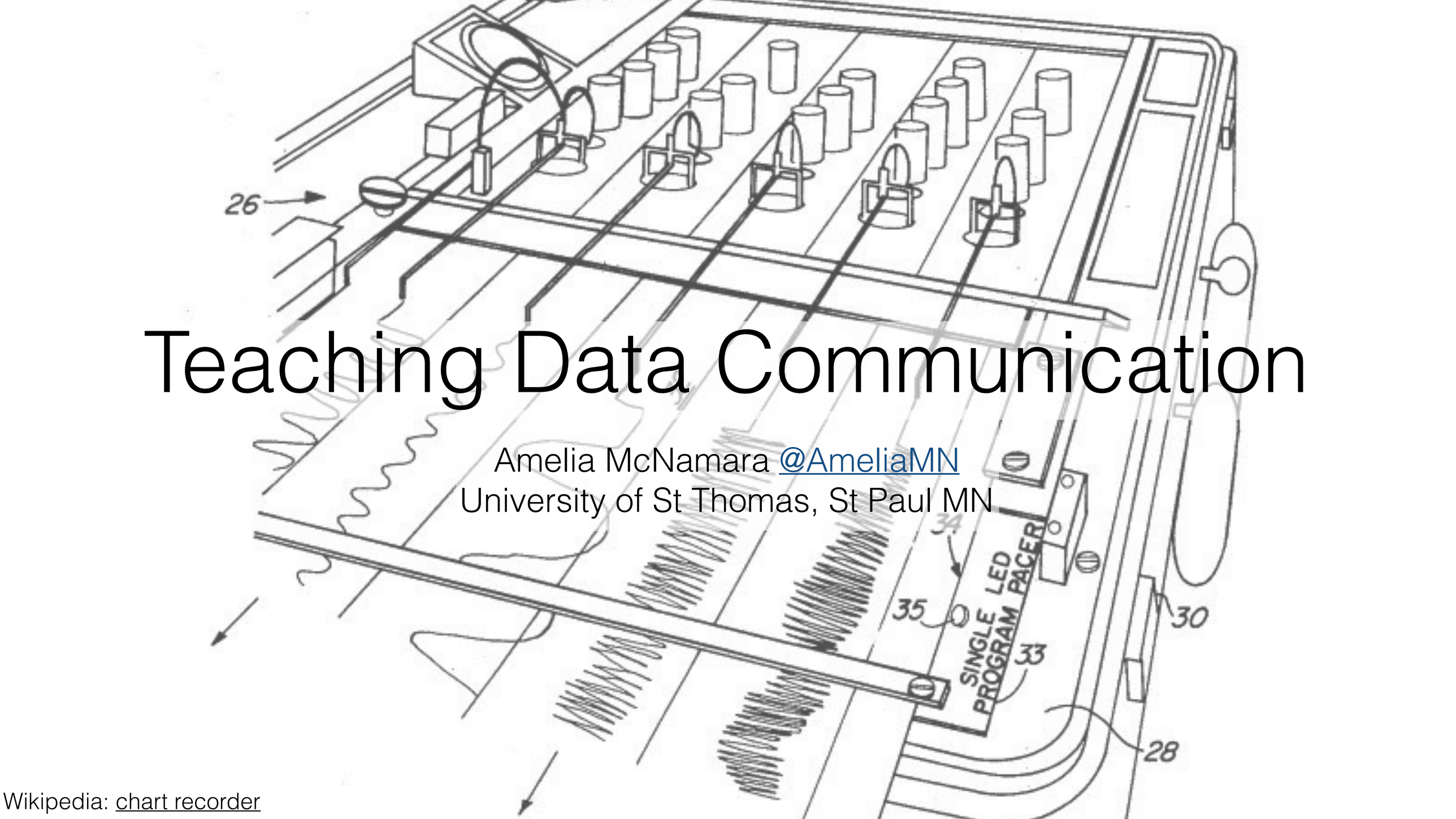


Teaching Data Communication

Amelia McNamara @AmeliaMN
University of St Thomas, St Paul MN

Wikipedia: chart recorder

We can all agree
communication is important

“Key concepts required to develop data acumen include mathematical foundations, computational foundations, statistical foundations, data management and curation, data description and visualization, data modeling and assessment, workflow and reproducibility, **communication**, domain-specific considerations, and ethical problem solving.”

