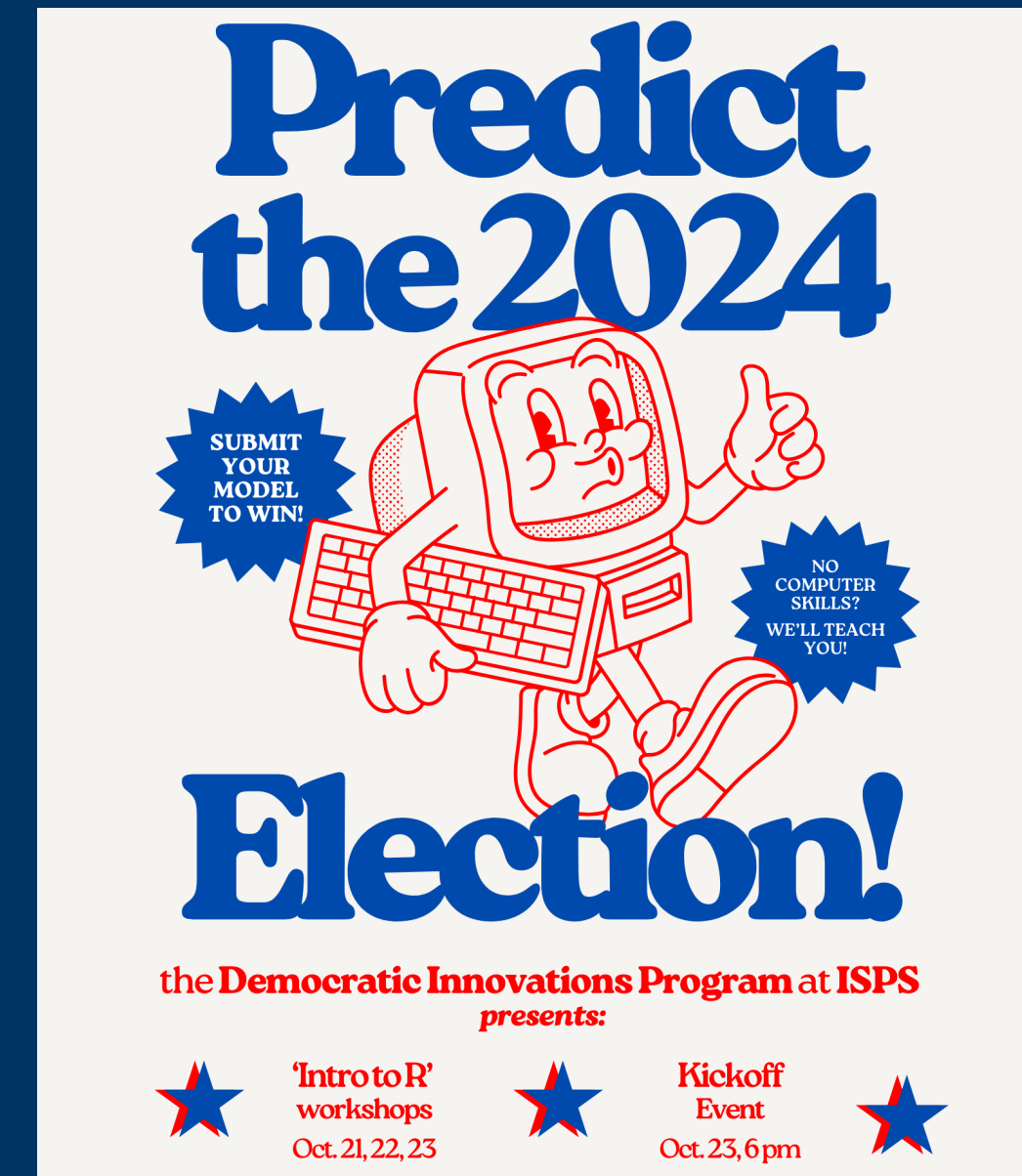# 2024 Election Prediction Competition

Joshua Kalla
Associate Professor, Political Science and Statistics & Data Science
Faculty Advisor, Democratic Innovations Undergraduate Research Program

Democratic Innovations Undergraduate Research Program

- Our mission is to identify and test innovative ideas for improving the quality of democratic representation and governance.

