Promoting Statistical and Data Science Literacy*

* Particularly the communication dimension of these literacies

{ DaSSWeb– 14 January 2025 }

A. John Bailer
Podcaster: Stats+Stories
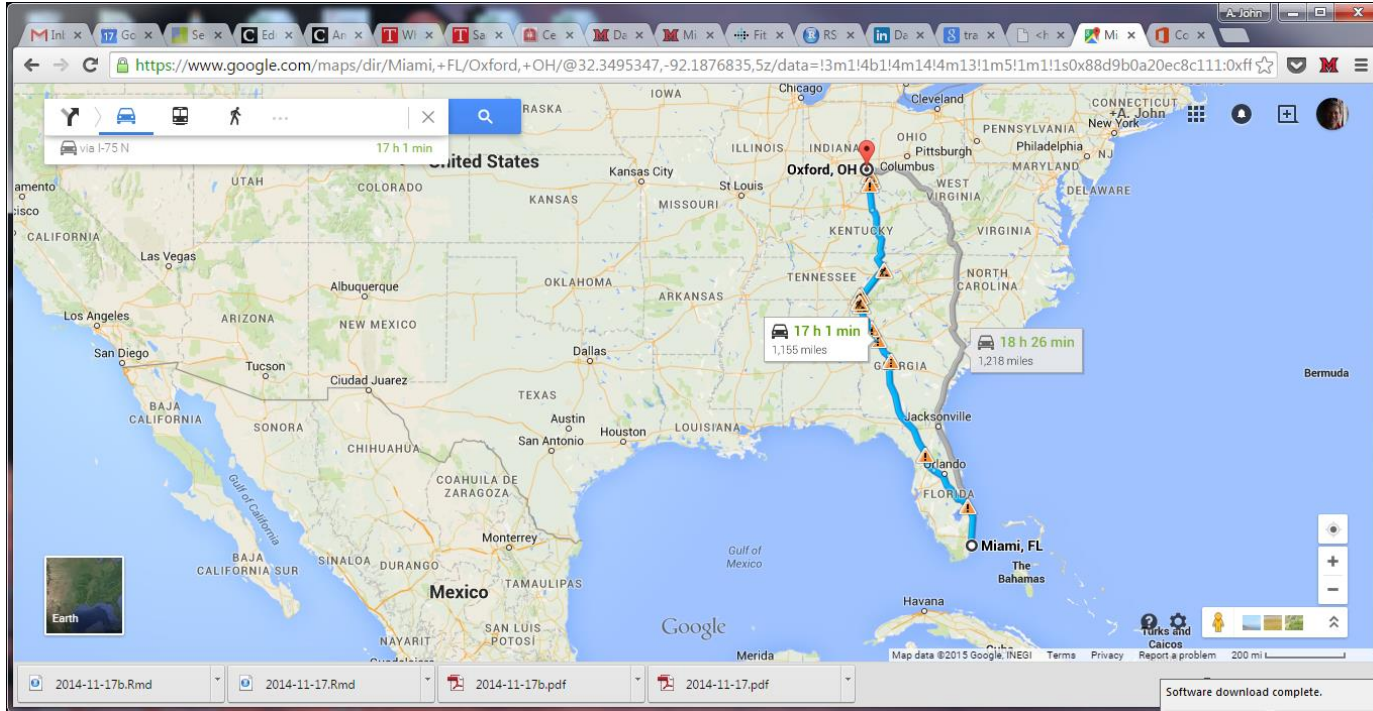### Professor Emeritus of Statistics, Miami University
baileraj@MiamiOH.edu

STATS + STORIES

A. JOHN BAILER
ROSEMARY PENNINGTON

STATISTICS BEHIND
THE HEADLINES

ASA   CRC Press

MIAMI UNIVERSITY

Thanks!

- Paula Brito and Maria Eduarda Silva for the invitation
- All attendees for joining this conversation
- My journalism colleagues and collaborators on the Stats+Stories podcast – Rosemary Pennington and Richard Campbell
- ASA for supporting the podcast
- Numerous colleagues over the years who have influenced my perspective and practice

Miami University? Oxford, Ohio?

