

# Political Analysis of Social Media Data

## Introduction

Instructor: Gregory Eady  
Office: 18.2.10  
Office hours: Fridays 13-15

# Today

1. About the course
2. The big questions in the study of social media
3. Trends in social media use
4. From “Liberation” to “Turmoil”
5. Class structure
6. Introduction to R

## About my own research

- Social media
  - Gender bias online
  - Political violence
  - Foreign interference campaigns
  - COVID & interest group behavior
  - Sharing of ideologically extreme political content
  - Echo chambers
- Quantitative methods
  - Measurement issues in social media
  - Sensitive beliefs

## About you

A small exercise...

[https://padlet.com/gregory\\_eady/SocialMedia](https://padlet.com/gregory_eady/SocialMedia)

### Create a post about the following:

1. Title: Your first and last name
2. Background in quantitative methods
3. Proficiency with statistical software (e.g. R, Stata, Python, etc.)
4. Your interests in the politics of social media
5. What you want to get out of this class

## Course website

<https://gregoryeady.com/SocialMediaDataCourse>

We won't use Absalon at all in this course:

- It's annoying
- Difficult to post videos and exercises together in a coherent way

## The aim of the course

1. Broad overview of the major questions in the study of social media
2. Detail the research designs used in current research
3. Provide you with the ability to reflect on the drawbacks and advantages of current research practices
4. Provide the technical tools to collect and analyze social media data

## 3-day take-home exam

1. Will provide you with new data
2. You will need to clean and re-structure those data
3. Then you will analyze them and interpret the results

## What are the most pressing issues concerning the politics of social media?



## What do you think would be the major empirical questions in the politics of social media?

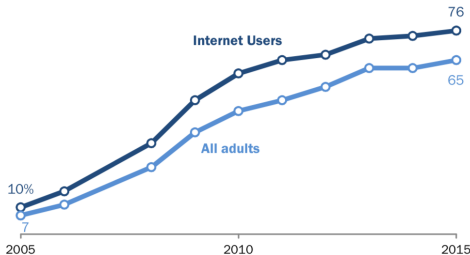
## A snapshot of social media

Site	Founding
4chan	2003
<b>Facebook</b>	<b>2004</b>
Reddit	2005
<b>YouTube</b>	<b>2005</b>
<b>Twitter</b>	<b>2006</b>
WhatsApp	2006
Tumblr	2007
Instagram	2010
8chan	2013
TikTok	2016
Gab	2017
Parler	2018

# A snapshot of social media

## Social Networking Use Has Shot Up in Past Decade

*% of all American adults and internet-using adults who use at least one social networking site*

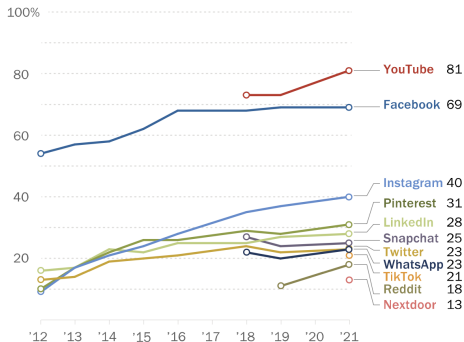


Source: Pew Research Center surveys, 2005-2006, 2008-2015. No data are available for 2007.

PEW RESEARCH CENTER

## A snapshot of social media

% of U.S. adults who say they ever use ...



Note: Respondents who did not give an answer are not shown. Pre-2018 telephone poll data is not available for YouTube, Snapchat and WhatsApp; pre-2019 telephone poll data is not available for Reddit. Pre-2021 telephone poll data is not available for TikTok. Trend data is not available for Nextdoor.

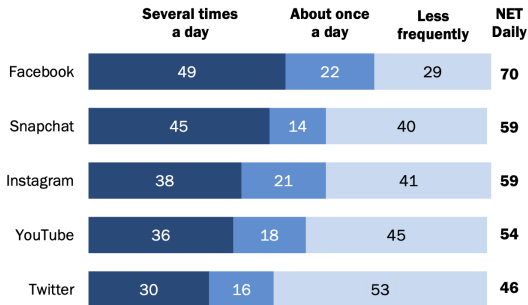
Source: Survey of U.S. adults conducted Jan. 25-Feb. 8, 2021.  
"Social Media Use in 2021"

PEW RESEARCH CENTER

## A snapshot of social media

### Seven-in-ten Facebook users say they visit site daily

Among U.S. adults who say they use \_\_\_, % who use that site ...



