

# Data and Society

## Election Models – Lecture 11

3/8/21

# Today (3/8/21)

- **Op-ed due 11:59 p.m. on March 15,**  
instructions in 3/1 Lecture
- **No class March 11 – Writing Day**
- Today:
  - Lecture / Discussion – Data and Elections 2
  - Presentations

# Speaker / Reading March 15

- Guest Speaker:  
**Alyssa Goodman,**  
Harvard
- Read  
<http://www.sci-news.com/astronomy/radcliffe-wave-07995.html>

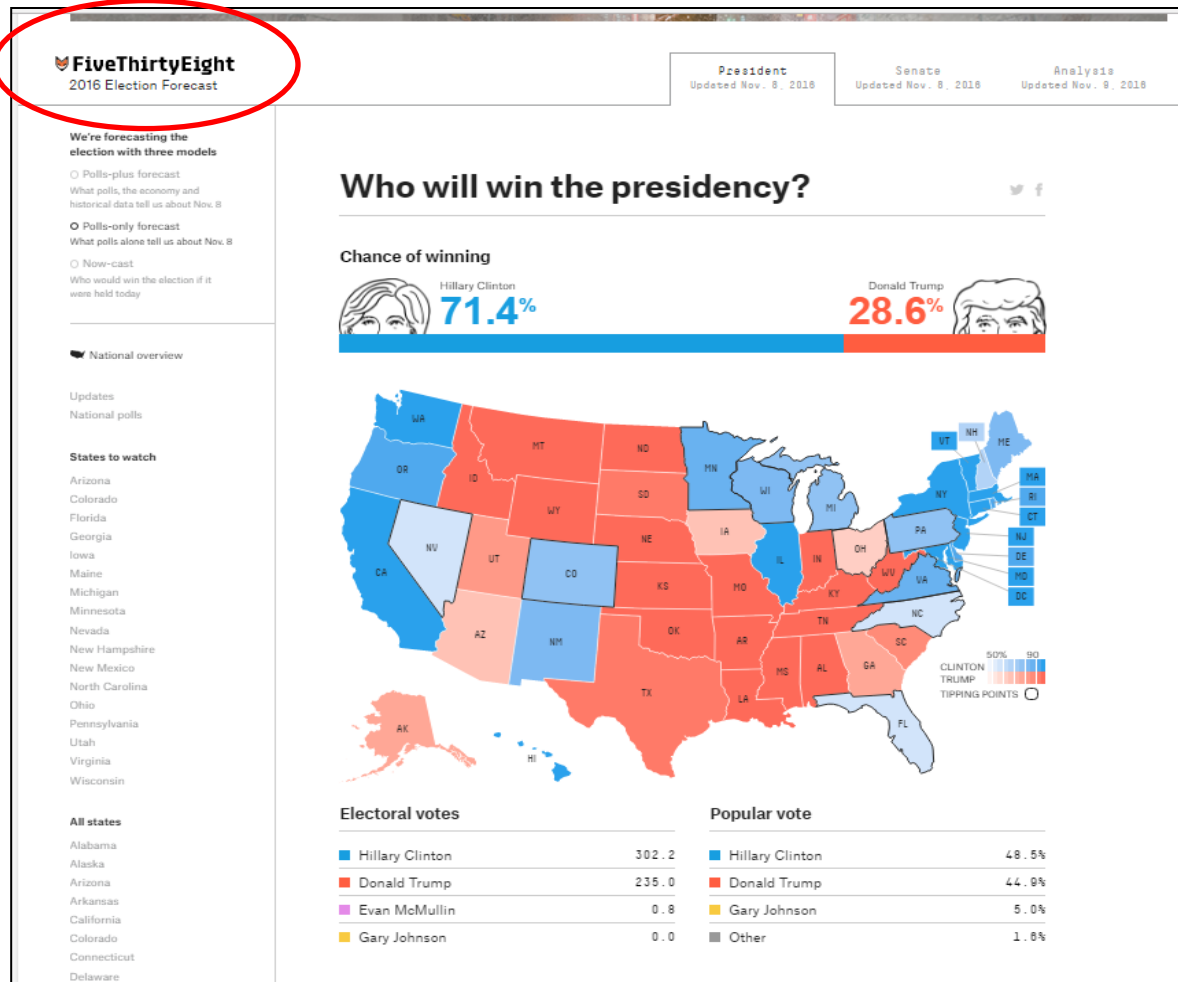


Date	Topic	Speaker	Date	Topic	Speaker
1-25	Introduction	Fran	1-28	The Data-driven World	Fran
2-1	Data and COVID-19	Fran	2-4	Data and Privacy -- Intro	Fran
2-8	Data and Privacy – Differential Privacy	Fran	2-11	Data and Privacy – Anonymity / Briefing Instructions	Fran
2-15	NO CLASS / PRESIDENT’S DAY		2-18	NO CLASS	
2-22	Legal Protections	Ben Wizner	2-25	Data and Discrimination 1	Fran
3-1	Data and Discrimination 2	Fran	3-4	Data and Elections 1	Fran
3-8	Data and Elections 2	Fran	3-11	NO CLASS / WRITING DAY	
3-15	Data and Astronomy (Op-Ed due)	Alyssa Goodman	3-18	Data Science	Fran
3-22	Digital Humanities	Brett Bobley	3-25	Data Stewardship and Preservation	Fran
