

Data Journalism

Session 1: What is data journalism?

Simon Munzert

Hertie School | GRAD-E1493

Table of contents

1. Welcome!
2. Class logistics
3. Origins of data journalism
4. Data journalism today
5. Discussion

Welcome!

Introductions

Course

 <https://github.com/data-journalism-26>

Parts of this course live on GitHub. You will find slides and additional materials there. Furthermore, we will use our GitHub organization to host projects that you are going to create during the semester. We also have Moodle, which is for everything else (literature, communication, grades).

Me

 I'm **Simon Munzert** [si'mən munsərt], or just Simon [saɪmən].

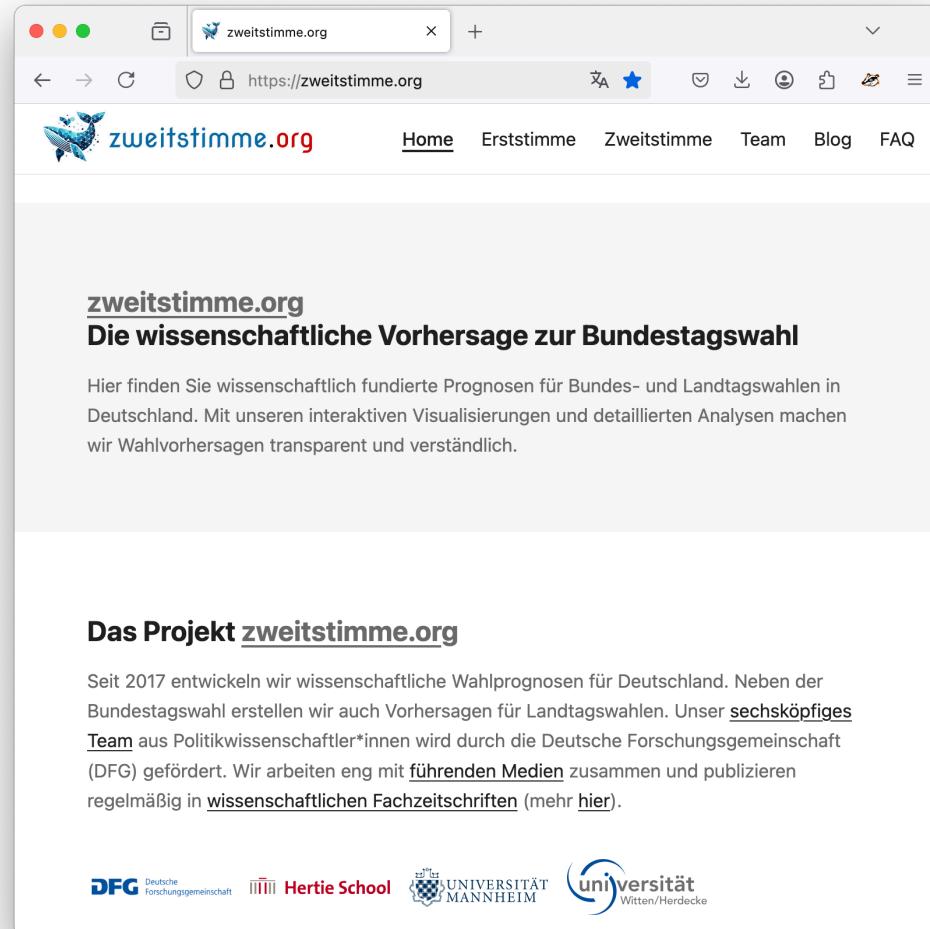
 munzert@hertie-school.org

 Professor of Data Science and Public Policy | Director of the Data Science Lab

You

What do you want to take away from this course? Let's collect your expectations! 

Where my interest in data journalism is coming from



The screenshot shows a web browser window with the URL <https://zweitstimme.org>. The page title is "zweitstimme.org" with a whale icon. The main content area features the heading "Die wissenschaftliche Vorhersage zur Bundestagswahl" and a brief description: "Hier finden Sie wissenschaftlich fundierte Prognosen für Bundes- und Landtagswahlen in Deutschland. Mit unseren interaktiven Visualisierungen und detaillierten Analysen machen wir Wahlvorhersagen transparent und verständlich." Below this, there's a section titled "Das Projekt zweitstimme.org" with a detailed paragraph about the project's history, funding by DFG, and collaboration with Hertie School and Universität Mannheim.

zweitstimme.org
Die wissenschaftliche Vorhersage zur Bundestagswahl

Hier finden Sie wissenschaftlich fundierte Prognosen für Bundes- und Landtagswahlen in Deutschland. Mit unseren interaktiven Visualisierungen und detaillierten Analysen machen wir Wahlvorhersagen transparent und verständlich.

Das Projekt zweitstimme.org

Seit 2017 entwickeln wir wissenschaftliche Wahlprognosen für Deutschland. Neben der Bundestagswahl erstellen wir auch Vorhersagen für Landtagswahlen. Unser sechsköpfiges Team aus Politikwissenschaftler*innen wird durch die Deutsche Forschungsgemeinschaft (DFG) gefördert. Wir arbeiten eng mit führenden Medien zusammen und publizieren regelmäßig in wissenschaftlichen Fachzeitschriften (mehr [hier](#)).

DFG Deutsche Forschungsgemeinschaft | **Hertie School** | **UNIVERSITÄT MANNHEIM** | **Universität Witten/Herdecke**



Where my interest in data journalism is coming from

Bundestagswahl 2025: Wie verlässlich sind Wahlumfragen?

GameStar Tech.

Bundestagswahl 2025: Mithilfe einer Website könnt ihr eurer Erststimme mehr Gewicht geben

www.zweitstimme.org zeigt euch unter anderem, welcher Politiker in eurem Landkreis die Nase vorne hat. Das ist wichtig für die Erststimme.

