

Mind the Gap: An incomplete picture of statistics, statisticians, & statistics education

Matthew Beckman
Penn State University

June 25, 2023
Maleny, Australia

A complete picture of statistics, statisticians, & statistics education

Matthew Beckman
Penn State University

June 25, 2023
Maleny, Australia

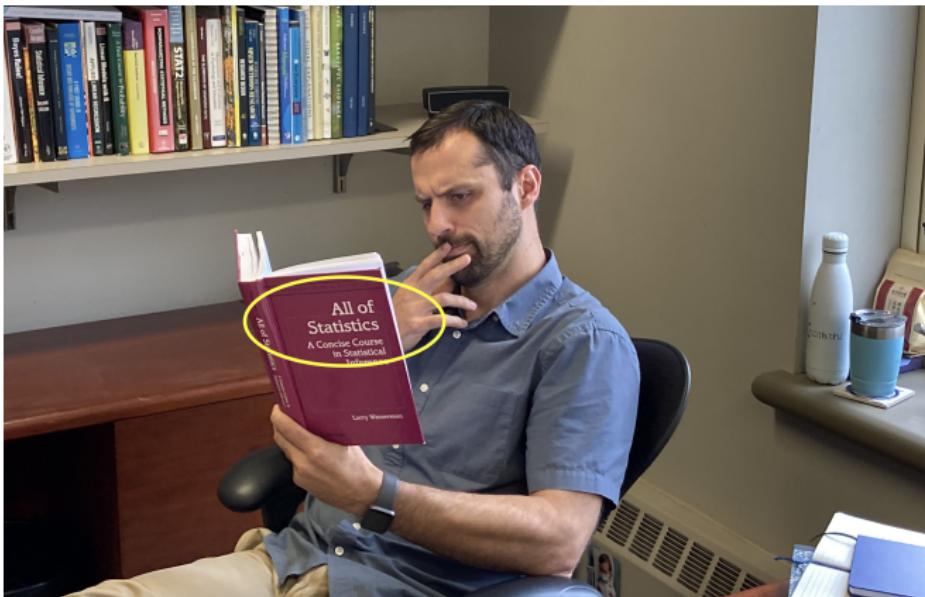


Figure 1: Getting ready to “Provide the perspective of the discipline . . .” as promised in the SRTL announcement.

- . . . on “reconceptualising data and data-ing”

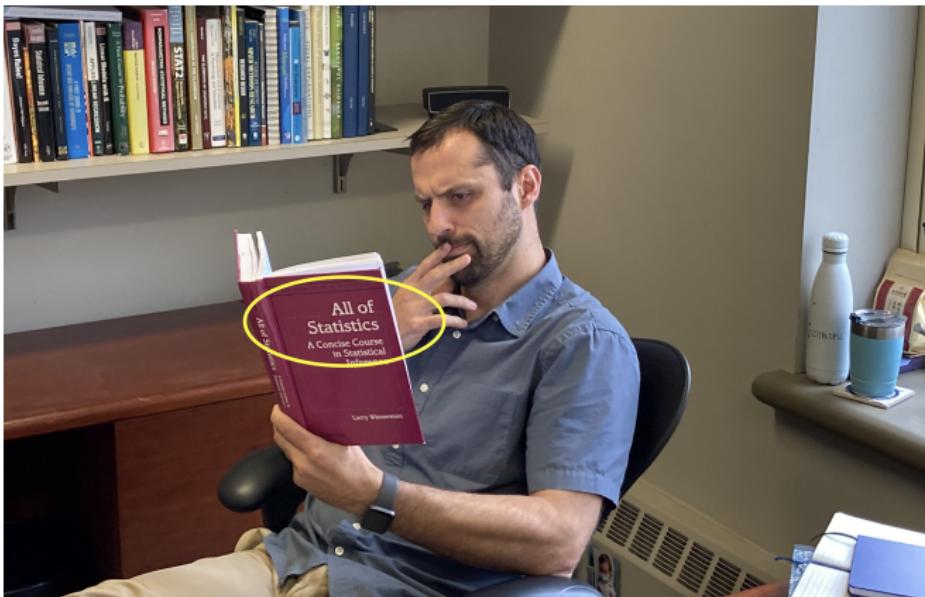


Figure 1: Getting ready to “Provide the perspective of the discipline . . .” as promised in the SRTL announcement.

