

# Elections, Polls, Campaigns

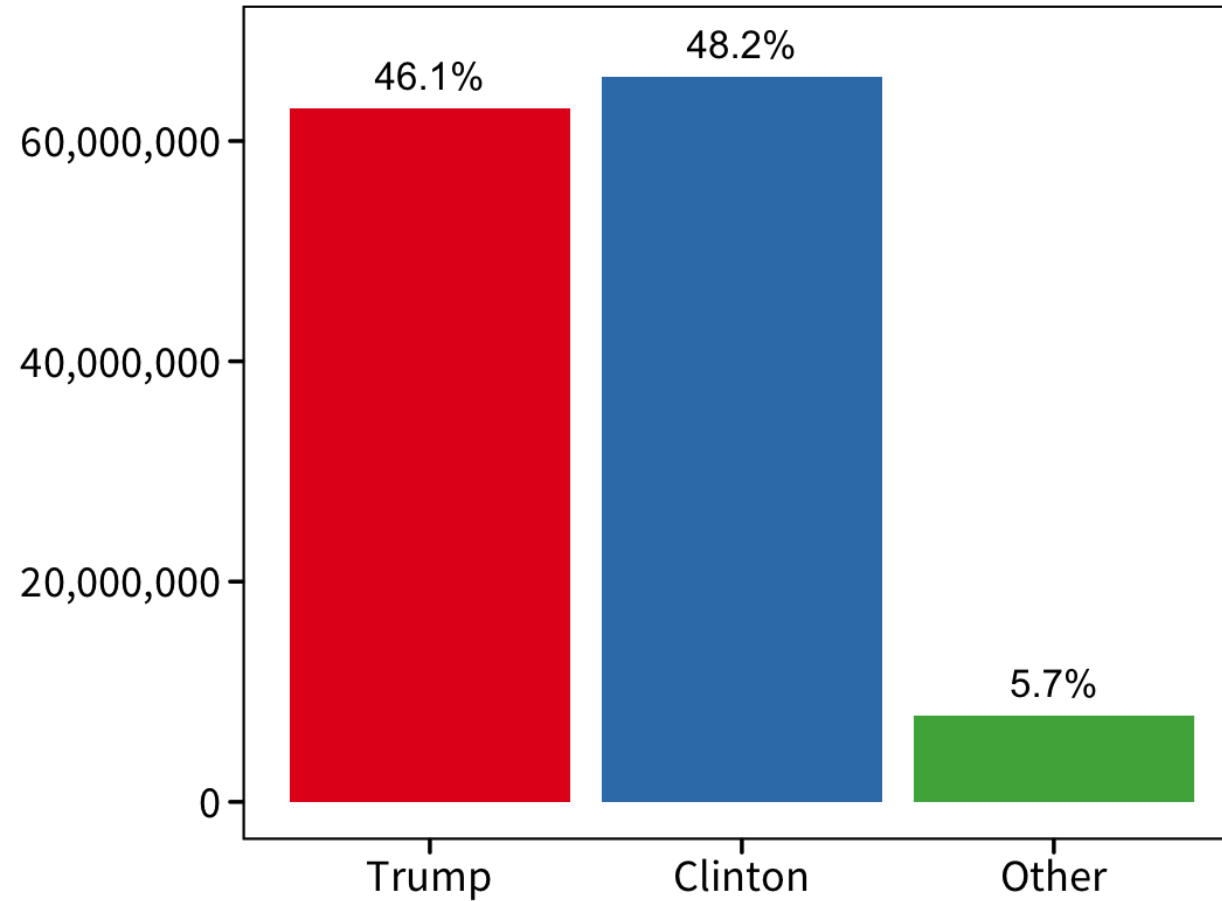
(or, “Being a better consumer of election news, or a producer of election news, depending on your career goals and whether you have enjoyed working with data so far”)

Understanding Political Numbers

April 10, 2019

# Understanding election data

# Votes for President



# Vote choices are *categorical* data

Individual data for votes

```
## [1] "Trump" "Trump" "Clinton" "Clinton" "Trump" "Clinton"
```

## Vote choices are *categorical* data

Individual data for votes

```
## [1] "Trump"  "Trump"  "Clinton" "Clinton" "Trump"  "Clinton"
```

What proportion voted for Trump?

