

Recreating, Understanding, and Visualizing FiveThirtyEight's Elections Forecast

Joseph Nelson, Matt Brems

PyData DC | 16 November 2018

Agenda

- * Intros
- * How to Forecast (Elections)
- * How FiveThirtyEight Forecasts Elections (and their not-so-secret sauce)
- * Creating a Dashboard with Python (Bokeh)
- * Simulating Results

Introductions

Joseph Nelson

@josephofiowa



Founded, bootstrapped, sold company
improving Congress's ability to process messages



Launched, taught 2000 hours, consults for
General Assembly's data science programs



Built data pipelines and dashboards
for the FB Government & Politics team



**Builds data science / machine learning
products. Empowers others to do the**

Matt Brems

@matthewbrems

Dedicated to putting the power of machine learning (responsibly) into the hands of others!



Fell in love with teaching; learned how to teach with the author of "Statistics for Dummies"



Fell in love with eating food; also built models sometimes.



Fell in love with FiveThirtyEight; also built models sometimes.



Fell in love with empowering others to do machine learning. (Also have taught over 3,500 hours - not that it's a competition...)



But who are **you**?

Raise your hand if...

- * You read FiveThirtyEight?
- * You've read FiveThirtyEight's electoral forecast methodology?
- * You've used Pandas ... SciPy ... NumPy ... Bokeh?
- * You feel comfortable with descriptive stats ... distributions ... simulation ... Bayesian stats?

How to Forecast (Elections)

**How do you forecast
an election?**

How do you forecast an election?

(This is not rhetorical. Shout out your answers.)

How do you forecast an election?

(This is not rhetorical. Shout out your answers.)

(Please. This awkward silence is awful.)

<https://goo.gl/eqW8Xr>

You said...

- * Horserace polls
- * National mood
- * Presidential sentiment
- * Turnout
- * Geography
- * Demographics
- * Expert opinions
- * Fundraising
- * Social media following
- * ...and more!

Much of this is correct. And we could list an infinite number of explanatory factors. But your time is not infinite.

Much of this is correct. And we could list an infinite number of explanatory factors. But your time is not infinite.

So, before we continue, a more fundamental question...

How do you forecast?

- * Brainstorm causal variables
- * Break the problem into achievable bits
- * Collect, clean data
- * Test
- * Accept done is better than perfect

**“All models are
wrong. Some are
useful.”**

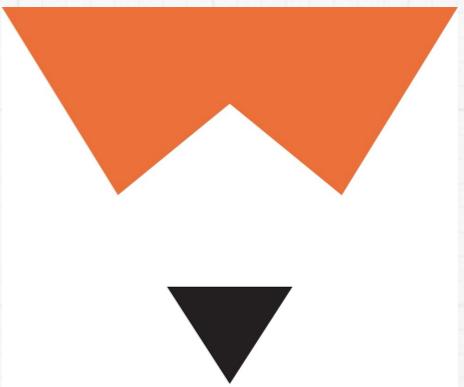
- George Box, 1976

How do you forecast an election?

- * Brainstorm causal variables
- * Break the problem into achievable bits
- * Collect, clean data
- * Test
- * Accept done is better than perfect

How do you forecast an election?

- * Brainstorm causal variables
- * Break the problem into achievable bits:
 - * Horse-racing polling
 - * Fundamentals
 - * Experts
 - * Noise



FiveThirtyEight

How do you forecast an election?

- * Brainstorm causal variables
- * Break the problem into achievable bits
- * Collect, clean data
- * Test
- * Accept ~~done~~ is better than perfect

To be
continued

How FiveThirtyEight Forecasts Elections

(And their not-so-secret sauce)

How FiveThirtyEight Forecasts Elections

TOP SECRET

LAST EDIT OCT. 17, 2018

How FiveThirtyEight's House, Senate And Governor Models Work

Filed under [Methodology](#)

[See our latest predictions](#)

the details

The principles behind our House, Senate and gubernatorial forecasts [should be familiar](#) to our longtime readers. They take lots of polls, perform various types of adjustments to them, and then blend them with other kinds of empirically useful indicators (what we sometimes call “the fundamentals”) to forecast each race. Then they account for the uncertainty in the forecast and simulate the election thousands of times. Our models [are probabilistic in nature](#); we do a *lot* of thinking about these probabilities, and the goal is to develop probabilistic estimates that hold up well under real-world conditions. For instance, when we launched the 2018 House forecast,

<https://goo.gl/hiZXdL>

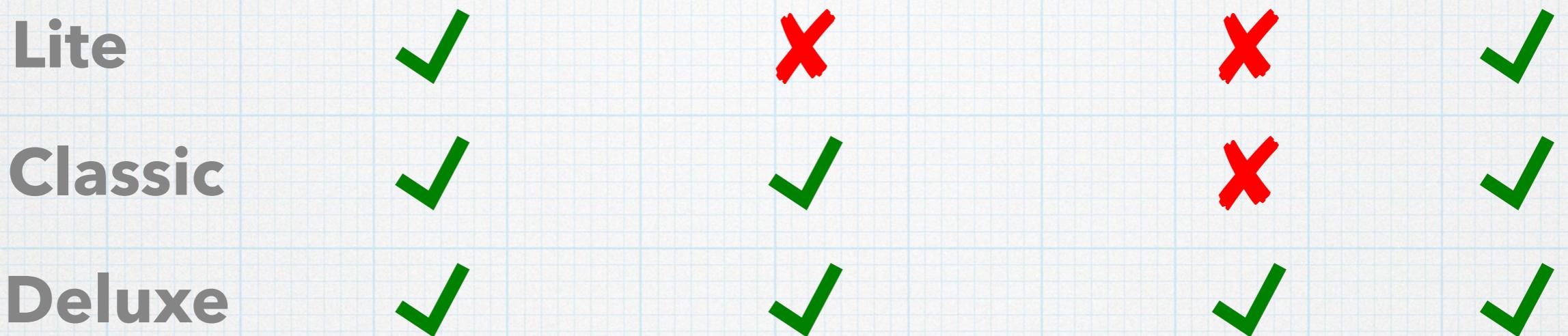
Forecast = Polling + Fundamentals + Experts + Noise

That's it. Mostly.