Maxe Schwind
22.02.2025 | 16:30 Uhr

Friedrich Merz

Das wären seine Koalitionen

Friedrich Merz könnte die Wahl gewinnen, aber kaum Optionen für eine Mehrheit haben. Aktuelle Daten zeigen: Linke, BSW und FDP verändern möglicherweise alles.

Von Gregor Aisch, Lisa Casper und Lisa Goldschmidtburg

German election 2025

Who is ahead in the race for Germany's next parliament?

Last updated on February 21st 2025

14:06 Signal Reels

Ihr könnt aber auf einer Seite, die da heißt "zweitstimme.org",

realreiners und neumeiermoritz Folgen realreiners - Original-Audio Was heißt Erst- und Zweitstimme? Weitere ...

14:04 Reels

Von ostdeutschem Alltag

Stand: 12. Februar 2025 natürlich gibt es auch N**is, die nicht in Ostdeutschland wohnen, z.B. Kanye West

hazelsgonnaheze Folgen Dem ist nichts hinzuzufügen #noafd ...

>> Auf zweitstimme.org/erststimme könnt ihr nachsehen, welche Kandidat*innen in eurem Wahlkreis auf dem Wahlzettel stehen, und wie viele Stimmen sie voraussichtlich erhalten werden. Kandidat*in, die sehr wahrscheinlich keine 5 % der Stimmen erhalten, werden auf der Website nicht einzeln aufgeführt. Diese Personen werden aber dennoch auf eurem Wahlschein am 23.2.2025 stehen.

r/politik • 8 days ago

Aromatic_Command_200

Warum eigentlich einigen sich die Parteien links der AfD nicht einfach auf einen gemeinsamen Direktkandidaten, um blaue Wahlkreise zu verhindern?

Frage

Es ist nur ein Gedankenexperiment, aber: In manchen Wahlkreisen, zB. in Leipzig, könnte die AfD mit nur 26% das Direktmandat holen, weil sich die demokratischen Parteien alle gegenseitig die Stimmen klauen, siehe: <https://zweitstimme.org/erststimme/>

Oh, and then there's this (pls don't share...)

K | MITTWOCH, 2. MAI 2012

POLITIK 13

Wie das Wetter die Politik beeinflusst

Uni untersucht Wettereffekte bei der Wahlbeteiligung: Regen kann vom Gang zur Wahlurne abschrecken

VON CLAUDIA RINDT

Konstanz – Wenn Experten und Politiker nach Erklärungen für eine besonders geringe Wahlbeteiligung suchen, dann berufen sie sich gern auf das Wahlwetter. Sie sagen dann, es sei eben zu warm oder zu kalt, zu schön oder zu schlecht gewesen, um die Bürger an die Wahlurne zu bringen.

Der Konstanzer Politik- und Verwaltungswissenschaftler Simon Munzert hat wissenschaftlich überprüft, was dran ist an der Volksthese vom wetterföhigen Wähler. Er hat dafür bei den 17 Bundestagswahlen seit 1949 das Wetter und die Wahlbeteiligung in den 5000 Wahlkreisen unter die Lupe genommen. Er kommt zu dem Schluss: das Wetter habe zwar Einfluss auf die Wahlbeteiligung, aber keinen so starken, dass ein Politiker sich darauf berufen könne, wenn er nach einer Erklärung für eine Wahlniederlage sucht.

Munzert arbeitet an der Universität Konstanz am Lehrstuhl für Umfrageforschung. Der typische Konstanzer Nebel im Herbst habe ihn auf die Idee gebracht, zu überprüfen, ob das Wetter tatsächlich Einfluss auf die Wahlbeteiligung hat. Gewählt wird in Deutschland fast immer im

Simon Munzert ist kein Wetterfrosch, sondern ein Politik- und Verwaltungswissenschaftler. Er hat herausgefunden, dass das Wetter Einfluss auf die Wahlbeteiligung hat. BILD: RINDT



Weather and Turnout in the 1949 - 2009 German Bundestag Elections

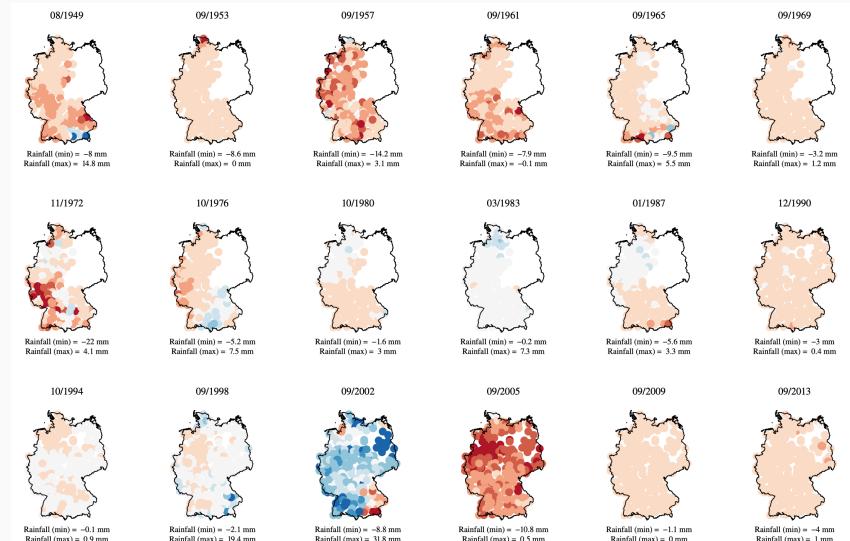
Simon Munzert

University of Konstanz
Department of Politics and Management

July 18, 2012

ABSTRACT

It is a popular belief that weather conditions are partly responsible for different levels of voter turnout. While there is some recently published evidence for and against this hypothesis for U.S. elections, studies for the German case are rare and suffer from poor data quality. I examine the effect of several weather characteristics on turnout in 17 German Bundestag elections from 1949 to 2009. By using meteorological information from many weather stations via geographical interpolation, I generate weather reports for all electoral constituencies. Bayesian hierarchical models which include district and election random effects reveal that precipitation reduces turnout by an average of 0.65% per 10mm, and temperature has a curvilinear effect, producing increasing turnout rates for daily mean temperatures below around 15 °C, and decreasing rates for temperatures above this value.