- Group Research Programs

  - Yale Youth Poll: https://youthpoll.yale.edu/

  - Citizen Assembly

  - Election Prediction Competition

- Individual Research Grants

- Contacts:

  - Prof. Josh Kalla: josh.kalla@yale.edu

  - Beata Fylkner: beata.fylkner@yale.edu

Rules

Election Prediction Competition Rules

- 7 prediction categories + overall winner

- Can submit in as many categories as you want

- Can work individually or in teams

- Can use any data or approaches out there — nothing is off limits

Category #1: Presidential Vote Share

2.1 Vote share

Predict the vote share of Donald Trump (votes cast for Trump divided by the total number of votes cast for US President, including for third-party candidates) in all 50 states, plus DC and each district in Nebraska and Maine. There will be 56 units of observation in total. The weighted MSE is given by

$$WMSE = \sum_{s=1}^{56} (\text{predicted share}_s - \text{actual share}_s)^2 \times (1 - \text{margin}_s)^2 \tag{1}$$

where the margin is the difference (in percentage points divided by 100) between the first and second place candidates. Your goal is to minimize MSE.

Category #2: Presidential Voter Turnout

Predict the number of votes cast for US president in all 50 states, plus DC and each district in Nebraska and Maine. There will be 56 units of observation in total. The normalized MSE is given by

$$NMSE = \sum_{s=1}^{56} \left(\frac{\text{predicted number of votes}_s - \text{actual number of votes}_s}{\text{actual number of votes}_s} \right)^2 \qquad (2)$$

Your goal is to minimize NMSE.

Category #3: Senate Vote Share

Predict the **difference in vote share** between Trump and the Republican candidate for Senate in all 34 Senate races. The vote share for Trump is calculated in the same way as in section 2.1. The vote share for the Republican Senate candidate is given by the number of votes cast for the candidate divided by the total number of votes cast for Senate, including for third-party candidates. The weighted MSE is given by

$$WMSE = \sum_{s=1}^{34}(\text{predicted difference}_s - \text{actual difference}_s)^2 \times (1 - \text{margin}_s)^2 \qquad (3)$$

Your goal is to minimize WMSE.

Category #4: Senate Voter Turnout

Predict the number of votes cast for Senate in all 34 Senate races **as a fraction** of votes cast for president. This is the number of votes cast for Senate in a state divided by the number of votes cast for the presidency in that state. The normalized MSE is given by

$$MSE = \sum_{s=1}^{34} \left(\text{predicted fraction}_s - \text{actual fraction}_s\right)^2 \tag{4}$$

Your goal is to minimize MSE.

Category #5: US House Vote Share

Predict the **difference in vote share** between Trump and the Republican candidate for the House of Representatives in all 435 house races. The vote share for Trump (respectively, the Republican House candidate) the number of votes cast for Trump (respectively, the Republican House candidate) divided by the total number of votes cast for US president (respectively, for all House candidates) in that district, including for third-party candidates. The weighted MSE is given by

$$WMSE = \sum_{s=1}^{34} (\text{predicted difference}_s - \text{actual difference}_s)^2 \times (1 - \text{margin}_s)^2 \qquad (5)$$

Your goal is to minimize WMSE.

Category #6: US House Voter Turnout

Predict the number of votes cast for in all 435 House of Representatives races **as a fraction** of votes cast for president in that district. This is the number of votes cast for House in a district divided by the number of votes cast for the presidency in that district. The normalized MSE is given by

$$MSE = \sum_{s=1}^{34} \left(\text{predicted fraction}_s - \text{actual fraction}_s\right)^2 \tag{6}$$

Your goal is to minimize MSE.

Category #7: Time State is Called

News agencies project a winner before all votes are fully counted. Predict the time it takes (in minutes) after the polls close for NBC News to project the winner of the electoral college votes in the following states: Arizona, Georgia, Michigan, Nevada, North Carolina, Pennsylvania, and Wisconsin. The log error is given by

$$\text{Log error} = \sum_{s=1}^{56} \ln\left(\left|\text{predicted number of minutes}_s - \text{actual number of minutes}_s\right| + 1\right)$$

The $+1$ is there to ensure error is always non-negative. We use log because some states may take several days to project a winner. Your goal is to minimize the log error.

Submission

Election Prediction Competition Submission

- We will email you `Election Prediction Competition Submission Spreadsheet.csv`

- Submit by **Saturday October 26 at 11:59pm ET**

- Email submissions to: josh.kalla@yale.edu; niklas.haehn@yale.edu; christopher.wu@yale.edu

- Late submissions will be discarded!