## Vote choices are *categorical* data

Individual data for votes

```
## [1] "Trump" "Trump" "Clinton" "Clinton" "Trump" "Clinton"
```

What proportion voted for Trump?

```
# logical, returns TRUE or FALSE  
# head returns first few elements  
head(vote_choice == "Trump")
```

```
## [1] TRUE TRUE FALSE FALSE TRUE FALSE
```

# Vote choices are *categorical* data

Individual data for votes

```
## [1] "Trump" "Trump" "Clinton" "Clinton" "Trump" "Clinton"
```

What proportion voted for Trump?

```
# logical, returns TRUE or FALSE  
# head returns first few elements  
head(vote_choice == "Trump")
```

```
## [1] TRUE TRUE FALSE FALSE TRUE FALSE
```

```
# add up Trump votes, divided by total  
sum(vote_choice == "Trump") / length(vote_choice)
```

```
## [1] 0.461
```

# Vote choices are *categorical* data

Individual data for votes

```
## [1] "Trump" "Trump" "Clinton" "Clinton" "Trump" "Clinton"
```

What proportion voted for Trump?

```
# logical, returns TRUE or FALSE  
# head returns first few elements  
head(vote_choice == "Trump")
```

```
## [1] TRUE TRUE FALSE FALSE TRUE FALSE
```

```
# add up Trump votes, divided by total  
sum(vote_choice == "Trump") / length(vote_choice)
```

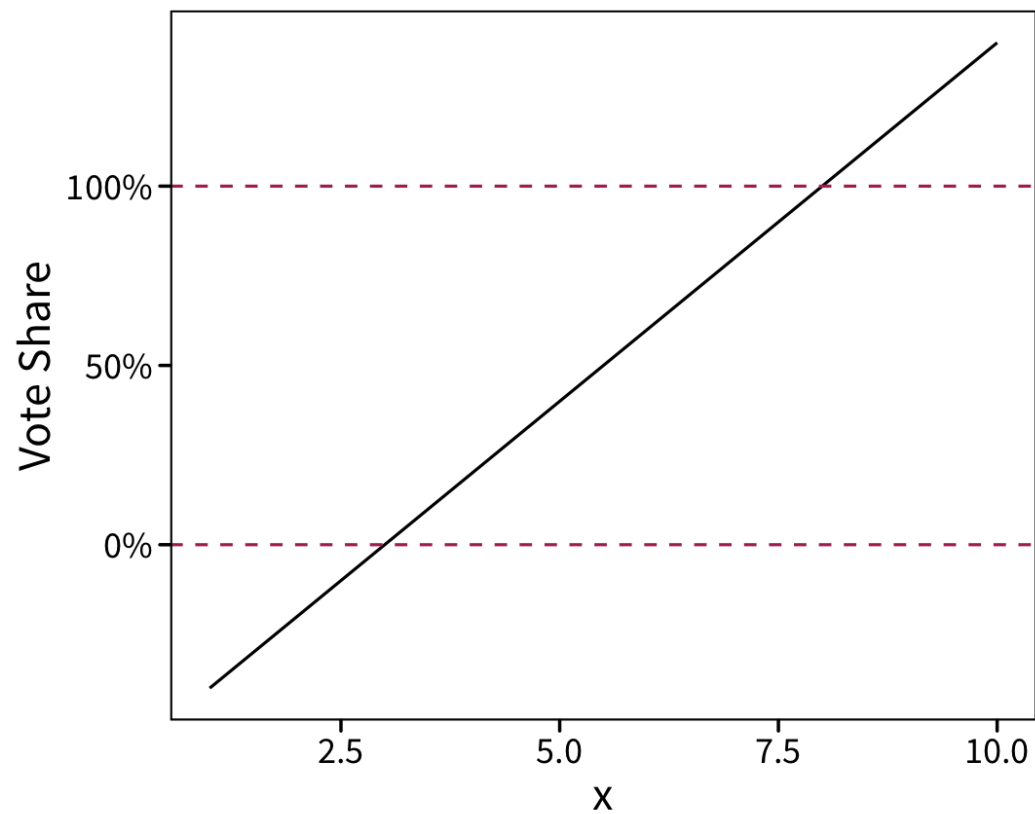
```
## [1] 0.461
```

```
# Or, proportions are means of a dummy variable  
mean(vote_choice == "Trump")
```

```
## [1] 0.461
```



If  $x$  affects votes (  $y$  )...



# Boundedness -> Nonlinearity

We have *linear* models but a *bounded* outcome...how?

# Boundedness -> Nonlinearity

We have *linear* models but a *bounded* outcome...how?

Unbounded representation of probability... "odds ratio"

$$\text{Odds} = \frac{p}{(1 - p)}$$

# Boundedness -> Nonlinearity

We have *linear* models but a *bounded* outcome...how?

Unbounded representation of probability... "odds ratio"

$$\text{Odds} = \frac{p}{(1 - p)}$$

"Two-to-one chance" that the Patriots win

# Boundedness -> Nonlinearity

We have *linear* models but a *bounded* outcome...how?

