

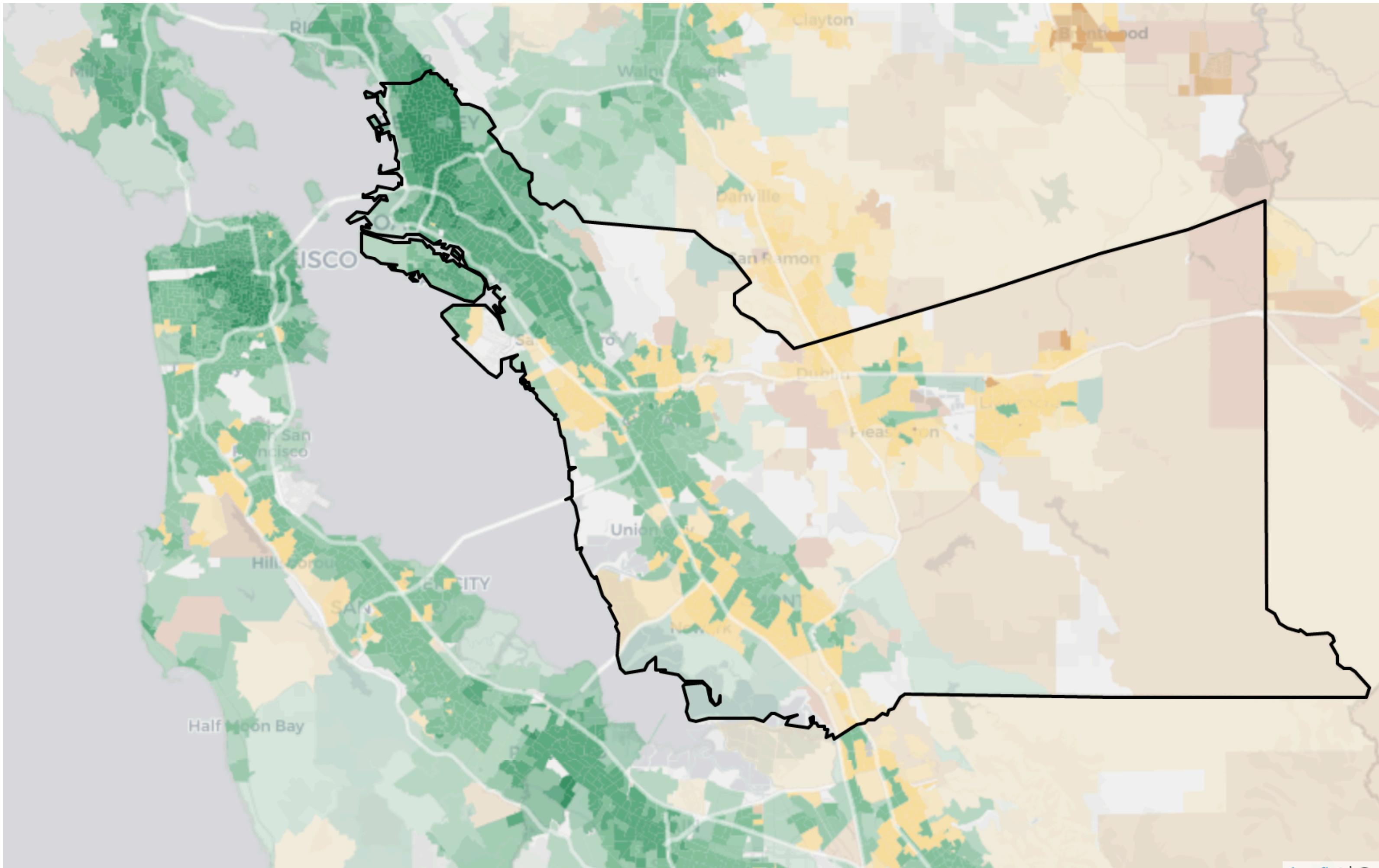
Data Science and Political Campaigns

Joshua Kalla

University of California, Berkeley

What kinds of data?

