Statistical Models of Presidential Elections

Bayesian inference using polls and the fundamentals

G. Elliott Morris
Data journalist
The Economist

October 9, 2020
Prepared for a talk at the IU-Bloomington Workshop on Methods

2020 presidential election forecast*

President Senate House *Last updated 23 hours ago*

National forecast
How this works

COMPETITIVE STATES
Arizona
Florida
Georgia
Iowa
Michigan
Nevada
New Hampshire
North Carolina
Ohio
Pennsylvania
Texas
Wisconsin

Right now, our model thinks **Joe Biden is very likely to beat Donald Trump** in the electoral college.

	Chance of winning the electoral college	Chance of winning the most votes	Predicted range of electoral college votes (270 to win)
 Joe Biden Democrat	**around 9 in 10** or 91%	**better than 19 in 20** or 99%	**224-424**
 Donald Trump Republican	**around 1 in 10** or 9%	**less than 1 in 20** or 1%	**114-314**

The probability of an electoral-college tie is <1%

**as of October 8 at 7:05 PM*

Our model

1. National economic + political fundamentals

2. Decompose into state-level priors

3. Polls

Uncertainty is propogated throughout the models, incorporated via MCMC sampling in step 3.

National Fundamentals

What fundamentals?

i) Index of economic growth (1940 - 2016)

- eight different variables, scaled to measure the standard-deviation from average annual growth

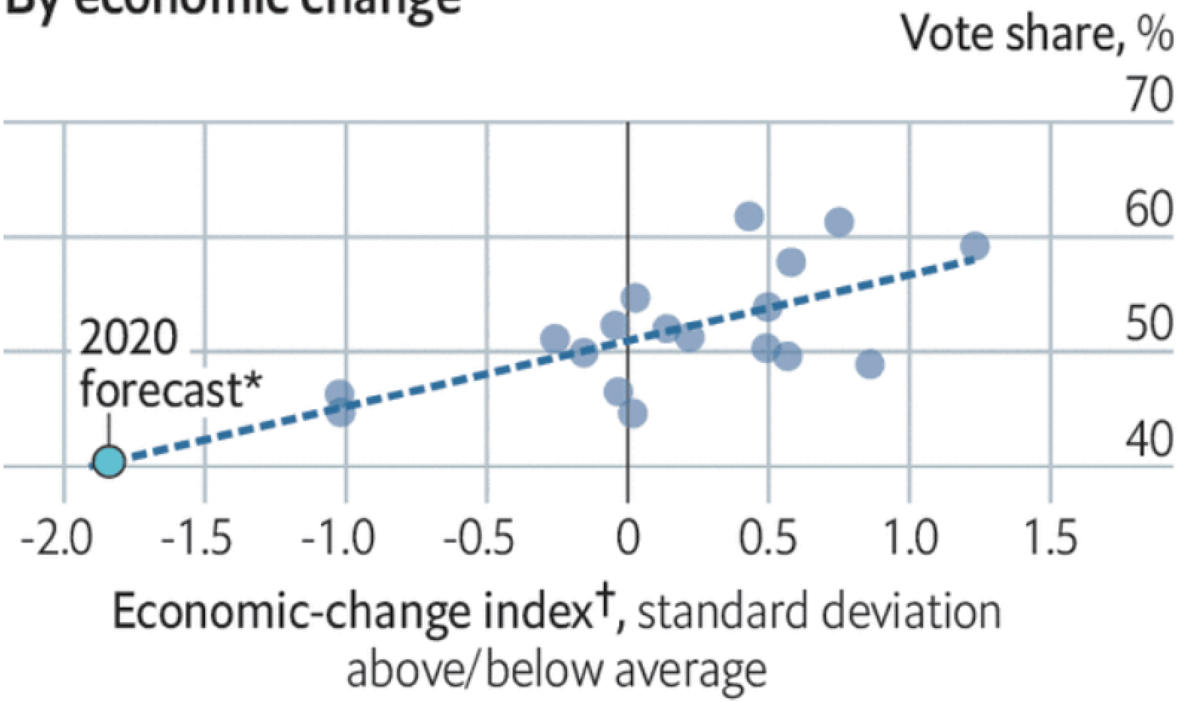
ii) Presidential approval (1948 - 2016)

iii) Polarization (1948 - 2016)