3-29	Data and the IoT	Fran	4-1	Data and Smart Farms	Rich Wolski
4-5	Data and Self-Driving Cars	Fran	4-8	Data and Ethics 1	Fran
4-12	Data and Ethics 2	Fran	4-15	Cybersecurity	Bruce Schneier
4-19	Data and Dating	Fran	4-22	Data and Social Media	Fran
4-26	Tech in the News	Fran	4-29	Wrap-up / Discussion	Fran
5-3	NO CLASS				

# Lecture and Discussion

- Election and Polling Models and Methods: 2016, 2020
- Campaigns and Data

# Why are election models inaccurate?

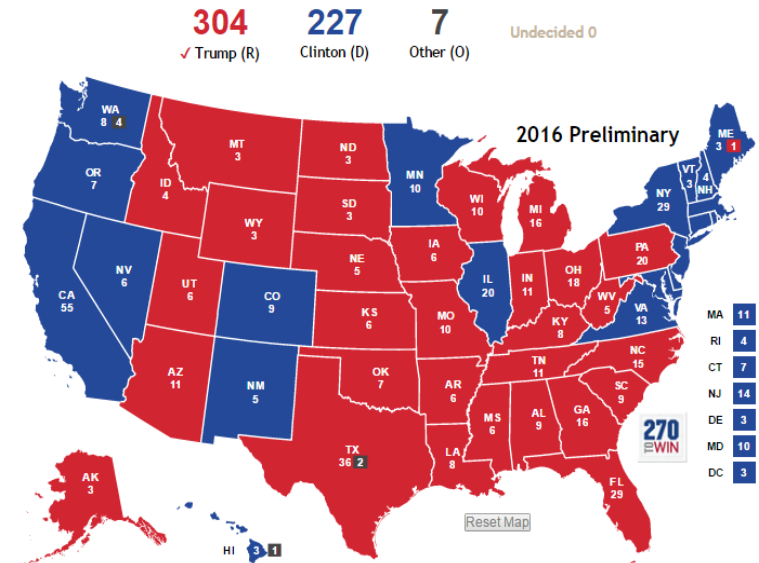


# Many possibilities for inaccuracy

- Models may not be representative of current election
- Interpretation of results may be incorrect
- Data may be faulty (low integrity, biased, not representative, etc.)
- Sampling / polling methods may not reflect voting population
- Results may be one of the low probability outcomes

## 2016 Presidential Election Interactive Map

270 electoral votes to win



For purposes of this temporary map, 'Other' is essentially a catch-all for any electoral votes not going to Clinton, Trump or Undecided.