Class logistics

Course schedule

	Session	Date
1	What is data journalism?	Feb 9
2	Global models of data journalism	Feb 16
3	How people understand numbers and visuals	Feb 23
4	Collecting data: web, APIs, open sources	Mar 2
5	Dealing with numbers, samples, and surveys	Mar 9
6	Mapping and geodata	Mar 16
7	Designing effective visualizations	Mar 23
	Spring break - 2 weeks!	
8	Interactive graphics and narrative structures	Apr 13
9	Writing with data	Apr 20
10	Engaging with experts I	Apr 27
11	Engaging with experts II	May 4
12	Final presentations	May 11

Grading and assignments

Component	Weight
2(3) × data bits	25% each
1 × final project	40%
class participation	10%

Data bits

- Short data journalism pieces.
- Focus on one technique or concept (e.g., scraping, use of survey or geographic data, building an interactive visualization).
- You identify a small, newsworthy angle, carry out the necessary technical work, and present your findings in a brief written piece with at least one visualization.
- All data bits have to be submitted by Session 10 (April 27).

Grading and assignments

Component	Weight
2(3) × data bits	25% each
1 × final project	40%
class participation	10%

Final data journalism project

- Produce a full data-driven article that demonstrates your ability to conceive, investigate, analyze, and present a compelling story grounded in evidence.
- You identify a topic, gather and clean the data, conduct the analysis, create the visualizations, and craft a narrative that is accurate, accessible, and engaging.
- You will workshop your idea, receive feedback from peers and instructors, and iterate through drafts.
- Your final submission should read like a near-publishable piece of data journalism, integrating rigorous data work with clear and effective storytelling.
- Deadline: May 15.

Grading and assignments

Component	Weight
2(3) × data bits	25% each
1 × final project	40%
class participation	10%

Participation

- Participate actively.
- Be prepared to discuss the readings.
- There will be opportunities to contribute during and outside class.

Setting expectations straight

What this course is about

1. Understanding the origins, evolution, and current state of data journalism as a field of practice
2. Getting hands-on experience with data journalism techniques and tools

What the course is not about

- Learning to code (but you will have plenty of opportunities to do so)
- Mastering specific software (but you will get introduced to a variety of tools)
- Data journalism as research subject (at least not in depth, but we will touch on it)



Origins of data journalism

What is data journalism?

Share your personal definition.

Excalidraw - excalidraw.com

What is data journalism?

A working definition

"Data journalism is journalism that uses data as the basis of its reporting. Data can be used to find stories, to tell stories, and to verify stories." - [Global Investigative Journalism Network](#)

"Obtaining, reporting on, curating and publishing data in the public interest" - [Jonathan Stray, 2011](#)

Key elements

- Data acquisition and cleaning
- Statistical analysis and visualization
- Storytelling with evidence
- Transparency and reproducibility

The definitive, two-part answer to "is data journalism?"

Written by [Adrian Holovaty](#) on May 21, 2009

It's a hot topic among journalists right now: Is data journalism? Is it journalism to publish a raw database? Here, at last, is the definitive, two-part answer:

1. Who cares?
2. I hope my competitors waste their time arguing about this as long as possible.

[Adrian Holovaty](#)

A short history of data in journalism

Journalism has always been about evidence

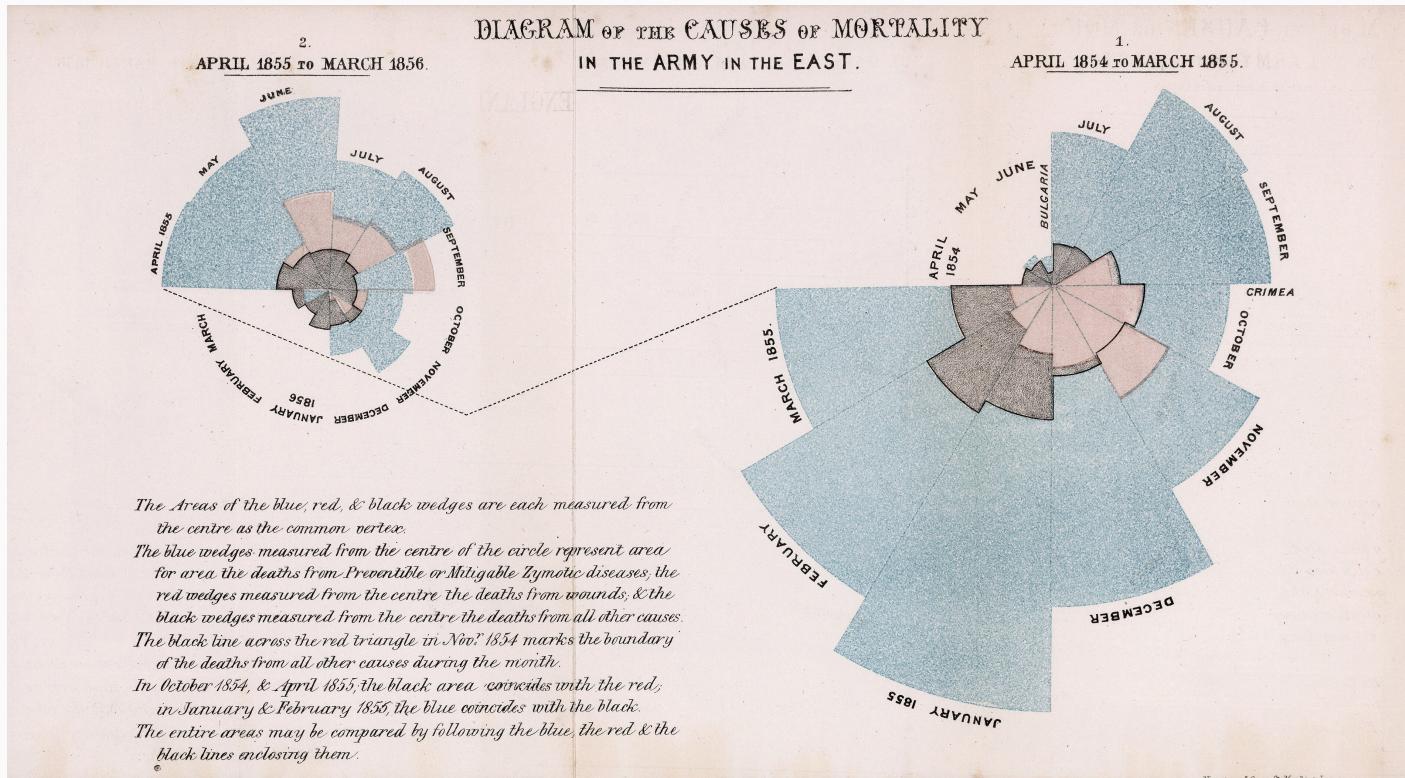
The use of data in journalism is not new. Journalists have long used numbers, records, and documents to hold power accountable.

