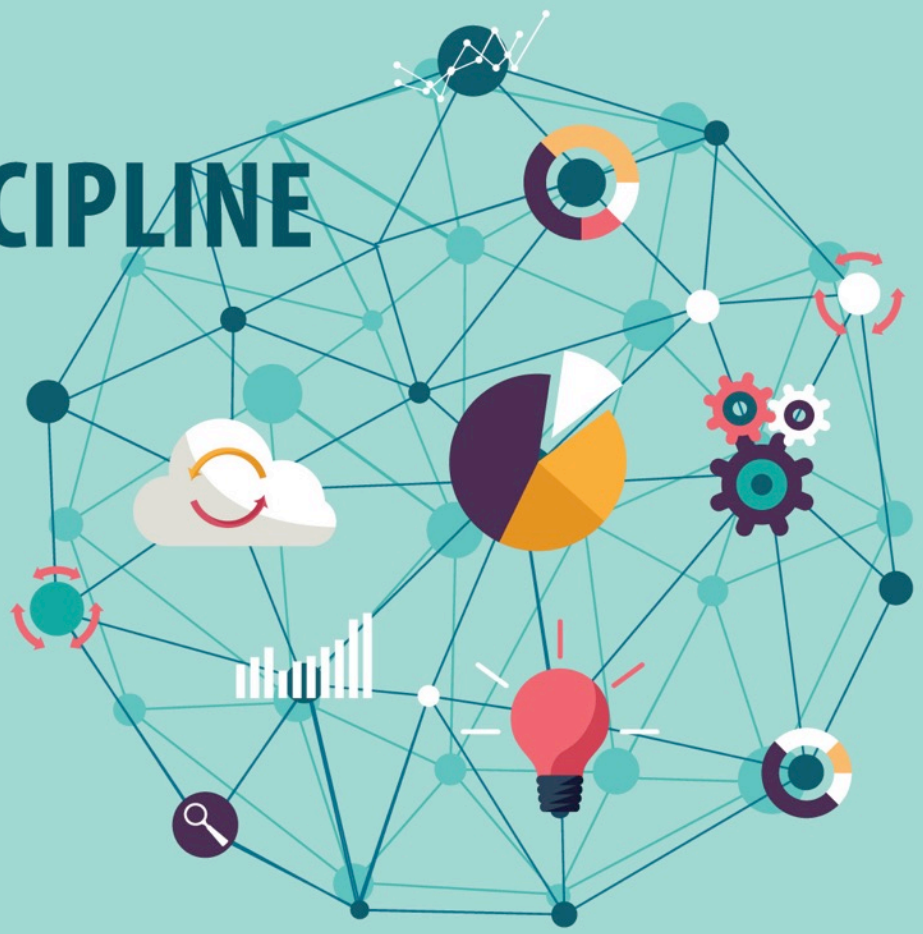
Envisioning the

DATA SCIENCE DISCIPLINE

The Undergraduate Perspective



Webinar Series
Fall 2017

The National Academies of SCIENCES ENGINEERING MEDICINE

nas.edu/EnvisioningDS

Envisioning the
DATA SCIENCE DISCIPLINE
The Undergraduate Perspective

9/12/17 – Building Data Acumen *(recording posted)*

9/19/17 – Incorporating Real-World Applications *(recording posted)*

9/26/17 – Faculty Training and Curriculum Development *(recording posted)*

10/3/17 – Communication Skills and Teamwork

10/10/17 – Inter-Departmental Collaboration and Institutional Organization

10/17/17 – Ethics

10/24/17 – Assessment and Evaluation for Data Science Programs

11/7/17 – Diversity, Inclusion, and Increasing Participation

11/14/17 – Two-Year Colleges and Institutional Partnerships

Provide input, download the interim report, and learn more about the study at www.nas.edu/EnvisioningDS

Envisioning the DATA SCIENCE DISCIPLINE

The Undergraduate Perspective

Communication Skills and Teamwork

Madeleine Claire Elish,
Data & Society Research Institute
Researcher, Intelligence & Autonomy Initiative

