

# Understanding Election Forecasts

Statistically but also Ethically

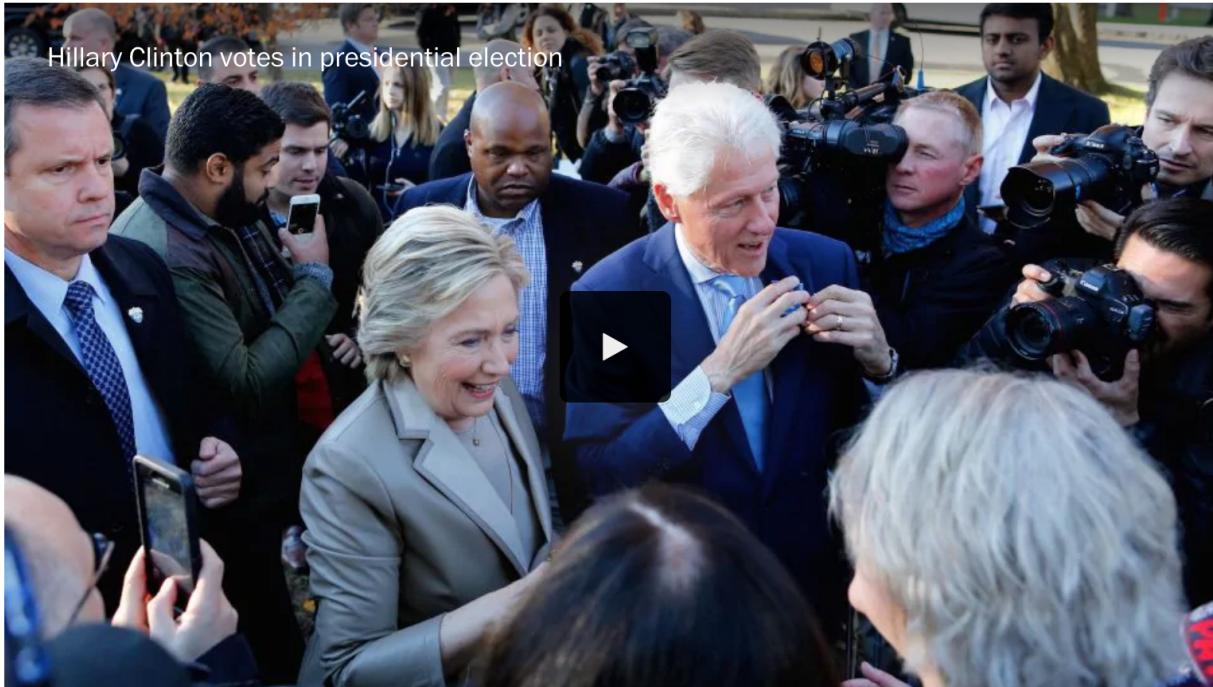
Understanding Political Numbers

April 22, 2019

Let's talk about the 2016 election (forecasts)

# A comprehensive average of election forecasts points to a decisive Clinton victory

---



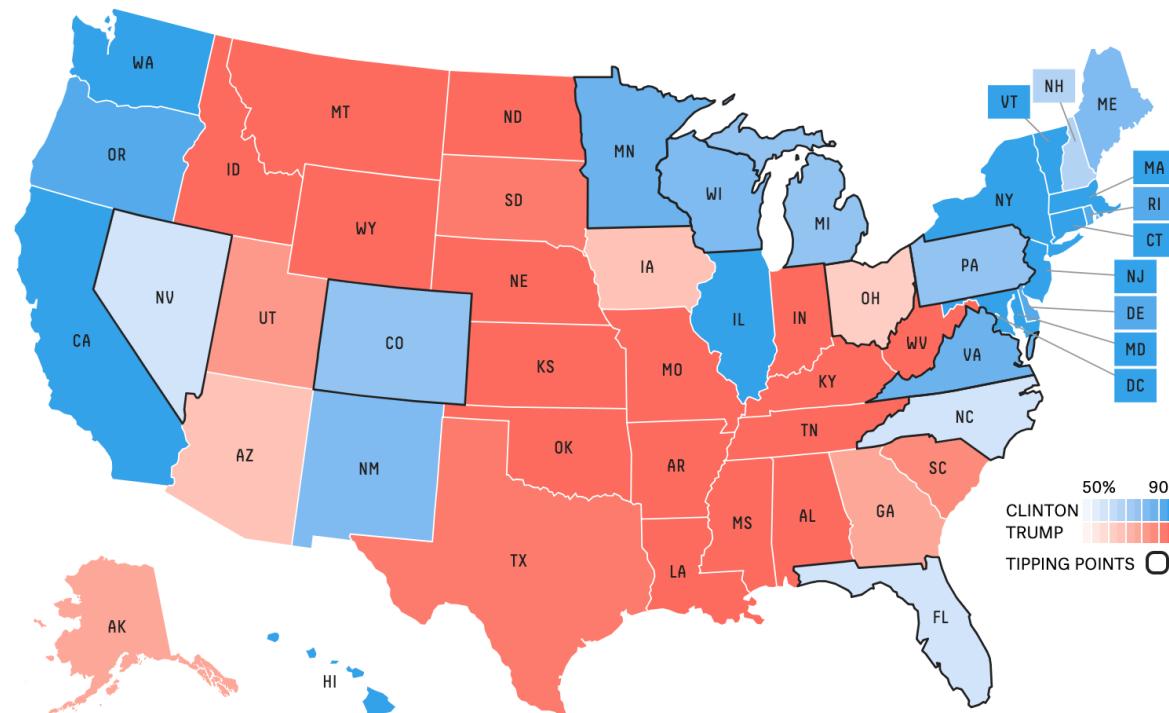
Democratic presidential candidate Hillary Clinton voted in Chappaqua, N.Y., Nov. 8, accompanied by her husband. (Photo: EDUARDO MUÑOZ ALVAREZ/The Washington Post)

# Who will win the presidency?



f

## Chance of winning



# Agenda

- (1) Probability and forecasting
- (2) How election forecasting works
- (3) Is forecasting a bad idea?