- . . . on “reconceptualising data and data-ing”
- “data-ing?”

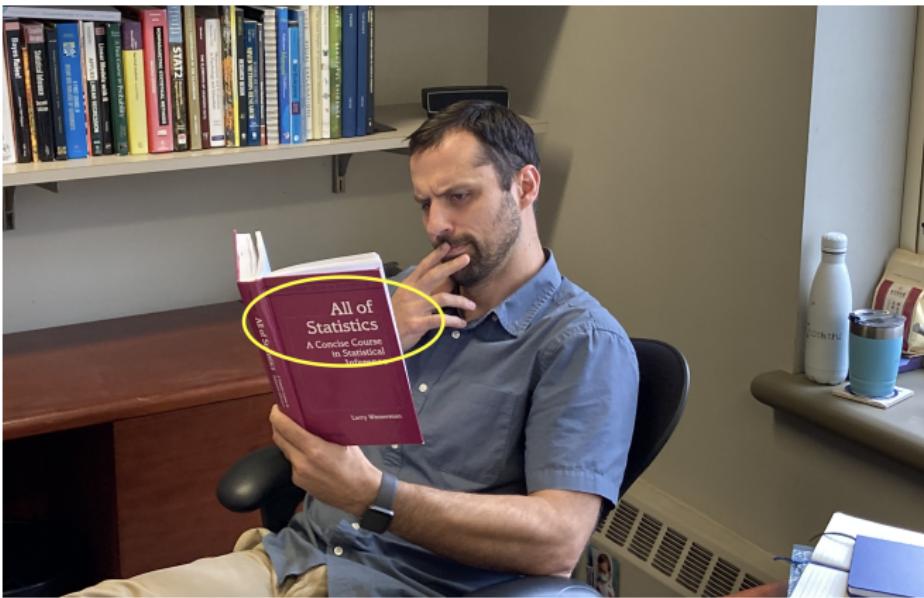


Figure 1: Getting ready to “Provide the perspective of the discipline . . .” as promised in the SRTL announcement.

- . . . on “reconceptualising data and data-ing”
- “data-ing?”
- that wasn’t in the book . . .

Mind the Gap: An incomplete picture of statistics, statisticians, & statistics education

Matthew Beckman
Penn State University

June 25, 2023
Maleny, Australia

“Data...ing?”

- I'm usually not one to “verb” my nouns (then use as a gerund)

“Data...ing?”

- I'm usually not one to “verb” my nouns (then use as a gerund)
- but I've heard this kind of thing happens on Twitter

ADULTING

What Is the Meaning of Adulting?

ADULTING

- Gerund that stems from the use of the word adult as a verb.
- Means behaving maturely or acting like an adult.
- A lighthearted term used to highlight the mundane or comical activities being mature and responsible can bring about.

grammarist.com

An illustration showing two people performing domestic tasks. On the left, a person in a yellow shirt and orange pants uses a red and blue vacuum cleaner. On the right, another person in a blue shirt and patterned pants carries a blue laundry basket filled with folded laundry. In the background, there is a blue washing machine and a bottle of laundry detergent. The scene is set against a light gray background with a white dashed border around the central text area.

The image shows two people performing domestic tasks. One person is vacuuming the floor, and the other is carrying a basket of laundry. This visual representation illustrates the concept of 'adulting' as mundane, responsible activities.

Adulting is a fairly new gerund that stems from the use of the word adult as a verb. A gerund is a verb form ending in -ing that acts as a noun. *Adulting* simply means behaving maturely or acting like an adult.