Adam Hughes,
Pew Research Center
Associate Researcher, Data Labs

Envisioning the

DATA SCIENCE DISCIPLINE

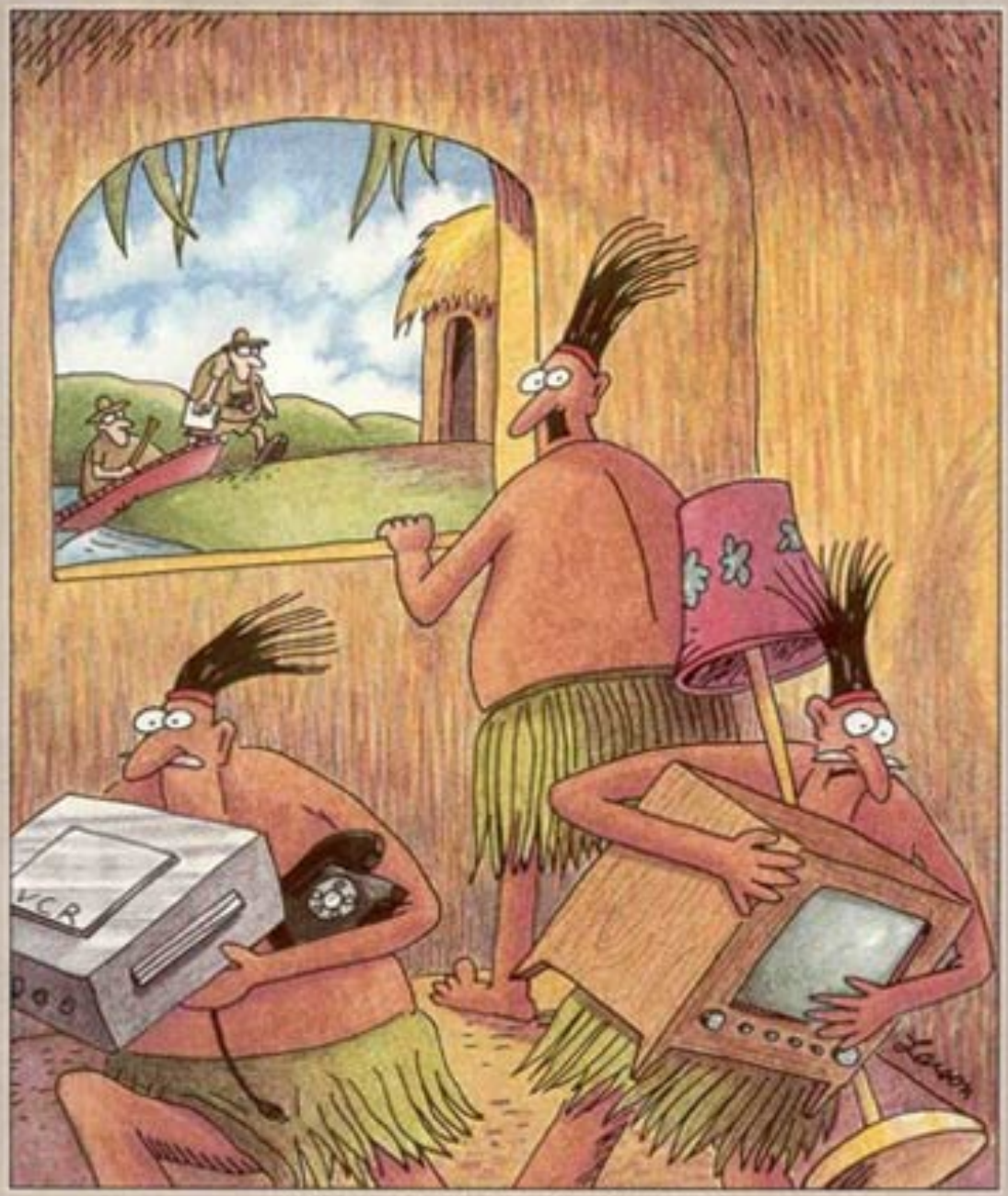
The Undergraduate Perspective

Communication Skills and Teamwork

Madeleine Claire Elish,
Data & Society Research Institute
Researcher, Intelligence & Autonomy Initiative

The Imperative of Interdisciplinarity in Data Science

Data&Society datasociety.net



"Anthropologists! Anthropologists!"

TECH 08/16/2014 11:18 am ET | **Updated** Aug 16, 2014

We May Soon Be Able To Charge Our Phone Using The Sound Of Our Voice

By Sara Gates

MASELKOO99 VIA GETTY IMAGES

What if you could charge your phone just by yelling at it?

That's the idea behind a recent collaboration between scientists at Queen Mary

TRENDING

Brian Karem Says Huckabee Sander Confrontation Was Time Coming'

Data & Society is a research institute in New York City that is focused on the social and cultural issues arising from data-centric technological development.

Data & Society is committed to identifying thorny issues at the intersection of technology and society, providing and encouraging research that can ground informed, evidence-based public debates, and building a network of researchers and practitioners who can anticipate issues and offer insight and direction.