Forecast = Polling + Fundamentals + Experts + Noise

| | | | | |
|----------------|---|---|---|---|
| Lite | ✓ | ✗ | ✗ | ✓ |
| Classic | ✓ | ✓ | ✗ | ✓ |
| Deluxe | ✓ | ✓ | ✓ | ✓ |

Forecast = Polling + Fundamentals + Experts + Noise



Within-sample accuracy of forecasting methods

Share of races called correctly based on House elections from 1998 to 2016

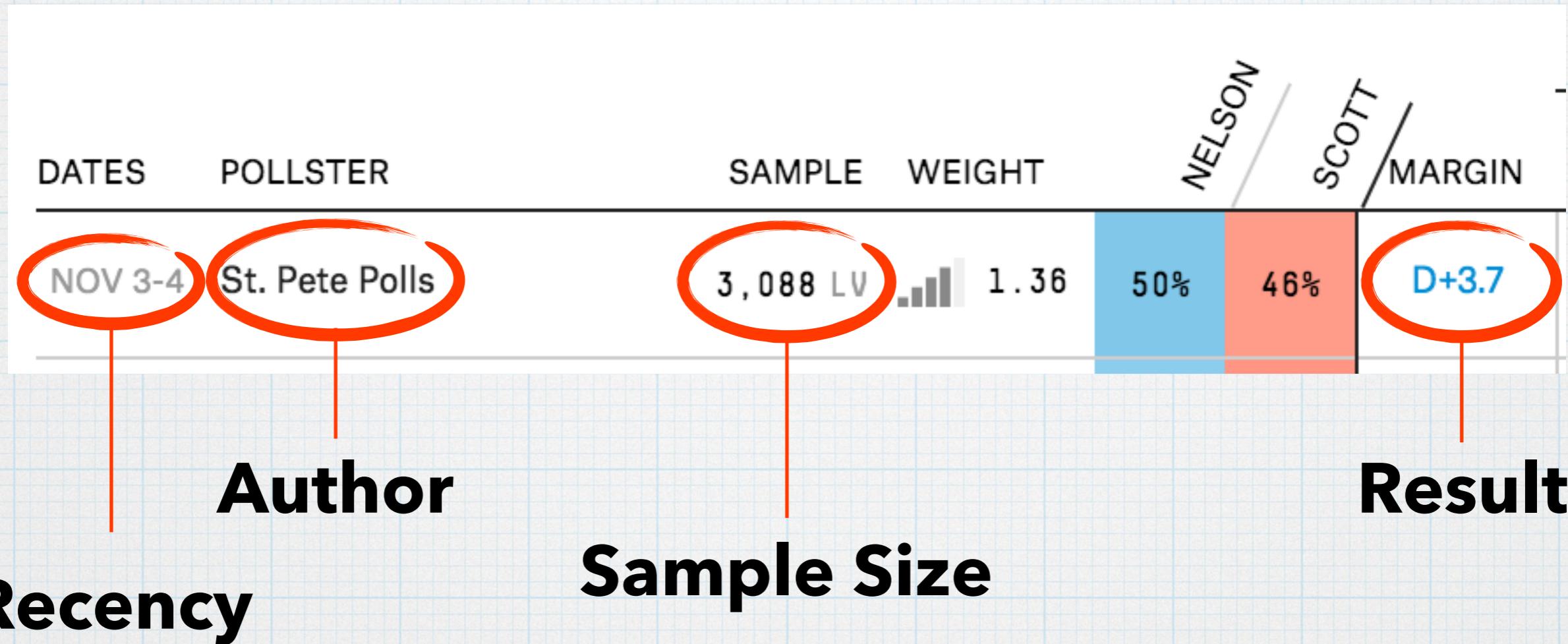
| FORECAST | 100 DAYS BEFORE ELECTION | ELECTION DAY |
|--|--------------------------|--------------|
| Lite model (poll-driven) | 94.2% | 96.2% |
| Fundamentals alone | 95.4 | 95.7 |
| Classic model (Lite model + fundamentals) | 95.4 | 96.7 |
| Expert ratings alone* | 94.8 | 96.6 |
| Deluxe model (Classic model + expert ratings) | 95.7 | 96.9 |

* Based on the average ratings from Cook Political Report, Inside Elections/The Rothenberg Political Report, Sabato's Crystal Ball and CQ Politics. Where the expert rating averages out to an exact toss-up, the experts are given credit for half a win.