Figure 2: image credit: Grammarist. URL:
<https://grammarist.com/new-words/adulting/>

But “Data” as a verb?

- I've heard of *data verbs*, just not “data” as a verb

But “Data” as a verb?

- I've heard of *data verbs*, just not “data” as a verb
- Amelia McNamara helpfully pointed out (to my surprise) this too has precedent!

But “Data” as a verb?

- I've heard of *data verbs*, just not “data” as a verb
- Amelia McNamara helpfully pointed out (to my surprise) this too has precedent!
- Who would do such a thing?

Jer Thorp



Roman Makhmutov

Jer Thorp is an artist, a writer, and a teacher. He was the first data artist in residence at *The New York Times*, is a *National Geographic Explorer*, and served as the innovator in residence at the Library of Congress in 2017 and 2018. He lives under the Manhattan Bridge with his family and his awesome dog, Trapper John, MD. *Living in Data* is his first book. You can sign up for email updates [here](#).

Coming around. . .

- “the pair data and data-ing refers to a similar conceptualization of the relation between sample and sampling, or model and modeling, where the first is the statistical concept and the second refers to the process of engaging or reasoning with this concept.”

Coming around. . .

- “the pair data and data-ing refers to a similar conceptualization of the relation between sample and sampling, or model and modeling, where the first is the statistical concept and the second refers to the process of engaging or reasoning with this concept.”
- *This* resonates with me. . .

Coming around. . .

- “the pair data and data-ing refers to a similar conceptualization of the relation between sample and sampling, or model and modeling, where the first is the statistical concept and the second refers to the process of engaging or reasoning with this concept.”
- *This* resonates with me. . .
- and it's starting to sound like something I've observed before

What perspective can I offer?

The birth of an academic



Figure 4: (Academic) offspring of Joan Garfield & Bob delMas

- if any of you know me, it's probably thanks to my parents PhD advisors Joan Garfield & Bob delMas

The birth of an academic



Figure 4: (Academic) offspring of Joan Garfield & Bob delMas

- if any of you know me, it's probably thanks to my parents PhD advisors Joan Garfield & Bob delMas
- there's plenty more to my upbringing prior to Statistics Education, including lots that has almost nothing to do with academics. . .