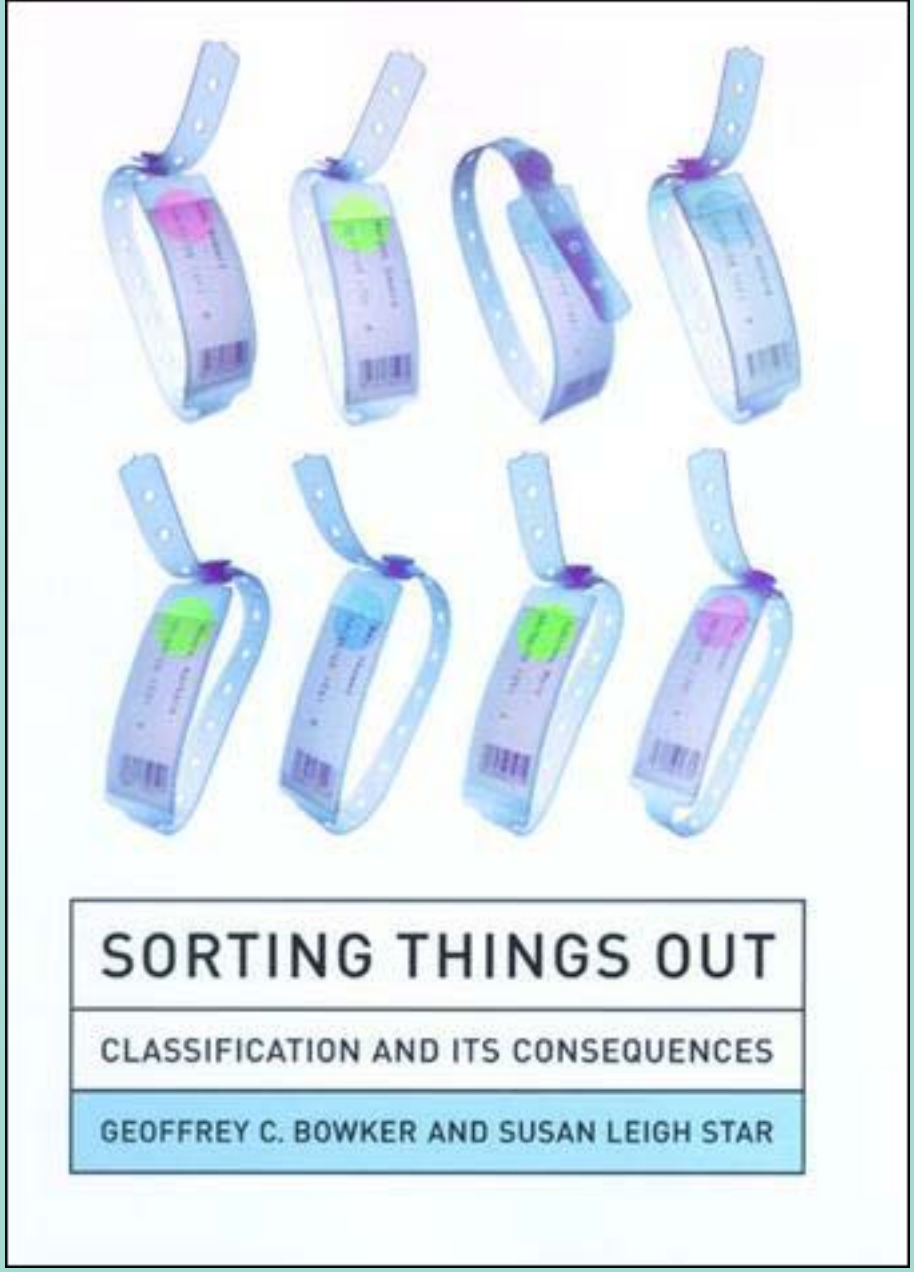
https://datasociety.net

Data&Society datasociety.net

Why is Data Science about more than numbers?