Note: Respondents who did not give an answer are not shown. "Less frequently" category includes users who visit these sites a few times a week, every few weeks or less often.

Source: Survey of U.S. adults conducted Jan. 25-Feb. 8, 2021.

"Social Media Use in 2021"

PEW RESEARCH CENTER

# A snapshot of social media

## Substantial 'reciprocity' across major social media platforms

% of \_\_ users who also ...

	Use Twitter	Use Instagram	Use Facebook	Use Snapchat	Use YouTube	Use WhatsApp	Use Pinterest	Use LinkedIn
Twitter	--	73%	90%	54%	95%	35%	49%	50%
Instagram	50	--	91	60	95	35	47	41
Facebook	32	47	--	35	87	27	37	33
Snapchat	48	77	89	--	95	33	44	37
YouTube	31	45	81	35	--	28	36	32
WhatsApp	38	55	85	40	92	--	33	40
Pinterest	41	56	89	41	92	25	--	42
LinkedIn	47	57	90	40	94	35	49	--

Source: Survey conducted Jan. 3-10, 2018.  
"Social Media Use in 2018"

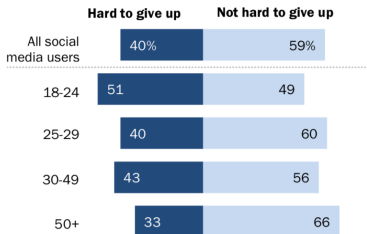
PEW RESEARCH CENTER

90% of LinkedIn users  
also use Facebook

## A snapshot of social media

### Majority of users say it would *not* be hard to give up social media

*Among U.S. social media users, the % of who say it would be \_\_\_ to give up social media*



Note: Respondents who did not give answer are not shown. "Hard to give up" include those saying it would be very or somewhat hard. "Not hard to give up" include those saying it would be not too hard or not hard at all.

Source: Survey conducted Jan. 3-10, 2018.

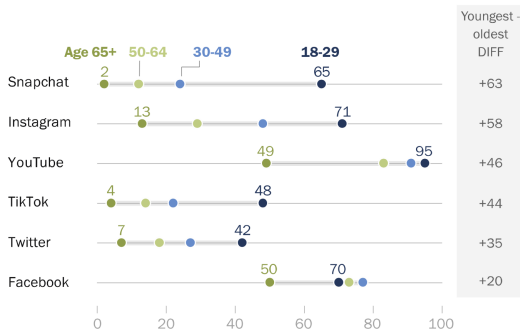
"Social Media Use in 2018"

PEW RESEARCH CENTER

## A snapshot of social media

### Age gaps in Snapchat, Instagram use are particularly wide, less so for Facebook

% of U.S. adults in each age group who say they ever use ...



Note: All differences shown in DIFF column are statistically significant. The DIFF values shown are based on subtracting the rounded values in the chart. Respondents who did not give an answer are not shown.

Source: Survey of U.S. adults conducted Jan. 25-Feb. 8, 2021.

"Social Media Use in 2021"

PEW RESEARCH CENTER



## From Liberation to Turmoil

Time Magazine on Mark Zuckerberg in 2010...

“[Mark Zuckerberg has turned] the lonely, antisocial world of random chance into a friendly world, a serendipitous world.”

## From Liberation to Turmoil

Social media was used to organize protests during the Arab Spring in early 2010s:

“We use Facebook to schedule the protests and [we use] Twitter to coordinate, and YouTube to tell the world.”

– Arab Spring activist (Mic, 2012)

## From Liberation to Turmoil

Yet social media as an information technology is double-edged.

It empowers various individuals, groups, and organizations:

- The disenfranchised and voiceless
- Anti-system forces (illiberal non-state individuals)
- Authoritarian regimes

## From Liberation to Turmoil

### Authoritarian regimes:

- Source of information to target dissidents
- Internet blackouts & slowdowns
- Selective censorship
- Biased algorithms
- Disinformation
- Flooding

## From Liberation to Turmoil

Solving the collective action problem is easier... but who benefits?