University of Miami

http://www.contacttopuniversities.com/blog/mainmenu/University%20Articles/top-universities-for-physical-therapy/umEntrance.jpg

4

Miami University

Outline

Part 1: We all believe that statistics and data science is important for understanding the world. Motivating examples; Context for communication – QL, SL, D(S)L; Preparing future data scientists and statisticians

Part 2: What might we do to help make these statistics and data science products / analyses understood?

MIAMI UNIVERSITY

Students, scientists and others may experience something like the following when we present the results of statistical and data science analyses in ways that don't connect to their current understanding.

In addition, some presentations may discourage an audience from wanting to engage with an analysis

(Credit to Gary Larson for the cartoon and Jim Oris for the customization)

MIAMI UNIVERSITY
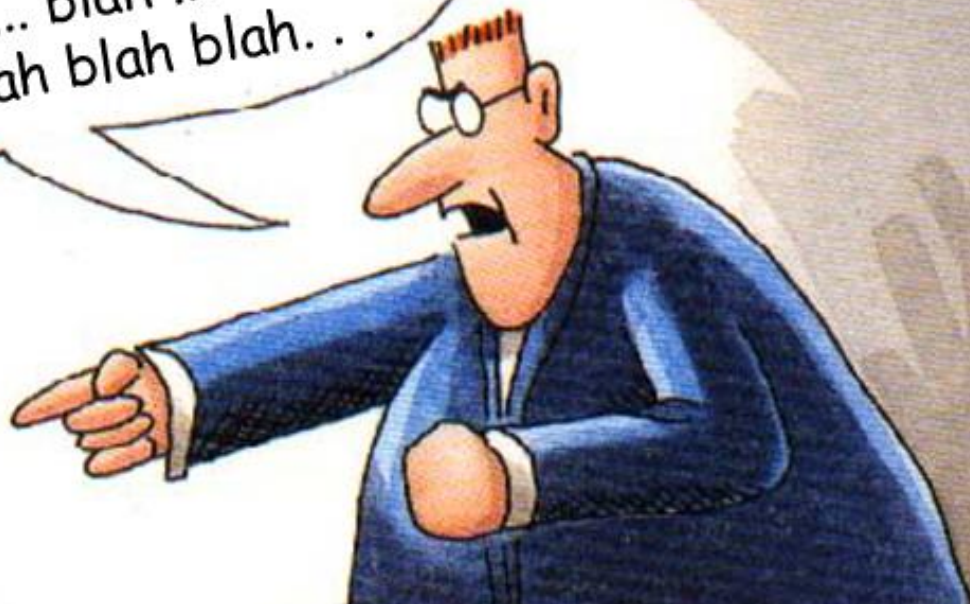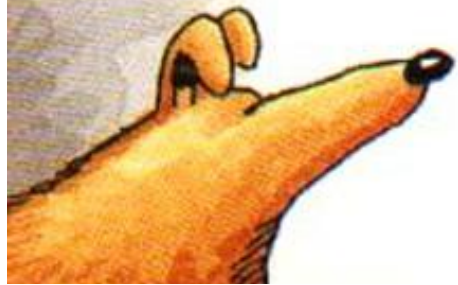
An actual conversation (paraphrased) …

Physician: with your level of cholesterol and other risk factors, you have a 10% risk of heart disease in the next 10 years.

Patient: what does that mean?

Physician : what do you mean "what does that mean?"

Patient :   is that an absolute risk, an relative risk, an odds ratio?

Patient :   can this level be controlled by exercise or diet or is a statin drug required?

Physician: I think it is time for my next patient.

Understanding the risk report leads to sense of urgency – absolute risk from 10% to 20% is serious; a relative risk shift from 1 in 10^4 to 1.1 in 10^4 ?

Next questions:

Can risk be mitigated with other actions (diet, exercise, etc.)? Statin drugs?

Are caregivers ready to have these conversations?