What has changed:

- **Scale:** The volume of available data has exploded
- **Tools:** Software puts advanced analyses within reach
- **Speed:** Real-time data enables (near-)live reporting
- **Reach:** Interactive graphics engage readers

Era	Key development
1820s	First data tables in newspapers
1850s	Statistical graphics for advocacy
1960s	Social science methods in reporting
1980s	Computer-assisted reporting
2000s	Proliferation of digital data
2010s	Dedicated data teams in newsrooms
2020s	AI and automation

Nightingale's rose (1858)



Advocacy through data viz

One of the most influential graphs is by [Florence Nightingale](#), statistician and founder of modern nursing.

Nightingale's Rose, a polar area diagram, illustrates seasonal sources of soldier mortality in the field hospital Nightingale managed during the Crimean War and highlights the role epidemic diseases.

Her work revolutionized points hygiene and other practices in hospitals, ultimately saving millions of lives.

An early data journalism disaster

The Literary Digest

NEW YORK

OCTOBER 31, 1936

Topics of the day

LANDON, 1,293,669; ROOSEVELT, 972,897

Final Returns in The Digest's Poll of Ten Million Voters

Well, the great battle of the ballots in the Poll of ten million voters, scattered throughout the forty-eight States of the Union, is now finished, and in the table below we record the figures received up to the hour of going to press.

These figures are exactly as received from more than one in every five voters polled in our country—they are neither weighted, adjusted nor interpreted.

Never before in an experience covering more than a quarter of a century in taking polls have we received so many different varieties of criticism—praise from many; condemnation from many others—and yet it has been just of the same type that has come to us every time a Poll has been taken in all these years.

A telegram from a newspaper in California asks: "Is it true that Mr. Hearst has purchased THE LITERARY DIGEST?" A telephone message only the day before these lines were written: "Has the Repub-

lian National Committee purchased THE LITERARY DIGEST?" And all types and varieties, including: "Have the Jews purchased THE LITERARY DIGEST?" "Is the Pope of Rome a stockholder of THE LITERARY DIGEST?" And so it goes—all equally absurd and amusing. We could add more to this list, and yet all of these questions in recent days are but repetitions of what we have been experiencing all down the years from the very first Poll.

Problem—Now, are the figures in this Poll correct? In answer to this question we will simply refer to a telegram we sent to a young man in Massachusetts the other day in answer to his challenge to us to wager \$100,000 on the accuracy of our Poll. We wired him as follows:

"For nearly a quarter century, we have been taking Polls of the voters in the forty-eight States, and especially in Presidential years, and we have always merely mailed the ballots, counted and recorded those

returned and let the people of the Nation draw their conclusions as to our accuracy. So far, we have been right in every Poll. Will we be right in the current Poll? That, as Mrs. Roosevelt said concerning the President's reelection, is in the 'lap of the gods.'

"We never make any claims before election but we respectfully refer you to the opinion of one of the most quoted citizens to-day, the Hon. James A. Farley, Chairman of the Democratic National Committee. This is what Mr. Farley said October 14, 1932:

"Any sane person can not escape the implication of such a gigantic sampling of popular opinion as is embraced in THE LITERARY DIGEST straw vote. I consider this conclusive evidence as to the desire of the people of this country for a change in the National Government. THE LITERARY DIGEST poll is an achievement of no little magnitude. It is a Poll fairly and correctly conducted."

In studying the table of the voters from

The statistics and the material in this article are the property of Funk & Wagnalls Company and have been copyrighted by it; neither the whole nor any part thereof may be reprinted or published without the special permission of the copyright owner.

Literary Digest's 1936 election poll

Background

- The weekly magazine *Literary Digest* had correctly predicted the outcomes of all presidential elections between 1920 and 1932 using straw polls.
- Their 1936 poll of 10m voters indicated that Republican candidate Alfred Landon would be the overwhelming winner.

Aftermath

- The poll was a disaster: Landon lost in a landslide to Franklin D. Roosevelt, who carried 46 out of 48 states and won 60.8% of the popular vote.
- The outcome was correctly predicted by George Gallup with sample of 50k people.
- The magazine went bankrupt in 1938.