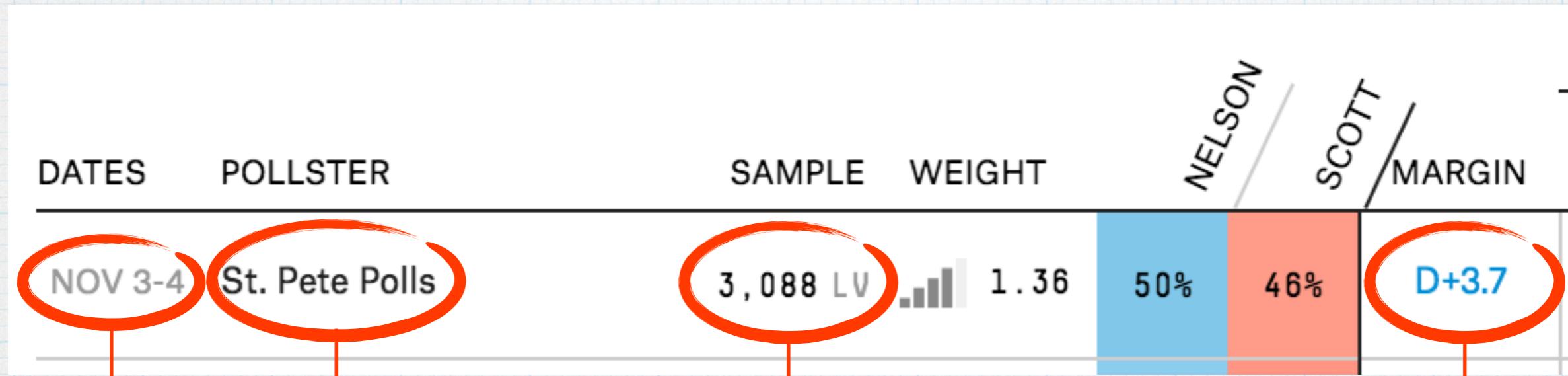
Forecast = Polling + Fundamentals + Experts + Noise

| DATES | POLLSTER | SAMPLE | WEIGHT | | | SCOTT | ADJUSTMENTS | | | |
|--------------|-----------------------|----------|--------|--------|--------|-------|--------------|------------------|---------------|-----------------|
| | | | | NELSON | MARGIN | | LIKELY VOTER | TIME-LINE | HOUSE EFFECTS | ADJUSTED MARGIN |
| NOV 3-4 | St. Pete Polls | 3,088 LV | 1.36 | 50% | 46% | D+3.7 | — | <0.1 ▶ | ◀ 1.0 | D+4.7 |
| NOV 2-4 | Harris Interactive | 600 LV | 0.69 | 45% | 49% | R+3.9 | — | <0.1 ▶ | ◀ 2.3 | R+1.6 |
| OCT 29-NOV 4 | Quinnipiac University | 1,142 LV | 2.26 | 51% | 44% | D+7.0 | — | ◀ <0.1 | 1.0 ▶ | D+6.1 |
| NOV 1-3 | Emerson College | 784 LV | 2.17 | 50% | 45% | D+5.0 | — | <0.1 ▶ | ◀ 0.6 | D+5.6 |
| NOV 1-3 | Harris Interactive | 600 LV | 0.21 | 45% | 47% | R+1.7 | — | <0.1 ▶ | ◀ 2.3 | D+0.6 |
| NOV 1-3 | Research Co. | 450 RV | 0.99 | 47% | 46% | D+1.0 | 0.2 ▶ | <0.1 ▶ | ◀ 0.5 | D+1.3 |
| NOV 1-2 | Change Research D | 1,236 LV | 1.33 | 50% | 48% | D+2.0 | — | <0.1 ▶ | 1.5 ▶ | D+0.5 |
| NOV 1-2 | St. Pete Polls | 2,733 LV | 0.96 | 48% | 49% | R+1.6 | — | <0.1 ▶ | ◀ 1.0 | R+0.6 |
| OCT 31-NOV 2 | Harris Interactive | 600 LV | 0.18 | 46% | 48% | R+2.6 | — | ◀ <0.1 | ◀ 2.3 | R+0.3 |
| OCT 30-NOV 2 | Marist College | 595 LV | 1.86 | 50% | 46% | D+4.0 | — | ◀ <0.1 | 0.5 ▶ | D+3.5 |
| | | | | | | | | Weighted average | D+3.0 | |

Forecast = Polling + Fundamentals + Experts + Noise



Forecast = Polling + Fundamentals + Experts + Noise



Author
Recency

Sample Size

Result

UPDATED MAY 30, 2018 AT 11:00 AM

FiveThirtyEight's Pollster Ratings

Based on the historical accuracy and methodology of each firm's polls.

[Read more](#) [Download the data](#) [See the latest polls](#)

| | | Ratings | Definitions | Search for a pollster | | | | | | | |
|--|-----------------|-----------------------------|------------------|-----------------------|----------------------|------------------------|-------------------------|-----------|---------------|--------------------|--|
| POLLSTER | METHOD | LIVE CALLER WITH CELLPHONES | NCPP/AAPOR/ROPER | POLLS ANALYZED | SIMPLE AVERAGE ERROR | RACES CALLED CORRECTLY | ADVANCED PREDICTIVE +/- | 538 GRADE | BANNED BY 538 | MEAN-REVERTED BIAS | |
| SurveyUSA | IVR/online/live | • | 777 | 4.6 | 90% | -1.1 | -0.9 | A | | D+0.1 | |
| Rasmussen Reports/Pulse Opinion Research | IVR/online | | 711 | 5.3 | 78% | +0.2 | +0.6 | C+ | | R+1.5 | |
| Zogby Interactive/JZ Analytics | Online | | 464 | 5.6 | 78% | +0.6 | +1.0 | C | | R+0.9 | |