MIAMI UNIVERSITY

Part 1: Case for Communication – QL,SL, DSL

Quantitative Literacy (QL)
Statistical Literacy (SL)
Data Science Literacy (DSL)

MIAMI UNIVERSITY

Part 1: Case for Communication – QL

Quantitative Literacy: "an aggregate of skills, knowledge, beliefs, dispositions, habits of mind, communication capabilities, and problem solving skills that people need in order to engage effectively in quantitative situations arising in life and work."

(definition from Numeracy Network, http://www.woodrow.org/nced/national_numeracy_network.html)

Aside: QL skills summary: understanding functional relationships, number sense / magnitude and statistical literacy

Aside: Bernie Madison advocated using news to teach quantitative literacy

MIAMI UNIVERSITY

Statistical literacy is portrayed as the ability to interpret, critically evaluate, and communicate about statistical information and messages. (Gal 2002 ISR)

"statistical literacy" refers broadly to two interrelated components, primarily

(a) people's ability to interpret and critically evaluate statistical information, data-related arguments, or stochastic phenomena, which they may encounter in diverse contexts, and when relevant

(b) their ability to discuss or communicate their reactions to such statistical information, such as their understanding of the meaning of the information, their opinions about the implications of this information, or their concerns regarding the acceptability of given conclusions.

Do we have a responsibility to help develop the statistical literacy skills of our audience?

MIAMI UNIVERSITY

Part 1: Case for Communication – D(S)L

Data literacy — the ability of a company's employees to understand and work with data to the appropriate degree (MIT Sloan, 2021) – includes the ability to

1. Read data
2. Work with data
3. Analyze data
4. Argue with data

In order to be data literate, "you have to be verbally literate, numerically literate, and graphically literate" M. Kazakoff (quoted in this article)
https://mitsloan.mit.edu/ideas-made-to-matter/how-to-build-data-literacy-your-company

Common QL, SL and D(S)L element: communicate (and/or argue) with data

MIAMI UNIVERSITY

ASA Workgroup on Master's degrees in statistics
http://magazine.amstat.org/wp-content/uploads/2013an/masterworkgroup.pdf

One Recommendation 2 (of 6) - based on survey of recent grads and employers):
Communication skills critical and should be developed and practiced throughout graduate programs.

ASA Undergraduate Guidelines Workgroup
http://www.amstat.org/education/pdfs/guidelines2014-11-15.pdf
Recommendation 3 (of 4) Ability to communicate
- Communicate complex methods to managers and other audiences
- Understand ethical standards
- Opportunities to practice and refine skills during studies

MIAMI UNIVERSITY

Learning outcomes for BS Statistics and Data Science students (undergraduate majors at Miami U):

1. Students shall be able to analyze and interpret data critically using statistical models and programming skills.
2. Students shall demonstrate understanding of the mathematical basis and theoretical foundations of statistics.
3. Students shall be able to effectively communicate, both orally and in written form, results of statistical analyses to both the expert and layperson.

16

MIAMI UNIVERSITY

Part 2: Strategy / framework to critical data-research consumption skills

"The job of the journalist is to ==make the significant interesting==." - Bill Kovach and Tom Rosenstiel's The Elements of Journalism.

Nick Kristoff: Reporters should not lead their reports with a lot of numbers and data but rather tell a story that grabs the reader and illustrates the big data to come later in the story. (RC)

Journalists are supposed to ==put facts into context== [and this] is the foundation of my work as a reporter and my work in the classroom as I teach journalism students how to write news. (R. Pennington)

What is the job of the statistician and data scientist?

MIAMI UNIVERSITY

'Job' of the Data Scientist and Statistician?

To conduct and to communicate the results of analyses to have impact.

"Grabbing the reader [audience]" - making the "significant interesting" – "putting facts in context" all seem like important for producing effective and impactful analyses

How often do we think about this?

What should we be communicating about Statistics and Research in the News?

Cohn and Cope *News & Numbers* – classic book for journalists (and interested statisticians and data scientists).

- Included questions journalists should ask about research AND statisticians should be ready to answer!

