

Unclear Politics

What's going on with Real Clear Politics' poll aggregator?





Background information

- Donald Trump's surprise victory in 2016 defied expected results in key battleground states
- Placed the spotlight on the reliability of polls & polling methods ("Can polls be trusted?")
- National polls were actually fairly accurate, and correctly predicted Clinton winning the national popular vote
- Polls in key battleground states were less accurate - unexpected shifts in voting patterns especially re: education level



Improving forecasting... but how?

- Many reputable pollsters have adjusted for this by weighting their polling sample by education levels
- However, the uncertainty caused by 2016 opened the door to...
 - Persistent narratives/questions about the reliability of polls
 - Alternative and sometimes... questionable methodological “innovations”
- E.g. Trafalgar’s “social desirability” approach



What about the meta?

- Because different pollsters rely on different methods/practices, over-reliance on one or several polls may not provide an accurate picture vs large-scale analysis
- Just as there are different methods employed by individual polls/pollsters, there are also different approaches to assessing the overall field of polls
- High sophistication: weighting the **pollsters** (FiveThirtyEight)
- Low sophistication: poll **aggregators** (Real Clear Politics)



Shifting Project Goals:

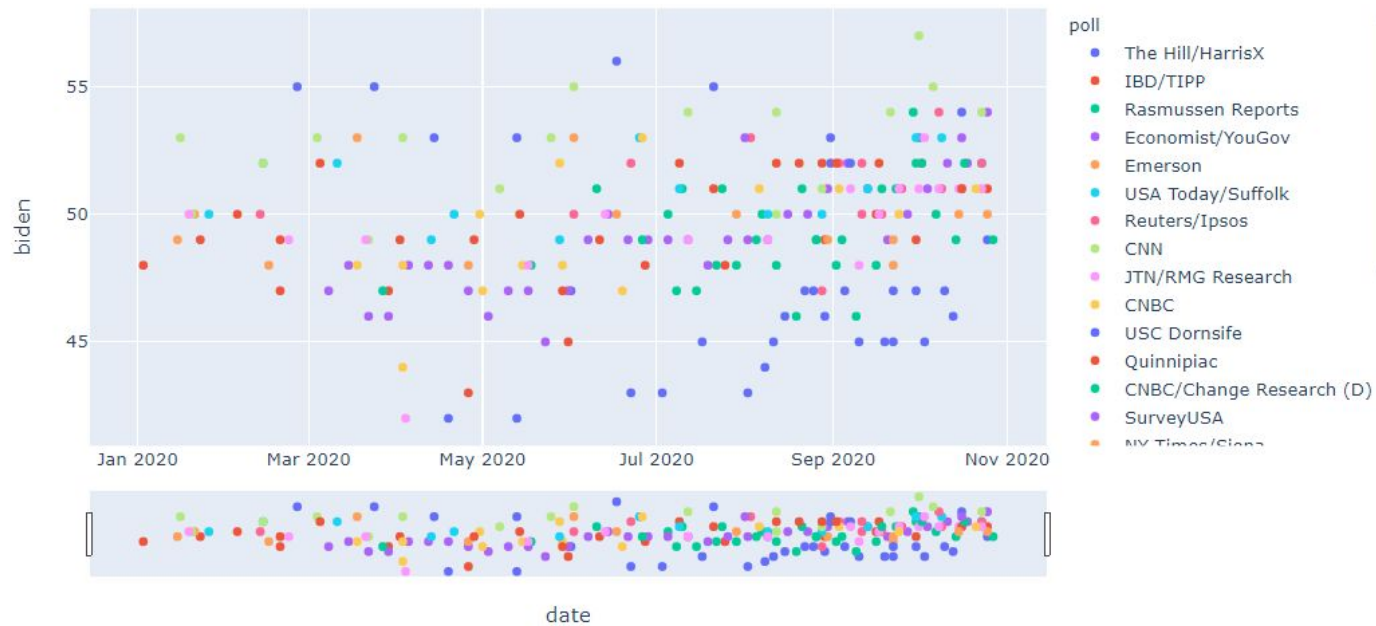
Original goal:

- Scrape data from RCP to perform a comparative analysis of the different pollsters + polls included in their aggregator
- Generate interactive tools/visualizations to allow user to include/exclude and compare user-selected pollsters from RCP's aggregator

Instead, several red flags about the data RCP uses & includes in its aggregator, and the process behind deciding what is/isn't included to build its "RCP Average," almost immediately shifted my focus to the aggregator itself