Probability  $\neq$  Vote Percentage

# Probability $\neq$ Vote Percentage

Popular vote  $\neq$  Electoral college vote

# Probability $\neq$ Vote Percentage

Popular vote  $\neq$  Electoral college vote

National vote  $\neq$  state vote

# Probability ≠ Vote Percentage

Popular vote ≠ Electoral college vote

National vote ≠ state vote

Not a "point prediction"—forecasts incorporate *uncertainty*

## The Media Has A Probability Problem

The media's demand for certainty — and its lack of statistical rigor — is a bad match for our complex world.

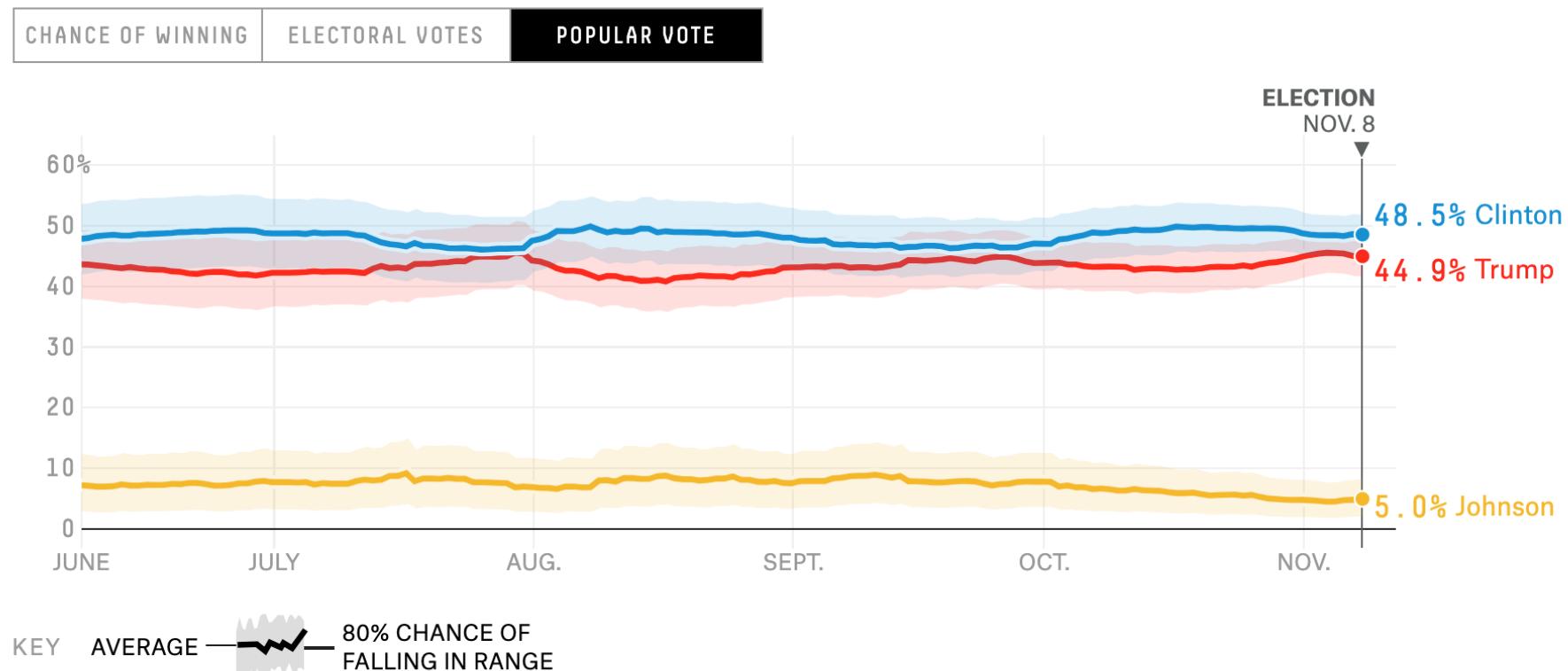
By [Nate Silver](#)

Filed under [The Real Story Of 2016](#)