- There is no such thing as raw data …

[Library of Missing Data Sets](#) (2016)
[Mimi Onuoha](#)

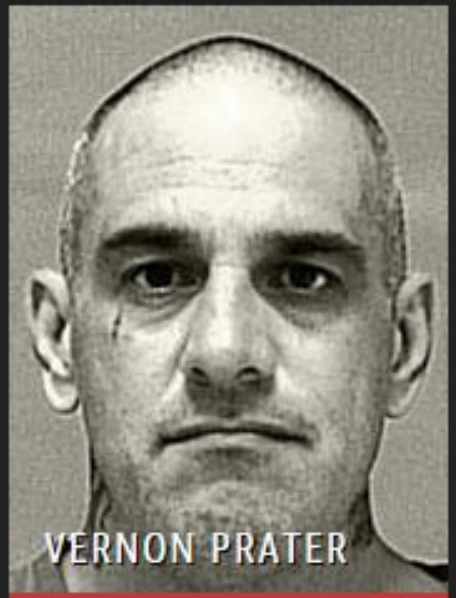

SORTING THINGS OUT
CLASSIFICATION AND ITS CONSEQUENCES
GEOFFREY C. BOWKER AND SUSAN LEIGH STAR

Why is Data Science about more than numbers?

- There is no such thing as raw data …

  or the perfect model.

Two Petty Theft Arrests



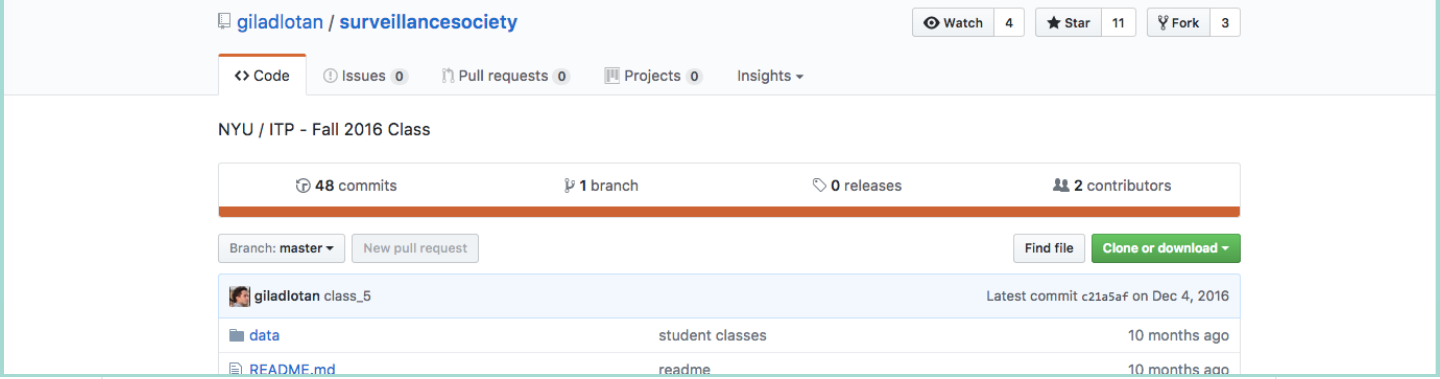
Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.

Two Petty Theft Arrests

Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.

Machine Bias, ProPublica (2016)
Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner

Data Science is an opportunity to mend the “two cultures” divide.



Final Project

This is your semester-long assignment. You should know this going into the class because you must agree to do this assignment to take the class AND you must agree to comply with the rules of engagement. This assignment is not just technically challenging - it's socially challenging. And that's the point of this class. It's one thing to talk about surveillance; it's another thing to engage in it.

On Day #1, your group will be handed a USB drive. It will contain a zipped file of data including the following three columns: latitude, longitude, and timestamp. Those are each geographic coordinates derived from the phone of your "Target." Together, this file reveals where your Target's phone has been over an extended period of time.

Your assignment in this class is two-part:

1. Identify your Target. (Goal: Dec 5)
2. Build a data portrait of your Target (to be displayed and explained on December 12).

We will help you develop the skills to do this. Ideally, you will use your sleuthing data-oriented skills to identify the full name and identity of your Target. But even if you don't get to that level of detail, you will develop an intimate understanding of your Target's activities and patterns. If you successfully identify your target, you should pull social media and web-based data about your target to help you build a portrait of your Target. You can use whatever medium makes you happy to build your portrait.

Each week, we will ask you to complete mini-assignments that show that you're working to understand your target. This is NOT an assignment that you can do last minute at the end of the semester so don't even try. You will be working on this assignment every week with your group, and we will discuss milestones.

Rules of Engagement

Your Target is a real person who has voluntarily given consent to participate in this class. We take the safety and wellbeing of the Targets seriously and you must do so as well.

Surveillance Society Course, ITP NYU (2016)

Today, Knight Foundation is announcing 17 winners of the Knight News Challenge on Data at a convening at Civic Hall in New York. Each of the winners will receive a share of $3.2 million to develop their project, which seeks to answer the question: How might we make data work for individuals and communities?