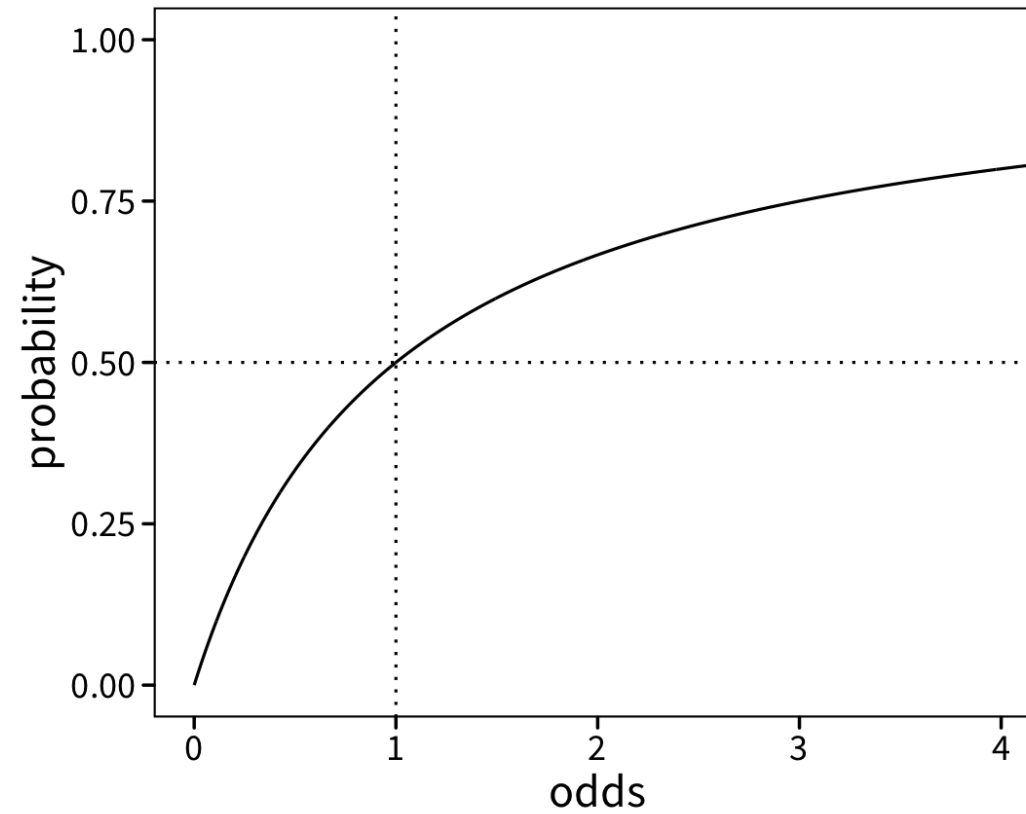
Unbounded representation of probability... "odds ratio"

$$\text{Odds} = \frac{p}{(1 - p)}$$

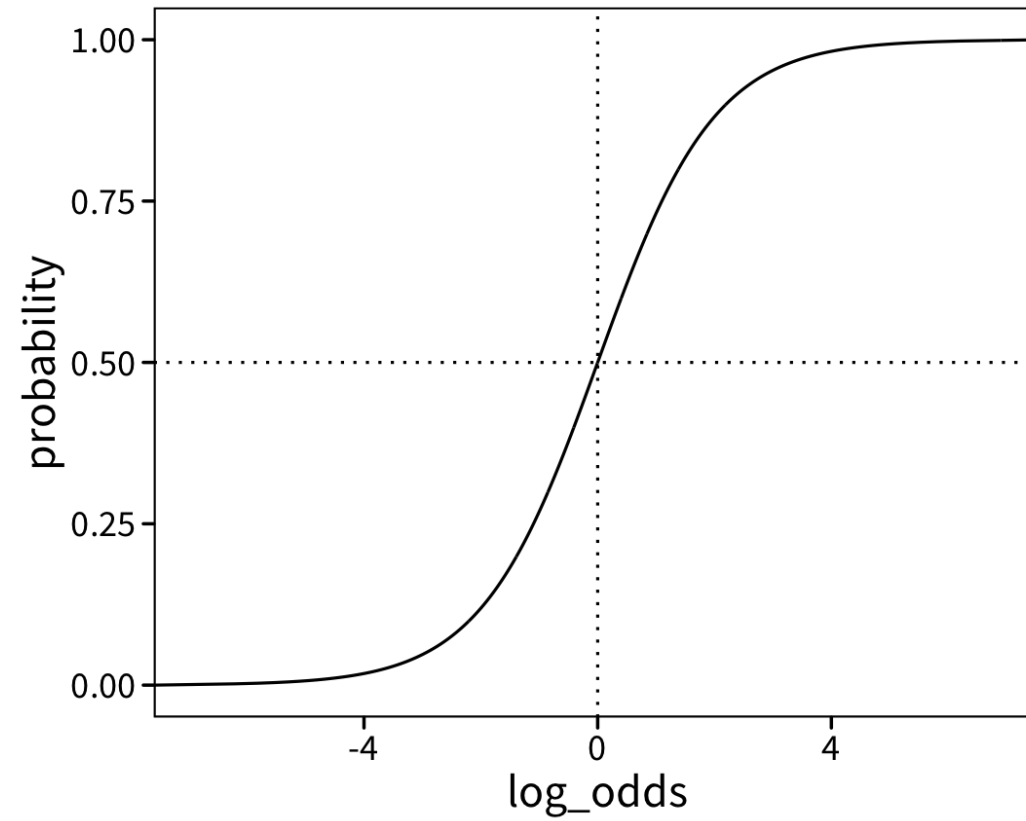
"Two-to-one chance" that the Patriots win