- measured as the share of swing voters in the electorate, per the ANES --- and interacted with economic growth

iv) Whether an incumbent is on the ballot

By economic change



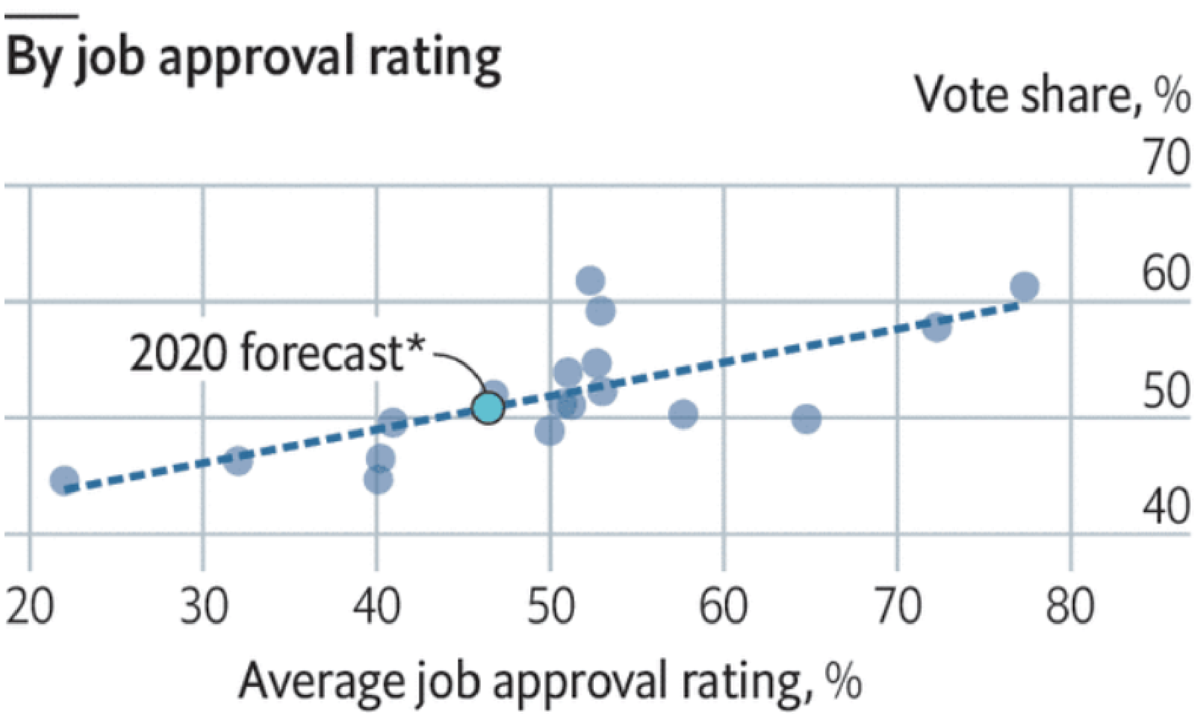
*224 days before the election
†Average of yearly change of nine economic indicators