So, what is one tactic to tell the statistics behind stories and stories behind statistics? Answer: a podcast

Stats+Stories podcast history

Years	"telling the statistics behind the stories and the stories behind the statistics"		
2009	News & Numbers course	QL FLC '07-'08	
2013	S+S episode (#1: Jim Albert)	3 episodes in '13	
2016	Rosemary-new moderator (#23)		
2017	1st Virtual (Zoom) Episode (#26DS)	1st S+SS: (#25)	ASA funding
2019	#100!	#200 (2021)	#300 (2023)
2022	*Statistics Behind the Headlines*	Audiobook (2024)	
2025	Becomes ASA podcast		

Statistics Behind the Headlines working list

1. What is claimed? and is it appropriate?
2. Who is claiming this?
3. Why is it claimed?
4. Is this a good measure of impact?
5. How is the claim supported?
   - What evidence is reported?
   - What is the quality / strength of the evidence?
6. Is the claim reasonable in itself?
   - Does prior belief impact my belief? Confirmation bias?
7. How does this claim fit with what is already known?
8. How much does this matter for me?
   - Comparison of population perspective vs. individual perspective?
   - Will I change my behavior as a consequence of this?

A. JOHN BAILER
ROSEMARY PENNINGTON

STATISTICS BEHIND
THE HEADLINES

ASA   CRC Press

MIAMI UNIVERSITY

Investigating Series Binge-Watching (SBtH Ch 5)

Photo by **Ketut Subiyanto** from **Pexels**
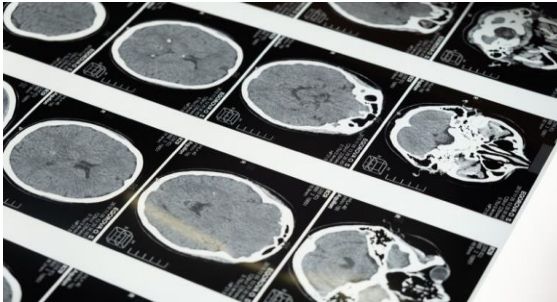(https://www.pexels.com/photo/interested-multiracial-family-watching-tv-on-sofa-together-with-dog-4545955/)

Photo by **cottonbro** from **Pexels** (https://www.pexels.com/photo/white-and-black-menu-board-5723883)

Headline; What Happens to your brain when you binge-watch a TV series – November 4, 2017 (by Danielle Page, NBC News) – https://www.nbcnews.com/better/health/what-happens-your-brain-when-you-binge-watch-tv-series-ncna816991

Data Cited in story:
Netflix Media Center Report: Netflix & Binge: New Binge Scale Reveals TV Series We Devour and Those We Savor.
https://media.netflix.com/en/press-releases/netflix-binge-new-binge-scale-reveals-tv-series-we-devour-and-those-we-savor-1

Government Report Citation: American Time Use Survey Summary
https://www.bls.gov/news.release/atus.nr0.htm

Presentation Scientific Meeting:
https://apha.confex.com/recording/apha/143am/mp4/free/4db77adf5df9fff0d3caf5cafe28f496/paper335049_1.mp4

MIAMI UNIVERSITY

- Exploration of how binge-watching may impact us (cognitively, emotionally)
- Stress alleviation although caution of let down
- Pointers for responsible binge-watching

Ideas Encountered in the Chapter

- Multiple data sources may provide the foundation of many stories
- Different data sources provide different types and amounts of information
- Quotes from experts are not the same thing as a systematic study or review of evidence
- Currency and novelty can drive news coverage of research

What is claimed? Who is claiming this?

- Brain-chemistry, psychological responses and stress reduction can be associated with binge-watching TV
- Caution to not binge to exclusion of human interaction
- Recommendations about responsible binging

Claimed by: NBC News writer who integrates various data sources for assertion about viewing habits and outcomes

Why is it being claimed?