Forecast = Polling + Fundamentals + Experts + Noise

| DATES | POLLSTER | SAMPLE | WEIGHT | NELSON | SCOTT | ADJUSTMENTS | | | |
|---------------|-----------------------|----------|--------|--------|-------|---------------------|-----------|---------------|-----------------|
| | | | | | | LIKELY VOTER MARGIN | TIME-LINE | HOUSE EFFECTS | ADJUSTED MARGIN |
| NOV 3-4 | St. Pete Polls | 3,088 LV | 1.36 | 50% | 46% | D+3.7 | — | <0.1 ▶ | 1.0 ▲ D+4.7 |
| NOV 2-4 | Harris Interactive | 600 LV | 0.69 | 45% | 49% | R+3.9 | — | <0.1 ▶ | 2.3 ▲ R+1.6 |
| OCT 29- NOV 4 | Quinnipiac University | 1,142 LV | 2.26 | 51% | 44% | D+7.0 | — | ◀ <0.1 | 1.0 ▶ D+6.1 |
| NOV 1-3 | Emerson College | 784 LV | 2.17 | 50% | 45% | D+5.0 | — | <0.1 ▶ | 0.6 ▲ D+5.6 |
| NOV 1-3 | Harris Interactive | 600 LV | 0.21 | 45% | 47% | R+1.7 | — | <0.1 ▶ | 2.3 ▲ D+0.6 |
| NOV 1-3 | Research Co. | 450 RV | 0.99 | 47% | 46% | D+1.0 | 0.2 ▶ | <0.1 ▶ | 0.5 ▲ D+1.3 |
| NOV 1-2 | Change Research D | 1,236 LV | 1.33 | 50% | 48% | D+2.0 | — | <0.1 ▶ | 1.5 ▶ D+0.5 |
| NOV 1-2 | St. Pete Polls | 2,733 LV | 0.96 | 48% | 49% | R+1.6 | — | <0.1 ▶ | 1.0 ▲ R+0.6 |
| OCT 31- NOV 2 | Harris Interactive | 600 LV | 0.18 | 46% | 48% | R+2.6 | — | ◀ <0.1 | 2.3 ▲ R+0.3 |
| OCT 30- NOV 2 | Marist College | 595 LV | 1.86 | 50% | 46% | D+4.0 | — | ◀ <0.1 | 0.5 ▶ D+3.5 |
| | | | | | | Weighted average | D+3.0 | | |

Forecast = Polling + Fundamentals + Experts + Noise

| MARGIN | ADJUSTMENTS | | | | HOUSE ADJUSTED MARGIN |
|--------|--------------|-----------|---------------|--------|-----------------------|
| | LIKELY VOTER | TIME-LINE | HOUSE EFFECTS | MARGIN | |
| D+3.7 | — | <0.1 ► | ◀ 1.0 | D+4.7 | |
| R+3.9 | — | <0.1 ► | ◀ 2.3 | R+1.6 | |
| D+7.0 | — | ◀ <0.1 | 1.0 ► | D+6.1 | |

real election result = β_0

+ β_1 "registered voter" result
+ β_2 time to election
+ β_3 pollster bias

Forecast = Polling + Fundamentals + Experts + Noise

What if polling is sparse?

| DATES | POLLSTER | SAMPLE | WEIGHT | | | MARGIN | ADJUSTMENTS | | | |
|------------------|---------------------------------|--------|--------|-----|-------|--------|--------------|-----------|---------------|-----------------|
| | | | | A | YOUNG | | LIKELY VOTER | TIME-LINE | HOUSE EFFECTS | ADJUSTED MARGIN |
| OCT. 29-NOV. 1 | Emerson College | 380 LV | 1.27 | 46% | 45% | D+0.9 | - | <0.1 | 0.6 | D+1.5 |
| OCT. 25-27 | Siena College/New York Times | 504 LV | 1.29 | 43% | 41% | D+2.2 | - | <0.1 | 0.3 | D+1.9 |
| SEP. 27-30 | Siena College/New York Times | 502 LV | 0.09 | 44% | 43% | D+1.3 | - | <0.1 | 0.3 | D+1.0 |
| SEP. 6-8 | Emerson College | 260 RV | 0.00 | 31% | 47% | R+16.0 | 0.2 | 0.1 | 0.6 | R+15.7 |
| SEP. 4-5 | DCCC Targeting Team D | 575 LV | 0.06 | 46% | 43% | D+3.0 | - | 0.1 | 2.9 | R+0.0 |
| JUL. 8-12 | Anzalone Liszt Grove Research D | 500 LV | 0.00 | 45% | 41% | D+4.0 | - | 0.6 | 4.0 | D+0.7 |
| Weighted average | | | | | | | | | | |
| D+1.6 | | | | | | | | | | |

KEY A = ADULTS RV = REGISTERED VOTERS V = VOTERS LV = LIKELY VOTERS R D = PARTISAN POLL

IA-03, one of the nation's most competitive House districts, recorded just six polls during the 2018 midterm cycle.

Forecast = Polling + Fundamentals + Experts + Noise

Sparse polling? Enter CANTOR.

Congressional Algorithm using Neighboring Typologies to Optimize Regression

Make a guess of what polling looks like in a given district based on similar districts.

CANTOR has two steps:

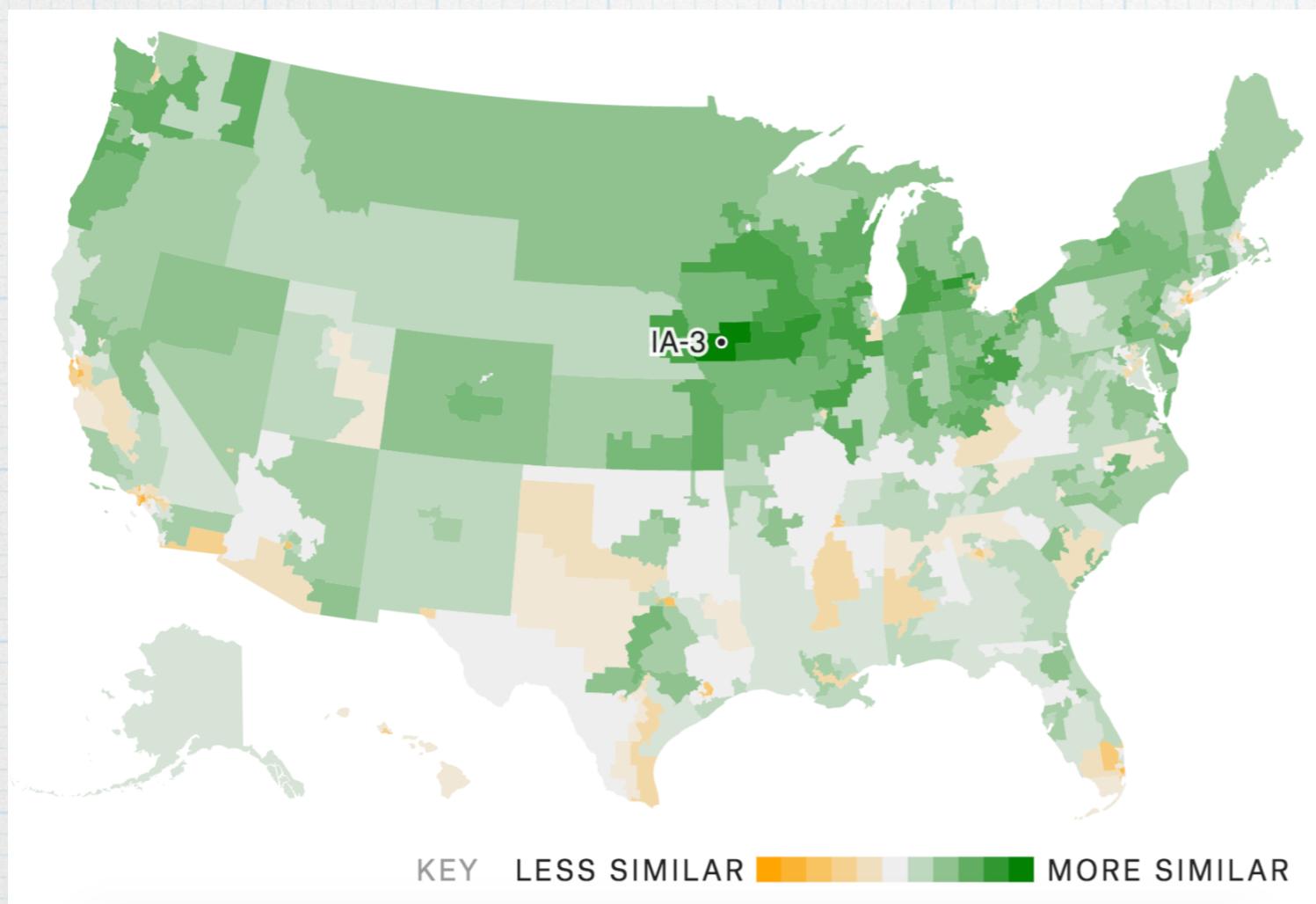
1. Initial race guesstimate based solely on FiveThirtyEight's partisan lean index
2. Adjust initial guesstimates based on KNN district-by-district similarity scores.

CANTOR is phased out as additional polls for a given race become available.

Forecast = Polling + Fundamentals + Experts + Noise

Sparse polling? Enter CANTOR.

Congressional Algorithm using Neighboring Typologies to Optimize Regression



**Districts most similar to the Iowa
3rd**

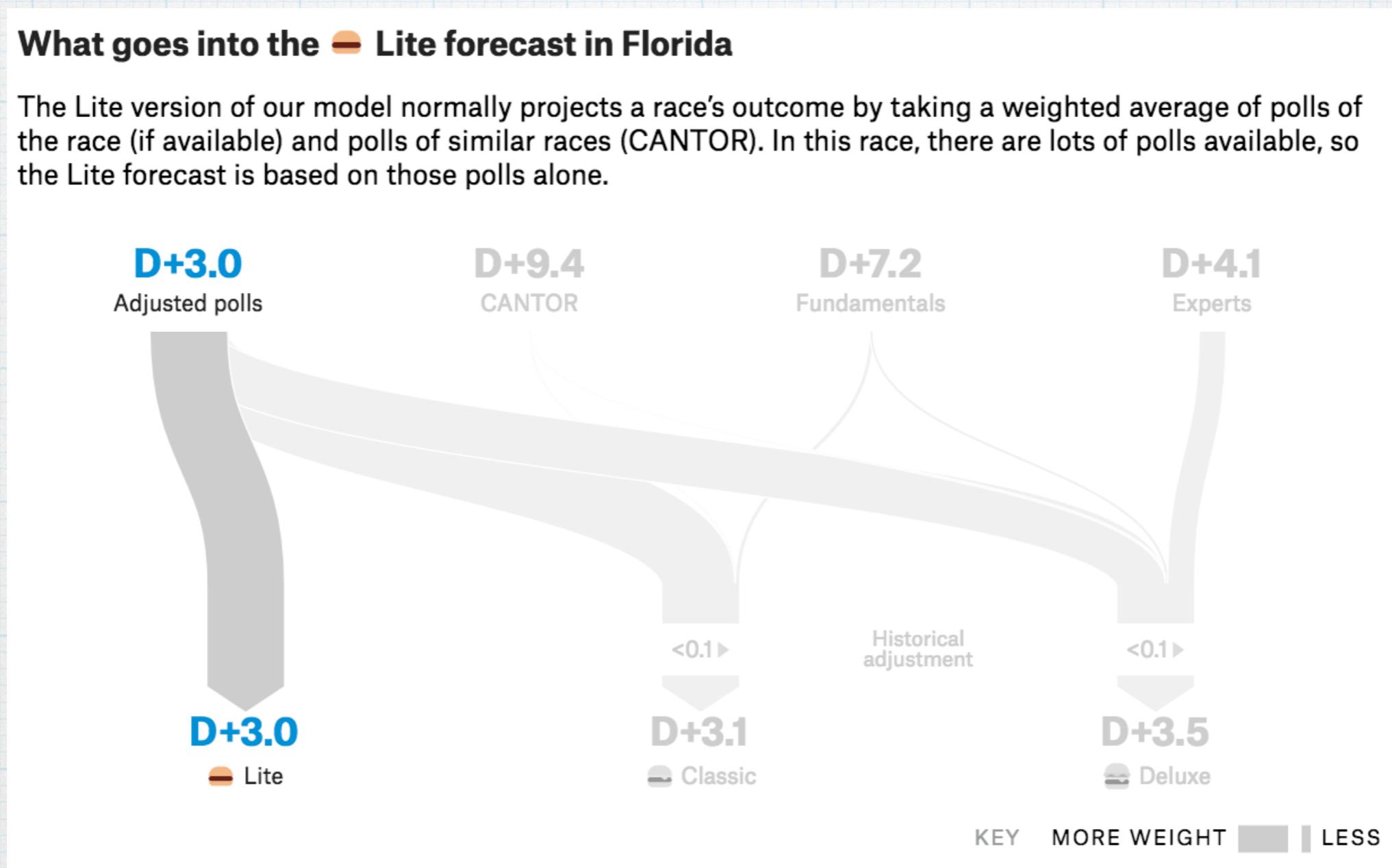
| | | SIM. SCORE | POLLING AVG. |
|---|-------|---------------|-----------------|
| 1 | IA-2 | 77 | D+13.3 |
| 2 | MI-8 | 72 | D+2.6 |
| 3 | IA-1 | 72 | D+8.8 |
| 4 | IL-13 | 71 | R+5.4 |
| 5 | MN-2 | 71 | D+6.1 |