Knight News Challenge: Data (2016)

The Glass Room, NYC (2016)
Tactical Technology Collective

Further reading & resources

- Six Provocations for Big Data (2011), danah boyd & Kate Crawford: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431&rec=1&srcabs=2477899&alg=1&pos=1
- Weapons of Math Destruction (2016), Cathy O'Neil: https://mathbabe.org/
- The point of collection (2016), Mimi Onuoha: https://points.datasociety.net/the-point-of-collection-8ee44ad7c2fa
- Big Data's Disparate Impact (2015), Solon Barocas & Andrew Selbst: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899
- Situating Methods in the Magic of Big Data & AI (2017), M.C. Elish & danah boyd https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3040201
- Doing Data Science: Straight Talk from the Frontline (2013), Cathy O'Neil & Rachel Schutt: http://shop.oreilly.com/product/0636920028529.do
- Pedagogical Approaches to Data Ethics (2017), Jake Metcalf, Kate Crawford, & Emily Keller: http://bdes.datasociety.net/council-output/pedagogical-approaches-to-data-ethics-2/
- The Human Insights Missing from Big Data (2017), Tricia Wang: https://www.ted.com/talks/tricia_wang_the_human_insights_missing_from_big_data/reading-list

Please be in touch!

mcelish@datasociety.net
@m_c_elish

Envisioning the DATA SCIENCE DISCIPLINE

The Undergraduate Perspective

Communication Skills and Teamwork

Madeleine Claire Elish,
Data & Society Research Institute
Researcher, Intelligence & Autonomy Initiative

The Imperative of Interdisciplinarity in Data Science

Q&A

Data&Society datasociety.net

Data Science Collaboration for Public-Facing Research

Adam Hughes,
Pew Research Center
Associate Researcher, Data Labs
(@aghpol)

Pew Research Center

Background: Pew Research Center Data Labs

- How can Pew Research Center incorporate new kinds of data?
- Focus on computational social science (not necessarily “big data”)
- Develop, deploy, and demonstrate new research tools and produce original research

Pew Research Center

Background: Labs team

- Me: political science PhD, focus on text analysis, large behavioral datasets, experimental designs
- Colleagues: academic backgrounds in economics, communications, data science, computer science
- Individual researchers report to managing director, collaborate on everything

Pew Research Center

Team communication

- Different ways of talking about the same topics
- Need to develop a common language accessible to team members, academics, journalists, the public
- Collaborative analysis = consistent communication
- Aspiration: no jargon, no acronyms

Pew Research Center

Workflow

- Use Amazon Web Services cloud computing
- Collect and analyze data via R and Python
- Collaborate via Jupyter Notebooks, RStudio Server, Github, Google docs, Slack
- Use Django databases for ongoing projects