- TV viewing habits have changed with streaming services
- Viewing options on different platforms (smart phones, tablets, computers, televisions)
- Typical consumption foundation of story [ATUS survey]
- Binge-watching behavior [Netflix survey]
- Drivers of binge-watching and explanations [interviews with psychologists / psychiatrists; conference presentations]
- Studies of brain chemistry and psychological responses
- Caution if social interactions lost due to binge-watching

MIAMI UNIVERSITY

Reading research behind stories – call outs

- There is a ==hierarchy of information== in stories – probability sample surveys and designed experiments are most likely to be representative and valid while convenience samples and expert opinions may be neither representative nor valid.
- ==Coding categories== of responses is a difficult task and the best research establishes precise rules for such assignments.
- ==Sampling variability== are captured in the ==margin of error== estimates in survey research
- It is fair to ask ==why a study was conducted==. This was produced by the Netflix Media Center and while it summarized some viewing behaviors, almost ½ of the report was devoted to listing series that might be binged.
- ==Experts== can provide great quotes but are they an expert in the area discussed and do they have evidence in ==support of their assertions==?
- A standard and uniform definition of response categories will increase confidence that respondents are answering the same question.

MIAMI UNIVERSITY

Engaging in communication: Contests - Better Bayes (first! Fall 2018)

Contest: Headline + Lead Sentence to explain Bayes (announced on episode 65) – contest inspired by an episode with Kerrie Mengersen

> *submit a headline and a lead explaining what Bayesian analysis is. Again, the way you might explain it to your grandma, to maybe your journalist colleague who has no idea what it means. We're not looking for a deep explanation, it's got to be the nuts and bolts of what Bayesian analysis is.*
>
> (Finalists and winner announced on episode 73)

Finalists included:

Lucy D'Agostino McGowan: - Headline: "Using Bayes to binge, A Netflix Original Series"
Lead: Deciding which Netflix series to binge based on prior shows you've enjoyed pick one. One episode in, decide if you like it. Not sure? Watch another episode. You're a Bayesian.

Rasmus Baath: - Headline: "Bayes' uncertainty modeling"
Lead: Use the language of probability to describe what you know and what is uncertain about a situation. Add data, voila! Bayes uses the new information to optimally reduce the uncertainty.

MIAMI UNIVERSITY

Better Bayes Contest (Fall 2018)

Winner: **Steve Ziliak (Haiku):** Headline: Better Bayes, found in linked haiku.

Bayesian methods
Making stuff you partly know
link that with what you don't.

Frequentist methods
Fear the null hypothesis
and large p-values

"It pays to go Bayes"
Epistemologically
Obvious in prior.

MIAMI UNIVERSITY

#MemeMedianMode contest (June 21, 2021)

Contest Announcement:

> *This week we're launching a contest in the run-up to our 200th episode.*

> *With any statistical meme of your creation, you can enter by replying to this tweet with #MemeMedianMode and following our account.*

> *The winner will get a chance to appear (virtually) on our 200th episode!*

Winners: Nynke Krol and Eric Daza were guests on 200th episode

MIAMI UNIVERSITY

#MemeMedianMode contest (June 21, 2021)

#MemeMedianMode contest (June 21, 2021)

#MyHeadlineIsBetter (250ᵗʰ episode)

#MyHeadlineIsBetter (250th episode)

Winner Dan Gaichas featured on episode: https://statsandstories.net/society1/my-headline-is-better-contest-winner

MIAMI UNIVERSITY

Data Visualization Contest: 300th episode

The **Stats+Stories 300th Episode Data Visualization Contest**
We have put out a data set containing various information on over 280 episodes of Stats+Stories recorded over the past decade, which we want you to use to create visualizations and tell a story with our data! … Static displays, dynamic dashboards, and insightful analyses are all invited to enter (the more unique the better) (https://statsandstories.squarespace.com/dataviz300)

https://statsandstories.net/resources > Stats+Stories Episode Dataset (for data set + finalists)

Winner: Nicole Mark (Tableau)
https://public.tableau.com/app/profile/nicole.mark/viz/StatsStories300thEpisodeDataVisualizationContest/StatsStories300thEpisode

MIAMI UNIVERSITY

Conclusion

Developing statistical and data science products that are used and impact decisions are a common goal.

Communicating these products in a way that is understood by a target audience is a critical key.