Literary Digest's 1936 election poll

The Literary Digest
NEW YORK NOVEMBER 14, 1936

Topics of the day

WHAT WENT WRONG WITH THE POLLS?

None of Straw Votes Got Exactly the Right Answer—Why?

In 1920, 1924, 1928 and 1932, THE LITERARY DIGEST Polls were right. Not only right in the sense that they showed the winner; they forecast the *actual popular vote* with such a small percentage of error (less than 1 per cent. in 1932) that newspapers and individuals everywhere heaped such phrases as "uncannily accurate" and "amazingly right" upon us.

Four years ago, when the Poll was running his way, our very good friend Jim Farley was saying that "no sane person could escape the implication" of a sampling "so fairly and correctly conducted."

Well, this year we used precisely the same method that had scored four bull's-eyes in four previous tries. And we were far from correct. Why? We ask that question in all sincerity, because *we want to know*.

"Reasons"—Oh, we've been flooded with "reasons." Hosts of people who feel they have learned more about polling in a few months than we have learned in more than a score of years have told us just where we were off. Hundreds of astute "second-guessers" have assured us by tele-

The following telegram was received by The Literary Digest: "With full and sympathetic appreciation of the rather tough spot you now

out of the 30,811 who voted returned ballots to us showing a division of 53.32 per cent. to 44.67 per cent. in favor of Mr. Landon. What was the actual result? It was 56.93 per cent. for Mr. Roosevelt, 41.17 per cent. for the Kansan.

In Chicago, the 100,929 voters who returned ballots to us showed a division of 48.63 per cent. to 47.56 per cent. in favor of Mr. Landon. The 1,672,175 who voted in the actual election gave the President 65.24 per cent., to 32.26 per cent. for the Republican candidate.

What happened? Why did only one in five voters in Chicago to whom THE DIGEST sent ballots take the trouble to reply? And why was there a preponderance of Republicans in the one-fifth that did reply? Your guess is as good as ours. We'll go into it a little more later. The important thing in all the above is that all this conjecture about our "not reaching certain strata" simply will not hold water.

Hoover Voters—Now for another "explanation" dimmed into our ears: "You got too many Hoover voters in your sample."

Well, the fact is that we've *always* got too big a sampling of Republican voters. That was true in 1920, in 1924, in 1928, and even in 1932, when we overestimated the Roosevelt popular vote by three-quarters of 1 per cent.

Literary Digest's 1936 election poll

Anatomy of a debacle

1. **Sampling frame:** (1) own readers, (2) registered automobile owners, (3) registered telephone users
2. **Data collection:** everyone was mailed a mock ballot and asked to return marked ballot
3. **Response rate:** 2.4m out of 10m

Selection bias as a consequence of **coverage** and
nonresponse bias: overrepresentation of wealthier individuals with a preference for Landon

Source Peverill Squire, 1988, Public Opinion Quarterly

Table 1. 1936 Presidential Vote by Car and Telephone Ownership (in Percent)

Presidential Vote	Car & Phone	Car, No Phone	Phone, No Car	Neither
Roosevelt	55	68	69	79
Landon	45	30	30	19
Other	1	2	0	2
Total N	946	447	236	657

SOURCE: American Institute of Public Opinion, 28 May 1937.

Table 2. Presidential Vote by Receiving *Literary Digest* Straw Vote Ballot or Not (in Percent)

Presidential Vote	Received Poll	Not Receive Poll	Do Not Know
Roosevelt	55	71	73
Landon	44	27	25
Other	1	1	3
Total N	780	1339	149

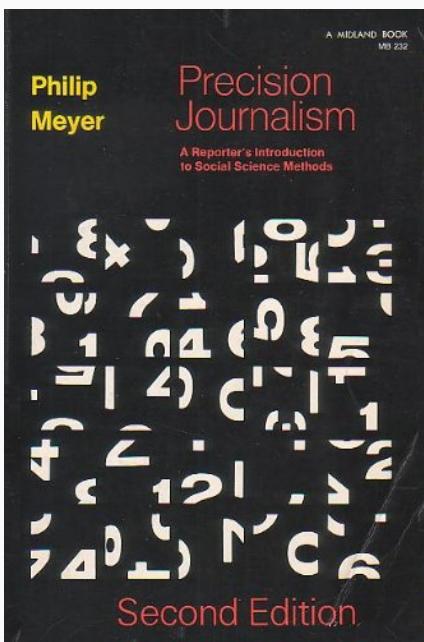
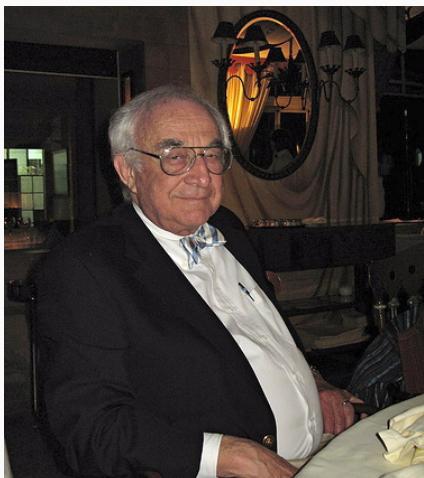
SOURCE: American Institute of Public Opinion, 28 May 1937.

Table 3. Presidential Vote by Returning or Not Returning Straw Vote Ballot (in Percent)

Presidential Vote	Did Return	Did Not Return	Do Not Know
Roosevelt	48	69	56
Landon	51	30	40
Other	1	1	4
Total N	493	288	48

SOURCE: American Institute of Public Opinion, 28 May 1937.