$$2 = \frac{\left(\frac{2}{3}\right)}{\left(\frac{1}{3}\right)}$$

# Odds and probability



# Odds and probability



# Regressions for proportions

Logistic model:

$$\log\left(\frac{p}{1-p}\right) = \alpha + \beta x + \epsilon$$



# Regressions for proportions

Logistic model:

$$\log\left(\frac{p}{1-p}\right) = \alpha + \beta x + \epsilon$$

Transform back to probability:

$$p = \frac{1}{1 + e^{-(\alpha + \beta x + \epsilon)}}$$

# Regressions for proportions

Logistic model:

$$\log\left(\frac{p}{1-p}\right) = \alpha + \beta x + \epsilon$$

Transform back to probability:

$$p = \frac{1}{1 + e^{-(\alpha + \beta x + \epsilon)}}$$

"Logit transformation" undone by the "inverse logit"

```
(x <- qlogis(0.75))
```

```
## [1] 1.098612
```

```
plogis(x)
```

```
## [1] 0.75
```

# Regressions for binary outcomes

Voted for Trump<sub>*i*</sub>  $\sim$  Bernoulli( $p_i$ )

# Regressions for binary outcomes

Voted for Trump  $_i \sim \text{Bernoulli}(p_i)$

$$\text{logit}(p_i) = \alpha + \beta x_i$$

# Regressions for binary outcomes

Voted for Trump<sub>*i*</sub>  $\sim$  Bernoulli( $p_i$ )

$$\text{logit}(p_i) = \alpha + \beta x_i$$

Hypothetical R example

```
glm(trump_vote ~ partisanship + education + income + race,  
    data = vote_data,  
    family = binomial(link = "logit"))
```

# Regressions for binary outcomes

Voted for Trump<sub>*i*</sub>  $\sim$  Bernoulli( $p_i$ )

$$\text{logit}(p_i) = \alpha + \beta x_i$$

Hypothetical R example

```
glm(trump_vote ~ partisanship + education + income + race,  
    data = vote_data,  
    family = binomial(link = "logit"))
```

GLM for "Generalized linear model": relationship between  $x$  and  $y$  isn't linear, but a transformation of  $x$  and  $y$  is linear

# Regressions for binary outcomes

Voted for Trump<sub>*i*</sub>  $\sim$  Bernoulli( $p_i$ )

$$\text{logit}(p_i) = \alpha + \beta x_i$$

Hypothetical R example

```
glm(trump_vote ~ partisanship + education + income + race,  
    data = vote_data,  
    family = binomial(link = "logit"))
```

GLM for "Generalized linear model": relationship between  $x$  and  $y$  isn't linear, but a transformation of  $x$  and  $y$  is linear

**A whole world of GLMs out there!**

Not requiring you to use them, but if you're interested, let's talk

Let's talk about the polls



# Don't trust the polls? You should. They've been fairly accurate so far this election cycle.



Analysis by [Harry Enten](#), CNN

Updated 4:38 PM ET, Wed May 2, 2018



ARTICLE

## Can We Trust the Polls?: It all depends

Michael W. Traugott · Sunday, June 1, 2003

MAY 14, 2018

## Can we still trust polls?

BY **COURTNEY KENNEDY**

## Survey: A majority of Americans don't believe polls are accurate

PHOTO BY GETTY IMAGES

## After 2016, Can We Ever Trust the Polls Again?

Data was a rare source of clarity in the 2008 and 2012 elections. In 2016, it failed catastrophically.

By **NICOLE NAREA** | December 14, 2016

Add to Pocket

# Don't trust the polls? You should. They've been fairly accurate so far this election cycle.



Analysis by [Harry Enten](#), CNN

Updated 4:38 PM ET, Wed May 2, 2018



ARTICLE

## Can We Trust the Polls?: It all depends

Michael W. Traugott · Sunday, June 1, 2003

MAY 14, 2018

## Can we still trust polls?

BY COURTNEY KENNEDY

## Survey: A majority of Americans don't believe polls are accurate

PHOTO: GETTY IMAGES

## After 2016, Can We Ever Trust the Polls Again?

Data was a rare source of clarity in the 2008 and 2012 elections. In 2016, it failed catastrophically.