Precinct Election Results: 2016 vote to repeal death penalty





Precinct Data is Messy

← → X 🔒 Secure | https://raw.githubusercontent.com/openelections/openelections-data-pa/master/2016/20161108_pa_general_precin...

Voter File

- The list of all registered voters are compiled by states and made available to campaigns (and academics!)
- Name, address, phone number, age, gender, race, political party registration, past voter turnout
 - Varies by state
- Append other data:
 - Commercial (e.g., consumer preferences)
 - Census (e.g., unemployment rate in a neighborhood)
 - Licensing (e.g. in some states, lists of who have hunting licenses are available)
 - Campaigns (e.g., field staff ID voters; donations)

Does all this data matter?

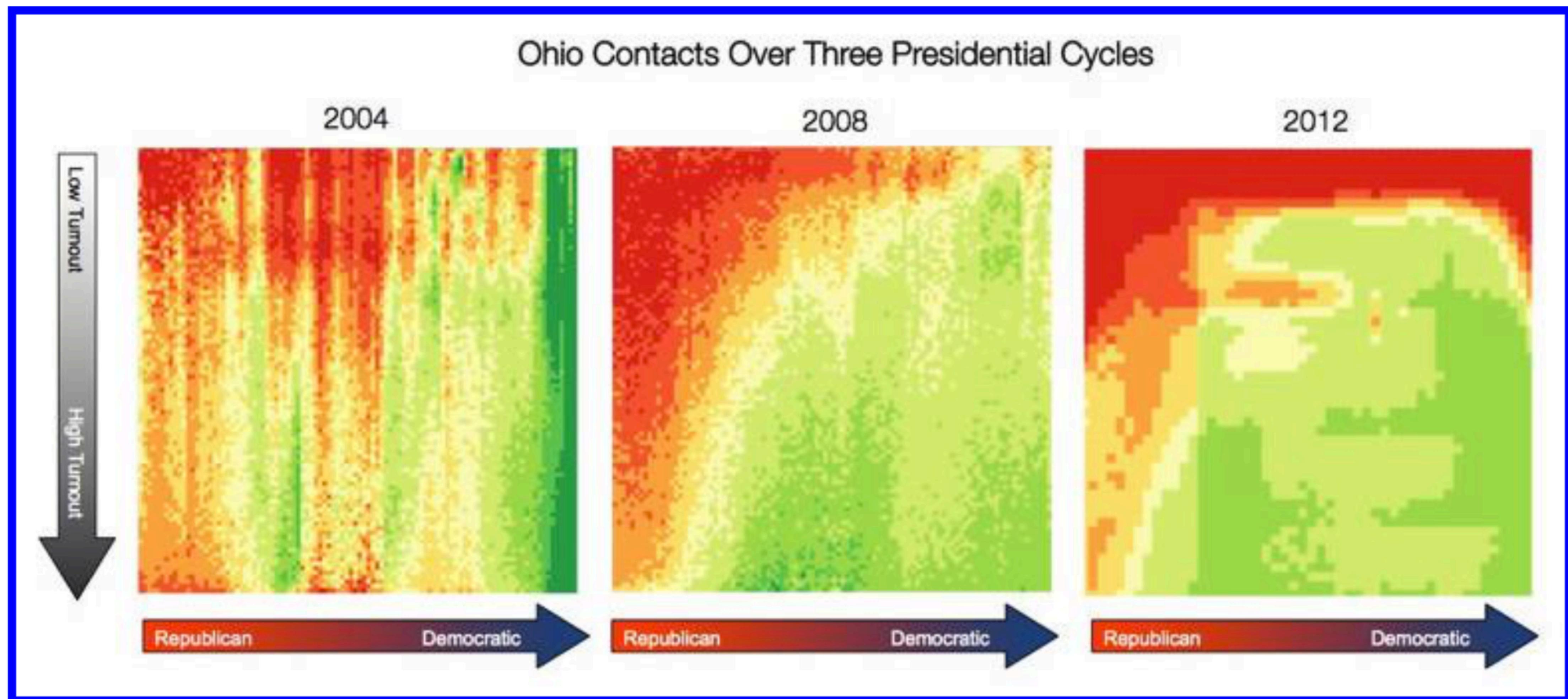
- Boat ownership is correlated with voting for a Republican
- But, so is being a white, male, registered Republican in a rich ZIP code
 - All of this is publicly available from Census and voter file
- Once I know your age, race, gender, income level, and the neighborhood you live in, consumer preferences tend to become irrelevant

How is this used?