Published Sep. 21, 2017

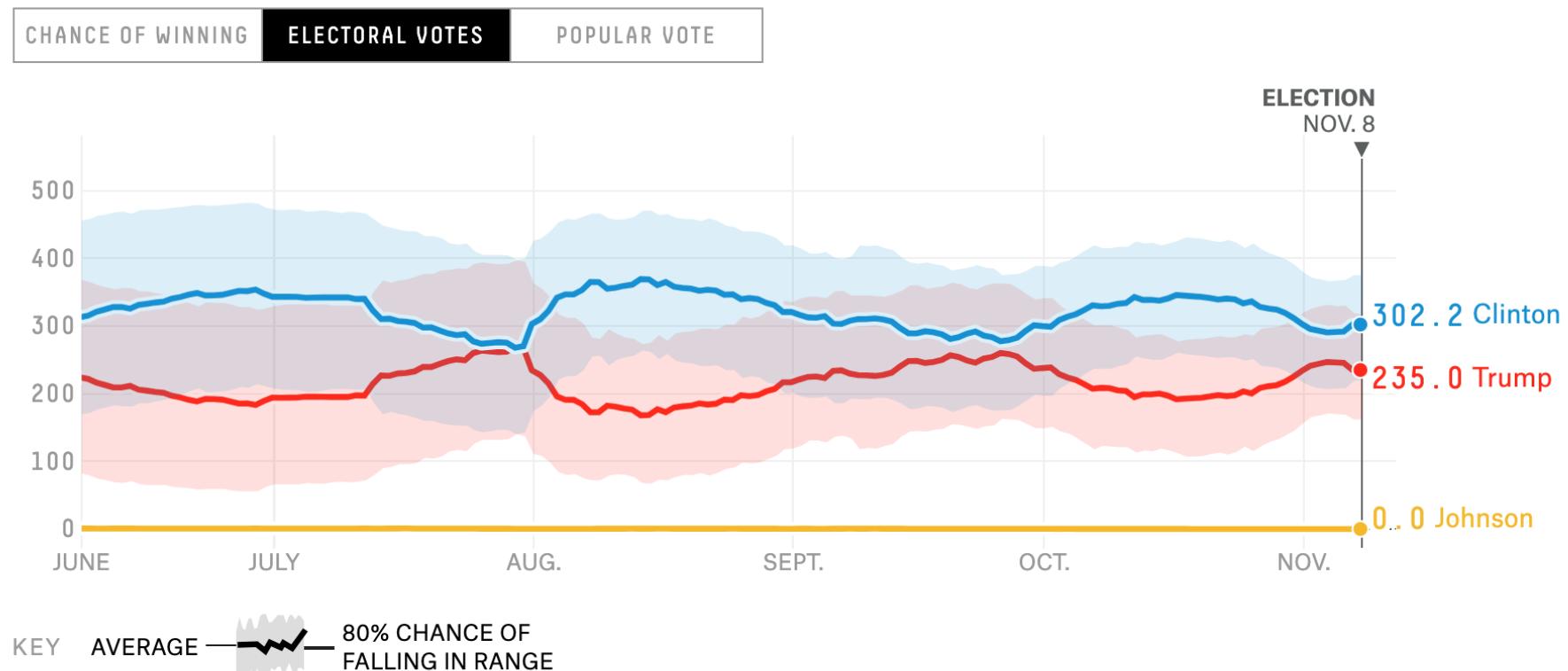
## How the forecast has changed

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



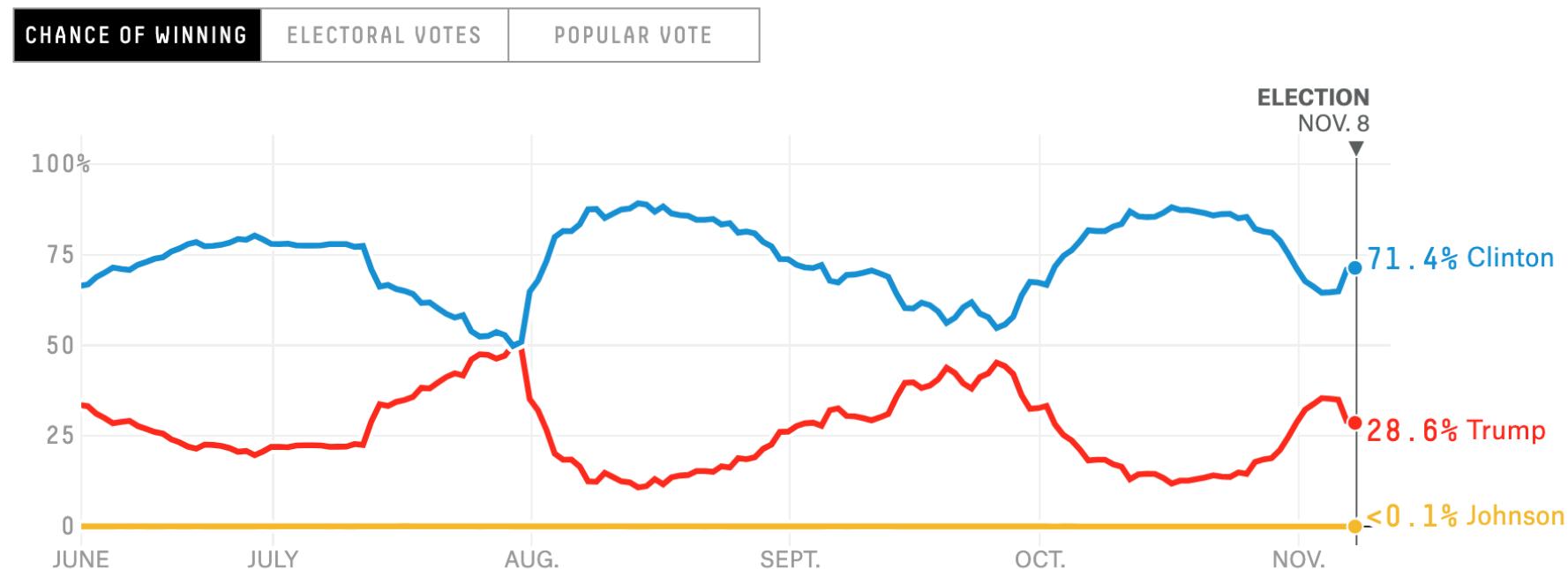
## How the forecast has changed

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



## How the forecast has changed

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



70-30 isn't certain

See for yourself

How can there be a probability? Either Clinton will win (100%), or she won't (0%)

How can there be a probability? Either Clinton will win (100%), or she won't (0%)

Out of all the possible scenarios that are consistent with our data, how often is Clinton the winner?