My actual Mum & Dad



Figure 5: Marian & Don Beckman

Socializing with friends



Rebellious phase

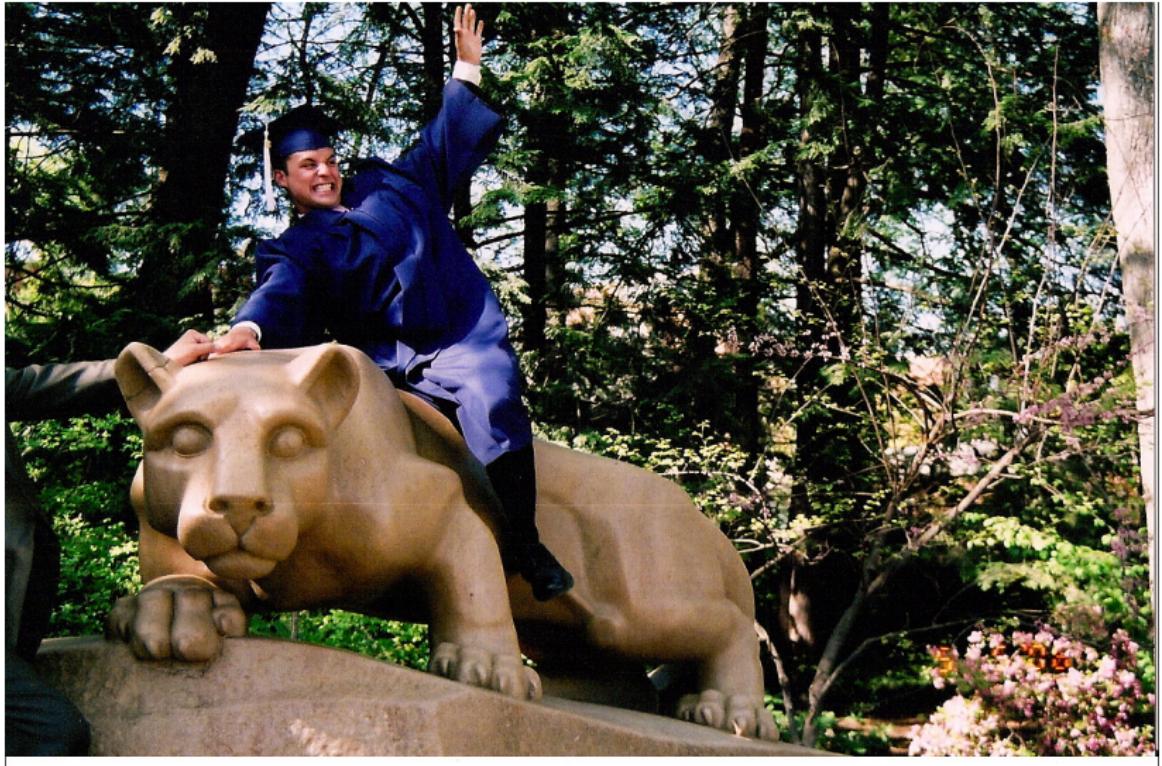


Figure 6: Who knows...

Got an education. . .



Getting more serious. . .

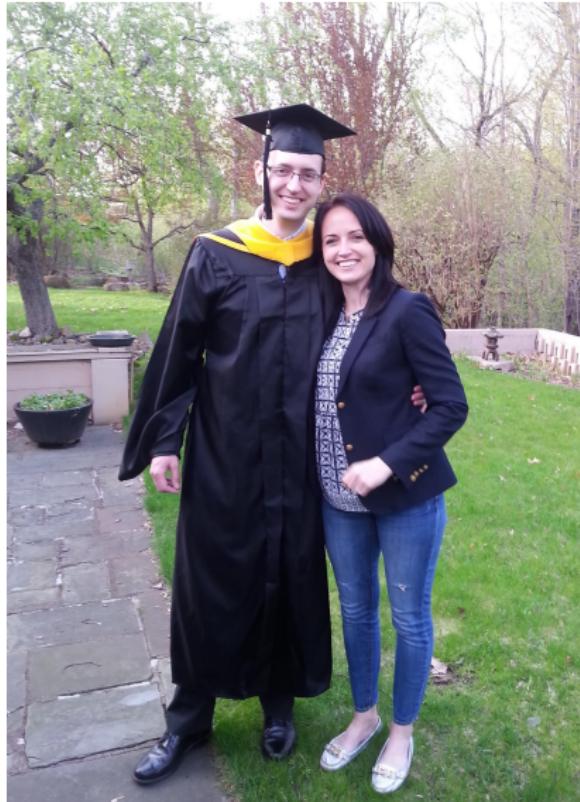


Figure 7: Met my wife in Minnesota and began a career in statistics.

Where the magic happens. . .