To Set Other (Faithless) Electors: 1) Add the state, via the drop down box, to the table below. 2) Reallocate the state's electoral vote. The Total must match the state's actual votes before you can click 'Update'.

Note that any state in the table will not be interactive on the map. Removing the state will set it to 'Undecided' and make it interactive.

Map from  
[http://www.270towin.com/2016\\_Election/interactive\\_map](http://www.270towin.com/2016_Election/interactive_map)

# How people voted: Exit Polls and Election Results

<https://www.nytimes.com/interactive/2020/11/03/us/elections/exit-polls-president.html>

<b>Election</b>	<b>Voting Age population (VAP)</b>	<b>% Turnout of VAP</b>
2008	229,945,000	57.1%
2012	235,248,000	53.8%
2016	249,422,000	54.8%
2020	257,605,088	62.0%



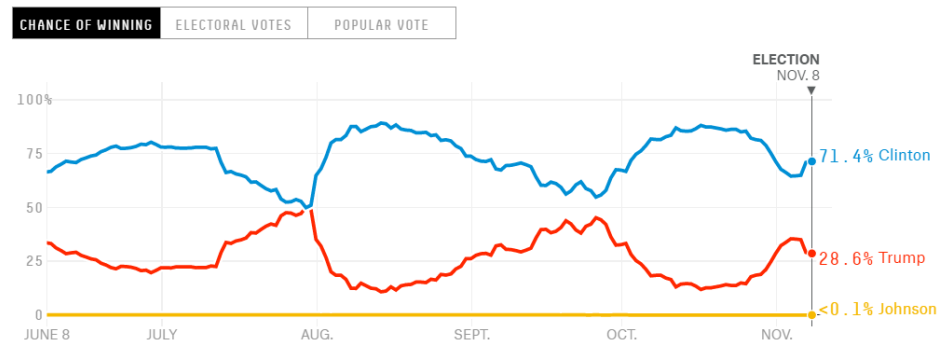
# Election models: Model and Interpretation Accuracy

## Many challenges in modeling and interpretation:

- Raw polling data supplemented by **estimates on how many people will vote and what undecided voters will do**
- **Historical inferences about past patterns** of turnout, demographics, economic conditions and party loyalty may not be accurate for present day
- If polls shows that candidate “wins” by a small margin within the **margin of error**, it is risky to interpret this as a “win”

### How the forecast has changed

We'll be updating our forecasts every time new data is available, every day through Nov. 8.

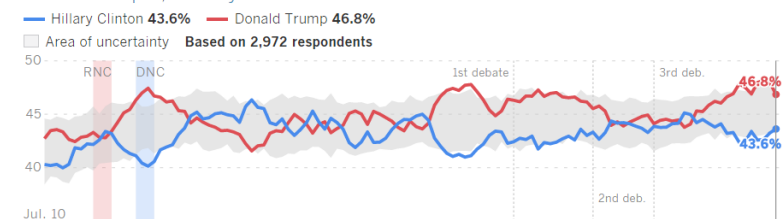


From: <https://projects.fivethirtyeight.com/2016-election-forecast/>

### The 2016 USC Dornsife/LA Times Election Poll

As of Nov. 08, 2016. Results lag events by several days.

[More from the poll, and why it differs from others.](#)



Note: Shaded gray area indicated the race was too close to call.

Sources: USC Dornsife/LA Times Presidential Election Daybreak Poll

From: <http://www.latimes.com/politics/la-na-pol-usc-latimes-poll-20161108-story.html>

# Data Integrity – Was 2016 poll data accurate?

- Many **suspected that people lied** about voting for Trump
- Trafalgar Group's approach to improving data accuracy -- Adjust numbers to account for people's hesitance to admit a Trump vote
  - Used robotic calls for which Trump voters seemed more comfortable
  - Added a “neighbor” question -- Who do you think your neighbors will vote for? – and checked to see if the numbers were different
  - Created a demographic of people who had not voted in 6+ years but planned to vote for Trump
- Trafalgar predicted Trump win in Pennsylvania and Michigan (but not all states)

# Sampling Accuracy

## Key sampling questions

- How representative is the sample of population?
- How big is the sample / what is the margin of error?
- How biased are the sampling vehicles – land lines, interviews, tweets, etc.?
- How representative is the sample of turnout? For eligible voters? For eligible voters who actually vote?
- How accurate is the data (are people lying)?