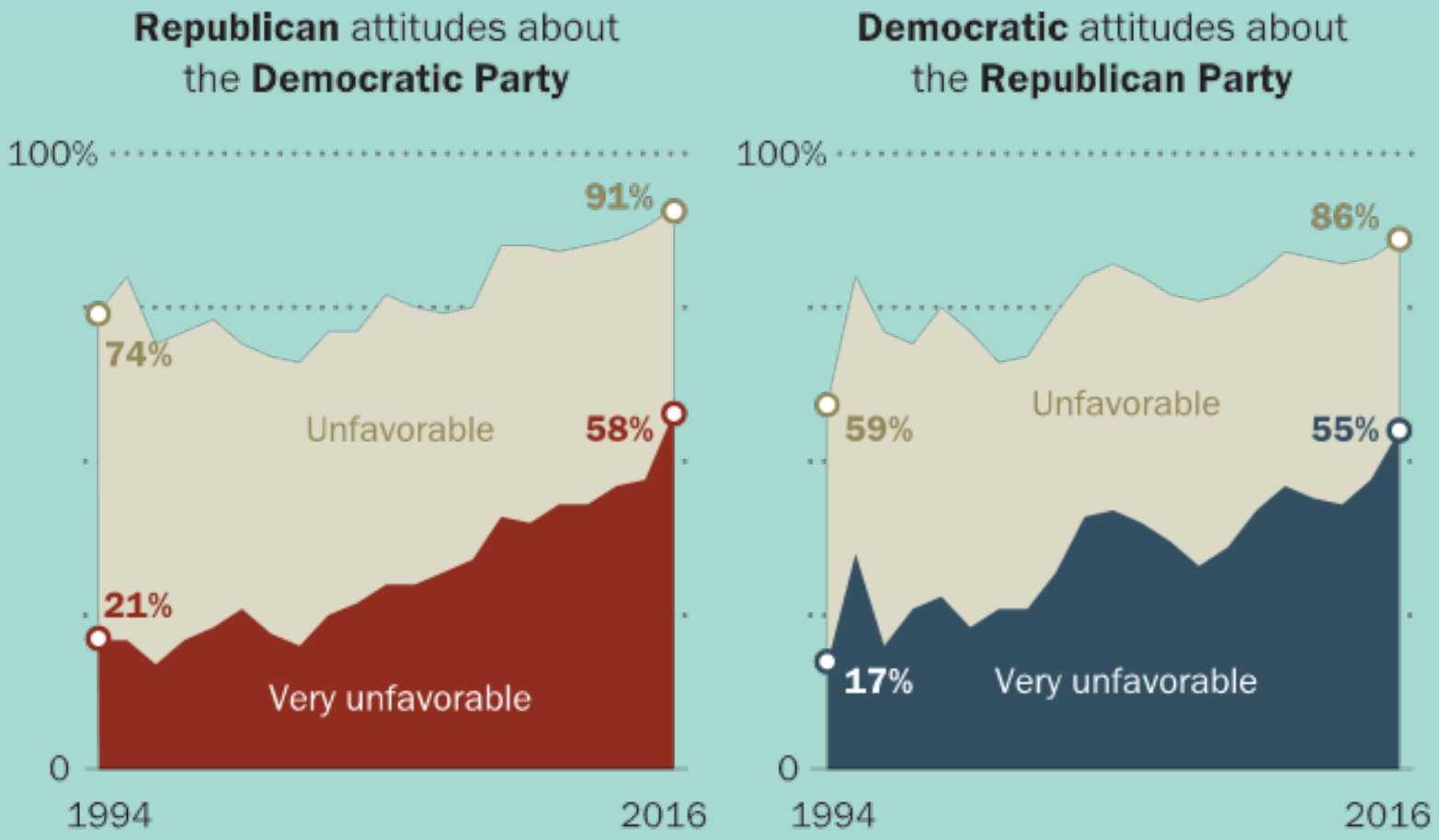
Pew Research Center

Storytelling

- Ask good research questions, then ask what you need to answer them
- Causal inference training is crucial when working with observational data
- Rely on editorial expertise from experienced writers

Pew Research Center

Surveys show deep partisan antipathy in the American public



Source: pewresearch.org, "Partisanship and Political Animosity in 2016"

Pew Research Center

Provide input and learn more about the study at www.nas.edu/EnvisioningDS

Similar pattern among elected officials?

- Collect 200,000 press releases and Facebook posts from 114th Congress (Jan 2015-Apr-2016)
- Measure cross-partisan negativity: human code, then machine learn
- Examine correlates of negativity
- Quantify audience response on Facebook