By NICOLE NAREA | December 14, 2016

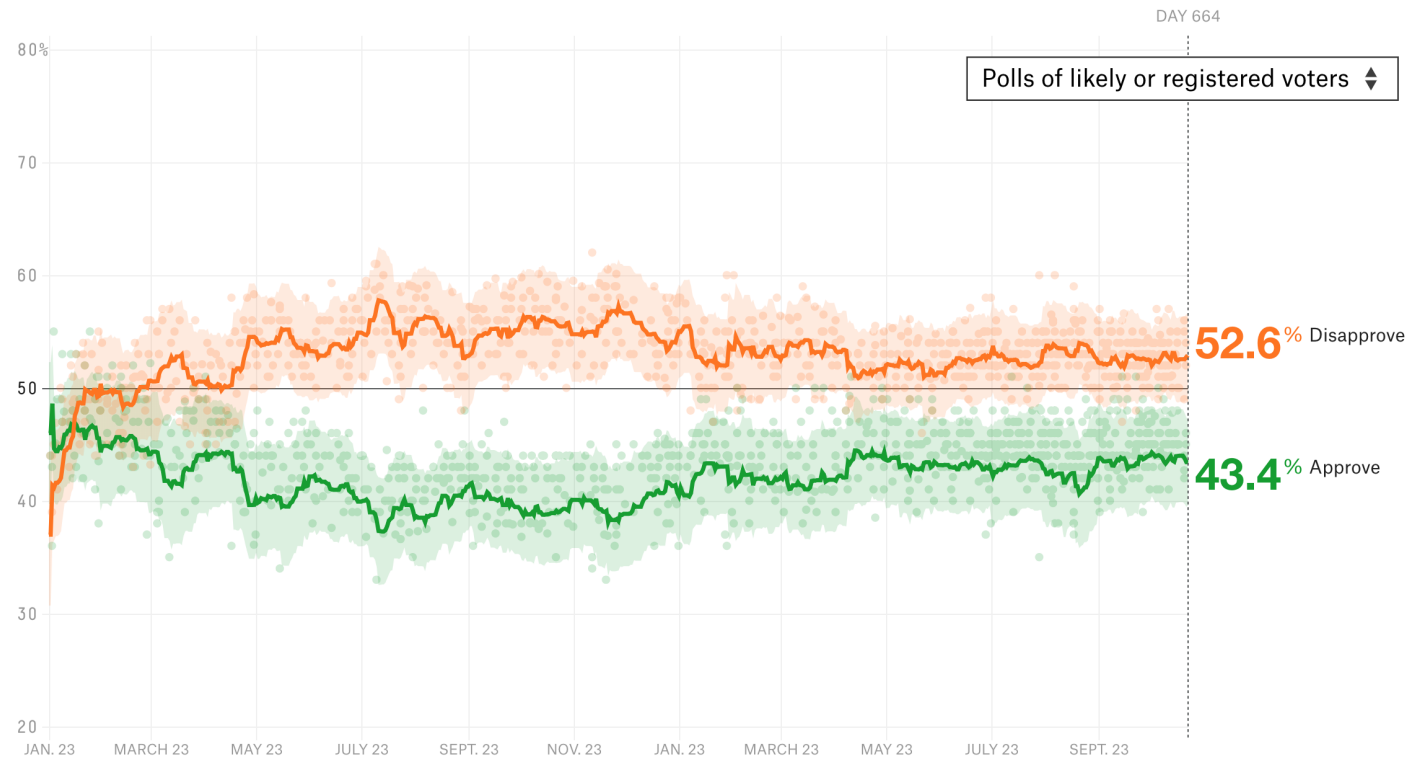
Add to Pocket

Polls aren't perfect (duh) but we're smart