How can there be a probability? Either Clinton will win (100%), or she won't (0%)

Out of all the possible scenarios that are consistent with our data, how often is Clinton the winner?

Out of all the possible scenarios...

How can there be a probability? Either Clinton will win (100%), or she won't (0%)

Out of all the possible scenarios that are consistent with our data, how often is Clinton the winner?

Out of all the possible scenarios...

- In each state, either Trump or Clinton will win
- How many ways can you configure the electoral college map?

How can there be a probability? Either Clinton will win (100%), or she won't (0%)

Out of all the possible scenarios that are consistent with our data, how often is Clinton the winner?

Out of all the possible scenarios...

- In each state, either Trump or Clinton will win
- How many ways can you configure the electoral college map?

...that are consistent with our data...

How can there be a probability? Either Clinton will win (100%), or she won't (0%)

Out of all the possible scenarios that are consistent with our data, how often is Clinton the winner?

Out of all the possible scenarios...

- In each state, either Trump or Clinton will win
- How many ways can you configure the electoral college map?

...that are consistent with our data...

- Not all scenarios are equally likely
- We learn about the plausibility of scenarios from history and polls

How can there be a probability? Either Clinton will win (100%), or she won't (0%)

Out of all the possible scenarios that are consistent with our data, how often is Clinton the winner?

Out of all the possible scenarios...

- In each state, either Trump or Clinton will win
- How many ways can you configure the electoral college map?

...that are consistent with our data...

- Not all scenarios are equally likely
- We learn about the plausibility of scenarios from history and polls

How often is Clinton the winner?

# How can there be a probability? Either Clinton will win (100%), or she won't (0%)

Out of all the possible scenarios that are consistent with our data, how often is Clinton the winner?

Out of all the possible scenarios...

- In each state, either Trump or Clinton will win
- How many ways can you configure the electoral college map?

...that are consistent with our data...

- Not all scenarios are equally likely
- We learn about the plausibility of scenarios from history and polls

How often is Clinton the winner?

- A poll has Clinton 1 pt. below Trump
- What could Clinton's "true vote" be?
- How many of those "true votes" are greater than Trump's?

How do we "learn" from polls? Bayes' Theorem

In how many scenarios does Clinton win?

$$p(\text{Clinton wins}) = ?$$

In how many scenarios *that are consistent with these polls* does Clinton win?

$$p(\text{Clinton wins} \mid \text{polls}) = \frac{p(\text{Clinton wins AND polls})}{p(\text{polls})}$$