–Data Science for Undergraduates: Opportunities and Options
National Academies, 2018

“Key Competencies for an undergraduate Data Science Major

Computational and Statistical Thinking
Mathematical Foundations
Model Building and Assessment
Algorithms and Software Foundation
Data Curation
Knowledge Transference – **Communication** and Responsibility”

–Curriculum Guidelines for Undergraduate Programs in Data Science
Park City Math Institute (PCMI)

But, who teaches communication, what does communication mean, and when should students learn it?

“Recommendation 2.1: Academic institutions should embrace **data science** as a vital new field that **requires specifically tailored instruction** delivered through majors and minors in data science as well as the development of a cadre of faculty equipped to teach in this new field.”

"As instructors rework individual classes based on outcomes and evaluation, it is likely that they will replace borrowed content from existing courses with **original materials** that fit together more naturally and better match personal educational styles or the culture of that institution or department."

–Data Science for Undergraduates: Opportunities and Options
National Academies, 2018

“Most institutions will implement a Data Science major from current courses in existing disciplines, **perhaps transitioning to more fully integrated courses** as outlined in the Appendix at a future date.”

"6.4. Related Courses
- Introduction to [Partner Discipline]
- Intermediate course in Discipline
- **Capstone Course with Data Experience and Projects**
- Two courses in writing, preferably one in technical writing.
- Public Speaking
- Ethics

[…] highlighed courses cover the bare necessities of the material required for a Data Science major"

–Curriculum Guidelines for Undergraduate Programs in Data Science
Park City Math Institute (PCMI)

Karen McGrane

@karenmcgrane

Following

Petition to rename "soft skills" to "power skills" who's with me

> **Dr. DeRionne Pollard** @DrPollard_MC
> Based on what I continuously hear, we really gotta drop the term “soft skills.” It’s diminutive and dismissive. Instead, we should be calling communication skills, problem-solving skills, critical thinking skills, etc “power skills” or “foundational skills.”

2:55 PM - 9 Dec 2018

232 Retweets **1,070** Likes

35 232 1.1K

Data communication

- Visualizing data
- Writing data
- Speaking data

University of Michigan: viz ✔ writing ✔ speaking ✔ (take three)

Smith College: viz ✔ writing ✔ speaking ✖ (take one)

Virginia Tech: viz ✔ writing ✖ speaking ✖

University of California, San Diego: viz ✔ writing ✖ speaking ✖

(example programs from National Academies report)