Educating ourselves and others to produce, as well as consume, these results is an aspiration.

Deconstructing news and research and the public scholarship of a podcast is one mechanism for doing this.

MIAMI UNIVERSITY

Thank you!

Contact information:

John Bailer

Email: baileraj@miamioh.edu
BlueSky: @johnbailer.bsky.social
LinkedIn: a-john-bailer

MIAMI UNIVERSITY

APPENDIX: Additional Slides with lists from …

Cohn and Cope: News & Numbers

Tim Harford: The Data Detective

David Spiegelhalter: The Art of Statistics

MIAMI UNIVERSITY

News & Numbers: Questions from *Testing the Evidence* chapter:

- How do you know?
- Have claims been subjected to any studies or experiments?
- If studies have been done, were they acceptable ones, by general agreement?
- Results fairly consistent with those from related studies, and with general knowledge in the field?
- Finding resulted in a consensus among other experts in the filed?
- Conclusions backed by believable statistical evidence? Degree of certainty or uncertainty?
- Reasonable theoretical plausibility to the findings?

MIAMI UNIVERSITY

Tim Harford- Financial Times index card (More or Less podcast and author of *The Data Detective*)

1. Observe your feelings - gut reaction
2. Understand the claim
   - What does it mean?
   - Causal?
   - What's being left out
3. Get the backstory
4. Put things in perspective
   - Is that a big number
   - What is the historical trend?
   - Beware 'statistical significance'
5. Embrace imprecision
6. Be curious
   - Go another click
   - Treat surprises as a mystery

MIAMI UNIVERSITY

David Spiegelhalter - 10 Questions to ask when confronted by a claim based on statistical evidence *The Art of Statistics*

Trustworthiness of numbers (1-3), source (4-6), interpretation (7-10)
1. How rigorously has the study been done?
2. What is the statistical uncertainty / confidence in the findings?
3. Is the summary appropriate
4. Reliability of the source?
5. Is the story being spun?
6. What am I not being told?
7. How does the claim fit with what else is known?
8. What's the claimed explanation for whatever has been seen?
9. How relevant is the story to the audience?
10. Is the claimed effect important?

Describing/Presenting research context

- Suppose you were involved in a project looking at the human impacts on a lake. A text description such as the following is clear and correct:

  "Aquatic toxicity among *C. dubia* exposed to phototoxic chemicals from motorized watercraft in Lake Tahoe, a Western Alpine lake, was studied."

MIAMI UNIVERSITY
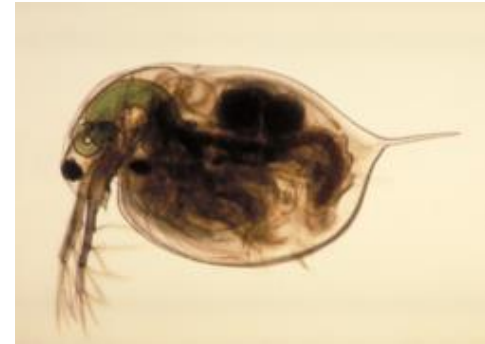
An alternative …

- Describing the study in the context of visualization

Green Pine Trees and Lake (pexels.com)

Man Riding on Green Personal Watercraft on Body of Water (pexels.com)

species.wikimedia.org

Abstract

How much statistics training does a manager of a company or a research scientist have? How about general members of the public? What role does statistics play in understanding the world and in the translation and communication of stories? The goals associated with quantitative literacy, statistical literacy and, now, data science literacy are explored as part of this talk.

The reporting of science and statistics in the news provides an important case study to think about statistical literacy. How do you learn about what's going on in the world? Did a news headline grab your attention? Did a news story report on recent research? What do you need to know to be a critical consumer of the news you read?

Strategies and a framework to inspire individuals to develop data self-defense and critical news consumption skills are discussed. The perspective presented reflects insights from a long-term collaboration with journalists to consider how a tactic promoting quantitative literacy led to a podcast exploring the statistics behind the stories and the stories behind the statistics and a book that did the same.