T. Bayes.

# Bayesian updating, in context

You have the disease :: Clinton wins

Imperfect diagnostic test :: polls

Taking more tests :: conducting more polls

# Bayesian updating, in context

You have the disease :: Clinton wins

Imperfect diagnostic test :: polls

Taking more tests :: conducting more polls

Location of ball #1 :: Clinton's EC vote share

Relative location of other balls :: polls

Throwing more balls :: conducting more polls

# How the forecasts work

# Main idea

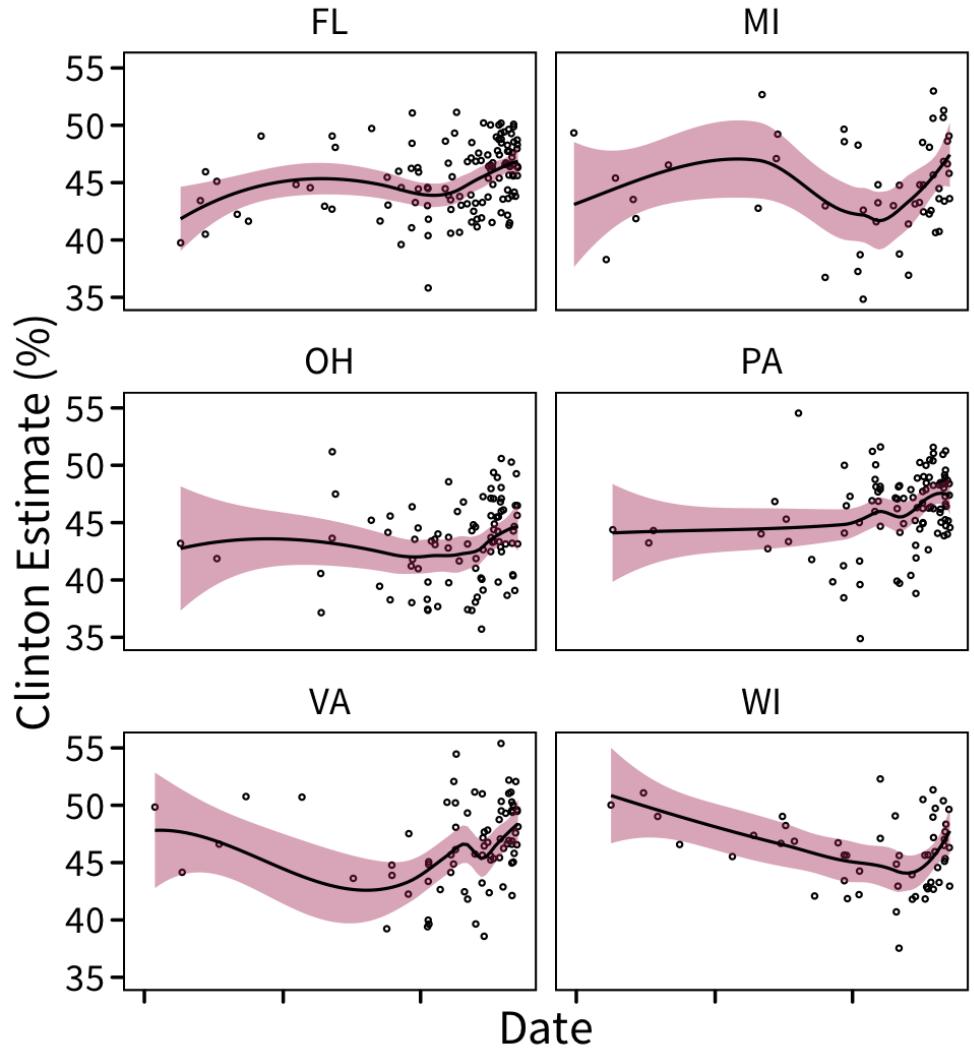
In regression, we use data to estimate *unknown parameters*

$$y = f(x, \alpha, \beta)$$

In forecasting, we use polls to estimate *unknown true support*

State poll today =  $f$  (True state vote today, Other)

Polls are samples from an underlying average, but...

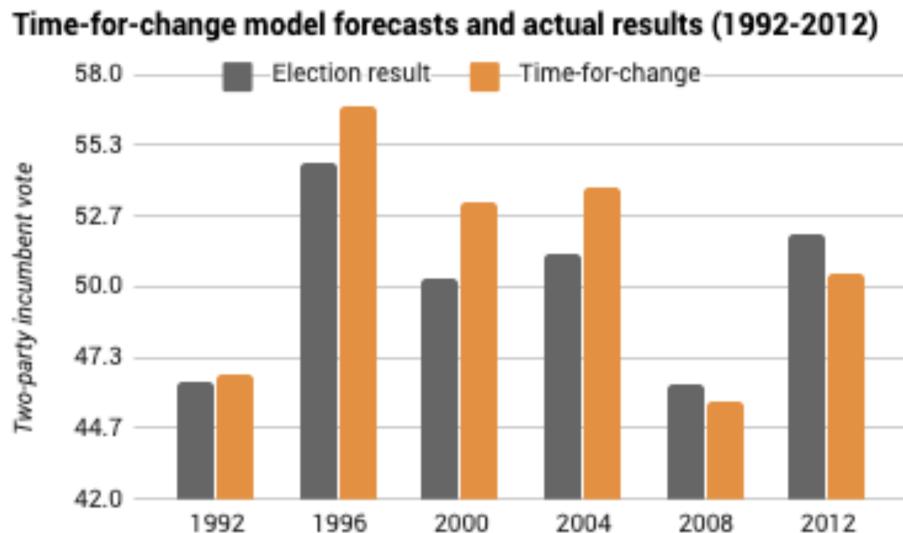


Uncertainty from *time*. (Answer: "the fundamentals")

# Uncertainty from *time*. (Answer: "the fundamentals")

Predicting an election > 6 months out:

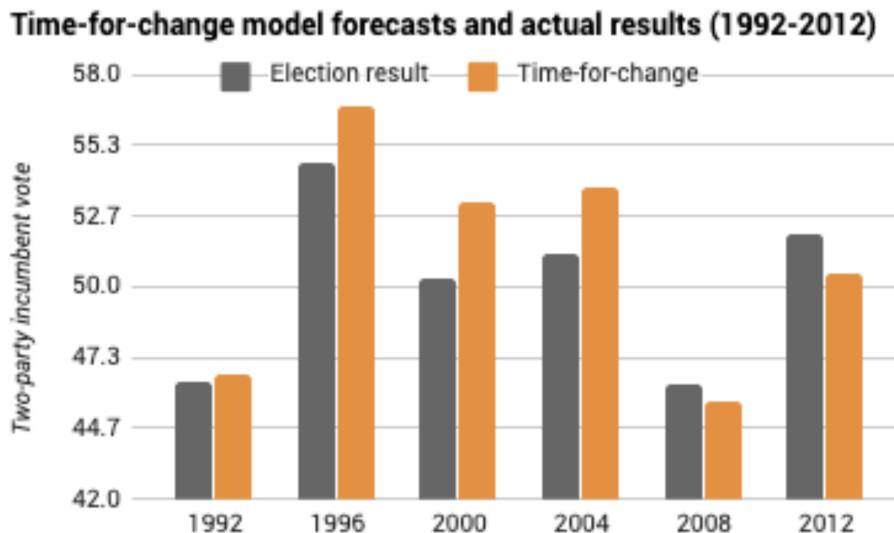
- Presidential approval
- Incumbency
- Economic Indicators