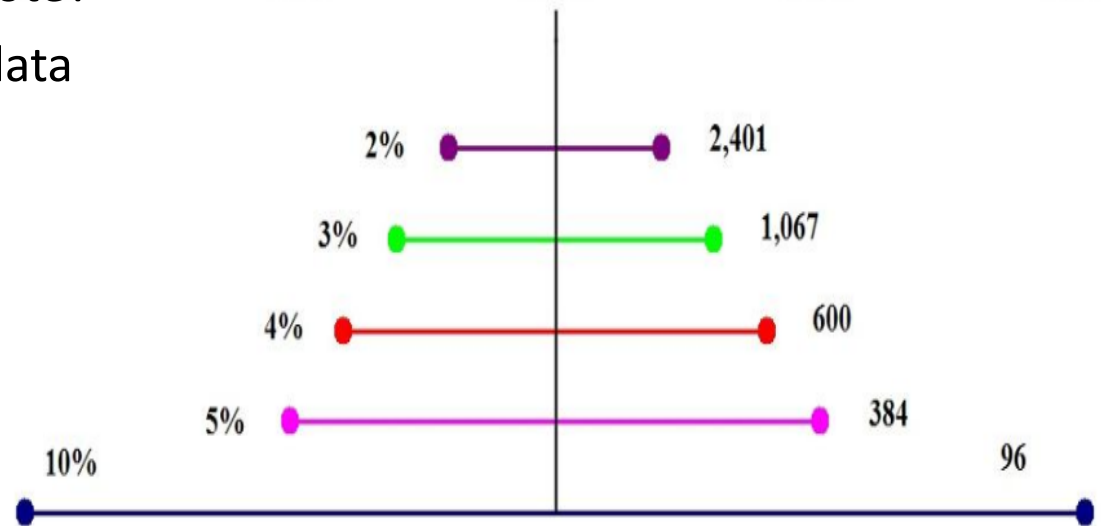


Figure from  
<http://www.forbes.com/sites/startswithabang/2016/11/09/the-science-of-error-how-polling-botched-the-2016-election/#75748a437da8>

*A visualization of how your statistical uncertainty drops as your sample size increases. Image credit: Fadethree at English Wikipedia.*

# Campaigns and data – predicting what will happen and what can be influenced (Nickerson and Rogers)

- Campaigns must perform cost-benefit analysis on all advertising, outreach, messaging, etc. for maximum effectiveness.
- Data from polls and surveys guide campaign strategies and expenditures.
- **Good data and good data analysis is a competitive advantage – can make the difference between winning and losing.**

# Campaign approaches

- **The “old days”**: citizen support gauged by party and the performance of their precincts, history of turnout; donors re-contacted; prior volunteer captains re-contacted; mail and TV advertising
- **Now**: Sophisticated data-typing, sampling techniques, greater use of statistical models, differentiated strategies; on-line outreach
- “... in a close political context, **data-driven campaigning** can have enough effect to make the difference between winning and losing.”

# Campaign data

- Accurate contact information needed on citizens, volunteers, donors
- Useful citizen data:
  - Whether they have donated (Fed. Election Commission requires disclosure of those giving \$200 or more)
  - Whether they have volunteered
  - Whether they have attended rallies
  - Whether they have signed petitions
  - Whether they have expressed support for candidates or particular issues
- **Predictive scores** include behavior scores, support scores, responsiveness scores

# Predictive scores

- **Behavior scores** – probability that citizens will turn out, donate, volunteer, and other forms of political activity
- **Support scores** – probability of support for candidate/issue based on statistical analysis (not cost-effective to poll everyone)
- **Responsiveness scores** – probability that a citizen will respond to campaign outreach strategy. Generally evaluated from fully randomized field experiments, used to detect and model heterogeneous treatment effects, and guide targeting decisions.