- Occupy Wall Street
- The Tea Party
- Black Lives Matters
- #MeToo

But also anti-system forces:

- White supremacists
- January 6 rioters
- Islamic State

## Large array of topics to research

- Social media censorship
- Role in protest behavior and organization
- Radicalization & rise of populism
- Racial, gender, and religious discrimination
- Echo chambers
- Polarization
- Misinformation/disinformation
- Public opinion
- Issue and voter mobilization
- Advertising
- Privacy
- Algorithms

## Social media data are useful for two broad reasons:

### 1. For measurement

- Might not care about social media *per se*, but allows us to measure attitudes and behaviors more easily and with high frequency

### 2. Social media is important in itself

## Social media for measurement purposes:

### Can Exposure to Celebrities Reduce Prejudice? The Effect of Mohamed Salah on Islamophobic Behaviors and Attitudes

ALA' ALRABABA'H *Stanford University*

WILLIAM MARBLE *Stanford University*

SALMA MOUSA *Yale University*

ALEXANDRA A. SIEGEL *University of Colorado Boulder*

**C**an exposure to celebrities from stigmatized groups reduce prejudice? To address this question, we study the case of Mohamed Salah, a visibly Muslim, elite soccer player. Using data on hate crime reports throughout England and 15 million tweets from British soccer fans, we find that after Salah joined Liverpool F.C., hate crimes in the Liverpool area dropped by 16% compared with a synthetic control, and Liverpool F.C. fans halved their rates of posting anti-Muslim tweets relative to fans of other top-flight clubs. An original survey experiment suggests that the salience of Salah's Muslim identity enabled positive feelings toward Salah to generalize to Muslims more broadly. Our findings provide support for the parasocial contact hypothesis—indicating that positive exposure to out-group celebrities can spark real-world behavioral changes in prejudice.



## Social media because it is important in itself:

### How Censorship in China Allows Government Criticism but Silences Collective Expression

GARY KING *Harvard University*

JENNIFER PAN *Harvard University*

MARGARET E. ROBERTS *Harvard University*

**W**e offer the first large scale, multiple source analysis of the outcome of what may be the most extensive effort to selectively censor human expression ever implemented. To do this, we have devised a system to locate, download, and analyze the content of millions of social media posts originating from nearly 1,400 different social media services all over China before the Chinese government is able to find, evaluate, and censor (i.e., remove from the Internet) the subset they deem objectionable. Using modern computer-assisted text analytic methods that we adapt to and validate in the Chinese language, we compare the substantive content of posts censored to those not censored over time in each of 85 topic areas. Contrary to previous understandings, posts with negative, even vitriolic, criticism of the state, its leaders, and its policies are not more likely to be censored. Instead, we show that the censorship program is aimed at curtailing collective action by silencing comments that represent, reinforce, or spur social mobilization, regardless of content. Censorship is oriented toward attempting to forestall collective activities that are occurring now or may occur in the future—and, as such, seem to clearly expose government intent.

## Structure of the course

- Two hour lectures
  - Also (very short) student presentations
- Video-recorded coding exercises

## Student presentations (start a month from now)

- Select an article that uses social media data
- Sign up for a time to present
- Create a slide deck for your presentation
  - Aim for a 10 minute presentation
- Add slides to your appendix for the question and answer

## Students without a presentation in a given week

- Skim *one* of the papers being presented
- Prepare one of the following:
  - A critical question about the paper
  - A criticism of the paper, to which the presenter(s) can respond
  - A clarifying question (e.g. “I don’t think I understand Figure 1. I think it shows [x], but shouldn’t it show [y]?”)
- Focus comments on the research design.
  - Comments about theory/implications are acceptable also

## This course is taught in R, but are many statistical programming languages

1. SPSS (social sciences)
2. Stata (economics)
3. Matlab (statistics, CS, natural sciences)
4. C++ (various)
5. **Python**
6. **R**