Pew Research Center

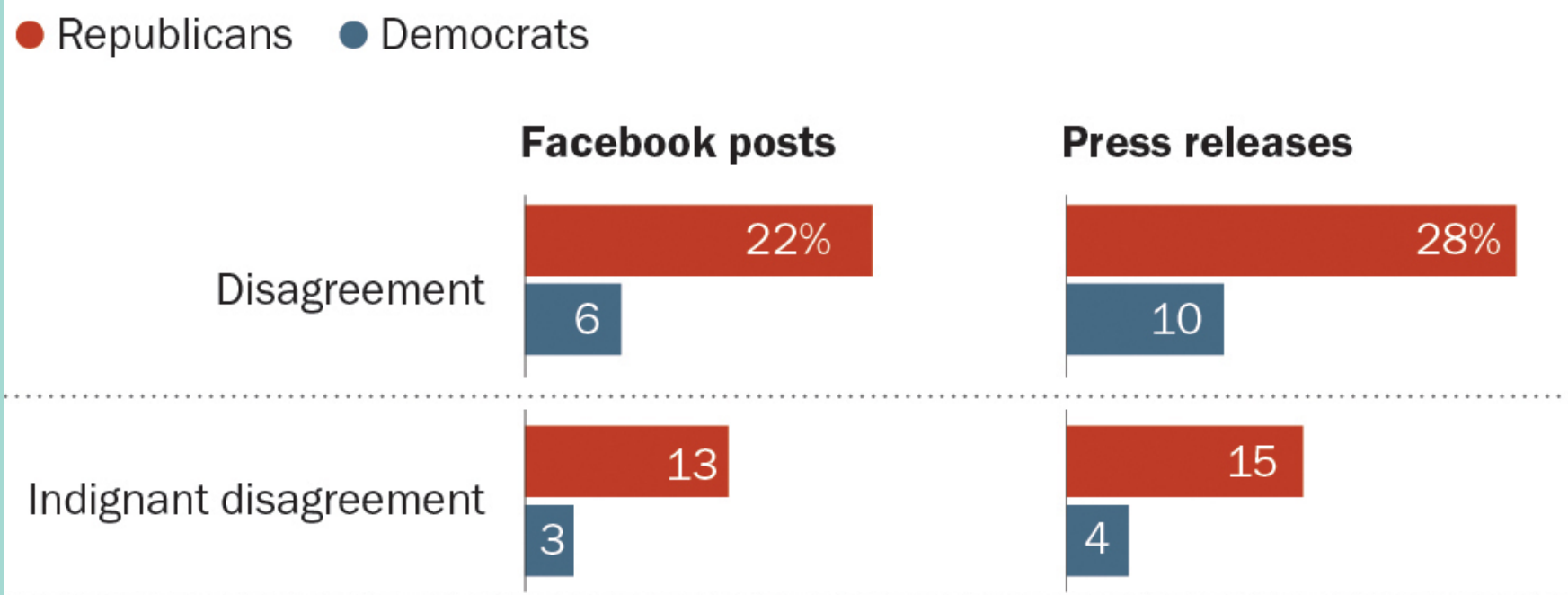
Measuring negativity

- Context *really* matters
- Focus on **targets** rather than sentiment alone
- Check validity of machine learning by comparing with human coders
- Count the number of statements that criticize someone in the other party, for every member

Pew Research Center

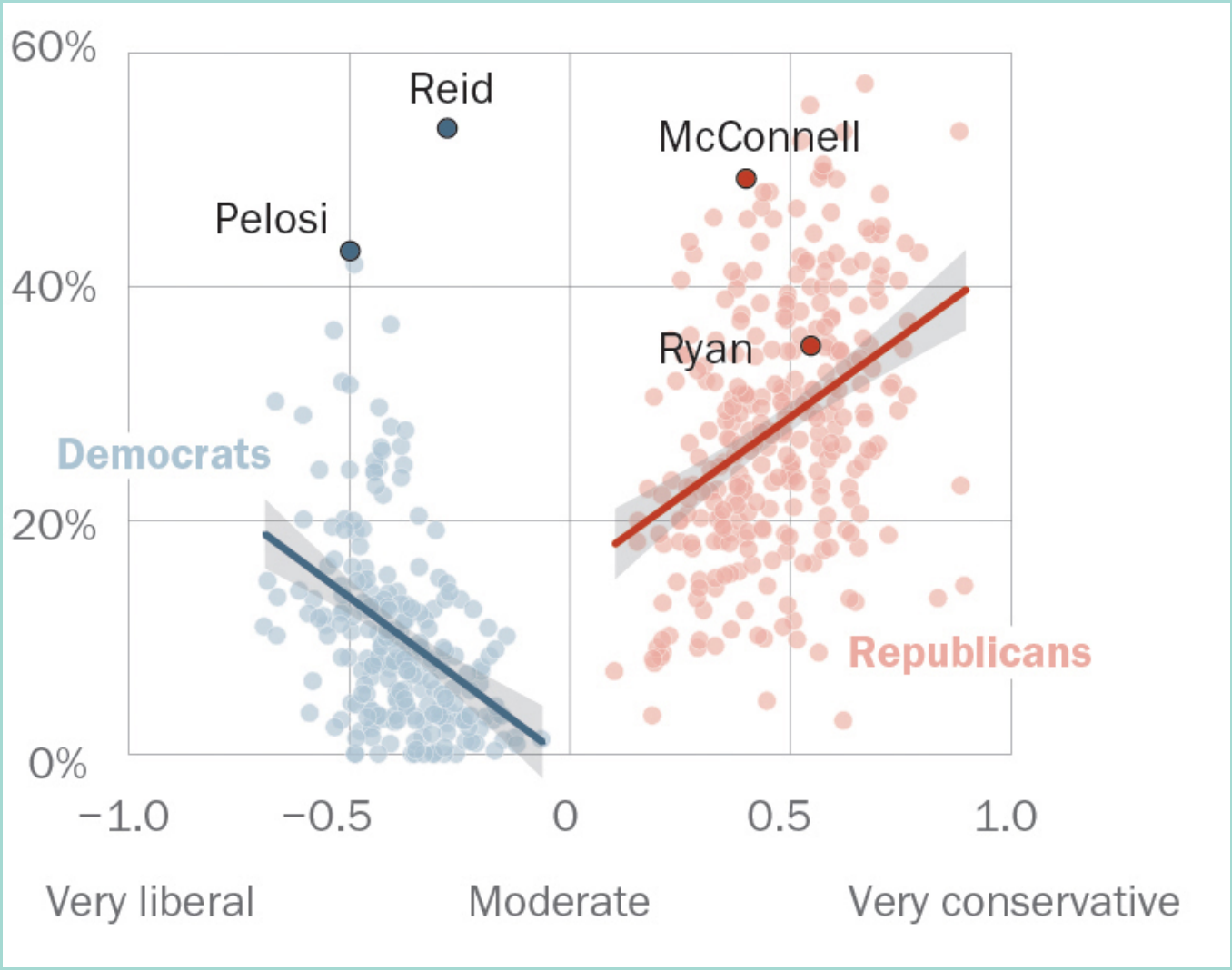
Under Obama, Democrats less likely than Republicans to criticize other side