enough to understand why

Any poll has *sampling error*

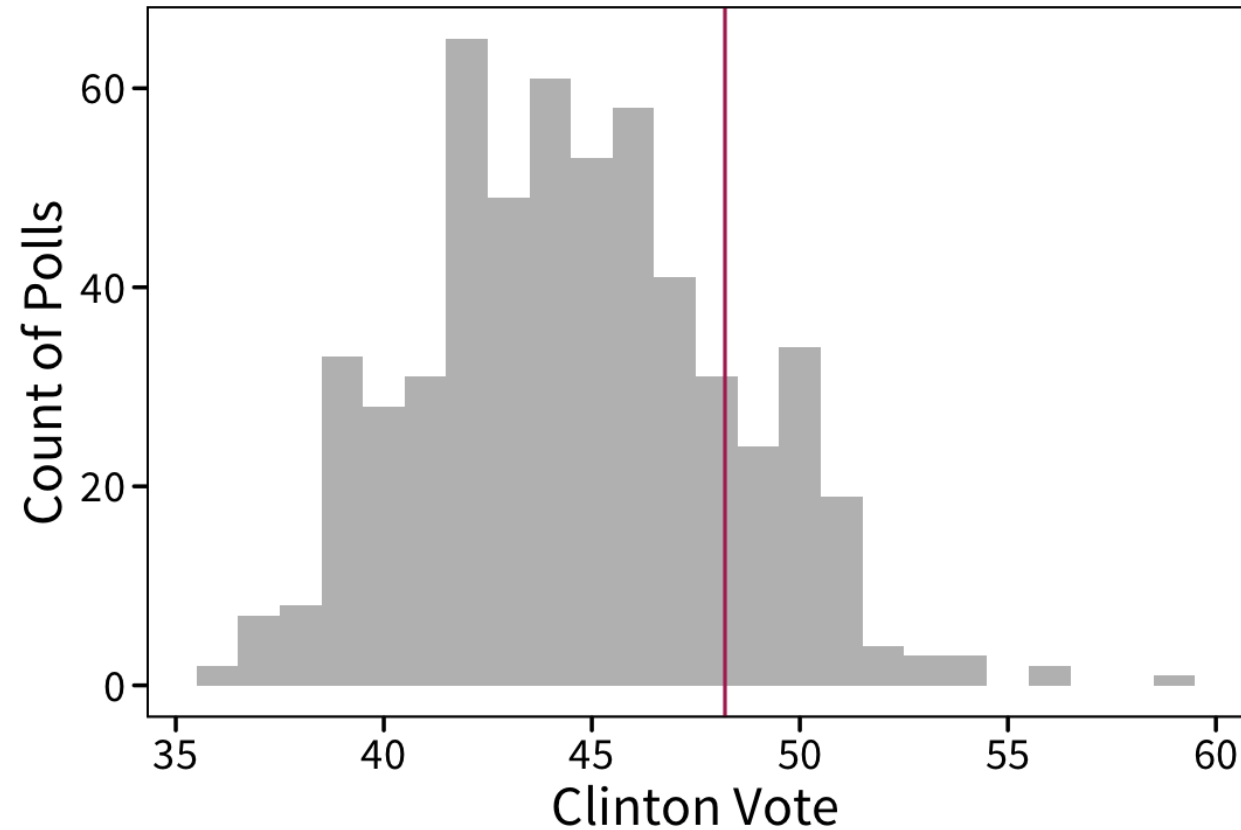
## How popular is Donald Trump?