The polls





Potential problems with poll aggregation

- Pure polling average is relatively unsophisticated, presents many potential issues
- Taking a pure average would be more accurate if aggregators selected polls that used similar...
 - Sampling methods
 - Sample sizes
 - Margins of error
 - Questions/wording of the polls themselves

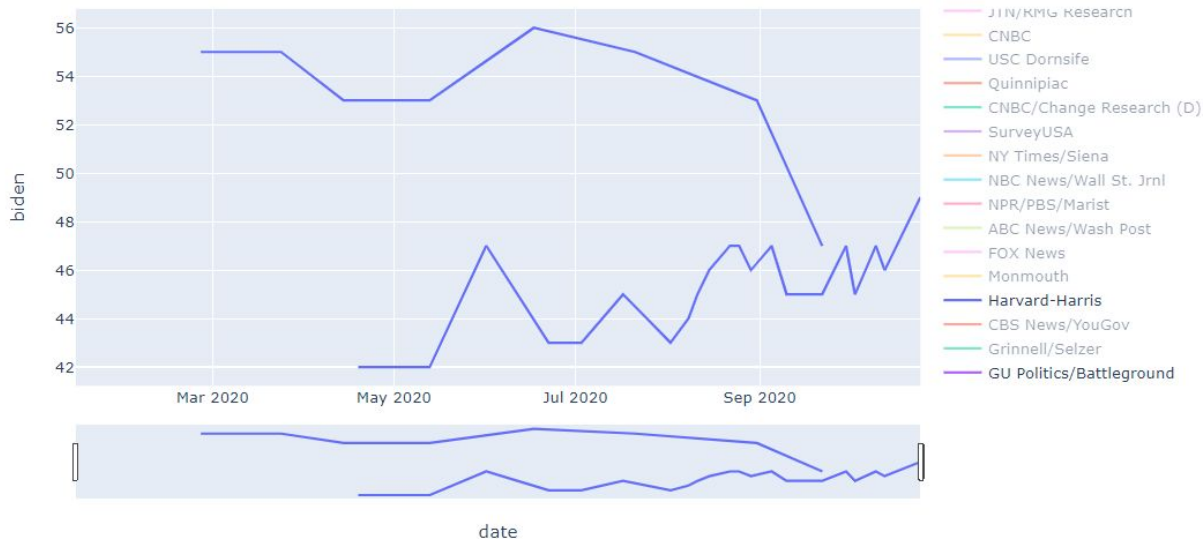
... But RCP doesn't



	date	poll	biden	trump	spread	sample	moe
0	2020-10-25	The Hill/HarrisX	49	45	Biden +4	2359 LV	2.0
1	2020-10-25	IBD/TIPP	51	44	Biden +7	959 LV ←	3.2
2	2020-10-27	Rasmussen Reports	49	46	Biden +3	1500 LV	2.5
3	2020-10-25	Economist/YouGov	54	43	Biden +11	1365 LV	3.2
4	2020-10-25	Emerson	50	45	Biden +5	1121 LV	2.8
...
214	2020-01-20	ABC News/Wash Post	50	46	Biden +4	880 RV	4.0
215	2020-01-15	LA Times/USC	49	40	Biden +9	4869 RV ←	2.0
216	2020-01-19	FOX News	50	41	Biden +9	1005 RV	3.0
217	2020-01-16	CNN	53	44	Biden +9	1051 RV	3.6
218	2020-01-03	IBD/TIPP	48	46	Biden +2	846 RV	3.3

What's going on with inclusion of extreme outlier polls/pollsters?

Some pollsters included in the aggregate have consistently produced outlier results/polls, esp. earlier in the campaign cycle



What's going on with weird deviation trends across pollsters AND candidates?

Standard Deviations

	biden	trump
poll		
ABC News/Wash Post	2.133910	1.414214
CBS News/YouGov	2.309401	1.272418
CNBC	2.258318	2.786874
CNBC/Change Research (D)	1.414214	1.000000
CNN	2.179449	1.900292
Economist/YouGov	2.026562	1.131308
Emerson	1.964971	3.000000
FOX News	2.828427	2.118700
GU Politics/Battleground	NaN	NaN
Grinnell/Selzer	1.414214	1.414214
Harvard-Harris	2.825269	1.187735
IBD/TIPP	2.174229	2.576114
JTN/RMG Research	1.133893	0.786796
LA Times/USC	NaN	NaN
Monmouth	1.767767	1.908627
NBC News/Wall St. Jnl	1.490712	1.632993
NPR/PBS/Marist	1.602082	0.752773
NY Times/Siena	0.577350	2.886751
Quinnipiac	1.301708	1.787301
Rasmussen Reports	1.581139	2.073802
Reuters/Ipsos	1.581139	0.927961
SurveyUSA	0.000000	0.000000
The Hill/HarrisX	1.856310	1.648822
USA Today/Suffolk	1.500000	1.825742
USC Dornsife	1.154701	0.000000