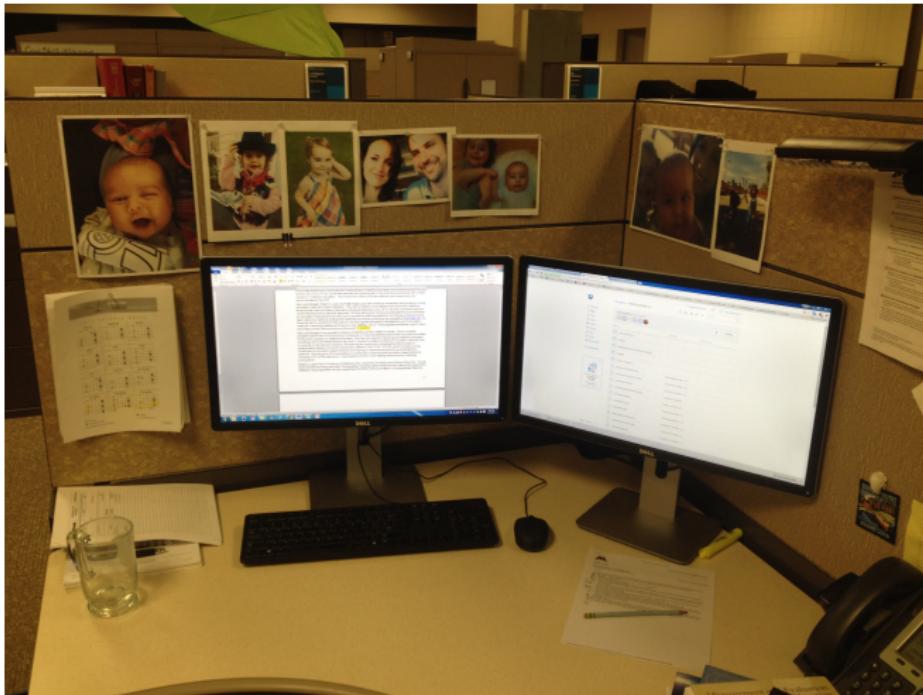


Figure 8: Working late. . .

Ecolab, Inc.



Figure 9: Image credit: Patrick Kennedy, Star Tribune²

- Started out at Ecolab working in R&D (+ Eng)
- Interned with a small team of staff statistical consultants
- Typical Data: Lots of design & analysis of experiments
- Assist teaching in-house statistics courses for Engineers, Mgmt, scientists, etc (Intro, DOE, MSA, SPC, RDSA)

²Kennedy, P., (6 Feb 2023). Ecolab now selling products at Home Depot — the first time available at retail stores. Star Tribune. URL: <https://www.startribune.com/ecolab-now-selling-products-at-home-depot-the->

8 yrs in Medical Technology



Figure 10: image credit:

<https://asiapac.medtronic.com/xp-en/about.html>

- Hired due to commitment to government regulators!
- Typical data: verbatim complaints from call center, manufacturing lines, clinical, sales & registration data, engineering diagnostics from in-house returned product analysis, lots more!
- Promoted to Senior Statistician; broader involvement including Quality, Manufacturing, R&D, Marketing, HR/Personnel, Six Sigma
- Regularly tasked with discussing/explaining statistical methodology and procedures to government regulators (e.g., post-market surveillance)

Nonin Medical. Inc



Figure 11: image credit: Michael Heisson³

- Senior Biostatistician (“only” statistician...)
- Internal & external collaborations (e.g., anesthesiology research, clinical trial design, etc)

³Lee, E., (4 Oct 2022). The Best Pulse Oximeter for Home Use. New York Times Wirecutter. URL: <https://www.nytimes.com/wirecutter/reviews/best-pulse-oximeter-for-home-use/>

Back to Medtronic!?

- Back to Medtronic for a lame duck session
- Goals: clean up special projects, train new statisticians (& business analysts) and help automate my own job away (!)
- Next Stop: Penn State!
- When I was a few years younger, people gave me puzzled looks when I described my background...

Back to Medtronic!?

- Back to Medtronic for a lame duck session
- Goals: clean up special projects, train new statisticians (& business analysts) and help automate my own job away (!)
- Next Stop: Penn State!
- When I was a few years younger, people gave me puzzled looks when I described my background. . .
- You know that look when someone is calculating your age while you talk with them?

Back to Medtronic!?

- Back to Medtronic for a lame duck session
- Goals: clean up special projects, train new statisticians (& business analysts) and help automate my own job away (!)
- Next Stop: Penn State!
- When I was a few years younger, people gave me puzzled looks when I described my background...
- You know that look when someone is calculating your age while you talk with them?
- *college grad at 21 yrs + 2 yrs (MS Statistics) + 8 yrs industry + 5 yrs (PhD) + yrs as Penn State faculty...*

Back to Medtronic!?