Three Types of Models

- **Voter Turnout Score:** use past behavior and demographic information to calculate probabilities a given person will vote in an upcoming election
- **Support Score:** contact a subset of voters and use their responses as data to develop models that predict the preferences of the rest of voters
- **Responsiveness Scores:** how will voters respond to campaign outreach (heterogenous treatment effects; persuadability)

Heatmap of Ohio Contacts over Three Presidential Cycles



Using experiments to build
responsiveness models

WORKING AMERICA

Community Affiliate of the AFL-CIO



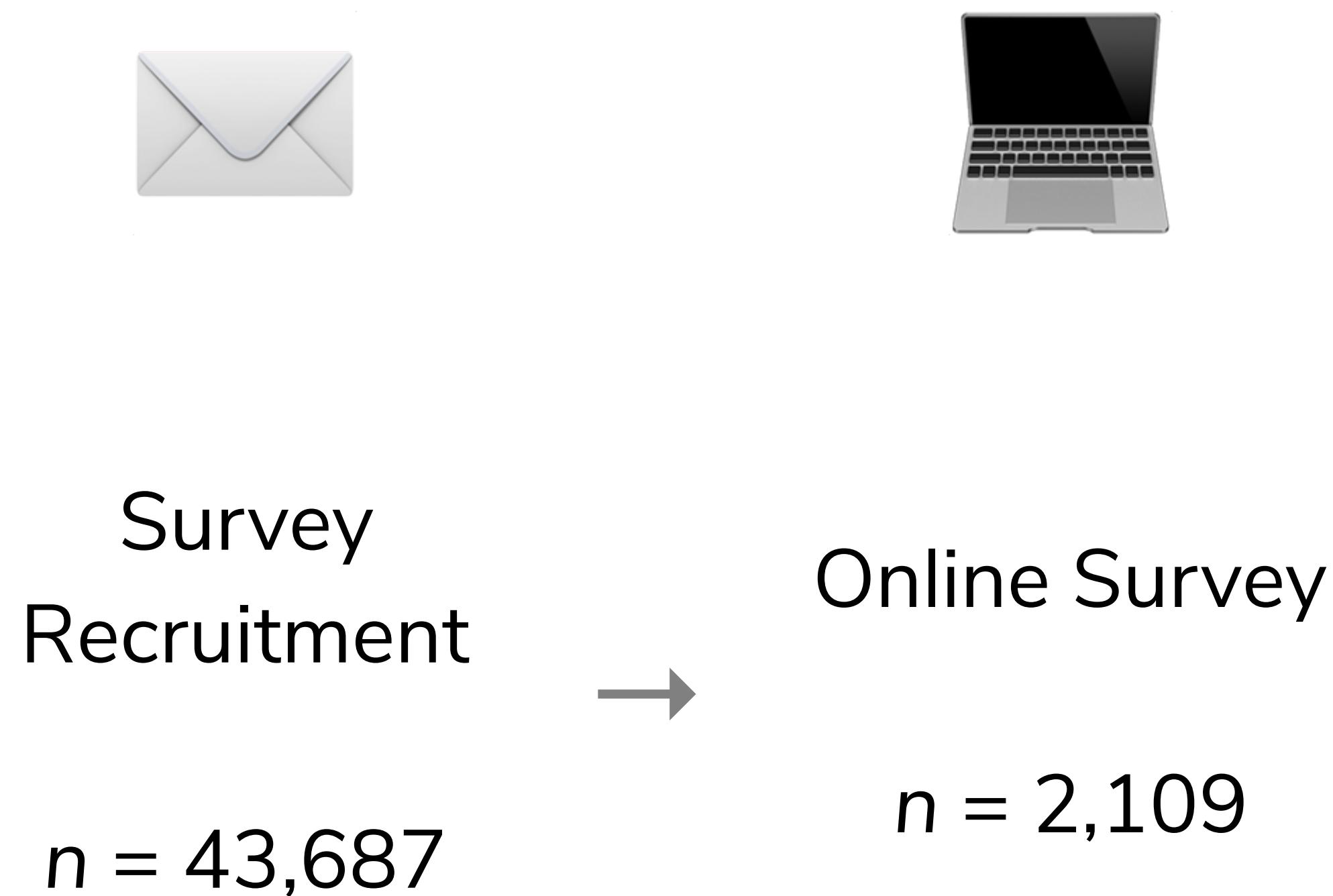
Experimental Design — Missouri Gubernatorial Race