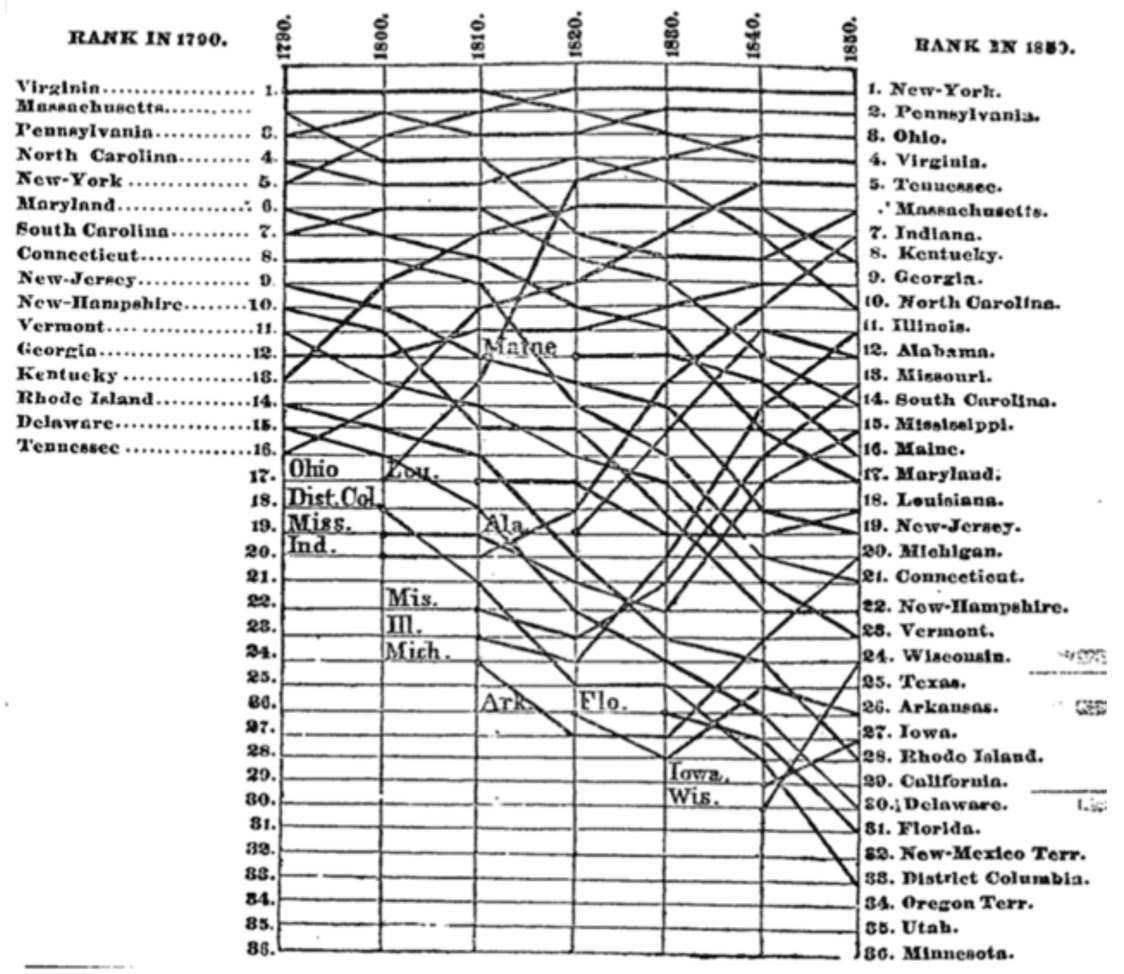
Visualizing data

Probably the least controversial of the courses/topics I am proposing.

Many institutions offer standalone data visualization courses (we're not sending students to the Art department for color theory).

Based on theory by Cleveland, Bertin, Wilkinson, Cook, Hofmann, etc.

Usually a blend of "critiquing" and "doing."



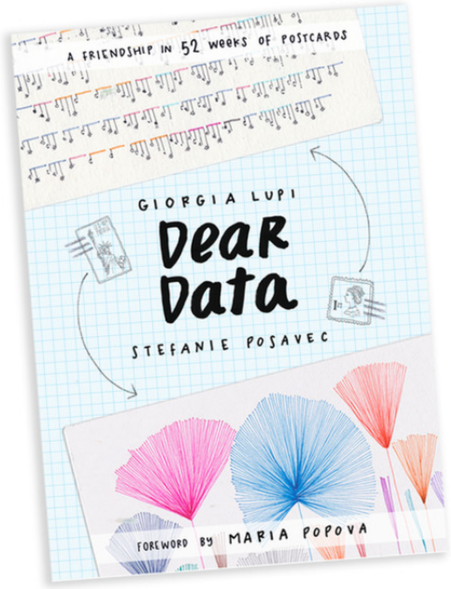
timesmachine via Scott Klein

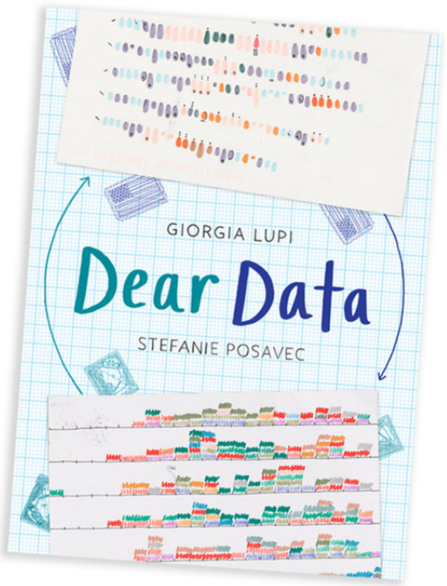
🔦 Handmade data viz

Begin with inspiration.
In my class, I started with visualization as art, and then moved on to art as visualization.

