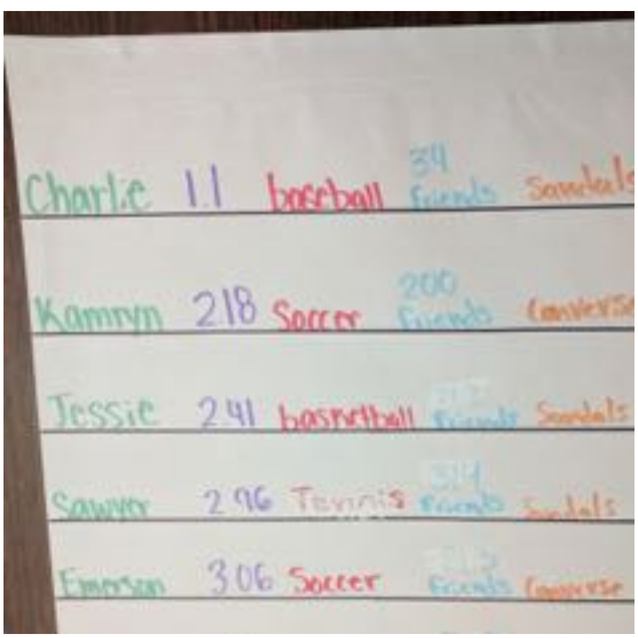
"Collect and record as much information as you can about these people"

"Organize this information on a poster any way that you feel is helpful"

Posters are displayed, and students discuss:

- what are similarities and differences in the ways the data were organized
- what information ('variables') is available?
- which organizations made it easiest to see the variables?





Konold, Finzer, Kreetong (2016)

- **"spreadsheet" format is not natural for many students (and their teachers)**
- **students need to develop the conception of "case"**
- **students have basically sound and solid notions of data**
- **students are comfortable and may even prefer hierarchical representations over spreadsheet representations**

Where do we go from here?

Questions for Educators

- What is "data literacy"?
- What is its relation to statistical literacy and statistical thinking?
- What do all citizens need to know about data?
- What do future researchers? Future statisticians?

What is data literacy? (And how does it relate to statistical thinking and literacy)

- We should replace the notion of "data literacy" with "data acumen", which lacks the connotations of "literacy" and has support of National Academies of Science.
- Statistical thinking and literacy are the foundations upon which data acumen is developed, but are insufficient unless strongly integrated with computational thinking. All of these are infused with mathematics.

What do all citizens need to know?

- A lot.
- <u>All students</u> (K-12) must develop appropriate levels of data acumen, and even higher levels if they intend to major in STEM.

What do future researchers need to learn?

- Hotelling said, in not so many words: Math and more math.
- But all K-12 students must also develop data acumen, and this raises the question: what's the relation between data acumen-building and mathematics preparation?

Where does DS fit in?

- Many states have situated Data Science within the mathematics curriculum, although some have Computer Science standards and have fit it within there.
- In California, Statistics and Data Science courses (and some computer science courses) count towards the math requirements for admission to the public university systems.

Cons of DS as Math

- Like Statistics, DS is mathematical but not necessarily part of mathematics. It has concepts and ideas that are not easily captured by mathematics. This means that sometimes, a derivation or proof does not aid in understanding. (Confidence intervals, anyone?)
- Math teachers require additional preparation, since inferential thinking and data analysis are often not included in their credentialling programs. Without this, DS classes can be taught in a very "rote", simplistic way.
- Students are (sometimes) limited in the amount of math they are allowed to take. This puts DS in direct competition with the calculus pathway. As a result, STEM-oriented students, students who most need to develop data acumen, will be least likely to take courses that can help.=

Pros of DS as a Math Course

- Mathematics is well situated to teach the universality of the methods and approaches used.
- DS is widely applicable and so offers an 'in' to every student. Well prepared instructors and well-crafted curricula can use this to "hook" students into further STEM study.
  - One study in California observed that among students who took Algebra I in 9th grade, only 18% made it past Algebra II. DS and Statistics can serve as a course for re-integrating students into mathematical and scientific preparation for college study.
  - Another found that, in California, about half of all seniors did not take any math. And in 2018, 40% of schools had no seniors at all enrolled in advanced math. Data Science can be a productive, rigorous, and important course for students who would otherwise not continue their studies.