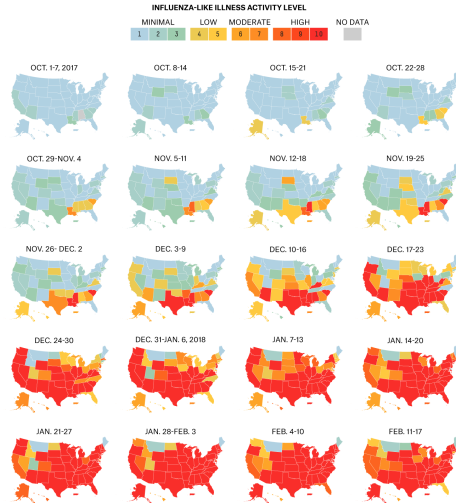
## Why R?

1. Extremely flexible
2. Workflow is easily replicable
3. Free
4. Massive R online community
5. Open-source libraries
6. Cutting-edge methods implemented first in R
7. Fantastic for data visualization

## About R



## About R



FiveThirtyEight

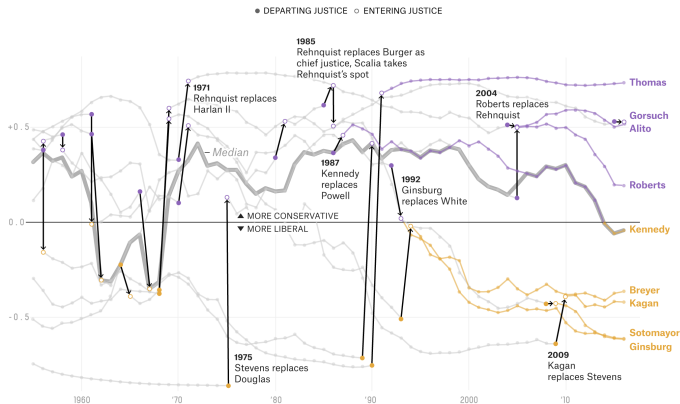
SOURCE: CENTERS FOR DISEASE CONTROL AND PREVENTION INFLUENZA DIVISION



# About R

## Supreme Court departures and replacements

Supreme Court justices from 1955-2016 by their ideological leanings, based on their Judicial Common Space scores



Years refer to Supreme Court terms, which run from October to September.

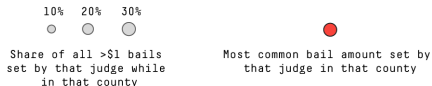
FiveThirtyEight

SOURCES: LEE EPSTEIN, CHAD WESTERLAND, THE JOURNAL OF LAW, ECONOMICS, AND ORGANIZATION

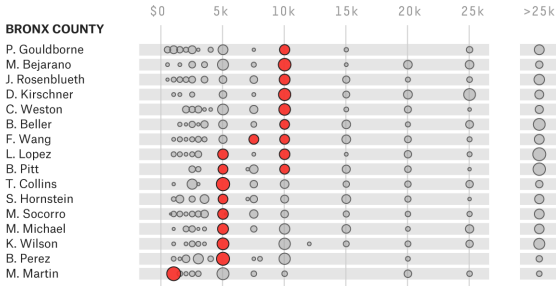
## About R

### Judges tend to set bail amounts in patterns

Bail amounts over \$1 set in felony arraignments handled by The Legal Aid Society in New York City in 2017, categorized by county and judge