Jose Duarte

Mona Chalabi





Giorgia Lupi and Stefanie Posavec

🔦 Handmade data viz

Curate a small dataset. I recommend fewer than 10 rows, but at least two variables.

Category	Percent
White	45
Unknown	8
Hispanic	10
Asian	12
Black	6
Multiracial	5
International	14

Smith student body diversity.
2017 data via Institutional Research

Name	Area	Max depth	Watershed area	Chain of lakes?
Bde Maka Ska	421 acres	89.9 feet	2,992 acres	Yes
Lake Harriet	353 acres	82. feet	1,139 acres	Yes
Lake Nokomis	204 acres	33.1 feet	869 acres	No
Cedar Lake	170 acres	50.9 feet	1,956 acres	Yes
Lake of the Isles	103 acres	30.8 feet	735 acres	Yes

Lakes in the city of Minneapolis
via Wikipedia

🔦 Handmade data viz

Bring craft supplies!

🔦 Handmade data viz

Encourage iteration. I recommend a quick sketching period to generate 4+ ideas of how to represent the data.

Georgia Lupi and Stefanie Posavec, Dear Data

🔦 Handmade data viz

Share out afterward

Writing data

In most programs, writing is either integrated into other courses (e.g., regression modeling) or outsourced to other departments

Data journalism provides a model for a course focused on writing about data

Again, a combination of critique and creation

flickr: lidor

🔦 One-number story

Get used to reading data journalism

In my course, we began every class with a discussion of current events, and the writing strategies used by the journalists who wrote articles highlighted by students.

🔦 One-number story

"Keep the number of digits in a paragraph below eight."

"You'd be over your allocation with a sentence like this:

The Office of Redundancy's budget rose 48 percent in 2013, from $700.3 million to $1.03 billion.

Think about how it could change:

Over the past year, the Office of Redundancy's budget grew by nearly half, to $1 billion."

- Sarah Cohen, Numbers in the Newsroom

🔦 One-number story

Focus on one number (but use more numbers to contextualize it!)

That number might be the mean, the median, the maximum, the total…

Use simple data tools— in my class, we use spreadsheets for this assignment (sort, summarize, pivot tables)