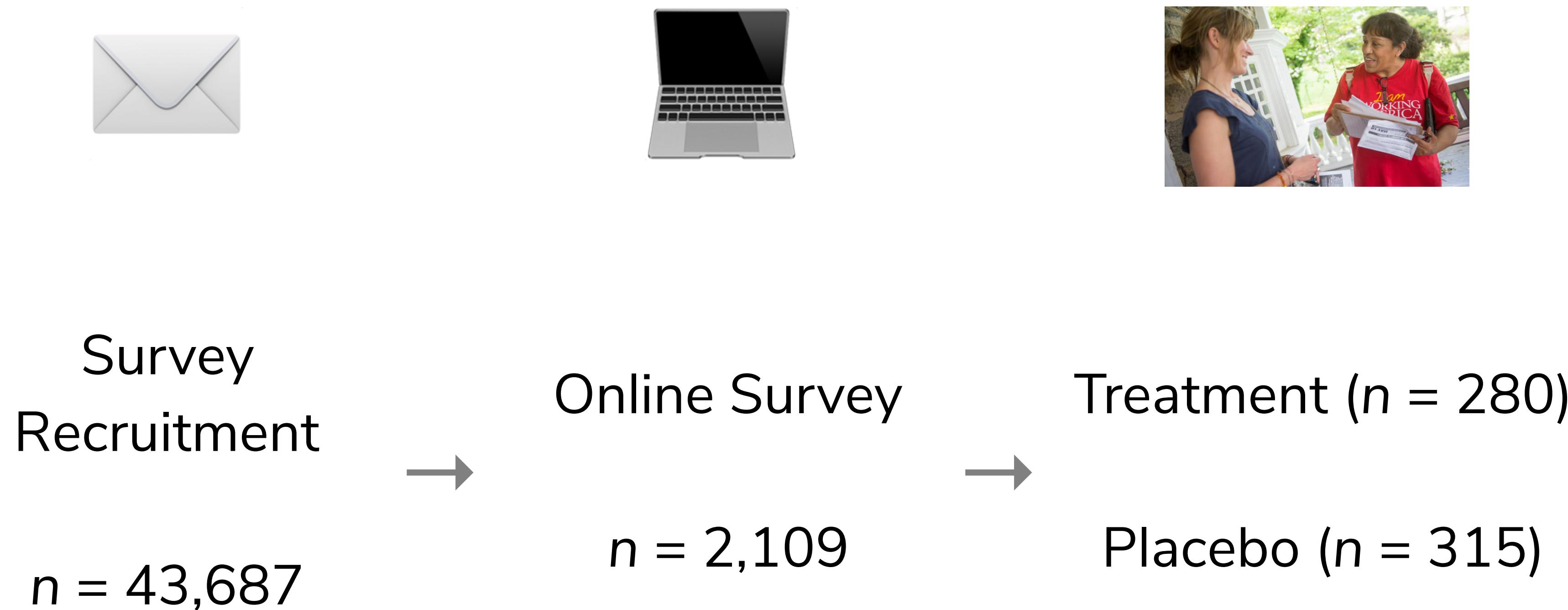
Survey
Recruitment

$n = 43,687$

Experimental Design — Missouri Gubernatorial Race



Experimental Design — Missouri Gubernatorial Race

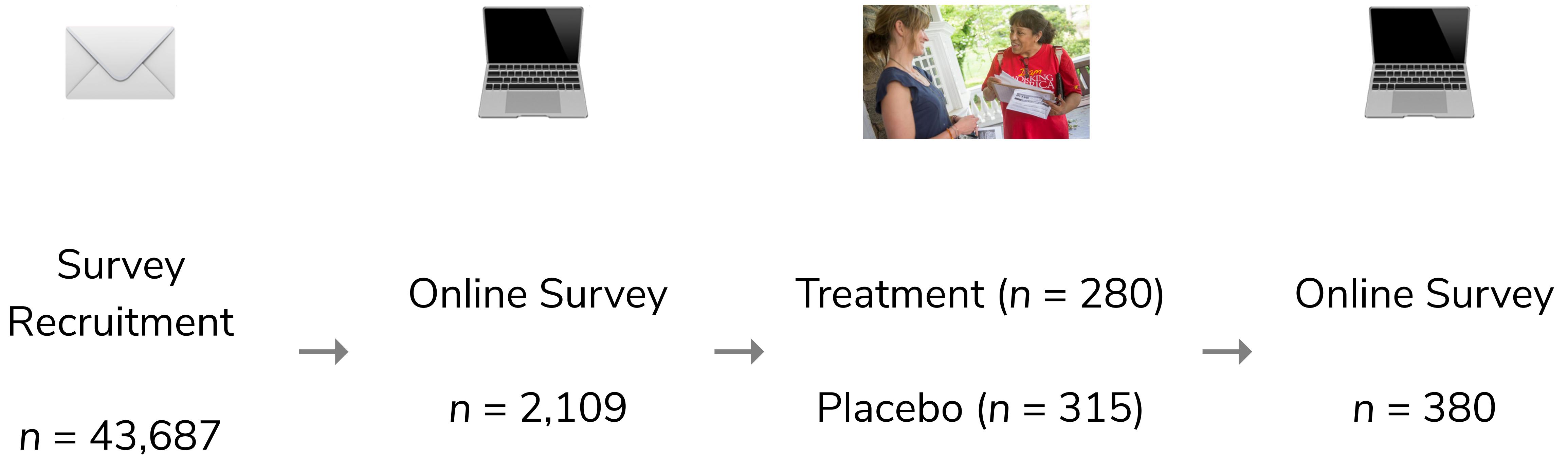


Placebo Script

Persuasion Script

- | | |
|-----------------------------------|--|
| | 1. Introduction: are you [NAME]? |
| 2. Source of news survey question | 2. Most important issue |
| 3. Have a good night! | 3. Intended vote choice |
| | 4. Persuasion based on relevant issue and candidate background |
| | 5. Have a good night! |