Precision journalism (1960s–1970s)



Philip Meyer and the Detroit riots

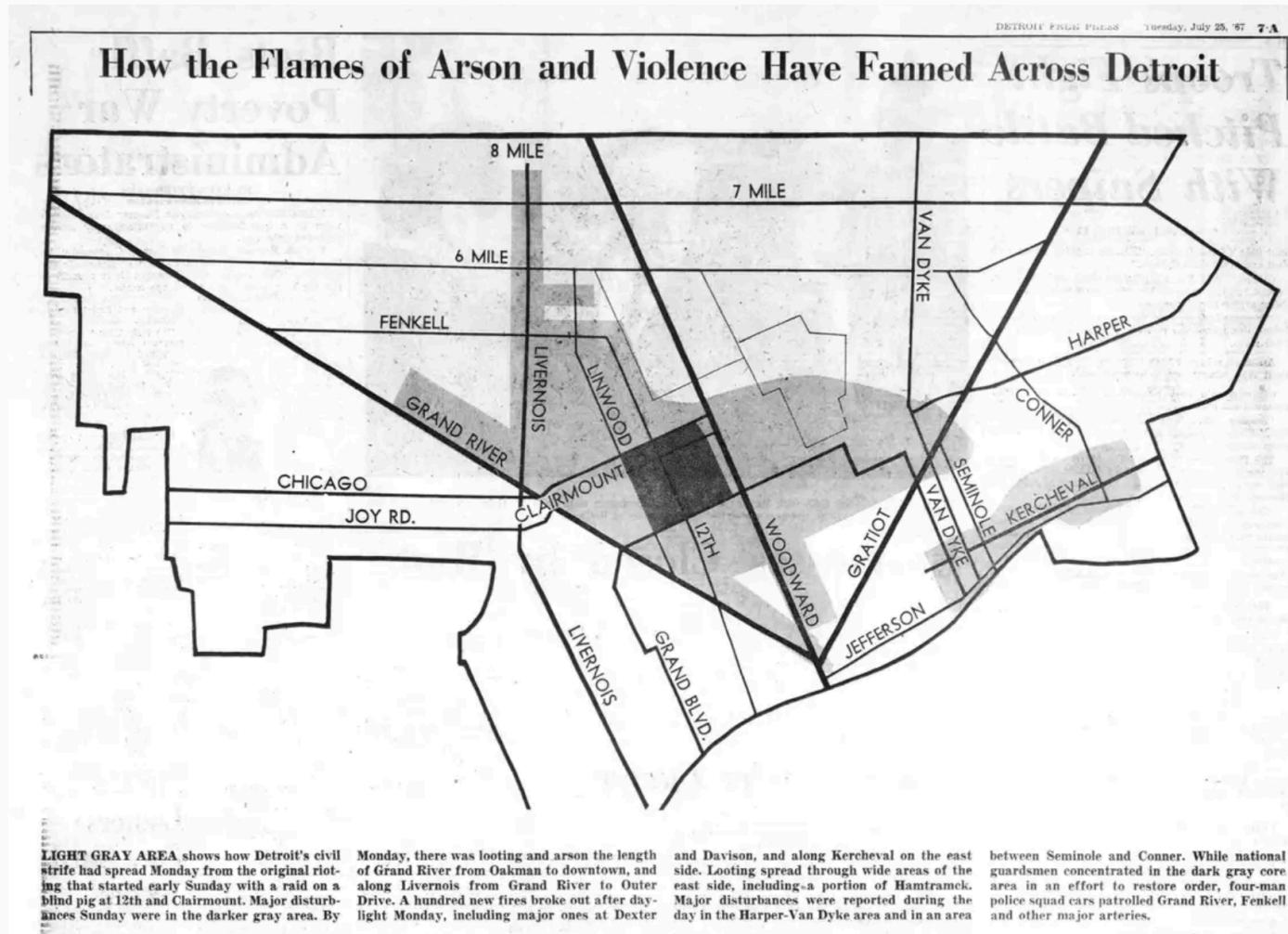
In 1967, [Philip Meyer](#) applied social science survey methods to investigate the Detroit riots for the Detroit Free Press.

Contrary to popular belief, rioters were not recent migrants from the South but long-term residents frustrated by lack of opportunity.

Meyer's book *Precision Journalism* (1973) argued that reporters should use the same rigorous methods as social scientists.

"Precision journalism is the use of social and behavioral science research methods to gather and analyze information for news stories."

Precision journalism (1960s–1970s)



Precision journalism (1960s–1970s)



Philip Meyer: Precision Journalism

Computer-Assisted Reporting (1980s–1990s)

From mainframes to newsrooms

The introduction of personal computers transformed investigative journalism.

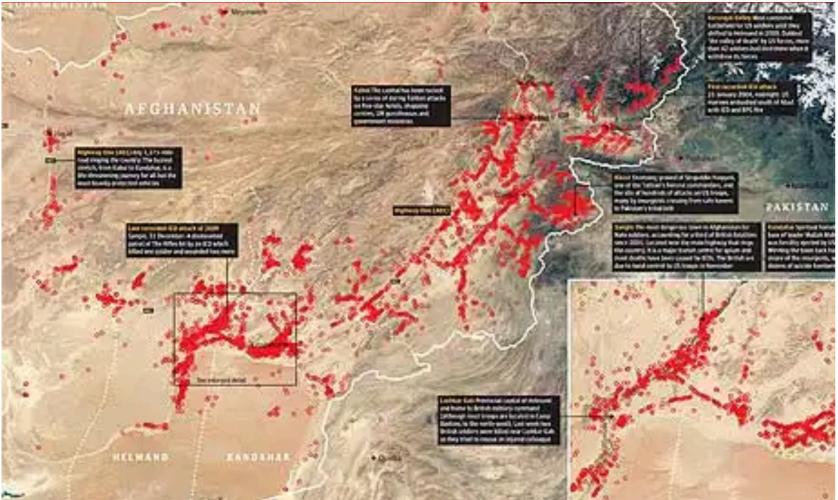
Milestones:

- 1989: [Elliot Jaspin](#) uses [database records of school bus drivers](#) to expose school bus safety violations
- 1992: The [Miami Herald](#) analyzes Hurricane Andrew damages
- 1993: Founding of [NICAR](#) (National Institute for Computer-Assisted Reporting)



The digital revolution (2000s)

Iraq: the war logs
Monday
31 January 2011



Source The Guardian, 2010

WikiLeaks and the new transparency

The 2010 release of the [Iraq War Logs](#) marked a turning point. For the first time, major news organizations collaborated to analyze massive document leaks, provided by [WikiLeaks](#).