Forecast = Polling + Fundamentals + Experts + Noise

So, what does polling say about FL's Senate race?

What goes into the 🍑 Lite forecast in Florida

The Lite version of our model normally projects a race's outcome by taking a weighted average of polls of the race (if available) and polls of similar races (CANTOR). In this race, there are lots of polls available, so the Lite forecast is based on those polls alone.



Forecast = Polling + Fundamentals + Experts + Noise

What does polling say about the IA-03 race?

What goes into the Lite forecast in the Iowa 3rd

The Lite version of our model projects a race's outcome by taking a weighted average of polls of a district (if available) and polls of similar districts (CANTOR).



Forecast = Polling + Fundamentals + Experts + Noise

U.S. House

- * Historic margin of victory
- * Incumbency polling
- * Generic Congressional ballot
- * Fundraising
- * FiveThirtyEight Partisan lean index (50% to 2016 election; 25% to 2012 presidential and state legislative elections)
- * Congressional approval ratings
- * Incumbent scandals, if relevant
- * Incumbent voting call record
- * Political experience of the challenger

U.S. Senate

- ▼ *Incumbent historic margin of victory*
- ▲ Incumbency polling
- * Generic Congressional ballot
- * Fundraising
- * FiveThirtyEight Partisan lean index (50% to 2016 election; 25% to 2012 presidential and state legislative elections)
- * Congressional approval ratings
- * Incumbent scandals, if relevant
- * Incumbent voting call record
- ▲ Political experience of the challenger
- * Changes in state partisanship

Governorships

- ▼ *Incumbent historic margin of victory*
- ▲ Incumbency polling
- * Generic Congressional ballot
- ▲ Fundraising
- * FiveThirtyEight Partisan lean index (50% to 2016 election; 25% to 2012 presidential and state legislative elections)
- * Congressional approval ratings
- * Incumbent scandals, if relevant
- ~~* Incumbent voting call record~~
- ▲ Political experience of the challenger
- * Changes in state partisanship

Forecast = Polling + Fundamentals + Experts + Noise

Lite



Classic



Deluxe



Forecast = Polling + Fundamentals + Experts + Noise

What do the fundamentals say about FL's Senate race?

The "fundamentals"

The Classic and Deluxe versions of our model use several non-polling factors to forecast the vote share margin in each seat.