# Uncertainty from *time*. (Answer: "the fundamentals")

Predicting an election > 6 months out:

- Presidential approval
- Incumbency
- Economic Indicators



FiveThirtyEight uses fundamentals (and historical data) to build "prior expectations" for Election Day

1. Anchors each state as the model projects to Election Day
2. How each state differs from the average
3. When state predictions are wrong, *which states are "wrong together"?*

New polls "update our prior expectations" about Election Day

# Projecting Forward ("Updating our Priors")

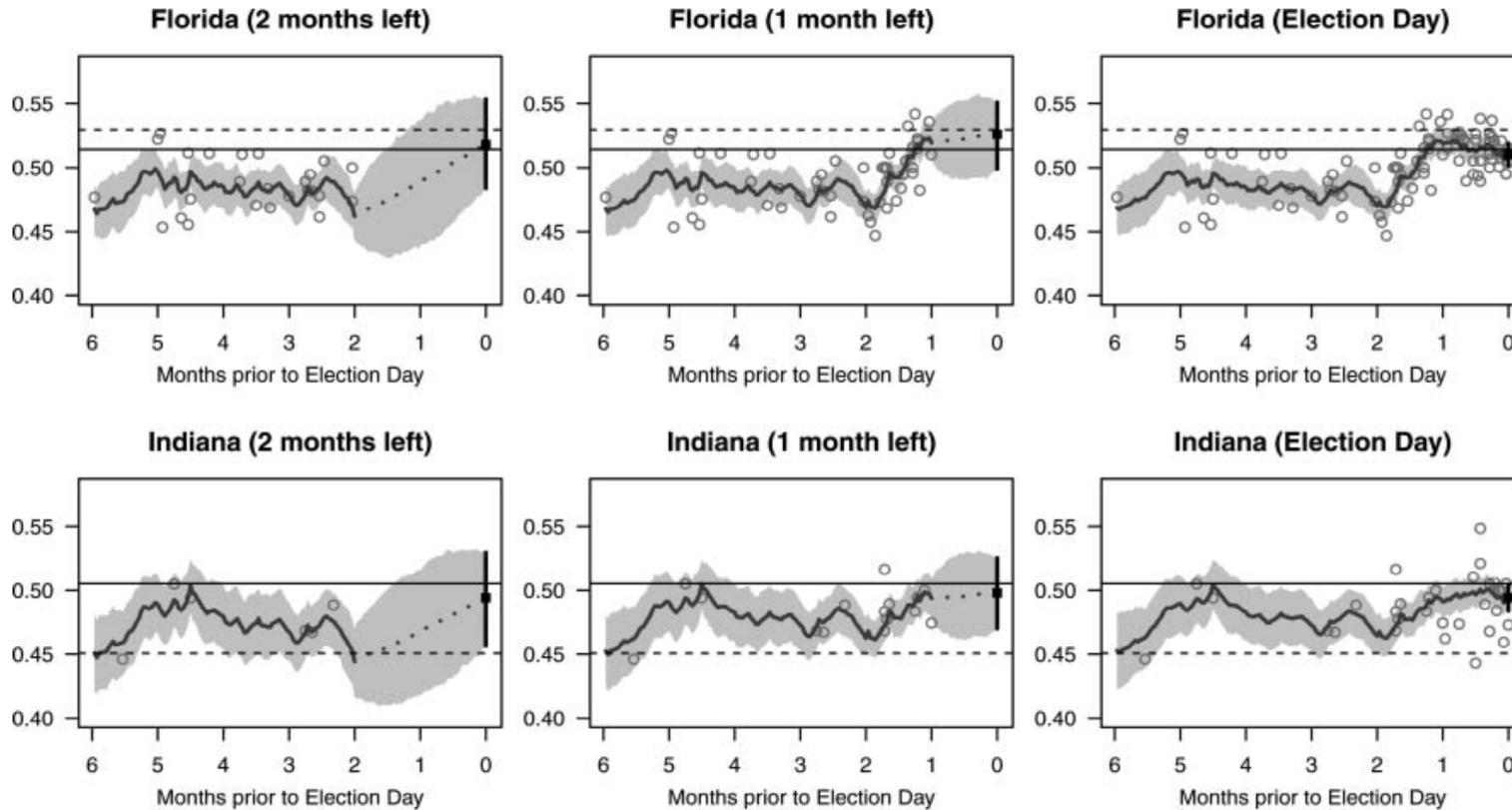


Figure 3. Forecasting the 2008 presidential election in real time. Results are shown for Florida and Indiana. The vertical axis is the percentage  
(Source: [Linzer 2013](#))

Polls are  $f(\text{truth} + \text{Other})$

## Three types of error

Each simulation accounts for three potential types of error and uncertainty:

- National error. The polls are systematically off throughout the country.
- Demographic and regional error. The polls are off in states that have demographic or geographic factors in common.
- State-specific error. The polls are off in a particular state, with no effect on other states.