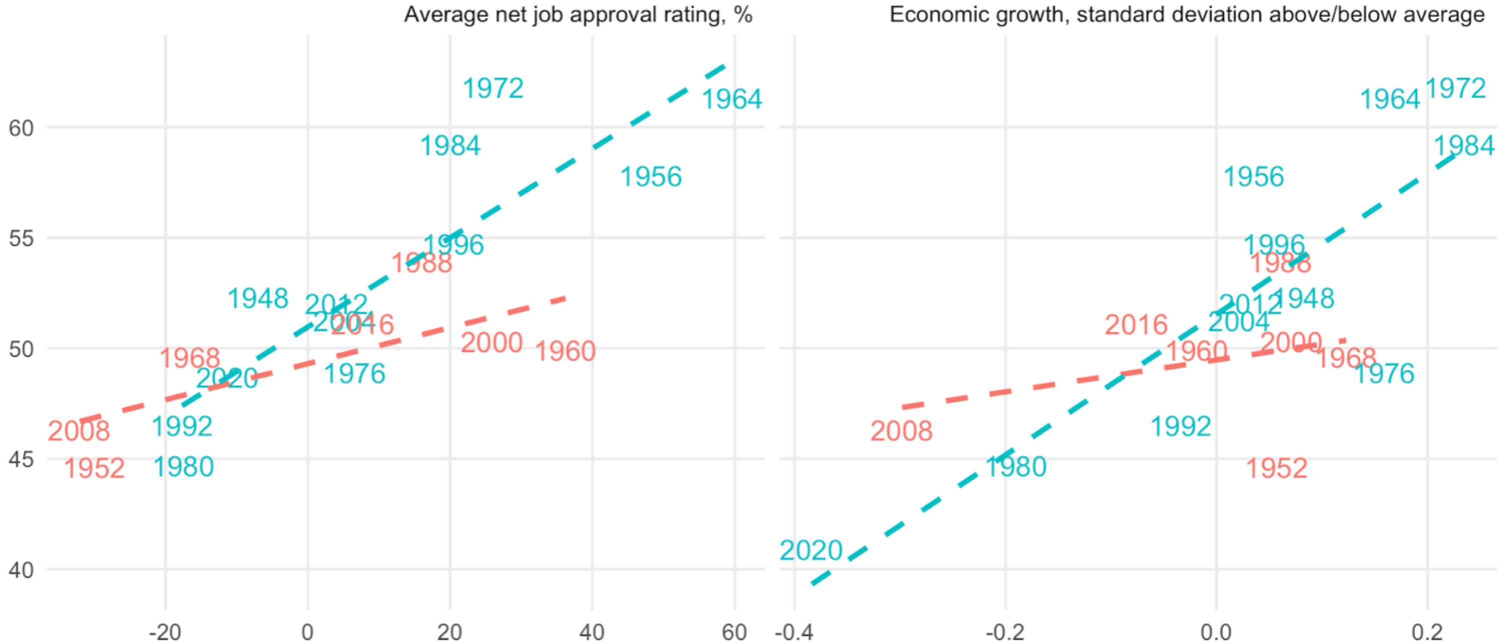
Sources: St Louis Fed; Yahoo; *The Economist*

The Economist

Leading indicators

United States, presidential elections, incumbent party's share of the major-party vote, %
1948-2016





*Values at 41 days before the election
Sources: St Louis Fed; Yahoo; The Economist

(blue = incumbents)

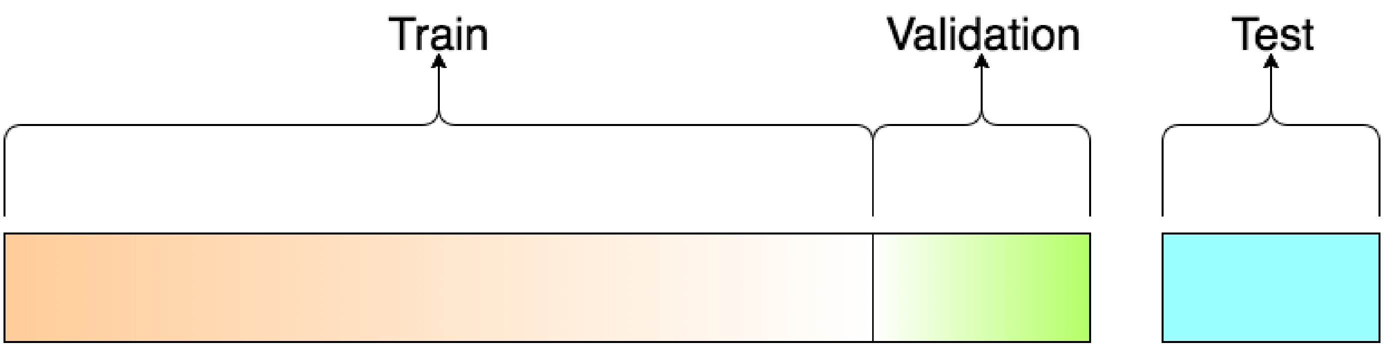
National fundamentals

Model formula:

vote ~ incumbent_running:economic growth:polarization + approval

Training

Model trained on 1948-2016 using elastic net regression with leave-one-out cross-validation



RMSE = 2.6 percentage points on two-party Democratic vote share

State-level prior

State-level prior

i) Train a model to predict the Democratic share of the vote in a state relative to the national vote, 1948-2016

- Variables are: lean in the last election, lean two elections ago, home state effects * state size, conditional on the national vote in the state