# Where data comes from

- **State voter files** (DOB, gender, electoral participation)
- **Census data** (average household income, children per household, ethnic distribution)
- **Precinct data**
- **Consumer data** [may be purchased] (current phone, contact info, years of education, mortgage information, home ownership status, etc.)
- **Data provided by the individual** (Donor and volunteer information, responses to surveys or telephone calls, on-line activities, responsiveness to campaign emails, etc.)
- **Exacerbates inequality in campaign communication and outreach between those who are politically engaged and those who are not ...**



# Campaign models

- Vary widely, generally constructed using statistical techniques (correlation not equal to causation); models customized to political environments.
- Models similar to psychographic analysis and prediction, may use supervised learning or other techniques.
- Good campaign models provide valuable details about where campaigns can be more successful and an indication of how specific strategies might work.

# A different campaign model -- Lichtman

- Alan Lichtman / American University historian

## Approach:

- Developed 13 T/F keys that predict election outcome. True favors incumbent party. If 6+ are false, change is predicted.
- Has worked in every election for the last 30 years.

- 2016 Prediction: Trump wins
- 2020 Prediction: Trump loses

- Lichtman's Keys:

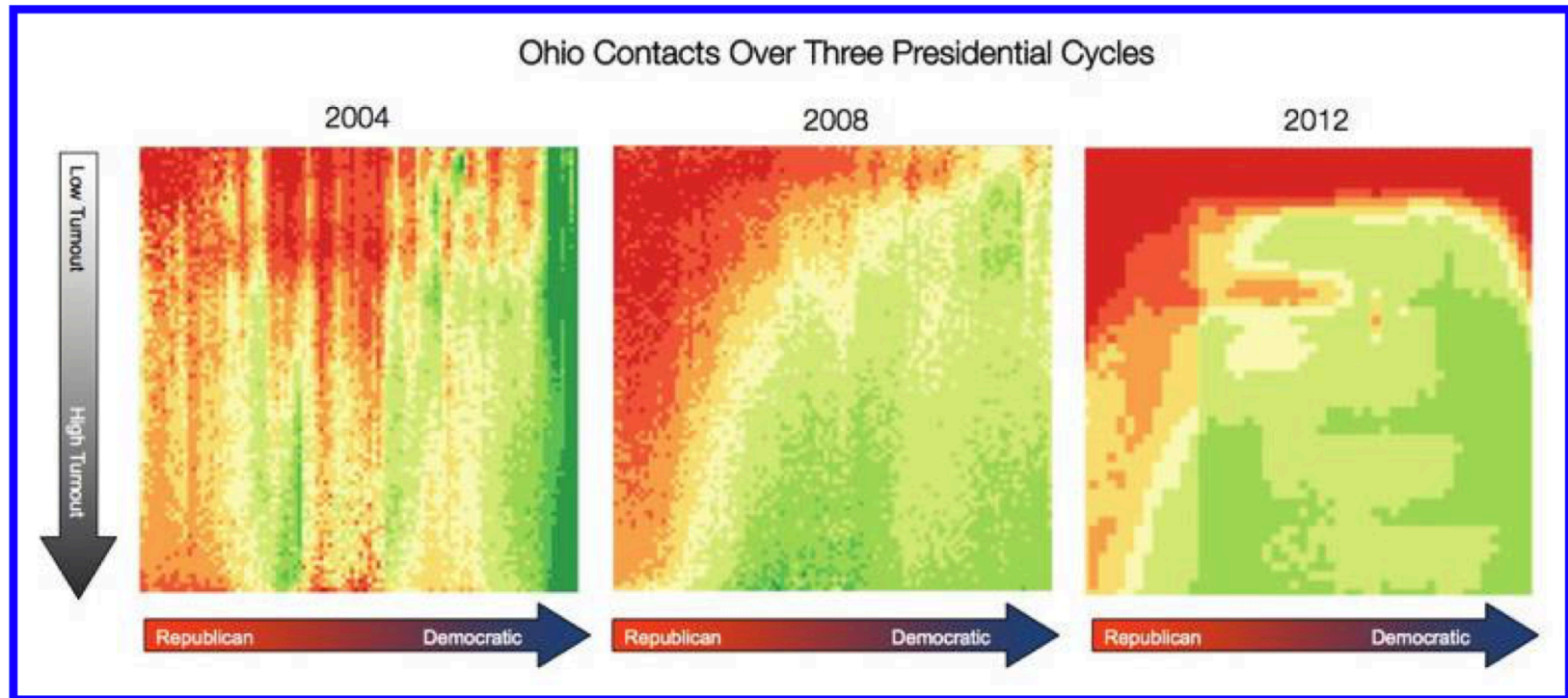
1. **Party Mandate:** After the midterm elections, the incumbent party holds more seats in the U.S. House of Representatives than after the previous midterm elections.
2. **Contest:** There is no serious contest for the incumbent party nomination.
3. **Incumbency:** The incumbent party candidate is the sitting president.
4. **Third party:** There is no significant third party or independent campaign.
5. **Short-term economy:** The economy is not in recession during the election campaign.
6. **Long-term economy:** Real per capita economic growth during the term equals or exceeds mean growth during the previous two terms.
7. **Policy change:** The incumbent administration effects major changes in national policy.
8. **Social unrest:** There is no sustained social unrest during the term.
9. **Scandal:** The incumbent administration is untainted by major scandal.
10. **Foreign/military failure:** The incumbent administration suffers no major failure in foreign or military affairs.
11. **Foreign/military success:** The incumbent administration achieves a major success in foreign or military affairs.
12. **Incumbent charisma:** The incumbent party candidate is charismatic or a national hero.
13. **Challenger charisma:** The challenging party candidate is not charismatic or a national hero.