- **Must submit your predictions, code, and a brief write-up of your modeling approach**

- Submission packages will be posted publicly on GitHub: https://github.com/joshuakalla/yale_2024_election_prediction_competition

Next Steps

Next Steps

- Discord Server: https://discord.com/invite/bx6qp3xb

- Hub: ISPS Library; 9a-5pm; Snacks!

- Submission: Saturday October 26 at 11:59pm ET

- Election Night Watch Party: ISPS Lower Level Classrooms from 6-10pm
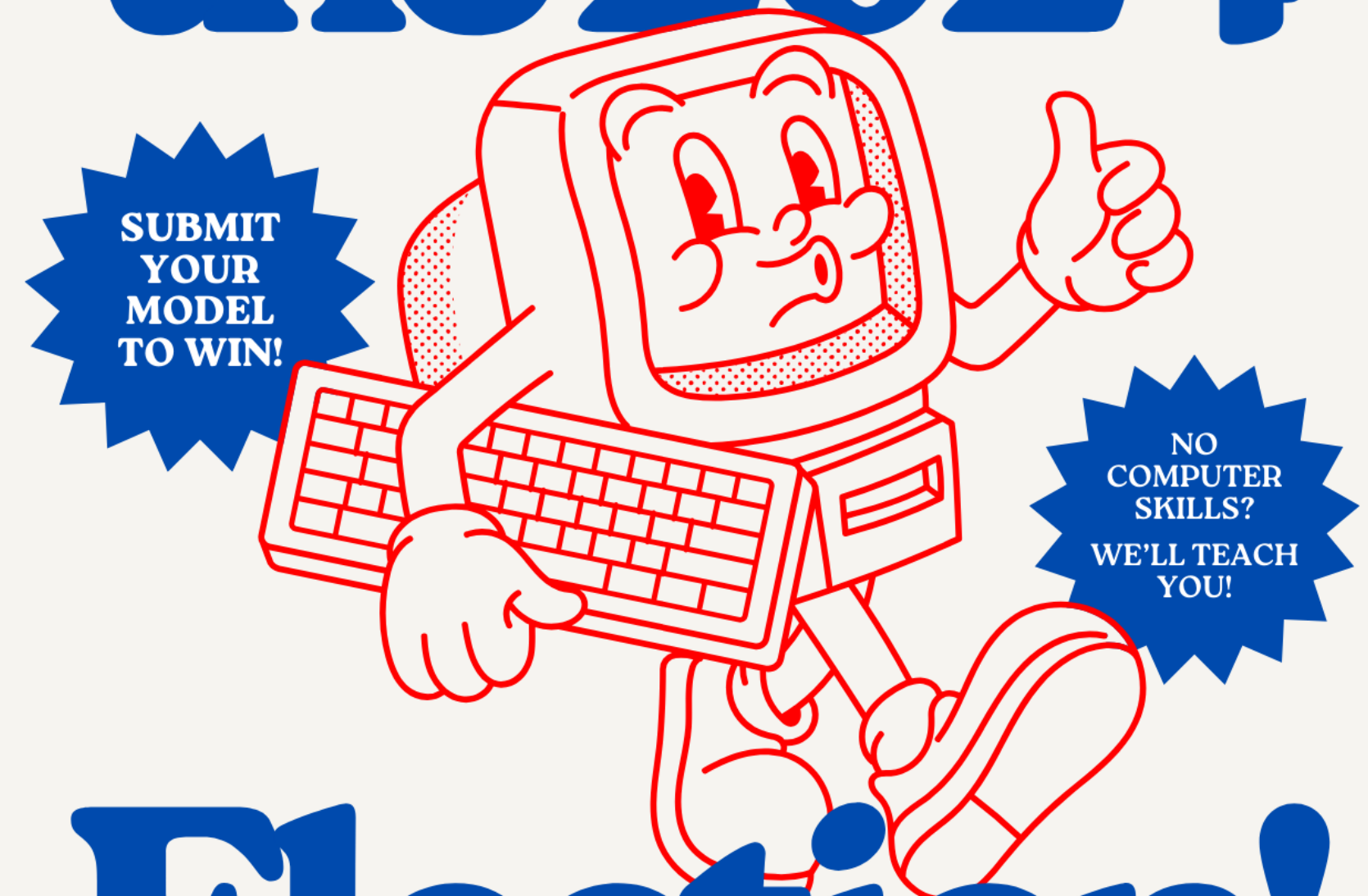
- Winners and prizes will be announced at some point…

Questions?

josh.kalla@yale.edu
niklas.haehn@yale.edu
christopher.wu@yale.edu