10 High Schools in Massachusetts had a Perfect Graduation Rate in 2016

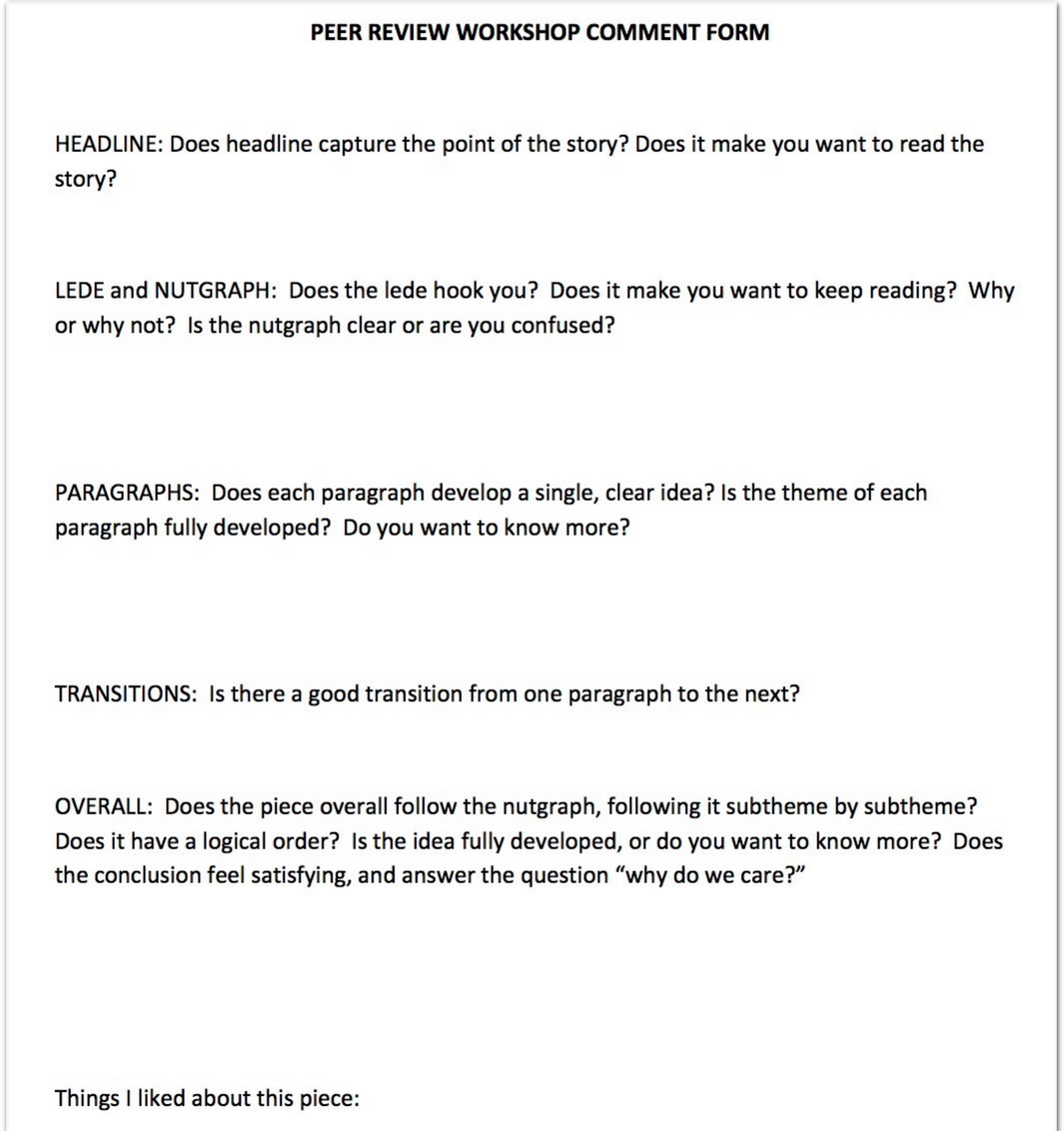
Boston Wins The High School Dropout Race

Massachusetts Academy of Math and Science Remains Atop the Podium

🔦 One-number story

Again, iteration is key

- First draft ➡️
- Peer editing in class ➡️
- Final draft ➡️
- Feedback from professor



PEER REVIEW WORKSHOP COMMENT FORM

HEADLINE: Does headline capture the point of the story? Does it make you want to read the story?

LEDE and NUTGRAPH: Does the lede hook you? Does it make you want to keep reading? Why or why not? Is the nutgraph clear or are you confused?

PARAGRAPHS: Does each paragraph develop a single, clear idea? Is the theme of each paragraph fully developed? Do you want to know more?

TRANSITIONS: Is there a good transition from one paragraph to the next?

OVERALL: Does the piece overall follow the nutgraph, following it subtheme by subtheme? Does it have a logical order? Is the idea fully developed, or do you want to know more? Does the conclusion feel satisfying, and answer the question "why do we care?"

Things I liked about this piece:

🔦 One-number story

Have students turn in hard copies, or print them out

Usually when we grade writing we ignore or give little weight to things like grammar, sentence structure, and awkward paragraphs.

This is the place to give feedback on those things.