Multiple Roles of K-12 Data Science

- DS for All: Develop data acumen for citizenship
  - We can't wait until college, since all students, not just college-bound, need data acumen.
- Develop necessary and powerful preparation for higher education, including those who might study STEM
- Provide rigorous, relevant mathematical, statistical and computational thinking skills for students who "fall off" the calculus pathway.
  - The debate is sometimes framed as if a data science math pathway might not well-prepare students for college STEM. But we must recognize that for a vast number of students, the traditional calculus pathway doesn't prepare them for college at all.

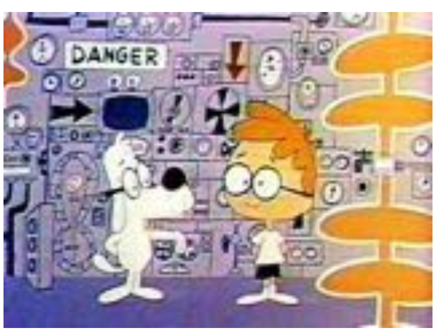
In answer to Hotelling's Question

- All students need data acumen
- Some students need calculus
- (Future statistics and data science researchers need both)

The Bridge Problem

- Imagine a group of students who are computer literate, strong programmers able to problem-solve algorithmically who have a strong understanding of the role of data in answering important social and scientific questions.
- But they haven't had Algebra II
- Is there a way to bring them into the profession?
  - "Compute-more", "Calu-less" Data Science: CSU East Bay
  - UCLA Life Sciences
  - Park City Math Institute: rethink mathematics (modeling, linear algebra)

The WABAC Machine



1948

2011

Now

2032

Incoming College Class of '36

All students have a well-developed, nuanced sense of data acumen.

Students who did not follow the Calculus path are able to take "on-ramp" math courses to pursue STEM, leveraging the particularly strong computational and statistical thinking skills they developed.

Students who did follow the Calculus path are able to more rapidly advance in their subjects because of their experience in working with data that can be applied to their field.

Thank you!

rgould@stat.ucla.edu

References

- Bob delMas (2004):"A Comparison of Mathematical and Statistical Reasoning, in The Challenge of Developing Statistical Literacy, Reasoning and Thinking, 79-95, Ben-Zvi and Garfield, eds)
- Rubin & Makar (2009):A Framework for thinking about informal statistical inference, SERJ, 8(1), 82-104.
- Wild, Pfannkuch, Regan, Horton (2011): J.R. Statist Soc. A , 2011, 174, Part 2, a pp. 247-295

- Kaplan (2007):Computing and Introductory Statistics. Technology Innovations in Statistics Education, 1(1). http://dx.doi.org/10.5070/T511000030 Retrieved from https://escholarship.org/uc/item/3088k195_
- Wickham (2014): "Tidy Data", Journal of Statistical Software, 59(1): 1-23.
- Erickson, Wilkerson, Finzer, Reichsman(2019): Data Moves. Technology Innovations in Statistics Education, 12(1). http://dx.doi.org/10.5070/T5121038001 Retrieved from https://escholarship.org/uc/item/0mg8m7g6
- Thoma, Deitrick, Wilkerson (2018):'It didn't really go very well': Epistemological framing and complexity of interdisciplinary computing activities, International Society of Learning Sciences, 2018, available at https://repository.isls.org/ bitstream/1/574/1/249.pdf.
- Erica Heinzman (2020):"Math is no longer a four-letter word: A mixed methods study of two non- traditional fourth-year mathematics classes", Ph.D. Dissertation, University of California, San Diego.

- Konold, Finzer, Kreetong (2014):“Students' methods of recording and organizing data,” Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Haldar, L. C, Wong, N., Heller, J. I, & Konold, C. (2018). Students Making Sense of Multi-level Data. Technology Innovations in Statistics Education, 11(1). http://dx.doi.org/10.5070/T5111031358

- Hicks & Peng (2019): "Elements and Principles of Data Analysis", arXiv:1903.07639v1\
- Arnold: (2013) Statistical investigative questions—An enquiry into posing and answering investigative questions from existing data. (Doctor of Philosophy), The University of Auckland, 2013, available at https://researchspace.auckland.ac.nz/handle/2292/21305.
    - R. Gould, A. Bargagliotti, and T. Johnson, An analysis of secondary teachers' reasoning with participatory sensing data, Stat. Educ. Res. J. 16(2) (2017), 305–334
    - D. Frischemeier and R. Biehler, Stepwise development of statistical literacy and thinking in a statistics course for elementary preservice teachers, in Proceedings of the Tenth Congress of the European Society for Research in Mathematics Education, T. Dooley and G. Gueudet, Eds.,