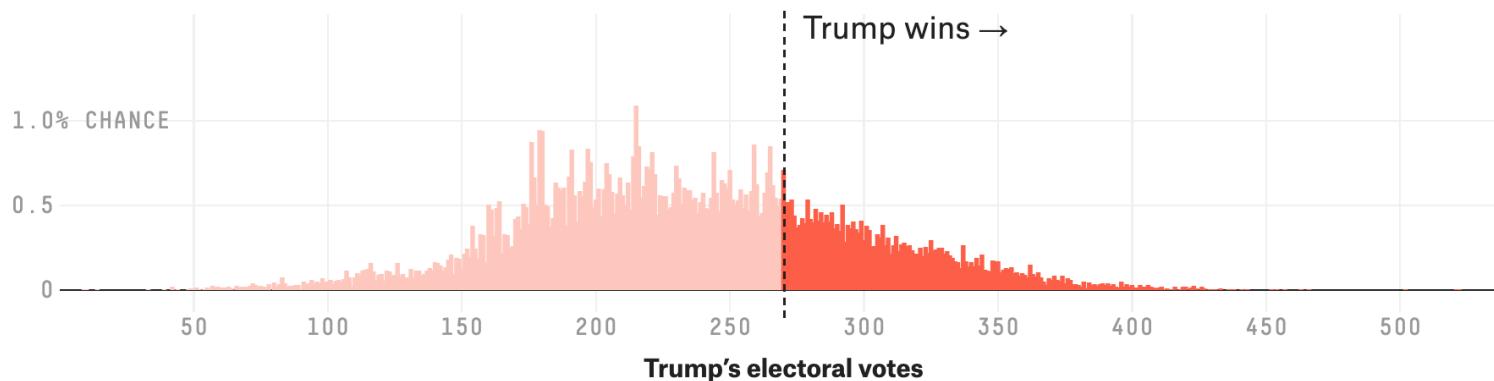
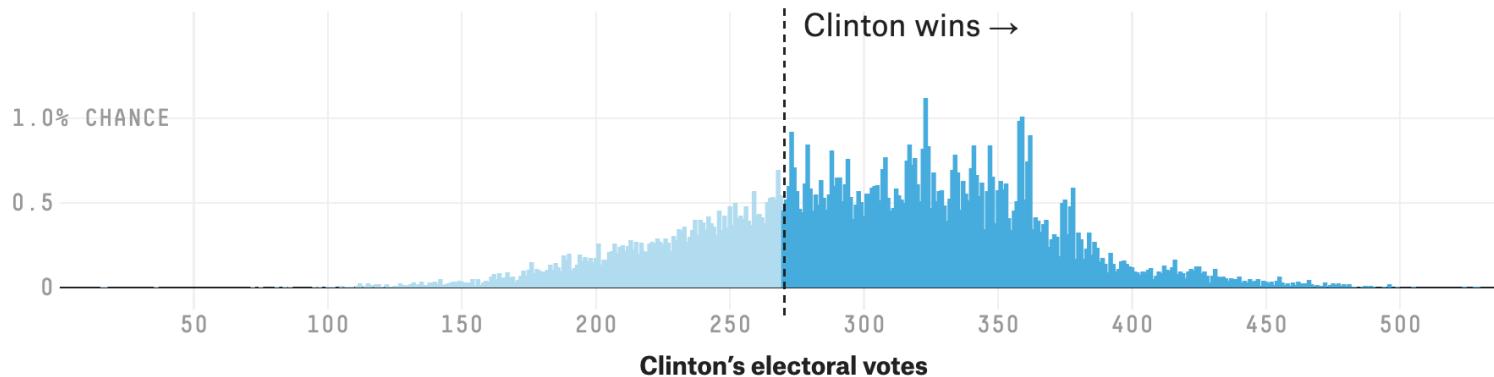
(source: [FiveThirtyEight](#))

# What scenarios are compatible with the data?

What scenarios give us data that look like this?

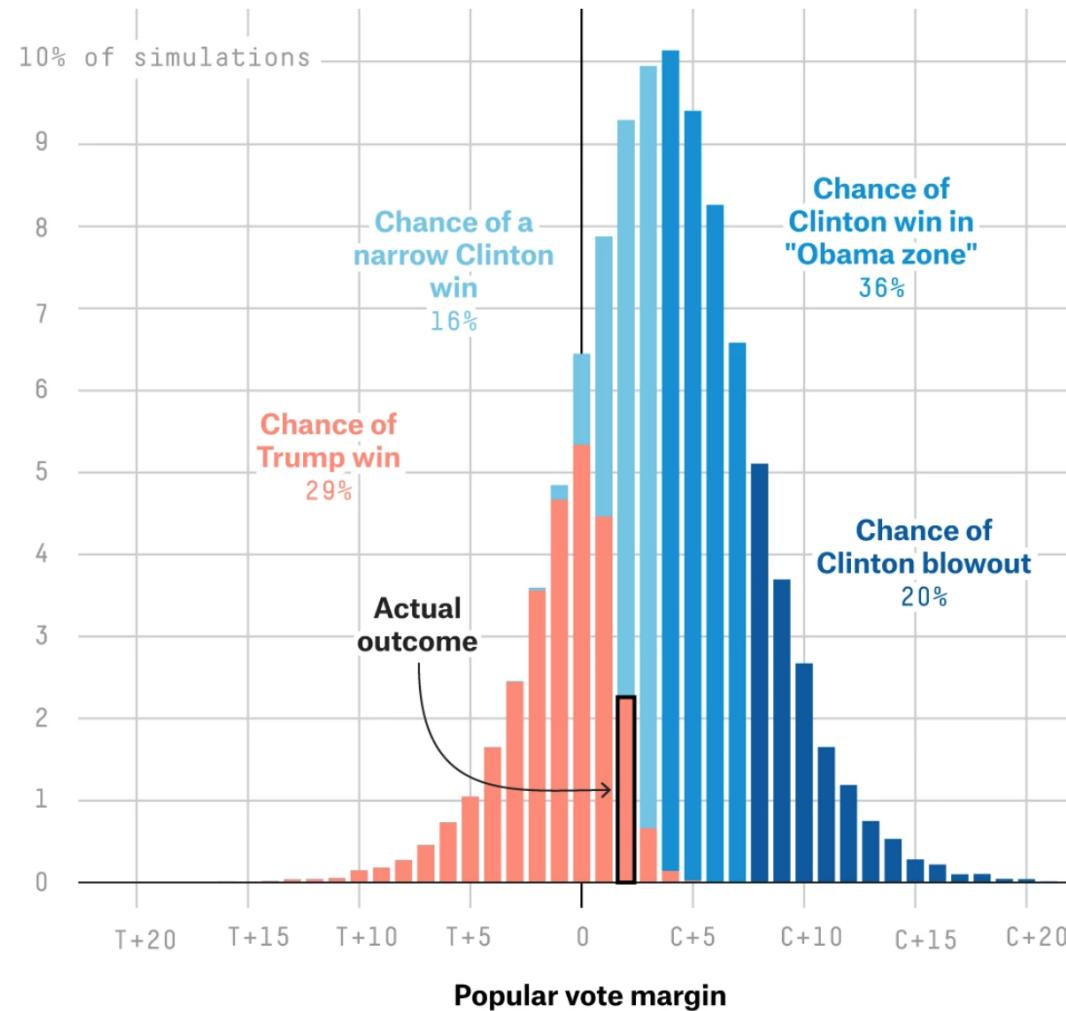
## What to expect from the Electoral College