Experimental Design — Missouri Gubernatorial Race

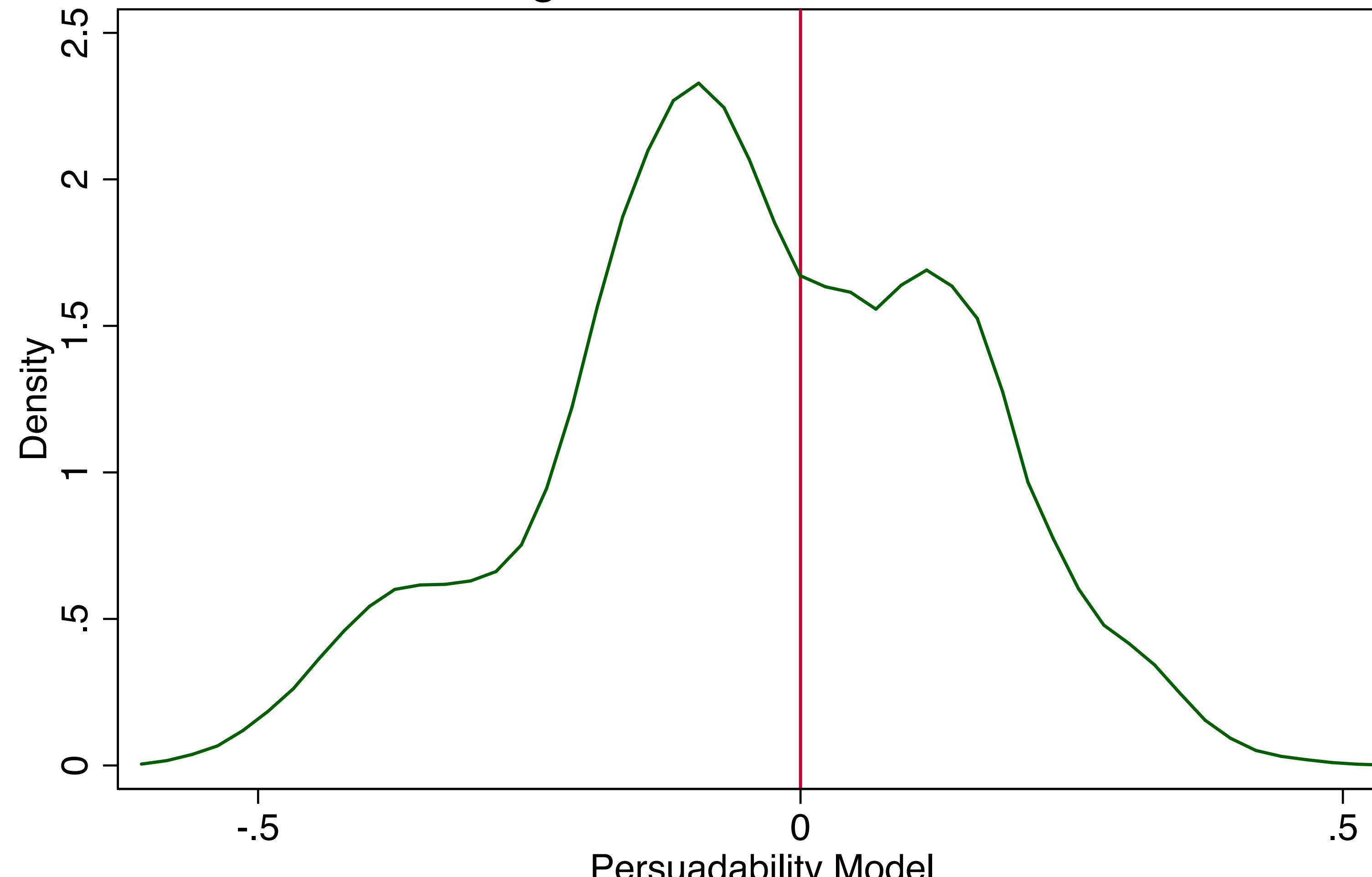


Missouri Gubernatorial Race Outcome

- If the election were held today between Republican Eric Greitens and Democrat Chris Koster, how do you think you would vote?
- Do you have a favorable or unfavorable opinion of Eric Greitens?
- Do you have a favorable or unfavorable opinion of Chris Koster?

Slight negative effect on average; campaign finds who is persuadable and targets similar people

Heterogeneous Treatment Effects



kernel = epanechnikov, bandwidth = 0.0278

Are campaigns actually using
this?

Many problems are organizational, not technical

- Lack of feedback loop in politics (only one election!)
- Incentives are often misaligned between candidates and consultants
- Intuition and experience are overvalued
 - Nobody ever got fired for buying IBM

Data science is more than just a technical problem

- Political: data often proves the “old guard” wrong
- Interpretable: people want an understanding of a model
- Actionable: findings need to be explained in a useful way to lay people

Thanks!
Questions? Comments?

josh.kalla@gmail.com