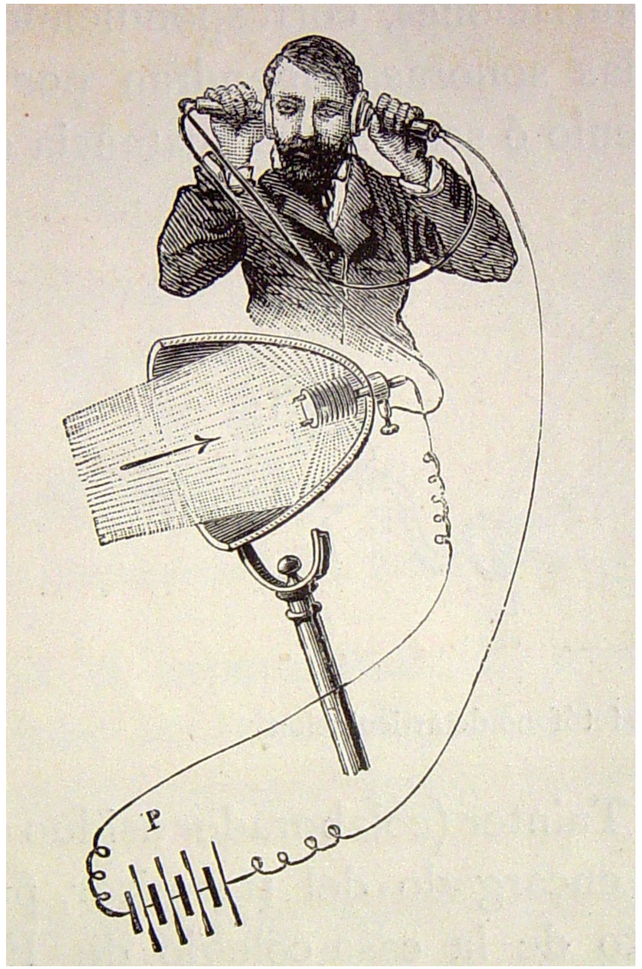
Speaking data

This is the category I think we have spent the least time formalizing.

When I teach Statistical Communication at St Thomas (Spring 2020) it will cover visualizing, writing, and speaking.

I'm looking for models of how to do it!

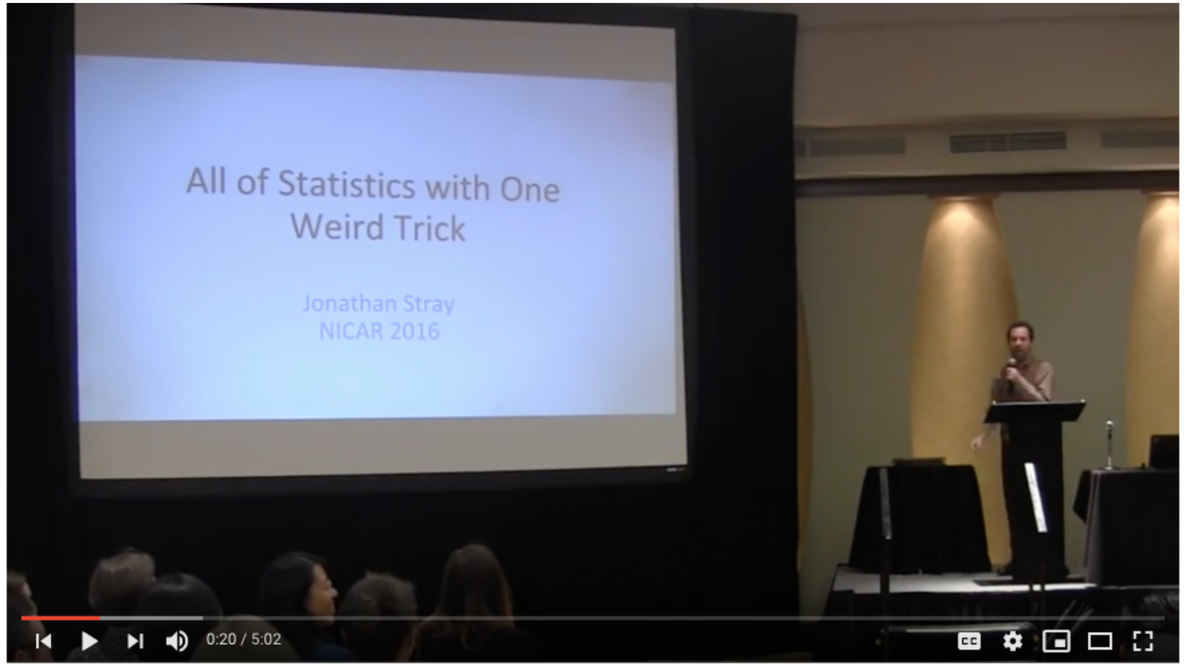
Probably: critique and create



flickr: fdctsevilla

🔦 Lightning talk

Begin from inspiration



Jonathan Stray. Solve Every Statistics Problem with One Weird Trick. NICAR 2016 Lightning talks