An updating calculation of the president's approval rating, accounting for each poll's quality, recency, sample size and partisan lean. [How this works »](#)



# Histogram of 2016 national polls

3.5pt error on average



# Sources of polling error

Random sampling error: estimate is randomly different from the population being sampled

# Sources of polling error

Random sampling error: estimate is randomly different from the population being sampled

Nonrandom biases: is the population being sampled the *population that matters*?

# Sources of polling error

Random sampling error: estimate is randomly different from the population being sampled

Nonrandom biases: is the population being sampled the *population that matters*?

Pollsters try to minimize bias, but there are issues

- Target population: voters
- Sampling frame: who we can contact
- Response rate: who takes the poll if contacted?
- Weighting: making the sample *representative*

# Sources of polling error

Random sampling error: estimate is randomly different from the population being sampled

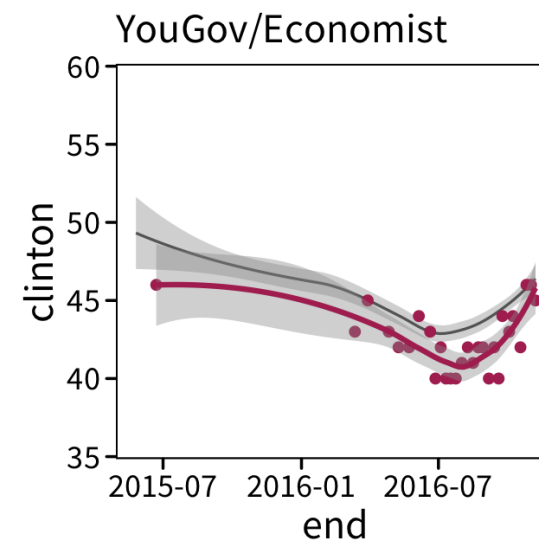
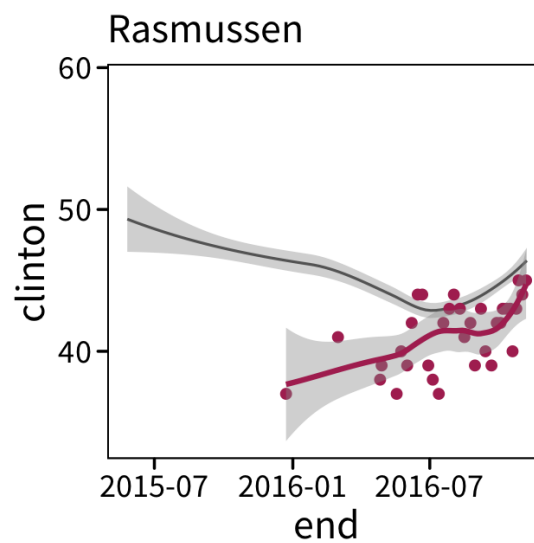
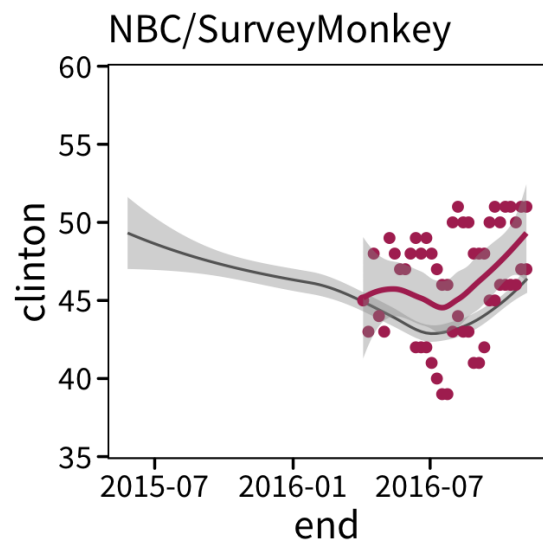
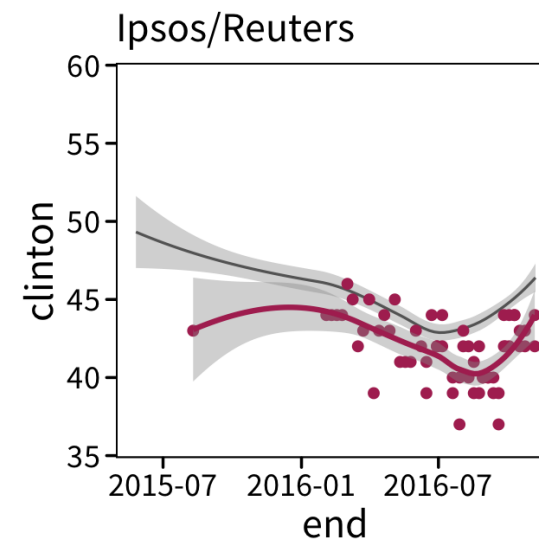
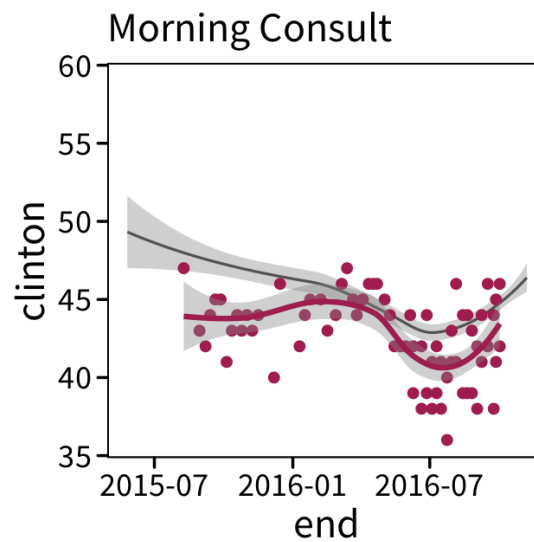
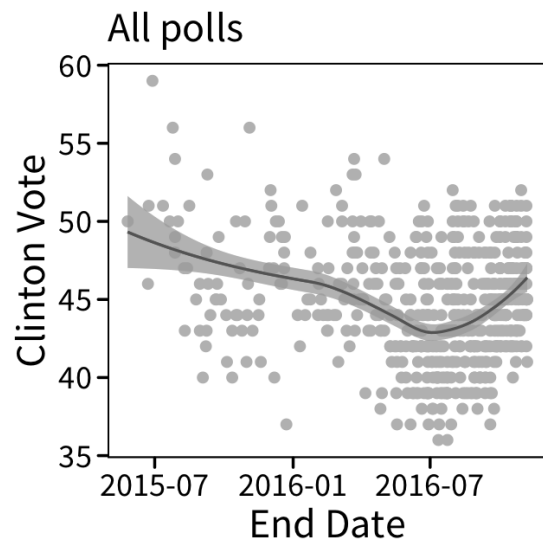
Nonrandom biases: is the population being sampled the *population that matters*?

Pollsters try to minimize bias, but there are issues

- Target population: voters
- Sampling frame: who we can contact
- Response rate: who takes the poll if contacted?
- Weighting: making the sample *representative*