- Back to Medtronic for a lame duck session
- Goals: clean up special projects, train new statisticians (& business analysts) and help automate my own job away (!)
- Next Stop: Penn State!
- When I was a few years younger, people gave me puzzled looks when I described my background...
- You know that look when someone is calculating your age while you talk with them?
- *college grad at 21 yrs + 2 yrs (MS Statistics) + 8 yrs industry + 5 yrs (PhD) + yrs as Penn State faculty...*
- *hmm... he doesn't look old enough for all that*

Back to Medtronic!?

- Back to Medtronic for a lame duck session
- Goals: clean up special projects, train new statisticians (& business analysts) and help automate my own job away (!)
- Next Stop: Penn State!
- When I was a few years younger, people gave me puzzled looks when I described my background...
- You know that look when someone is calculating your age while you talk with them?
- *college grad at 21 yrs + 2 yrs (MS Statistics) + 8 yrs industry + 5 yrs (PhD) + yrs as Penn State faculty...*
- *hmm... he doesn't look old enough for all that*
- *... but I don't think he's a boy-genius that finished college at 16 either*

Back to Medtronic!?

- Back to Medtronic for a lame duck session
- Goals: clean up special projects, train new statisticians (& business analysts) and help automate my own job away (!)
- Next Stop: Penn State!
- When I was a few years younger, people gave me puzzled looks when I described my background...
- You know that look when someone is calculating your age while you talk with them?
- *college grad at 21 yrs + 2 yrs (MS Statistics) + 8 yrs industry + 5 yrs (PhD) + yrs as Penn State faculty...*
- *hmm... he doesn't look old enough for all that*
- *... but I don't think he's a boy-genius that finished college at 16 either*
- Life hack: I earned my PhD *while* working full-time

Personal reflections from my industry experience

- I began to notice gaps between my expectations/assumptions for my contributions at work and the reality of my experiences on the job...
- *Expectation:* UMN Statistics Dept filled my toolbox with advanced methods & fancy models... that's what they'll expect me to do at work.
- *Mind the gap:* At work, I generally used 10-20% of the fancy things I learned in those courses

Personal reflections from my industry experience

- I began to notice gaps between my expectations/assumptions for my contributions at work and the reality of my experiences on the job...
- *Expectation:* UMN Statistics Dept filled my toolbox with advanced methods & fancy models... that's what they'll expect me to do at work.
- *Mind the gap:* At work, I generally used 10-20% of the fancy things I learned in those courses
- Lots of the fancy methods that I needed at work, I *learned* at work. This is a common refrain among professional statisticians.

Personal reflections from my industry experience

- I began to notice gaps between my expectations/assumptions for my contributions at work and the reality of my experiences on the job...
- *Expectation:* UMN Statistics Dept filled my toolbox with advanced methods & fancy models... that's what they'll expect me to do at work.
- *Mind the gap:* At work, I generally used 10-20% of the fancy things I learned in those courses
- Lots of the fancy methods that I needed at work, I *learned* at work. This is a common refrain among professional statisticians.
- *Mind the gap:* The statisticians on my team were regularly requested to engage with myriad issues that were **not at all statistical** in nature.

Personal reflections from my industry experience

- I began to notice gaps between my expectations/assumptions for my contributions at work and the reality of my experiences on the job...
- *Expectation:* UMN Statistics Dept filled my toolbox with advanced methods & fancy models... that's what they'll expect me to do at work.
- *Mind the gap:* At work, I generally used 10-20% of the fancy things I learned in those courses
- Lots of the fancy methods that I needed at work, I *learned* at work. This is a common refrain among professional statisticians.
- *Mind the gap:* The statisticians on my team were regularly requested to engage with myriad issues that were **not at all statistical** in nature.