Carin Fishel. Date-ah. Eyeo 2016 Ignite!

🔦 Lightning talk

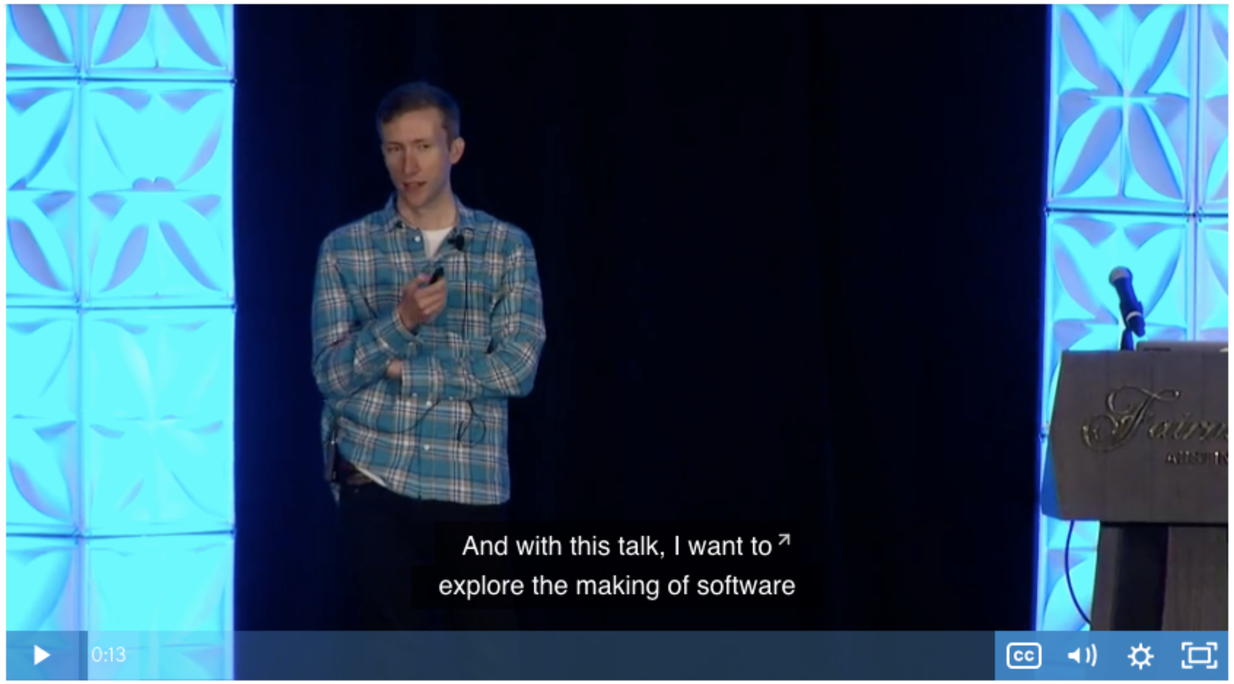
Give students lots of time to prepare

"How long does it take you to prepare one of your speeches?" asked a friend of President Wilson not long ago.

"That depends on the length of the speech," answered the President. "If it is a ten-minute speech it takes me all of two weeks to prepare it; if it is a half-hour speech it takes me a week; if I can talk as long as I want to it requires no preparation at all. I am ready now."