What changed:

- Scale of data required new tools and methods
- Collaboration across newsrooms became essential
- Visualization helped make sense of complexity
- Questions of security and ethics became central

Outlets like The Guardian, New York Times, and Der Spiegel built interactive databases allowing readers to explore the data themselves.

The digital revolution (2000s)

The Panama Papers (2016)

- 11.5 million documents leaked from Mossack Fonseca
- 400 journalists from 80 countries collaborated
- Exposed offshore holdings of world leaders, celebrities, and criminals
- Led to resignations, investigations, and policy reforms worldwide

Technical innovations:

- Machine learning for document classification
- Network analysis to map company relationships
- Interactive tools for readers to explore connections

Source ICIJ, 2016

OFFSHORE LEAKS DATABASE

Find out who's behind more than 810,000 offshore companies, foundations and trusts from the [Pandora Papers](#), [Paradise Papers](#), [Bahamas Leaks](#), [Panama Papers](#) and [Offshore Leaks](#) investigations.

The screenshot shows the homepage of the Offshore Leaks Database. At the top, there is a search bar with the placeholder "Search the full Offshore Leaks database" and a red "SEARCH" button. Below the search bar, there is a text input field with the placeholder "Example: British Virgin Islands | Entity Name LLC | Lima, Peru". Underneath the search area, there is a navigation menu with links: "Explore the investigations", "Pandora Papers >", "Paradise Papers >", "Panama Papers >", "Bahamas Leaks >", and "Offshore Leaks >".

Source Offshore Leaks, ICIJ

The rise of data teams (2010s)

Dedicated data journalism units

Major news organizations established specialized teams. These teams combine journalists, developers, designers, and statisticians.

Newsroom	Team	Founded
The Guardian	Datablog	2009
New York Times	The Upshot	2014
FiveThirtyEight	—	2008/2014
ProPublica	Data Team	2008
The Economist	Data Team	~2015
ZEIT Online	Data Journalism	~2015

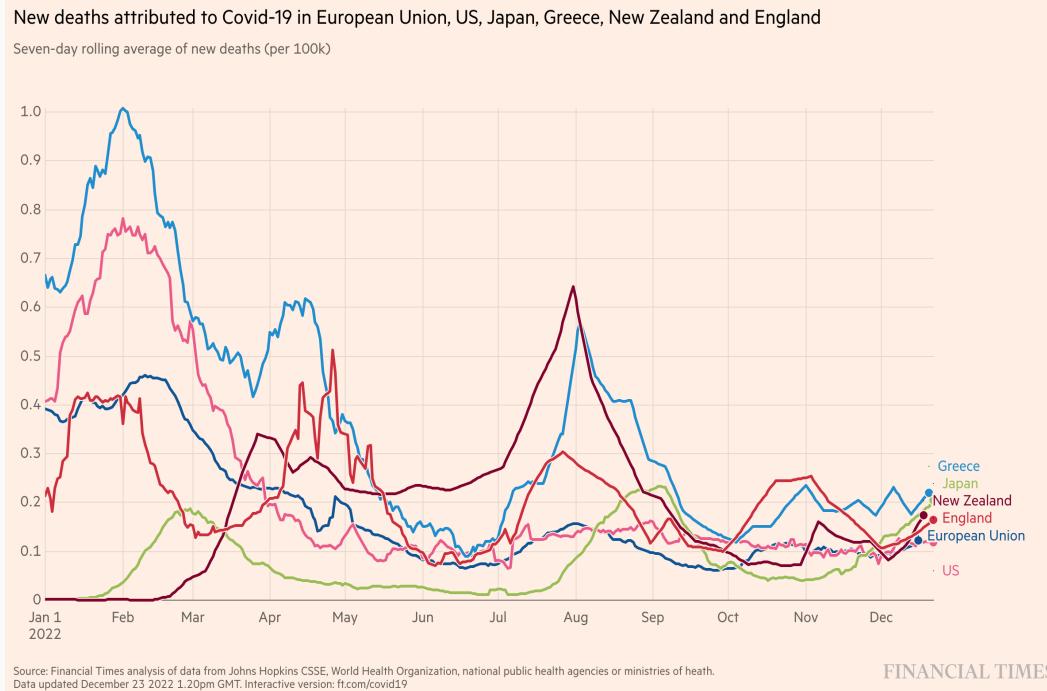


PROPUBLICA

The
Economist

Data journalism today

Data journalism today



Source [Financial Times COVID Tracker](#)

COVID-19: A stress test for data journalism

The pandemic demonstrated both the power and limitations of data journalism:

Successes:

- Real-time tracking dashboards (Johns Hopkins, FT)
- Explaining exponential growth to general audiences
- Holding governments accountable for data reporting

Challenges:

- Data quality varied wildly across countries
- Uncertainty was difficult to communicate
- Misinformation spread faster than corrections

The value proposition of data journalism

Why data matters for journalism

- **Accountability:** Data enables systematic investigation of institutions
- **Scale:** Stories impossible to tell through traditional reporting
- **Precision:** Quantitative evidence strengthens claims
- **Engagement:** Interactive graphics invite reader exploration
- **Trust:** Transparent methods build credibility

The fundamental tension

Data journalism must balance:

- Accuracy vs. accessibility
- Complexity vs. clarity
- Speed vs. verification
- Engagement vs. manipulation

Data journalism as research subject

An emerging field of study

Data journalism has become an object of scholarly inquiry in **communication** and **journalism studies**.