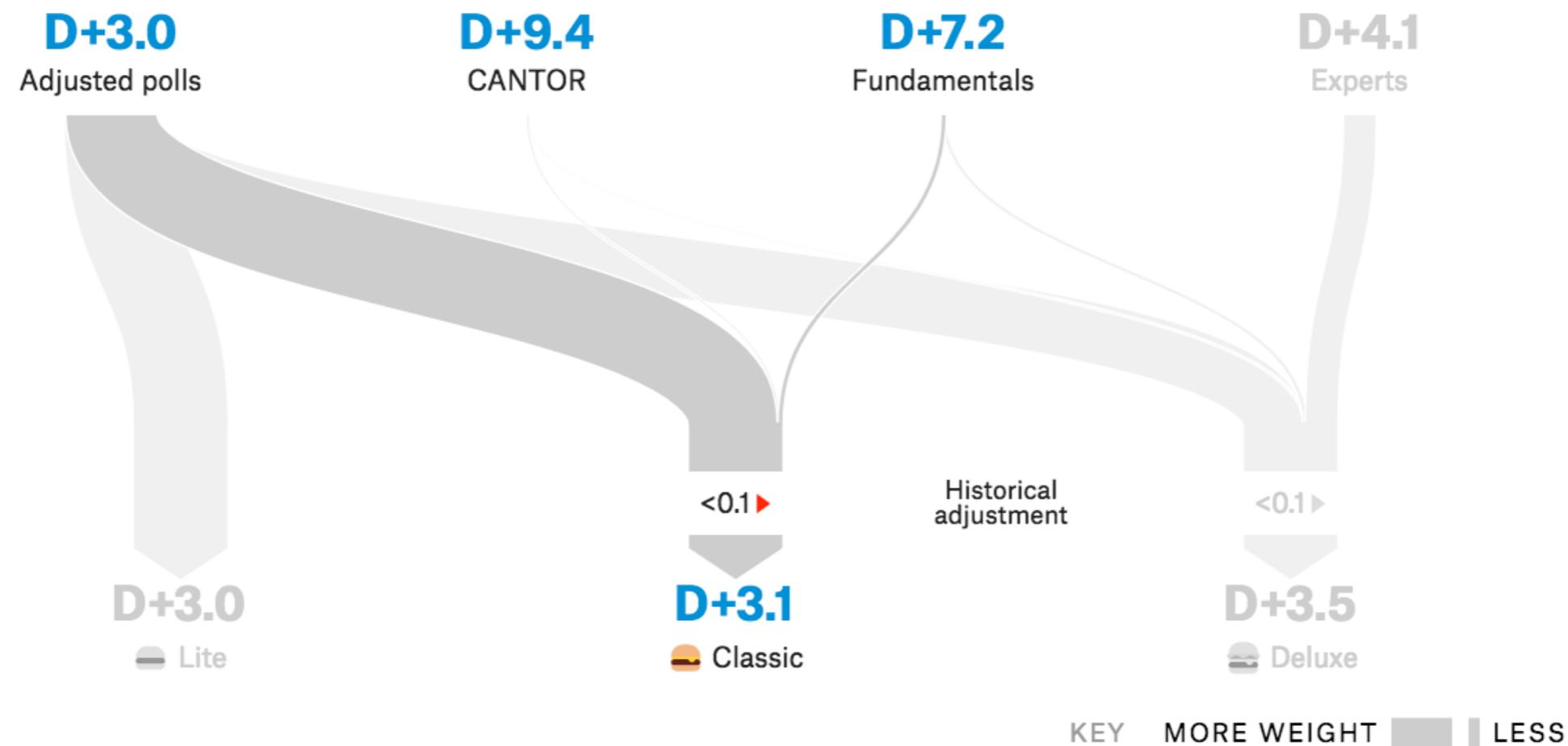
| FACTOR | IMPACT | EXPLANATION |
|---------------------------------------|--------|---|
| Incumbency | ← 5.9 | Bill Nelson has been elected to 3 terms. Congress has only a 20.1% approval rating, reducing the incumbency advantage. The incumbency advantage is largest in small states with distinctive demographic patterns. Florida is a large state with somewhat distinctive demographics, so incumbency won't help Nelson as much there. |
| State partisanship | 1.9 → | Florida is 5.4 percentage points more Republican-leaning than the country overall, based on how it has voted in recent presidential and state legislative elections. It voted for Trump in 2016 and Obama in 2012. Florida has been considerably more Republican in state legislative elections than in presidential elections. |
| Incumbent's margin in last election | + 2.8 | Nelson won by 13.0 percentage points in 2012. |
| Generic ballot | ← 7.2 | Democrats lead by an average of 8.6 percentage points in polls of the generic congressional ballot. |
| Fundraising | ← 1.6 | As of Oct. 17, Nelson had raised \$22,880,000 in individual contributions (59% of all such contributions to the major-party candidates); Rick Scott had raised \$15,994,000 (41%). |
| Incumbent's voting record in Congress | 0.2 → | Nelson has voted with Democrats 80% of the time in roll-call votes in recent sessions of Congress. |
| Challenger experience | 8.1 → | Scott's highest previous elected office is governor, which ranks in our top tier of elected offices. (Nelson's highest elected office is U.S. senator, but this is accounted for in our incumbency calculation.) |
| Scandals | 0.0 | Neither candidate is involved in a scandal. |
| Total | D+7.2 | |

Forecast = Polling + Fundamentals + Experts + Noise

How do the fundamentals affect FL's Senate race?

What goes into the 🍔 Classic forecast in Florida

The Classic version of our model projects a race's outcome by taking a weighted average of polls of the race (if available), polls of similar races (CANTOR) and non-polling factors (fundamentals). It is then reverted toward a mean based on long-term trends in midterms and presidential approval ratings.



Forecast = Polling + Fundamentals + Experts + Noise

What do the fundamentals say about IA-03?

The “fundamentals”

The Classic and Deluxe versions of our model use several non-polling factors to forecast the vote share margin in each district.