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.



## FiveThirtyEight's final forecast for 2016

Likelihood of popular vote outcomes according to FiveThirtyEight's polls-only model at 9:35 a.m. on Election Day 2016. Based on 20,000 simulations.



## Crazy and not-so-crazy scenarios

Here are the chances we'll see these election outcomes.

Electoral College deadlock <small>no candidate gets 270 electoral votes</small>	1 . 0 %
Electoral College 269-269 tie	0 . 5 %
Recount <small>at least one decisive state within 0.5 ppt</small>	8 . 3 %
<b>Clinton wins popular vote</b>	<b>81 . 4 %</b>
Trump wins popular vote	18 . 6 %
<b>Clinton wins popular vote but loses Electoral College</b>	<b>10 . 5 %</b>
Trump wins popular vote but loses Electoral College	0 . 5 %
Johnson wins at least one electoral vote	0 . 3 %
McMullin wins at least one electoral vote	13 . 5 %
Clinton majority <small>wins at least 50 percent of the vote</small>	28 . 7 %
Trump majority <small>wins at least 50 percent of the vote</small>	2 . 3 %
Clinton landslide <small>double-digit popular vote margin</small>	6 . 1 %
Trump landslide <small>double-digit popular vote margin</small>	0 . 3 %
Map exactly the same as in 2012	0 . 2 %
Clinton wins at least one state Mitt Romney won in 2012	71 . 6 %
Trump wins at least one state President Obama won in 2012	85 . 0 %

Forecasts...are they bad?

# Forecasts are confusing

"The public has difficulty reasoning about the probability of a candidate's victory. [...] When one candidate is ahead, win-probabilities convey substantially more confidence that she will win compared to vote share estimates."

[...] Small differences in the election metric most familiar to the public—vote share estimates—generally correspond to very large differences in the probability of a candidate's chance of victory, and a high degree of statistical sophistication is required to map between the two.

Projecting confidence: How the probabilistic horse race confuses and demobilizes the public

Sean J. Westwood<sup>1</sup>, Solomon Messing<sup>2</sup>, and Yphtach Lelkes \*<sup>3</sup>

<sup>1</sup>Program in Quantitative Social Science, Dartmouth College

<sup>2</sup>Data Labs, Pew Research Center

<sup>3</sup>Annenberg School of Communication, University of Pennsylvania

February 6, 2018

# Forecasts are confusing

People tend to think in qualitative terms about the likelihood of events [...]; if candidate A has an 85% chance of victory, they see victory the likely outcome (this may help explain why after the 2016 election, so many criticized forecasters for "getting it wrong" [...] ).

[...] one-off event probabilities—candidate A has an 85% chance of winning—are often misunderstood [...] compared to statements such as "if the election was repeated 1,000 times, candidate A would win 850 times[...]"

Projecting confidence: How the probabilistic horse race confuses and demobilizes the public

Sean J. Westwood<sup>1</sup>, Solomon Messing<sup>2</sup>, and Yphtach Lelkes \*<sup>3</sup>

<sup>1</sup>Program in Quantitative Social Science, Dartmouth College

<sup>2</sup>Data Labs, Pew Research Center

<sup>3</sup>Annenberg School of Communication, University of Pennsylvania

February 6, 2018

# Forecasts distract us from what really matters

(The issues)

Can the press "mediate a campaign?"

# Is forecasting undemocratic?

Lowering turnout, feeding back into the system

Forecasts → Over-confidence in Outcome → Election Outcomes

Or...Are forecasts *necessary*

# Coming up next...

Two final lectures about *science in practice*