Personal reflections from my industry experience

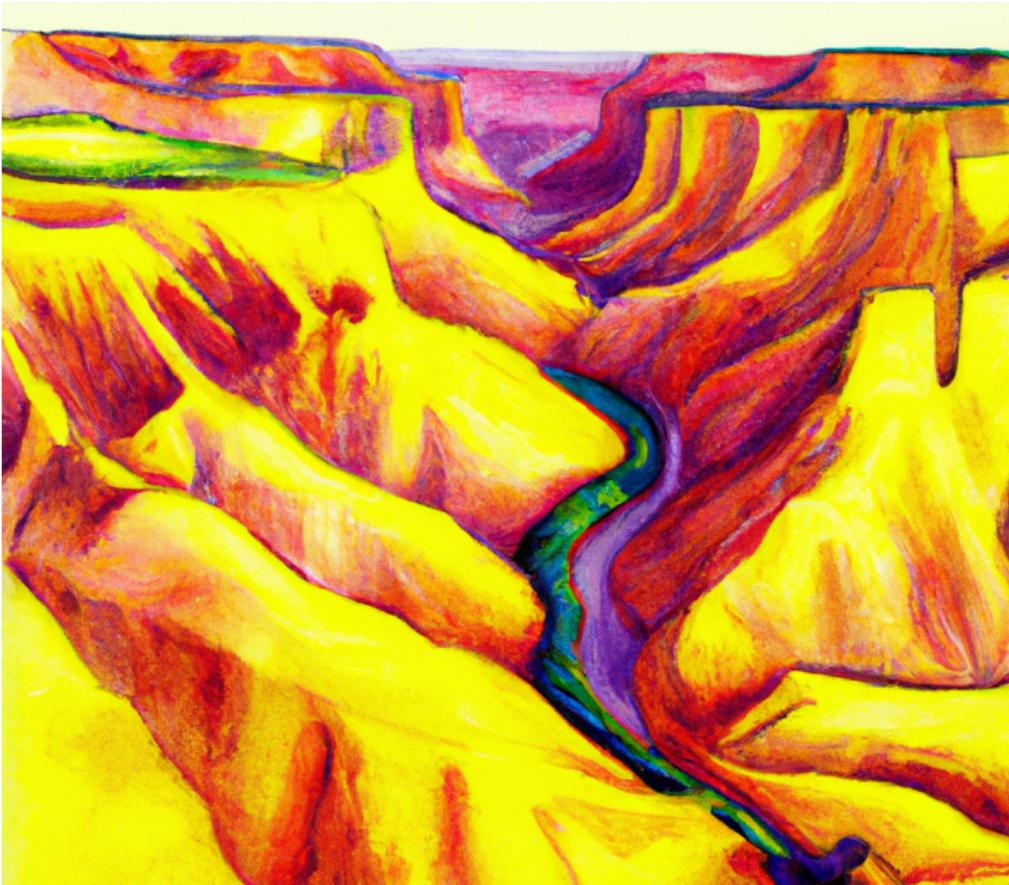
- I began to notice gaps between my expectations/assumptions for my contributions at work and the reality of my experiences on the job...
- *Expectation:* UMN Statistics Dept filled my toolbox with advanced methods & fancy models... that's what they'll expect me to do at work.
- *Mind the gap:* At work, I generally used 10-20% of the fancy things I learned in those courses
- Lots of the fancy methods that I needed at work, I *learned* at work. This is a common refrain among professional statisticians.
- *Mind the gap:* The statisticians on my team were regularly requested to engage with myriad issues that were **not at all statistical** in nature.

Mind the Gap: An incomplete picture of statistics, statisticians, & statistics education

Matthew Beckman
Penn State University

June 25, 2023
Maleny, Australia

Mind the gap



Mind the gap: In the public

statistics vs Statistics

Mind the gap: EDA

Perceived value vs potential contribution of EDA (exploratory data analysis)

Mind the gap: EDA

- My solution?