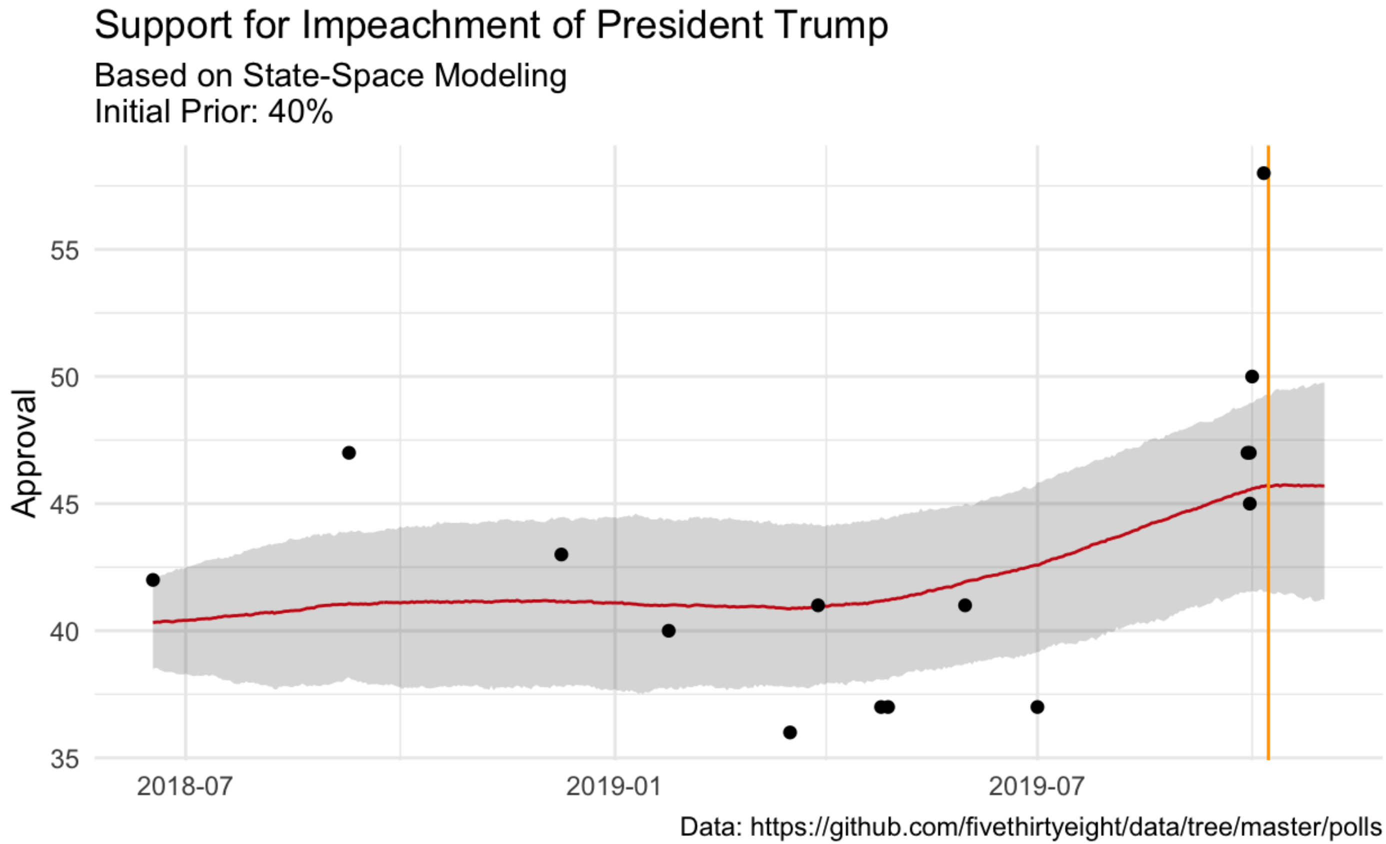
ii) Use the covariates to make predictions for 2020, *conditional on the national fundamentals prediction for every day*

ii) Simulate state-level outcomes to extract a mean and standard deviation

- Propogates uncertainty both from the LOOCV RMSE of the national model and the state-level model

Pooling the polls

It's just a trend through points...