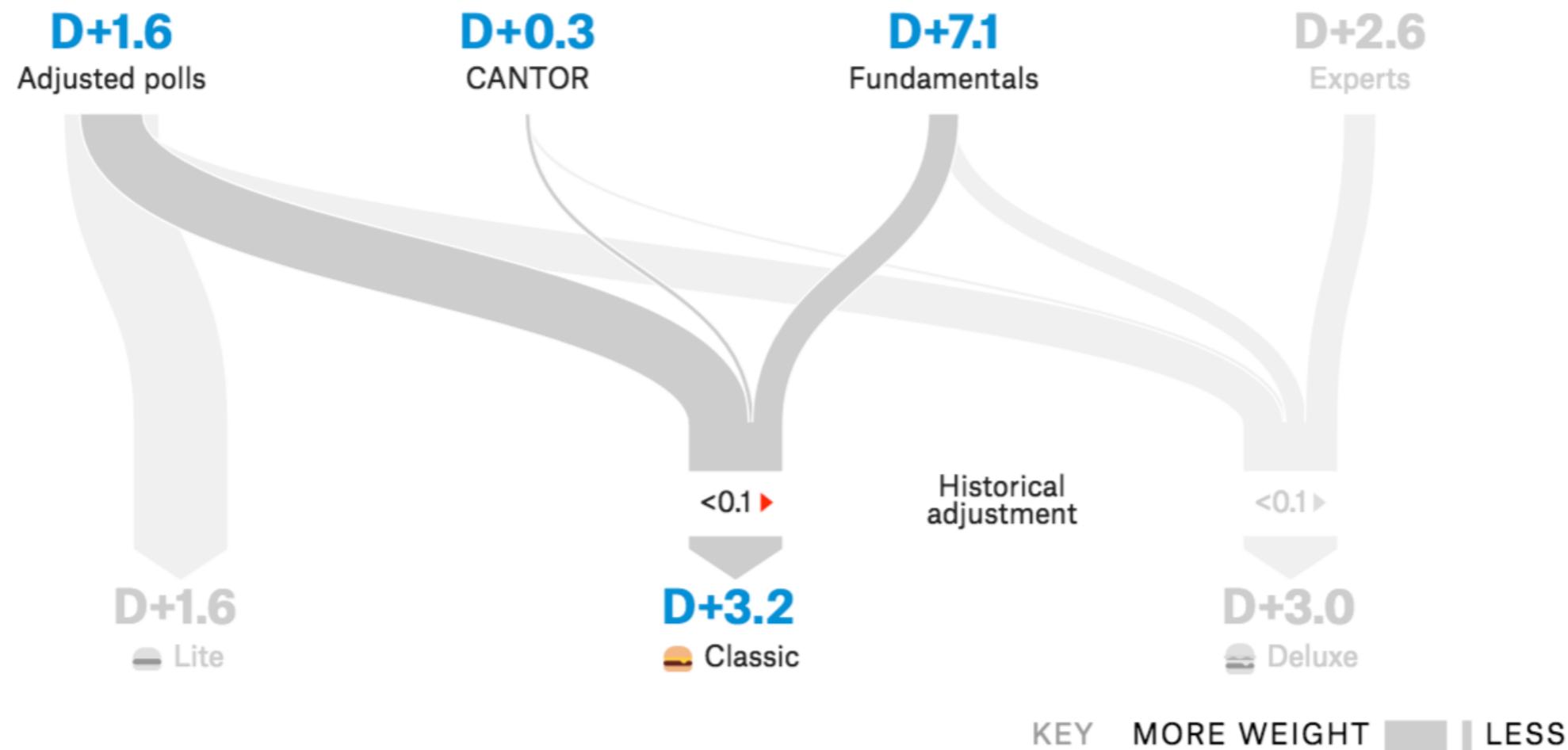
| FACTOR | IMPACT | EXPLANATION |
|---------------------------------------|--------|---|
| Incumbency | 3.1 ➔ | David Young has been elected to 2 terms. Congress has only a 20.1% approval rating, reducing the incumbency advantage. |
| District partisanship | 0.8 ➔ | IA-3 is 2.4 percentage points more Republican-leaning than the country overall, based on how it has voted in recent presidential and state legislative elections. It voted for Trump in 2016 and Obama in 2012. |
| Incumbent's margin in last election | 5.0 ➔ | Young won by 13.7 percentage points in 2016. |
| Generic ballot | ↔ 6.7 | Democrats lead by an average of 8.6 percentage points in polls of the generic congressional ballot. |
| Fundraising | ↔ 8.3 | As of Oct. 17, Cindy Axne had raised \$3,558,000 in individual contributions (84% of all such contributions to the major-party candidates); Young had raised \$695,000 (16%). |
| Incumbent's voting record in Congress | ↔ 1.7 | Young has voted with Republicans 85% of the time in roll-call votes in recent sessions of Congress. |
| Challenger experience | 0.7 ➔ | Axne has never held elected office. |
| Scandals | 0.0 | Neither candidate is involved in a scandal. |
| Total | D+7.1 | |

Forecast = Polling + Fundamentals + Experts + Noise

How do the fundamentals affect IA-03?

What goes into the 🍔 Classic forecast in the Iowa 3rd

The Classic version of our model projects a race's outcome by taking a weighted average of polls of a district (if available), polls of similar districts (CANTOR) and non-polling factors (fundamentals). It is then reverted toward a mean based on long-term trends in midterms and presidential approval ratings.



Forecast = Polling + Fundamentals + Experts + Noise

| | | | | |
|----------------|---|---|---|---|
| Lite | ✓ | ✗ | ✗ | ✓ |
| Classic | ✓ | ✓ | ✗ | ✓ |
| Deluxe | ✓ | ✓ | ✓ | ✓ |

Forecast = Polling + Fundamentals + Experts + Noise

Incorporating Experts is intuitive. Evaluate their historic performance, weight them accordingly, and shift projections.

Forecast = Polling + Fundamentals + Experts + Noise

Incorporating Experts is intuitive. Evaluate their historic performance, weight them accordingly, and shift projections.

The experts

- * Cook Political Report
- * Inside Elections
- * Sabato's Crystal Ball

The weights

What do ratings like "lean Republican" really mean?

| EXPERT RATING | AVERAGE MARGIN OF VICTORY |
|---------------------------------|---------------------------|
| Toss-up | 0 points |
| "Tilts" toward candidate | 4 points |
| "Leans" toward candidate | 7 points |
| "Likely" for candidate | 12 points |
| "Solid" or "safe" for candidate | 34 points |

Based on House races since 1998.

Forecast = Polling + Fundamentals + Experts + Noise

Lite



Classic



Deluxe



Forecast = Polling + Fundamentals + Experts + Noise

As with all models, we recognize our forecasts aren't going to be perfect. (Remember George Box?)

We can try to improve our model's performance by **quantifying the noise** that might exist.

Forecast = Polling + Fundamentals + Experts + Noise

| Within-Race Error (75%+) | Between-Race Correlation (<= 25%) |
|-------------------------------------|---|
| Quality of Polls | Regional/Demographic Correlation (i.e. Rust Belt) |
| Number of Polls | Incumbency (i.e. feelings toward Congress) |
| Elasticity of District | National Swing (i.e. systematic polling error) |

To the code!

Repo: <http://bit.ly/pydatadc18-code>