# Campaign Economics

- All campaigns have limited budgets – what is the best way to spend money? (Way that will get the most and the most important votes?)
  - **Persuasive communication** should be targeted to voters most likely to be influenced positively
  - **Voter mobilization** should be targeted to voters most likely to vote for candidate and less likely to do so without help/provocation
  - **Door to door canvassing** and **direct mail** should target persuadable voters
  - **Social media outreach** should increase support and mobilization
  - Want to mobilize voters and support in “**battleground**” **areas** – areas where campaign efforts could change the outcomes (winner vs. loser)
  - Want to optimize the **number of contacts from the campaign** for voters (few if it’s hopeless or they are strong supporters, more if they are persuadable and in an important target cohort, not too many to be annoying)

Figure 1

## Heatmap of Ohio Contacts over Three Presidential Cycles



Source: Derived from Catalist, LLC.

Notes: The x-axis is likelihood of supporting a Democratic candidate over a Republican candidate, ranging from 0 (left) to 100 (right). The y-axis is likelihood of voting ranging, from 100 (low) to 0 (high). Colors (or in grayscale, shade) of each cell indicate how many direct contacts were made by a particular campaign. In the grayscale version of the heatmap, darker means more contacts. In the color version, dark red represents the least contacts and dark green the most contacts. Readers can see the color heatmap in the online version of this paper.

# Discussion

- How do you vote?
  - Where do you get your information?
  - How persuadable are you?
- Would you be a good target for more effort by a campaign?
  - What could a campaign do that would change your mind?
- Do you see data-driven opportunities with the potential to improve elections, voting, turnout, citizen engagement in the political process?