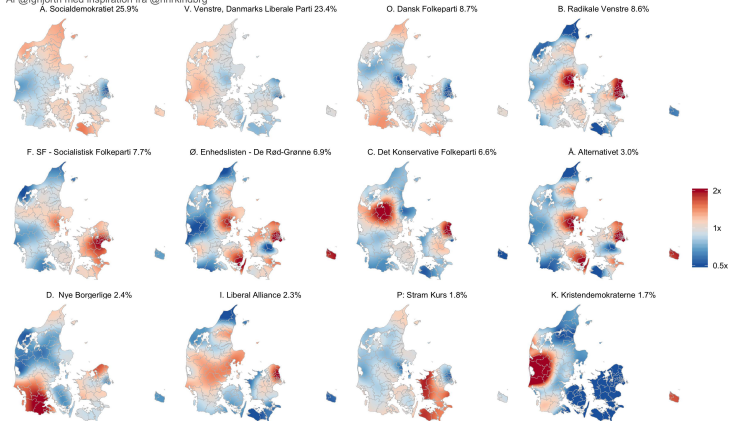
### Bail amounts by county and judge



# About R

## Partiernes Danmarkskort ved Folketingsvalget 2019

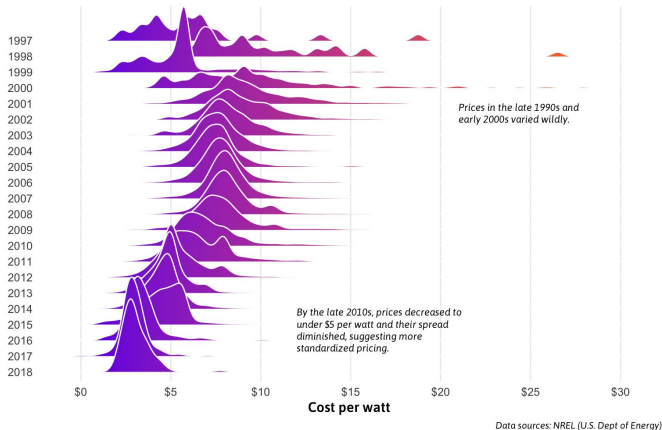
Af @fghjorth med inspiration fra @hnrkIndbrg



# About R

## Solar panels have become steadily cheaper

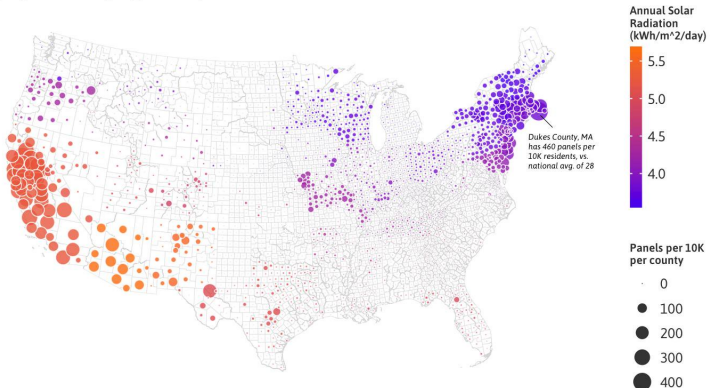
When accounting for a solar panel's generative capacity (using cost-per-watt), solar panel costs have decreased and become more consistent over the last 20 years.



## About R

### Large panels-per-capita in the Northeast, despite low annual solar radiation

While the Northeast region receives the least solar irradiance in the US, it also has counties with disproportionately high solar panel installation rates.



Data source: NREL (U.S. Dept of Energy)

Note: HI and AK are omitted due to lack of comparable solar resource data available.

## About R

R does have some drawbacks:

- No GUI (but are IDEs)
- Cryptic error messages
  - “Error in xxx: object of type ‘closure’ is not subsettable”
- Some people have strong preferences for R versus Python
  - These are actually complements
  - You should try to learn both

## The coming weeks

### R, Data, and Experiments

---

- Week 2 **R Basics II & Visualization**
  - Week 3 **Overview of social media research & Representativeness**
  - Week 4 **Data collection & Experiments**
  - Week 5 **Text Analysis I: Basics & Presentations begin**
-

## The coming weeks

### Data Analysis I

---

Week 7	<b>Echo Chambers</b>
Week 8	<b>Ideology</b>
Week 9	<b>Mid-course exercise on data collection</b>
Week 10	<b>Text Analysis II: Topic models</b>

---



## About the course

### Data Analysis II

---

Week 11	<b>Text Analysis III: Sentiment analysis</b>
Week 12	<b>Text Analysis IV: Supervised Learning</b>
Week 13	<b>Keyword expansion</b>
Week 14	<b>Event study models</b>

---

## Final note

Walling off of the Twitter API makes gaining hands-on experience with social media data difficult.

- Focus more on principles of quantitative analysis
- May teach the broader set of tools for analyzing quantitative data
- May not only use social media data