🔦 Lightning talk

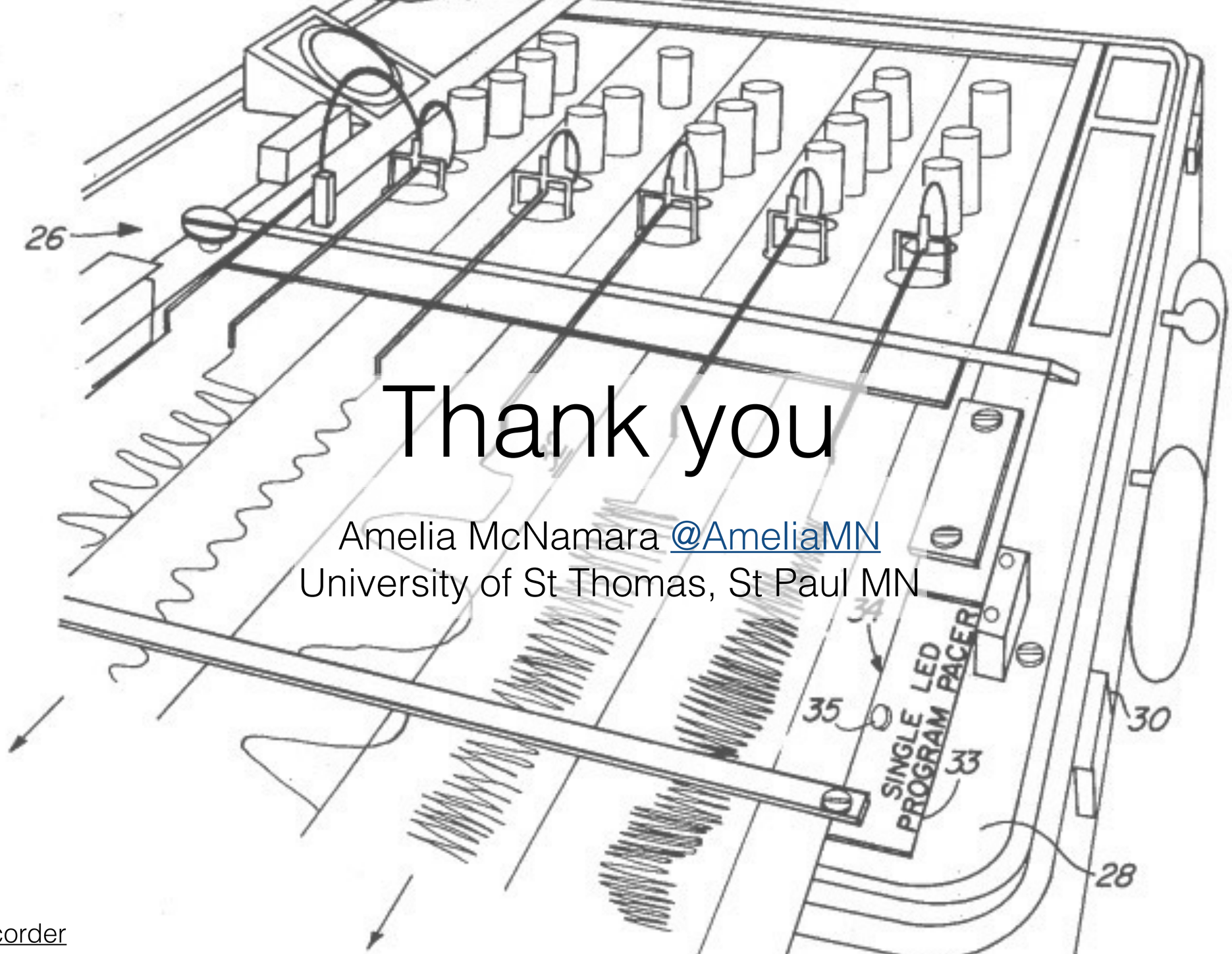
Have students do the talk twice.
(Should I tell them ahead of time?)



Miles McBain.
Our colour of magic: The open sourcery of fantastic R packages.
rstudio::conf 2019

Commonalities for teaching data communication (visual, written, or oral)

- Begin with inspiration
- Start small and simple
- Iterate and give feedback



Thank you

Amelia McNamara @AmeliaMN
University of St Thomas, St Paul MN

Wikipedia: chart recorder