Mind the gap: EDA

- My solution?
- my students need a framework for a careful EDA

Mind the gap: EDA

- My solution?
- my students need a framework for a careful EDA
- with an acronym to help remember it

EDA Framework (2019)

Get in “B-E-D” with your data

- Become acquainted with the data

EDA Framework (2019)

Get in “B-E-D” with your data

- **B**ecome acquainted with the data
- **E**xplore intuition for your research question(s)

EDA Framework (2019)

Get in “B-E-D” with your data

- **B**ecome acquainted with the data
- **E**xplore intuition for your research question(s)
- **D**iscover features in the data that impact modeling decisions

EDA Framework (2nd attempt)

... not a great acronym

Let's just reuse "EDA" instead

- **Examine the data source(s):** data provenance, variable types, coding, missingness, summary statistics/plots;

EDA Framework (2nd attempt)

... not a great acronym

Let's just reuse "EDA" instead

- **Examine the data source(s):** data provenance, variable types, coding, missingness, summary statistics/plots;
- **Discover features that influence modeling decisions:** investigate potential outliers, consideration for recoding variables (e.g., numeric data that's functionally dichotomous), evaluate correlation structure (e.g., autocorrelation, hierarchy, spatial/temporal proximity);

EDA Framework (2nd attempt)

... not a great acronym

Let's just reuse "EDA" instead

- **Examine the data source(s):** data provenance, variable types, coding, missingness, summary statistics/plots;
- **Discover features that influence modeling decisions:** investigate potential outliers, consideration for recoding variables (e.g., numeric data that's functionally dichotomous), evaluate correlation structure (e.g., autocorrelation, hierarchy, spatial/temporal proximity);
- **Address research questions:** build intuition and note preliminary observations/conclusions related to each research question. Also, note observations that prompt you to refine your research questions or add new questions to investigate

EDA Framework (2nd attempt)

... not a great acronym

Let's just reuse "EDA" instead

- **Examine the data source(s):** data provenance, variable types, coding, missingness, summary statistics/plots;
- **Discover features that influence modeling decisions:** investigate potential outliers, consideration for recoding variables (e.g., numeric data that's functionally dichotomous), evaluate correlation structure (e.g., autocorrelation, hierarchy, spatial/temporal proximity);
- **Address research questions:** build intuition and note preliminary observations/conclusions related to each research question. Also, note observations that prompt you to refine your research questions or add new questions to investigate
- p.s. Lonneke's work included a concise summary of literature that I think aligns quite closely with the "E" step!

Mind the gap: Opportunity for research

Mind the gap: Opportunity for research

International Statistical Review (1999), 67, 3, 223–265, Printed in Mexico
© International Statistical Institute

Statistical Thinking in Empirical Enquiry

C.J. Wild and M. Pfannkuch

Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand

This paper had its genesis in a clash of cultures. Chris Wild is a statistician. Like many other statisticians, he has made impassioned pleas for a wider view of statistics in which students learn “to think statistically” (Wild, 1994). Maxine Pfannkuch is a mathematics educator whose primary research interests are now in statistics education. Conception occurred when Maxine asked “What is statistical thinking?” It is not a question a statistician would ask. Statistical thinking is the touchstone at the core of the statistician’s art. But, after a few vague generalities, Chris was reduced to stuttering.

- item 1
- item 2

Box

- ① item
- ② item

- Cite stuff people have already done⁵
- Link to URL's: <https://en.wikipedia.org/wiki/URL>

⁵American Statistical Association Undergraduate Guidelines Workgroup (2014). *Curriculum guidelines for undergraduate programs in statistical science.*

Here's a table

Col1	Col2
1	One
2	Two
3	Three
4	Four
5	Five
6	Six

Acknowledgments

- So many to thank

References

- ① American Statistical Association Undergraduate Guidelines Workgroup (2014). 2014 Curriculum guidelines for undergraduate programs in statistical science. Alexandria, VA: American Statistical Association.
<http://www.amstat.org/education/curriculumguidelines.cfm>

Q & A

Mind the Gap: An incomplete picture of statistics, statisticians, & statistics education

Matthew Beckman
Penn State University

June 25, 2023
Maleny, Australia

Include URL for gh-pages site for presentation

Backup slide