Research questions:

- How do data teams operate within newsrooms?
- What skills define the data journalist?
- How does data journalism differ across countries?
- What ethical challenges are unique to data work?

Key outlets:

- *Digital Journalism*
- *Journalism Studies*
- *Journalism Practice*
- *Journalism*



Effects of data journalism

Does data journalism work?

Researchers study how audiences respond to data-driven content.

Potential benefits:

- Increased credibility and trust
- Better comprehension of complex issues
- Reduced misperceptions
- Higher engagement

Potential risks:

- Overconfidence in uncertain forecasts
- Demobilization effects from probabilistic reporting
- Misinterpretation of statistics

Source (right) Westwood, Messing & Lelkes (2020)

Projecting Confidence: How the Probabilistic Horse Race Confuses and Demobilizes the Public

Sean Jeremy Westwood, Dartmouth College

Solomon Messing, Acronym

Yphtach Lelkes, University of Pennsylvania

Recent years have seen a dramatic change in horse-race coverage of elections in the United States—shifting focus from late-breaking poll numbers to sophisticated meta-analytic forecasts that emphasize candidates' chance of victory. Could this shift in the political information environment affect election outcomes? We use experiments to show that forecasting increases certainty about an election's outcome, confuses many, and decreases turnout. Furthermore, we show that election forecasting has become prominent in the media, particularly in outlets with liberal audiences, and show that such coverage tends to more strongly affect the candidate who is ahead—raising questions about whether they contributed to Trump's victory over Clinton in 2016. We bring empirical evidence to this question, using American National Election Studies data to show that Democrats and Independents expressed unusual confidence in a decisive 2016 election outcome—and that the same measure of confidence is associated with lower reported turnout.

I don't know how we'll ever calculate how many people thought it was in the bag, because the percentages kept being thrown at people—"Oh, she has an 88% chance to win!"
—Hillary Clinton quoted in Traister (2017)

Political information about electoral competition is central to the study of political behavior. It can alter the strategic calculus used to decide whether or not to show up to the polls (e.g., Ansolabehere and Iyengar 1994; Delli Carpini 1984; Mutz 1998); after all, why should a voter take hours off work and arrange a trip to their polling place if they are certain one side will win or lose? Horse-race coverage may play an outsized role in this calculus as it is widely available and dominates coverage of substantive issues in American elections (Iyengar, Norpoth, and Hahn 2004; Patterson 2016).

Yet as we show, the dynamics between horse-race coverage and voter behavior are shifting because of a form of horse-race coverage that has emerged in recent elections: the probabilistic forecast. In contrast to traditional horse-race coverage that often focuses on unusual polls (Searles, Ginn, and Nickens 2016) or speculates about a candidate's "paths to victory" (Silver 2017) these forecasts aggregate polling data

into a concise probability of winning, providing far more conclusive information about the state of a race.

In this article, we show that probabilistic forecasts have fundamentally altered the political information environment, because they are (1) widely available in the media, (2) lead voters to different assessments of electoral competition and whether their vote matters (pivotality) compared to traditional vote share estimates, and (3) affect potential supporters of one political party more than another. We first show that probabilistic forecasts are highly salient in the mainstream media and provide evidence of their importance by documenting downstream effects on markets. We also show that they are more prominent in media outlets with left-leaning audiences. Using a survey experiment, we show that not only do these forecasts confuse some potential voters, they also lower perceptions that an election is competitive. Finally, we present an original behavioral game that simulates elections, which shows that probabilistic forecasts reduce voting

Sean Jeremy Westwood (sean.j.westwood@dartmouth.edu) is an assistant professor at Dartmouth College, Hanover, NH 03750. Solomon Messing (solomon.messing@gmail.com) is chief scientist at Acronym, Washington, DC 20036. Yphtach Lelkes (yphtach.lelkes@asc.upenn.edu) is an assistant professor at the University of Pennsylvania, Philadelphia, PA 19104. Authors are listed in reverse alphabetical order.

Experimental protocols were reviewed by an institutional review board. No external funding supported this work. Data and supporting materials necessary to reproduce the numerical results in the article are available in the JOP Dataverse (<https://dataverse.harvard.edu/dataverse/jop>). An online appendix with supplementary material is available at <https://doi.org/10.1086/708682>.

The Journal of Politics, volume 82, number 4. Published online August 5, 2020. <https://doi.org/10.1086/708682>
© 2020 by the Southern Political Science Association. All rights reserved. 0022-3816/2020/8204-0024\$10.00

Discussion

Class discussion: Shaping our semester together

Data journalism is learned by doing. Let's discuss how we can make this semester more **hands-on** and **collaborative**.

Possible activities

- Build a **class blog** – publish our work and advertise it
- Develop a **common brand** – logo, style guide, identity
- Co-run the **talk series** – moderate sessions
- Organize a **field visit** – tour a newsroom data team
- **Weekly critique** – analyze one DJ piece together

Discussion questions

1. Which activities interest you most?
2. What skills would you like to develop this semester?
3. Are there other ideas we should consider?

Coming up next week

Global Models of Data Journalism

1. Read Yuliya Ilyuk's *Journalistic investigations in the digital age of post-truth politics* to make yourself familiar with the MH17 case and the rise of Bellingcat.
2. From Mutsvairo et al.'s "*Data journalism in the Global South*", choose two chapters and be prepared to discuss them.
3. Engage with a data journalism project (to be shared).