Model Plan + Sketches

- goal: to forecast the percentage of support Harris and Trump may have in the week leading up to the election

① Initial Analysis

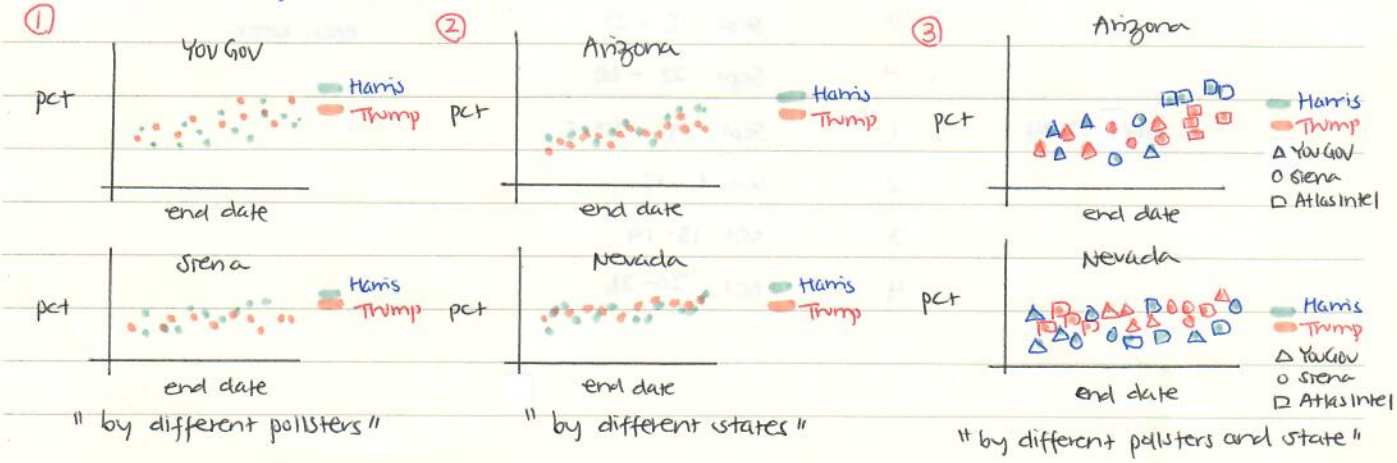
- want to understand how the percentage of support for both candidates has changed over time
- consider pollsters with a ① numeric grade 2.7 or above, (IV) ② likely voter population, polls that did not ask about ③ hypothetical match-ups, ④ states, Harris and Trump ⑤ pct, and ⑥ start and ⑦ end dates
  - ↳ want polls from July 7, 2024 to October 26, 2024 because this is after the Presidential debate that sparked discussions around President Biden's run and roughly one week before the election

(data) (table)

want to filter the raw polls data to consider the variables above

pollster	numeric grade	state	population	start date	end date	hypothetical	candidate_name	pct
Siena	2.8	Arizona	lv	8/12/24	8/14/24	FALSE	Donald Trump	47
Atlas Intel	2.7	Nevada	lv	9/3/24	9/6/24	FALSE	Kamala Harris	48.2
YouGov	3	Nevada	lv	10/3/24	10/5/24	FALSE	Kamala Harris	46.7
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

(graph) (showing how pct for both candidates has changed across poll end dates of all pollsters) - maybe show pct vs. end date by ① different pollsters or by ② state or ③ both?



" by different pollsters "

" by different states "

" by different pollsters and state "

② Model → voters' opinions vary across time, so to account for this variability, will use seasonal indexes and linear regression to forecast (Book) Keller & Warrack, 2003

↓

account for the degree to which seasons differ from one another

- each "season" in this analysis will a week starting from July 7 (because this is after the debate that sparked discussion around President Biden's campaign and some support may have been leaning towards Harris at this time) to October 26, 2024 (because this leaves roughly a week before the election)
  - ↳ want to pool polls to get an average percentage of support for both candidates in each week
  - ↳ weeks are as follows:

Month	week #	week dates
July 2024	1	July 7 - 13
	2	July 14 - 20
	3	July 21 - 27
	4	July 28 - August 3
August 2024	1	Aug. 4 - 10
	2	Aug. 11 - 17
	3	Aug. 18 - 24
	4	Aug. 25 - 31
September 2024	1	Sept. 1 - 7
	2	Sept. 8 - 14
	3	Sept. 15 - 21
	4	Sept. 22 - 28
October 2024	1	Sept. 29 - Oct. 5
	2	Oct. 6 - 12
	3	Oct. 13 - 19
	4	Oct. 20 - 26

* there are 16 weeks in total and the model will be forecasting week 17 which is from Oct. 27 - Nov. 2. (about a week before the election)

* need to consider all states as each state doesn't have enough poll data for each week.

for both candidates

4 for each week, will find polls that start and end in that week and pool their pct results to get a "precision-weighted average" of their estimates (this will be done by using a pooling method outlined by Jackman (2005) with the formula:

$$\hat{\alpha}_{AB} = \frac{P_A \hat{\alpha}_A + P_B \hat{\alpha}_B}{P_A + P_B}$$

$\hat{\alpha}_A$ and $\hat{\alpha}_B$: estimated pct from poll A and poll B

$\hat{\alpha}_{AB}$: pooled estimate for pct (i.e. precision-weighted average).

P_A, P_B: precision of each poll

$$P_A = \frac{1}{\left(\sqrt{\frac{\hat{\alpha}_A(1-\hat{\alpha}_A)}{n_A}}\right)^2}$$

* same for P_B and poll B.

n_A: sample size of poll A

this method can be used because the polls that will be pooled occurred at a similar time BUT this method assumes that polls are unbiased - which is often not the case so while this may be a limitation, the method still provides more precision than a single poll

- after getting precision-weighted averages, will perform a regression analysis with $\hat{y}$ = pct, w = week 1, 2, ..., 16 for both candidates

$$\begin{cases} \hat{y}_H = b_0 + b_1 w & \text{for Harris} \\ \hat{y}_T = b_0 + b_1 w & \text{for Trump} \end{cases}$$

$\hat{y}$ = precision-weighted average

- for each week, will compute the ratio, $\frac{y}{\hat{y}}$, to remove some trend variation
- for each week in each month (i.e. for all week ones across July, August, September, and October), will compute the average of ratios to remove some of the random variation (get a measure of seasonality)
- all these steps should give the following tables: (each candidate should have these two tables).

①

Month	week	y	$\hat{y}$	$\frac{y}{\hat{y}}$
July	1	46.7	45.9	$\frac{46.7}{45.9}$
	2	48.2	50.1	$\frac{48.2}{50.1}$
	⋮	⋮	⋮	⋮

②

Month	week 1	2	3	4
July	$y/\hat{y}$	-	-	-
August	⋮	⋮	⋮	⋮
Sept.	⋮	⋮	⋮	⋮
Oct.	⋮	⋮	⋮	⋮
average of ratios	○	○	○	○

- using this, will forecast pct for week 17
  - for $w = 17$, calculate $\hat{y} = b_0 + b_1 w$, multiply by seasonal index for week 1
  - this will give us pct for Harris in the week leading up to the election

- can present the results of the model (ie. the forecasts for week 17 for both Harris and Trump) with a graph similar to: