

# Politics, data and journalism

## Assorted questions and answers

**G. Elliott Morris**

Data journalist

*The Economist*

October 18, 2019

**What is a "data journalist"?**

# What is a "data journalist"?

A "data journalist" is just like a "regular" journalist, but we rely on our own skills in empiricism to tell a story.

## Empiricism?

- 1. Find a story**
- 2. Find a data-driven angle in said story**
- 3. Analyze data with statistics programs (Excel, STATA, Python, R)**
- 4. Convey information (with words and graphics)**

**Example: What if everyone  
voted?**

# **Scientific process:**

- 1. Ask a question**
- 2. Form a hypothesis**
- 3. Test that hypothesis**
- 4. Make a conclusion**

# Guiding questions

## **1. How many Democrats and Republicans are there?**

Given data constraints, we're really asking: How many Clinton and Trump voters are there?

## **2. How are they distributed geographically?**

The answer lets us assign Electoral College votes.

# Data

## **1. Cooperative Congressional Election Study (CCES): A survey of 64,000 Americans**

Includes demographic data and 2016 vote choice for 40,000+ validated voters

## **2. American Community Survey (ACS): A Census Bureau survey of 175,000 Americans**

Includes the same demographic data as the CCES 32,640 “cells”

# Tools:

## The R statistical programming language

Someone asked about this beforehand...

Why?

- Full-suite development: data wrangling, statistical analysis, visualization and more
- With easy-to-learn tools (like the `tidyverse` packages and `ggplot`)
- An active online community for answering questions, suggesting methods, etc.

## Data visualization

- The outcome is electoral college results, which we can show to people at the state level



# Method

## 1. Train a predictive model on CCES data

- Multi-level logistic regression
- Predict vote choice with: age, gender, race, education, region and interactions between them

## 2. Use the model to predict voting habits for every eligible American

Via “post-stratification” on the ACS

# ACS Post-stratification

## 1. Each "type" of person gets their own "cell":

- One cell for white men ages 18-30 without college degrees who live in the Northeast
- Another for white men ages 18-30 without college degrees who live in the South
- Another for non-white men ages 18-30 without college degrees who live in the Northeast
- etc.

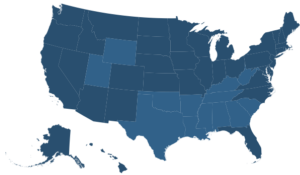
## 2. We know how many voters in that "cell" live in each state

## 3. So we can say that x and y% of each "cell" vote for Clinton or Trump, then add up

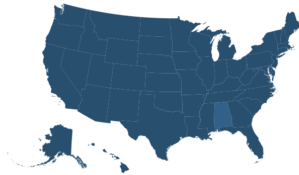
- For example, a Latino female age 18-30 with a college degree in Texas is 85% likely to vote for a Democrat for president (White man 65+ is 80% Republican)

# Results

Black, Non-Hispanic  
HS or less



Black, Non-Hispanic  
Some college



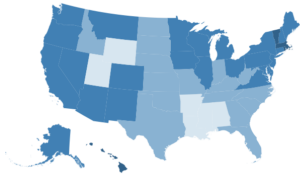
Black, Non-Hispanic  
College



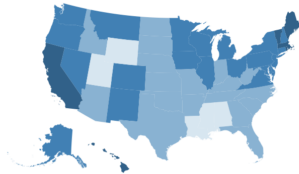
Black, Non-Hispanic  
Post-grad



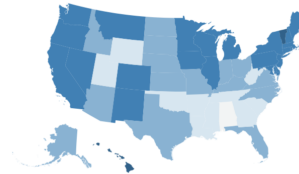
Hispanic  
HS or less



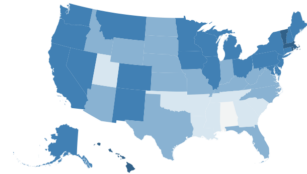
Hispanic  
Some college



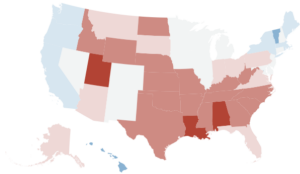
Hispanic  
College



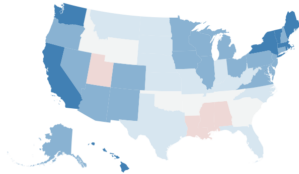
Hispanic  
Post-grad



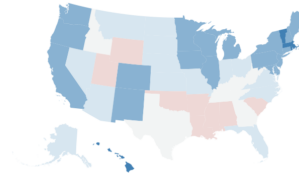
Other, Non-Hispanic  
HS or less



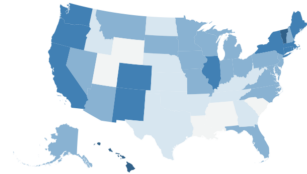
Other, Non-Hispanic  
Some college



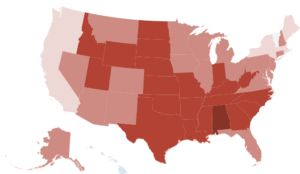
Other, Non-Hispanic  
College



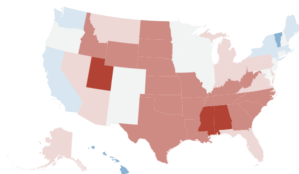
Other, Non-Hispanic  
Post-grad



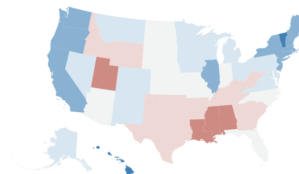
White, Non-Hispanic  
HS or less



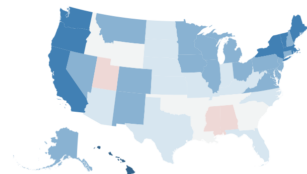
White, Non-Hispanic  
Some college



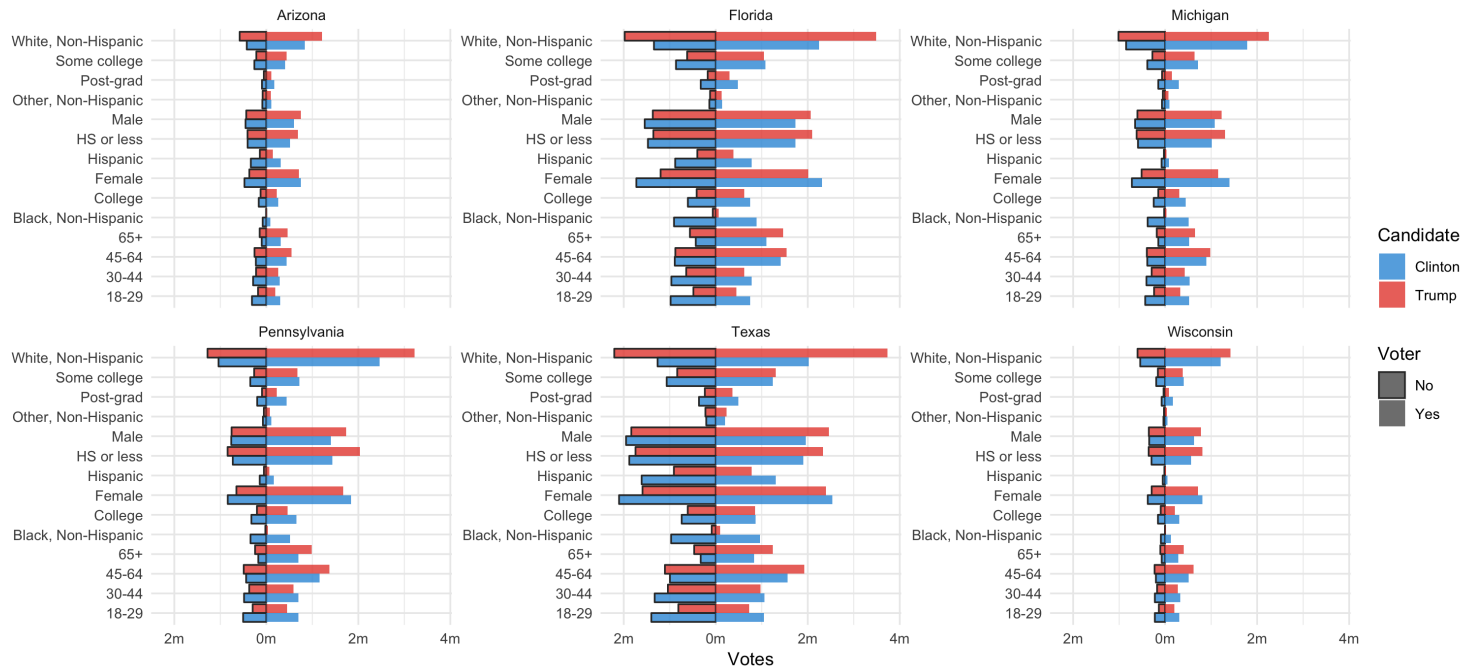
White, Non-Hispanic  
College



White, Non-Hispanic  
Post-grad



# Results



# **Results: If everyone voted**

# Election forecasting: How do we do it, and why?

# **Election forecasting**

## **Some of your questions:**

**How do you build a forecast?**

**How do you iterate on other scholarship?**

**Do you use data besides polls?**

**Is this different in other countries?**

# How to build an election forecasting model:

## What's in a model?

1. Start with historical data: a measurement variable and outcome variable
  - Like polls and presidential vote share, for example
2. Build a statistical model that predicts the outcome variable given some value of the measurement variable
3. Add more measurement variables (but not too many!)
  - like GDP or presidential approval "fundamentals"

## Election forecasting

Fundamentals + polls -> vote share + simulation -> win probabilities!



# Forecasting the 2018 mid-terms

Build off of Bafumi, Erikson and Wlezien (2010; 2014) with a Bayesian framework.

## 1. Create a "prior" prediction for each seat

- Regress 2016 vote on previous House and presidential vote and whether an incumbent is running

## 2. Create a polling average for each seat

- Simply a weighted average where recent polls receive more weight in the average

## 3. Combine them and simulate 10,000 elections

- Using Bayesian statistics; a weighting of each predictor in proportion to its variance

# When forecasts go wrong

- **Misspecified models**
- **Bad training data**
- **Surprise outcomes — 2016**
  - Shy Trump (Tory) effect? No evidence.
  - Late deciders? Slow-moving averages? Yes.
- **Miscommunicating uncertainty**

# Miscommunicating uncertainty

## The dangers:

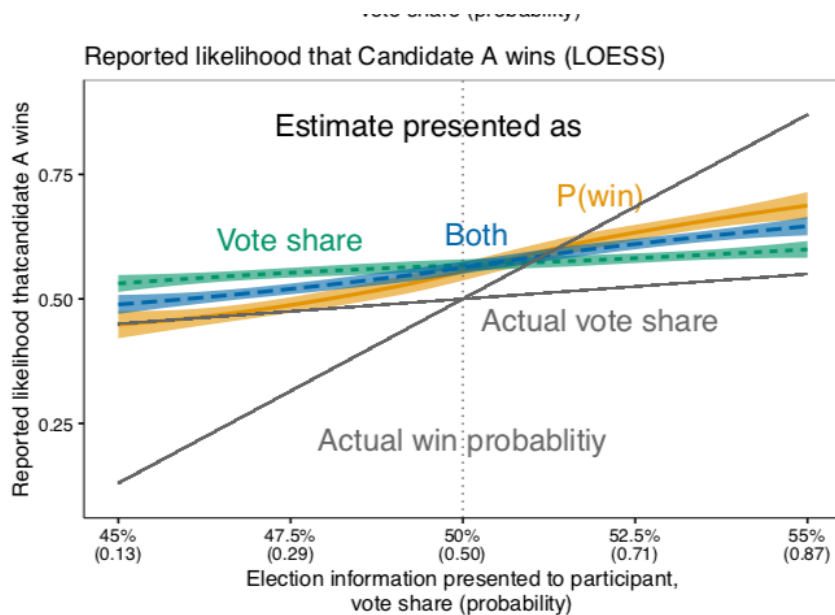
- False certainty can bias media narratives (especially when combined with reporters' political biases)
- False certainty can lead to severe consequences
- False certainty betrays our real understanding or how often "unlikely" election outcomes can happen (see: Trump 2016)

James Comey, 2018, "A Higher Loyalty":

"It is entirely possible that because I was making decisions in an environment **where Hillary Clinton was sure to be the next president**, my concern about making her an illegitimate president by concealing the restarted investigation bore greater weight than it would have if the election appeared closer or if Donald Trump were ahead in all polls."

# Miscommunicating uncertainty: probability

Readers have the best understanding of the horse race when presented with probabilities



Source: Westwood, Messing and Lelkes (2019)

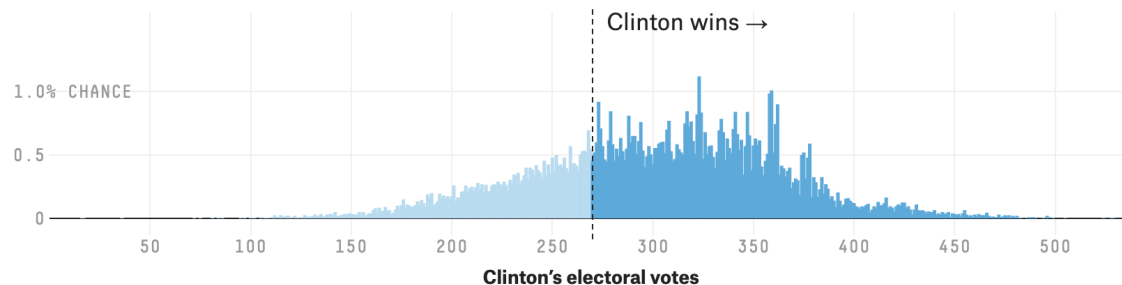
# Probability: a better way

## Point projections don't matter, distributions do...

- If we are not giving readers a sense of our certainty, we are lying to them.
- The best way to convey our certainty is to produce a distribution of possible outcomes for the election, combining confidence intervals with our point projections to transform them into probabilities

### What to expect from the Electoral College

In each of our simulations, we forecast the states and note the number of electoral votes each candidate wins. That gives us a distribution for each candidate, where the tallest bar is the outcome that occurred most frequently.



# "Trump will lose" in 2020?

Opinion

## Why Trump Will Lose in 2020

The president is running hard on a strategy of riling up his base. But by doing that, he riles up the Democratic base, too, and that one is bigger.

**By Rachel Bitecofer**

Dr. Bitecofer is a professor of political science at Christopher Newport University.

Jan. 24, 2019

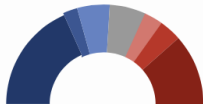
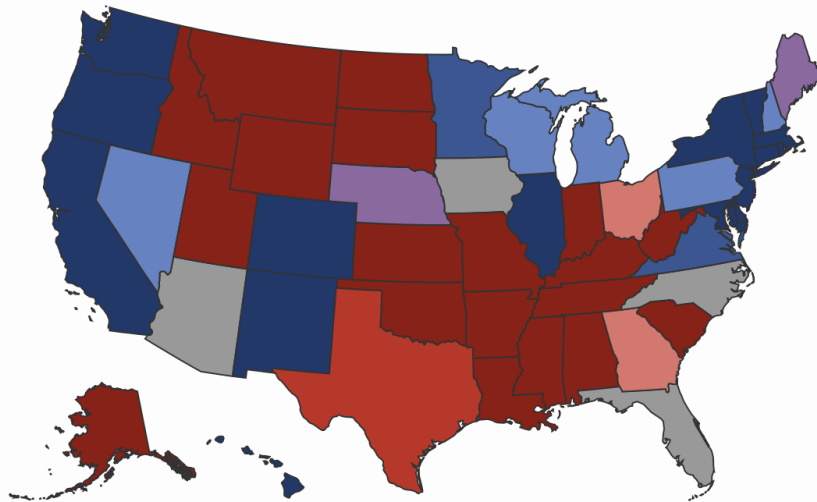


*Source: Rachel Bitecofer; New York Times (2019)*

- How can she know for sure?

# Forcing uncertainty

## 2020 Bitecofer Model Electoral College Predictions



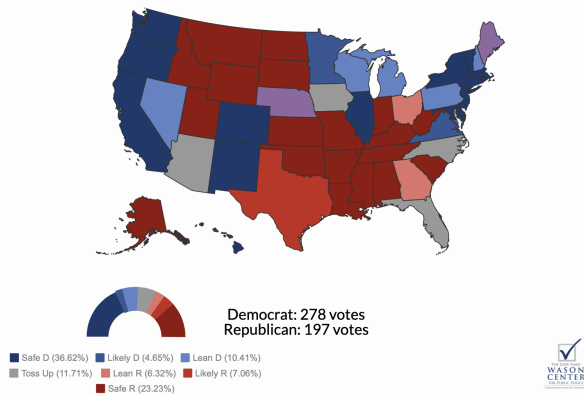
Democrat: 278 votes  
Republican: 197 votes

Safe D (36.62%) Likely D (4.65%) Lean D (10.41%)  
Toss Up (11.71%) Lean R (6.32%) Likely R (7.06%)  
Safe R (23.23%)



# Forcing uncertainty

2020 Bitecofer Model Electoral College Predictions

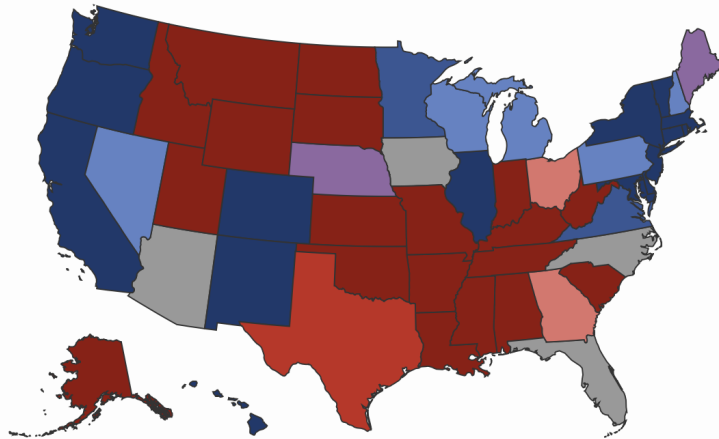


- Expected probabilistic outcome: 300 Democratic electoral votes
- If you assume a root mean squared error of 45 EVs (half of the 2016 error):
- Distribution of outcomes (95% confidence interval): 212 - 388 Democratic electoral votes
- Or just a 75% chance of Democratic victory
- (If you assume 2016 error is normal for the forecast, then it's just a 64% chance)



# Forcing uncertainty

## 2020 Bitecofer Model Electoral College Predictions



Democrat: 278 votes  
Republican: 197 votes

■ Safe D (36.62%) ■ Likely D (4.65%) ■ Lean D (10.41%)  
■ Toss Up (11.71%) ■ Lean R (6.32%) ■ Likely R (7.06%)  
■ Safe R (23.23%)



*Source: Rachel Bitecofer; Wason Center (2019)*

Hardly looks like "Trump will lose" once you look under the hood

# Questions?

(Volunteers get priority!)

"How do you combine your interest in history (or generally qualitative/theoretical knowledge) with quantitative techniques that you can use? For example, if my main interest is in political theory, in what ways could quantitative techniques make me better than “pure” political theorists?"

- History is often helpful for framing our articles — I studied the founding
  - History also provides journalists with some amusing leads for our stories
  - And it's also fun!
- What you're really asking about is how numbers can help your qualitative study, right?
  - William Petty set out in the 1600s in his work “Political Arithmetick” to prove that England's government was better than France or Holland's. Boom, political theory + data
  - For this she used data on shipping, trade, acres farmed, population and territories governed
  - The ability to test our hypotheses against data — our ability to be rational thinkers, as Thomas Hobbes and Francis Bacon would call it — is how you make political theory better with numbers
- And on the subject of political arithmetic....
  - Yes, political scientists study math and programming
  - They help us do things like test hypotheses and even to write fancy simulations to forecast elections

"I would like to know Elliott's opinion on the NC electoral college and how it may affect the upcoming election."

- North Carolina leans slightly more Republican than other toss-up states like Florida
- Democrats won the state house and senate popular vote there in 2018
- But 2020 is unlikely to be such a wave year
- So I'd say NC is a toss-up state, but unlikely to be the tipping point or focus of much intense activity from Democrats

"do you think that investment from the DNC is worthwhile in opposition to groups like Engage Texas? Additionally, do you think that likeliness for a Democratic win in 2020 can be measured similarly to the midterm election, considering the high disapproval rating of Trump versus the lower disapproval rating of Ted Cruz in the months preceding the 2018 election?"

- Whatever I may report, it's clear to me that Democrats think Texas is competitive, and so they'll spend resources there
- There are some reasons to think Texas is competitive, everything else disregarded:
  - Trump is unpopular
  - The state quickly trended D from 2012 to 2016 — could be that large swings in partisan loyalty only happen in POTUS election years
- But again, it leans 11-13 points to the right, depending on your formula
- That's very red!

Besides Texas, "do you think there are other states that the Democratic Party could be focusing this energy towards?"

- Offense:
  - Arizona and especially Georgia are much more likely to flip than Texas
- Toss-up:
  - And the 2020 election looks like it very well could be very close, so they should obviously focus resources on Wisconsin, Pennsylvania and Michigan
- Defense:
  - New Hampshire, the forgotten swing state
- ??
  - Iowa: Hurt by tariffs and unclear where it stands politically; looked more like 2012 voting in 2018 mid-terms and has high Trump disapproval

" Now that the Brexit movement is floundering, with a no-deal Brexit off the table, do you think the traditional powerhouse parties will regain their popularity? Or has the whole situation changed the party landscape in the UK more permanently?"

- First, it's not clear to me that a no-deal Brexit is actually "off the table". Markets give it a 13% chance of happening in 2019.
- But nevertheless, the Brexit Party has lost strength
- This seems likely due to the Tories electing Boris Johnson as their party leader and PM
- This is a realignment!
- The Tories chose to keep their party label, but by handing the reins to Johnson, fully oriented the party toward a no-deal/independence position
- I cannot emphasize enough how fluid this scenario is
- Because the country uses first past the post elections, if either of the minor parties— Liberal Democrats or Brexit Party—can prove their ability to win the majority, they will quickly move ahead
- In this way, voting in the UK is tactical—but also psychological!
- Don't rule out a Lib Dem or even BP Premiership — though I think they are unlikely (the latter more than the former — Lib Dems have worked with the government before to form a working majority)

■ "Who will win the primary?"

- Based on polls, Warren has something like a 30% chance to win. Biden has maybe 25%.
- That leaves a 45% chance that anyone else wins the nomination.
- I wouldn't bet on anyone individual given those odds. (But if I had to, I'd have to pick Warren)



■ "And how do they win the presidency?"

- I've argued for The Economist that Democrats can find much future success by orienting their politics around class — rich v poor, Wall Street v Main Street
- Notable that a wealth tax and “economic patriotism” (American manufacturing) are popular among all Americans
- This strategy de-emphasized the rhetorics and politics of race and immigration that Donald Trump capitalized upon to win
- But he knows that — so it's a tough gamble. Trump can probably keep immigration on the table so long as he talks about it; amplification via Fox News and a willingness from all GOP Reps to echo his positions and agenda setting

"In your article on the urban-rural divide in American politics you allude to both passive demographic change and intentional party realignment as means for the Democratic Party to court rural communities. Do you believe the party will make inroads and if so, which method do you believe will have a greater impact?"

- I think you've got this backwards (or I misunderstand the question):
  - Demographic change in cities—from white labor centers to multi-ethnic immigrant communities—has forced Democrats to embrace diversity or face total electoral annihilation
- They are facing some consequences from the geographic nature of US politics because of this
- To stay competitive in the US Senate, the Democrats may soon have to consider new strategies for courting rural voters
- Maybe better farm policy? Job training programs in Middle America?
- Or, again, class-based politics that de-emphasize race without decreasing black turnout

# Thank you!

## G. Elliott Morris

Data journalist, *The Economist*

Website: [gelliottmorris.com](http://gelliottmorris.com)

Email: [elliott@thecrosstab.com](mailto:elliott@thecrosstab.com)

Twitter: [@gelliottmorris](https://twitter.com/gelliottmorris)

---

*These slides were made with the `xaringan` package for R from Yihui Xie. They are available online at <https://www.thecrosstab.com/slides/2019-10-18-gw/>*