# Lecture 11 Sources 1

- **“Exit polls and election results – what we learned”**. The Guardian, <https://www.theguardian.com/us-news/2016/nov/12/exit-polls-election-results-what-we-learned>
- **“The Science of Error: How Polling Botched The 2016 Election”**, Forbes, <http://www.forbes.com/sites/startswithabang/2016/11/09/the-science-of-error-how-polling-botched-the-2016-election/#75748a437da8>
- **“The trouble is not with polling but with the limits to human interpretation of data,”** Quartz, <http://qz.com/832908/confirmation-bias-is-why-we-couldnt-predict-a-trump-victory/>
- **“There are Many Ways to Map election Results. We’ve Tried Most of Them.”**, NY Times, [http://www.nytimes.com/interactive/2016/11/01/upshot/many-ways-to-map-election-results.html?\\_r=0](http://www.nytimes.com/interactive/2016/11/01/upshot/many-ways-to-map-election-results.html?_r=0)
- **“Trump’s win isn’t the death of data – it was flawed all along,”** Wired, <https://www.wired.com/2016/11/trumps-win-isnt-death-data-flawed-along/>
- **“2016 Election Oracles: These People Predicted Trump Would Win”**, Heavy, <http://heavy.com/news/2016/11/2016-final-election-results-predictions-helmut-norpoth-abramowitz-michael-moore-nate-silver-vote-count-turn-out-electoral-college-maps-donald-trump-hillary-clinton-polls-forecasting-pennsylvania-michi/>

# Lecture 11 Sources 2

- **“No, one 19-year-old Trump supporter probably isn’t distorting the polling averages all himself”**, LA Times, <http://www.latimes.com/politics/la-na-pol-daybreak-poll-questions-20161013-snap-story.html>
- **“How IBD Accurately Gauged Voter Enthusiasm and Got the Polls Right,”** Townhall, <http://townhall.com/tipsheet/cortneyobrien/2016/11/11/how-ibd-got-the-polls-right-n2244109>
- **“Political Campaigns and Big Data”**, Nickerson and Rogers, Journal of Economics Perspectives

# Presentations





# Upcoming Presentations

## March 15

- **“Hunting for a giant black hole, astronomers found a nest of darkness,”** New York Times, <https://www.nytimes.com/2021/02/26/science/astronomy-black-hole-ngc6397.html> (Greg S.)
- **“NASA’s Webb telescope will be the world’s premier space science observatory – Here’s what those those powerful capabilities mean for astronomy,”** SciTechDaily, <https://scitechdaily.com/nasas-webb-telescope-will-be-the-worlds-premier-space-science-observatory-heres-what-those-powerful-capabilities-mean-for-astronomy/> (Angelina M.)

## March 18

- **“Why so many data science projects fail to deliver”**, MIT Sloan Management Review, <https://sloanreview.mit.edu/article/why-so-many-data-science-projects-fail-to-deliver/?og=Home+Editors+Picks>
- **“Using big data to measure environmental inclusivity in cities,”** EOS, <https://eos.org/articles/using-big-data-to-measure-environmental-inclusivity-in-cities>

## Need Volunteers – Presentations for March 22

- **“New open source database tracks data on slaves, slavers, and allies”**, Harvard Gazette, <https://news.harvard.edu/gazette/story/2021/03/new-open-source-database-tracks-data-on-slaves-slavers-allies/> (Eric X.)
- **“Robo-writers: the rise and risks of language-generating AI”**, Nature, <https://www.nature.com/articles/d41586-021-00530-0> (Josh M.)

# Presentations for Today

## March 8

- **“Election forecast models are worth more attention than polls”**, Bloomberg Opinion, <https://www.bloomberg.com/opinion/articles/2020-11-22/election-forecast-models-have-more-potential-than-simple-polling> (Chris P.)
- **“Which 2020 election polls were most – and least – accurate?”**, Washington Post, <https://www.washingtonpost.com/politics/2020/11/25/which-2020-election-polls-were-most-least-accurate/> (Isaac L.)