(...but with some fancy extra stuff)

```
mu_b[:,T] = cholesky_ss_cov_mu_b_T * raw_mu_b_T + mu_b_prior;
for (i in 1:(T-1)) mu_b[:, T - i] = cholesky_ss_cov_mu_b_walk * raw_mu_b[:, T - i] + mu_b[:, T + 1 - i];
national_mu_b_average = transpose(mu_b) * state_weights;
mu_c = raw_mu_c * sigma_c;
mu_m = raw_mu_m * sigma_m;
mu_pop = raw_mu_pop * sigma_pop;
e_bias[1] = raw_e_bias[1] * sigma_e_bias;
sigma_rho = sqrt(1-square(rho_e_bias)) * sigma_e_bias;
for (t in 2:T) e_bias[t] = mu_e_bias + rho_e_bias * (e_bias[t - 1] - mu_e_bias) + raw_e_bias[t] * sigma_rho;
//*** fill pi_democrat
for (i in 1:N_state_polls){
  logit_pi_democrat_state[i] =
    mu_b[state[i], day_state[i]] +
    mu_c[poll_state[i]] +
    mu_m[poll_mode_state[i]] +
    mu_pop[poll_pop_state[i]] +
    unadjusted_state[i] * e_bias[day_state[i]] +
    raw_measure_noise_state[i] * sigma_measure_noise_state +
    polling_bias[state[i]];
}
```

Poll-level model

i. Latent state-level vote shares evolve as a random walk over time

- Pooling toward the state-level fundamentals more as we are further out from election day

ii. Polls are observations with measurement error that are debiased on the basis of:

- Pollster firm (so-called "house effects")
- Poll mode
- Poll population

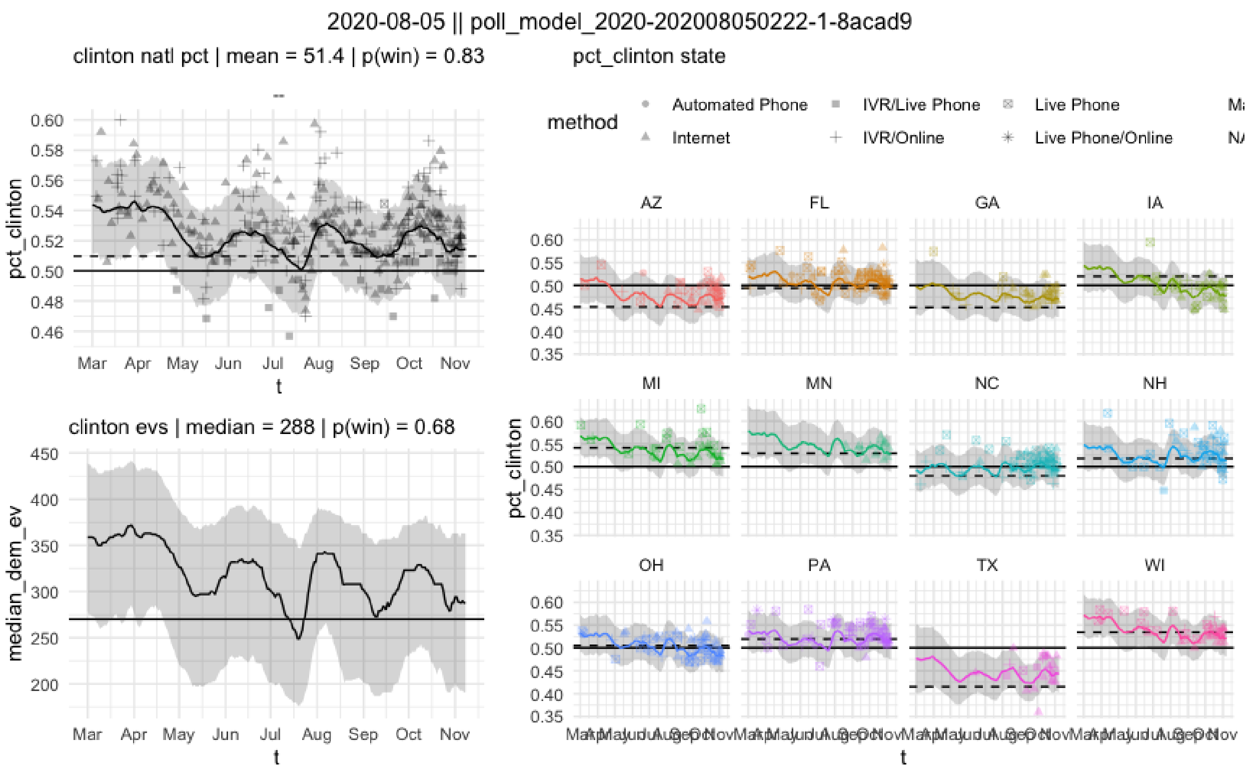
iii. Correcting for partisan non-response