"Total Survey Error" sampling error *plus everything else*

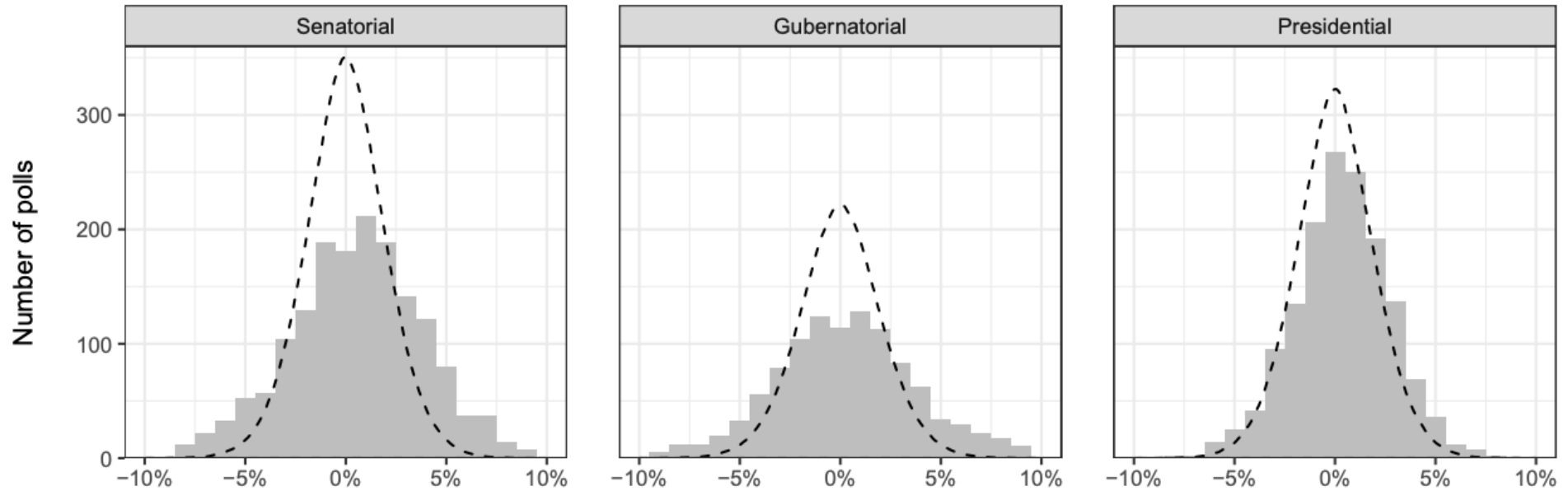




Grading the polls ([click here](#))

More grading ([click here](#))

Difference between poll results and election outcomes



Total Survey Error ([click here](#))

# Campaigns using data

Campaign tactics ([click here](#))

---

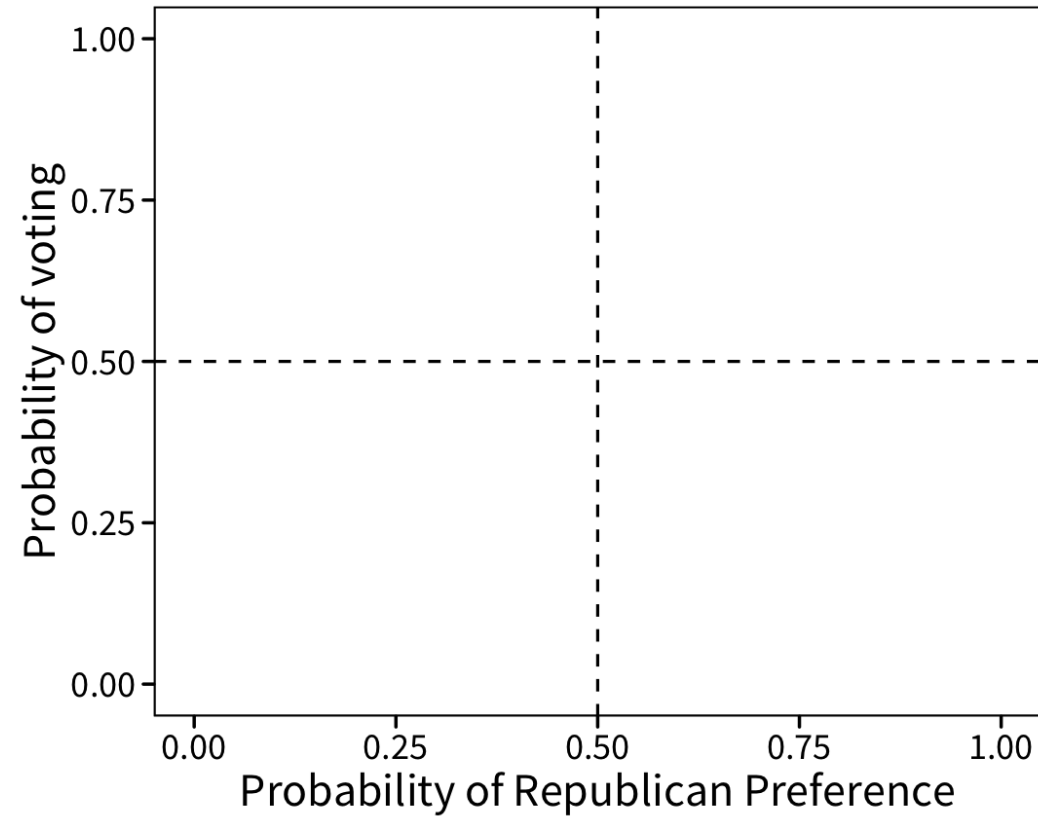
Intelligent Machines

---

# How Obama's Team Used Big Data to Rally Voters

How President Obama's campaign used big data to rally individual voters.

# Mobilization and Persuasion



# How?

## Data

- voter file, credit reports, web/ad data

## Modeling questions

- what is your party ID?
- Race?
- Ideology?

## Gathering data

- Surveys, experiments

## Interventions:

- TV, mail, web traffic, issue agenda



# Looking ahead

Check the syllabus for cool readings and podcasts and stuff

Next week: judiciary and paper guidelines

In section: regression practice (this week) and then presentations