Average % of members' communications containing or mentioning…



Pew Research Center

Very liberal, very conservative members most likely to disagree

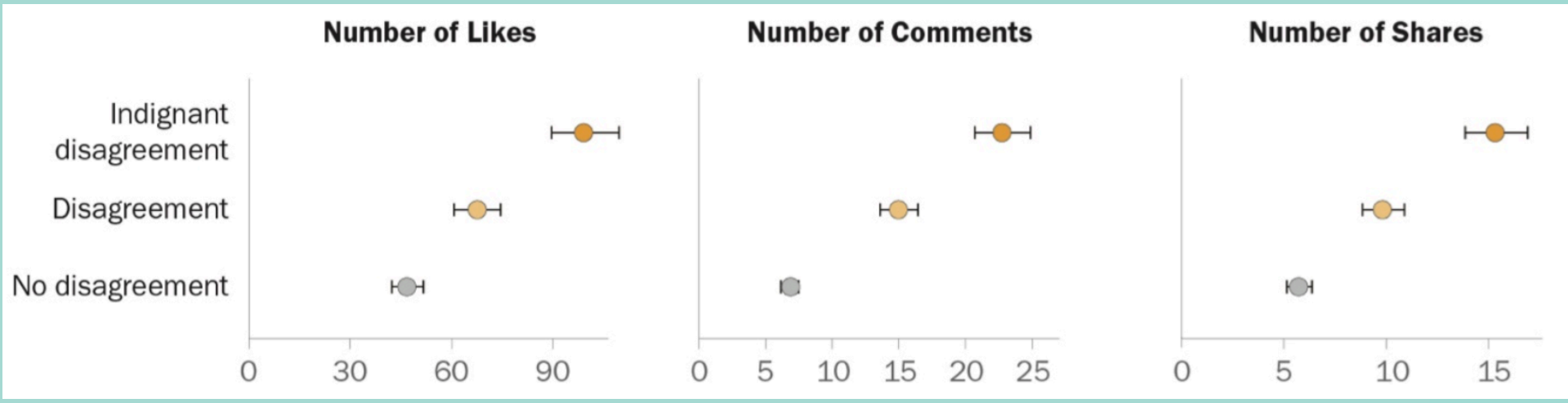


Average % of press releases containing disagreement

Pew Research Center

Negativity sees more engagement

Estimated Facebook engagement for a post containing…



Pew Research Center

Lessons learned

- Clearly define concepts, check for measurement problems early
- Build on each researcher's specialized knowledge but interpret in general ways
- Share results widely within organization to improve accessibility

Pew Research Center

Thanks!

AHughes@pewresearch.org

Pew Research Center

Envisioning the DATA SCIENCE DISCIPLINE

The Undergraduate Perspective

Communication Skills and Teamwork

Madeleine Claire Elish,
Data & Society Research Institute
Researcher, Intelligence & Autonomy Initiative

Adam Hughes,
Pew Research Center
Associate Researcher, Data Labs

Envisioning the
DATA SCIENCE DISCIPLINE
The Undergraduate Perspective

9/12/17 – Building Data Acumen *(recording posted)*

9/19/17 – Incorporating Real-World Applications *(recording posted)*

9/26/17 – Faculty Training and Curriculum Development *(recording posted)*

10/3/17 – Communication Skills and Teamwork

10/10/17 – Inter-Departmental Collaboration and Institutional Organization

10/17/17 – Ethics

10/24/17 – Assessment and Evaluation for Data Science Programs

11/7/17 – Diversity, Inclusion, and Increasing Participation

11/14/17 – Two-Year Colleges and Institutional Partnerships

Provide input, download the interim report, and learn more about the study at www.nas.edu/EnvisioningDS