- Whether a pollster weights by party registration or past vote
- Incorporated as a residual AR process

Tying it all together

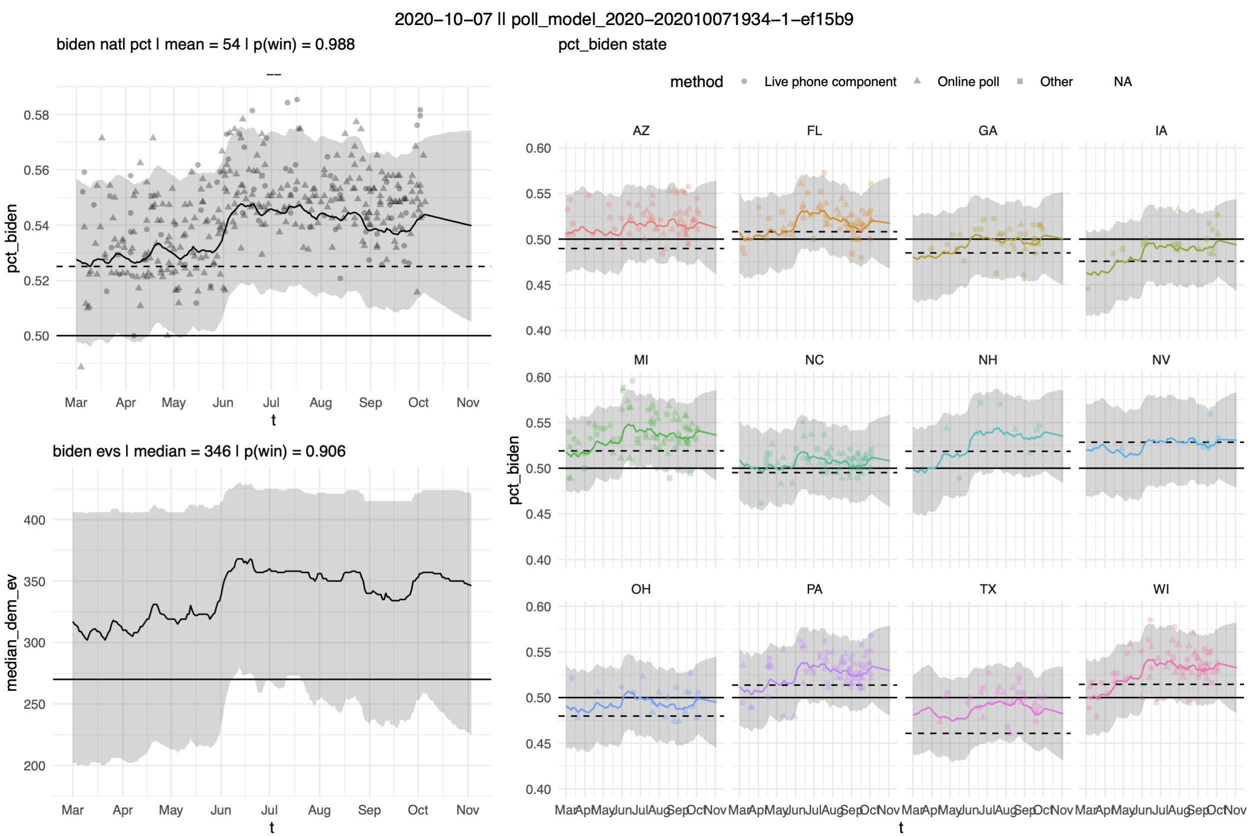
Tying it all together

1. 2016 election-day forecast:



Tying it all together

2. 2020 forecast*:



Q&A

Thank you!

Website: gelliottmorris.com

Email: elliott@thecrosstab.com

Twitter: @gelliottmorris

These slides were made with the xaringan package for R from Yihui Xie. They are available online at https://www.gelliottmorris.com/slides/2020-10-09-indiana-bloomington-methods-workshop/