- Some pollsters demonstrate bizarrely high variance/deviation in poll results throughout the year. Others, very low.
- Because of extremely stable (dis)approval ratings for Trump and firmly cemented public opinion, we should assume that lower deviation is more likely to be accurate -- although it's hard to say to what extent, or where the line is
- Why are some pollsters' deviations wildly different across **candidates**? Shouldn't these be more strongly correlated?
 - (planning on assessing this further by plotting SD lines for each pollster)

What's going on with random one-off inclusions?

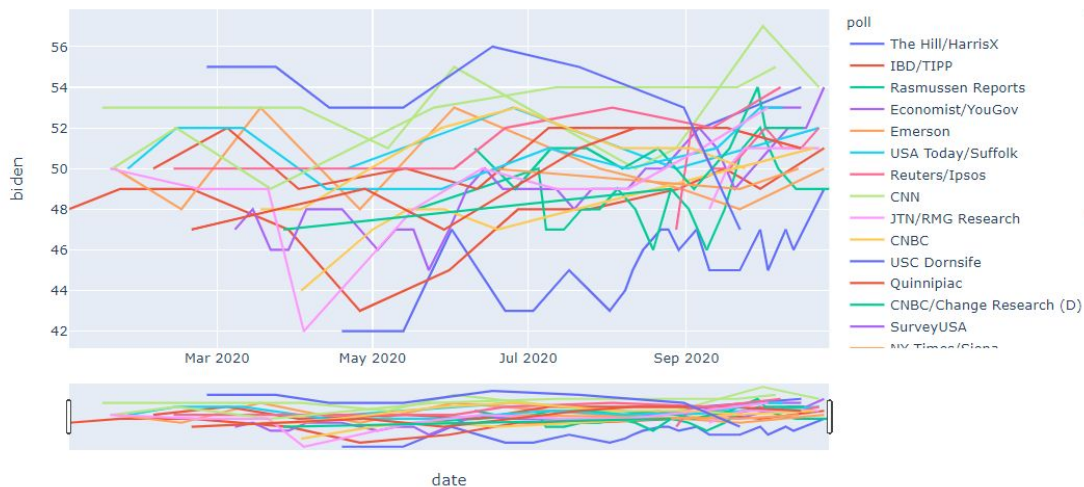
Why include these pollsters when RCP otherwise seemed to prioritize those who polled many times throughout the year?





What's going on with poll herding?

Results from different pollsters start herding/trending inwards toward each other as we near the end of an election season





Unclear patterns of poll inclusion/exclusion by quality

RCP includes some highly partisan / flawed / etc. pollsters (Change Research, Rasmussen, etc.) while excluding more reliable polls (Cygnal, GQRR, PRRI, University of Massachusetts Lowell, Marquette, etc.)

Search for a pollster

POLLSTER	METHOD	LIVE CALLER WITH CELLPHONES	NCPP/ AAPOR/ ROPER	POLLS ANALYZED	SIMPLE AVERAGE ERROR	RACES CALLED CORRECTLY	ADVANCED +/-	PREDICTIVE +/-	538 GRADE	BANNED BY 538	MEAN-REVERTED BIAS
SurveyUSA	IVR/online/live	●	●	787	4.7	89%	-1.1	-0.8	A		D+0.1
Rasmussen Reports/ Pulse Opinion Research	IVR/online			722	5.3	78%	+0.2	+0.8	C+		R+1.5
Zogby Interactive/JZ Analytics	Online			473	5.4	77%	+0.4	+0.9	C+		R+0.6
Mason-Dixon Polling & Strategy	Live	●		433	5.1	87%	-0.6	-0.3	B+		R+0.6
Public Policy Polling	IVR/text			423	5.0	80%	-0.4	+0.1	B		D+0.3
YouGov	Online			416	4.9	88%	-0.2	+0.3	B		D+0.4
Research 2000	Live*			280	5.5	88%	-0.1	+0.3	F	✗	D+1.3
American Research Group	Live	●	●	273	7.4	75%	+0.3	+0.2	B		R+0.2
SurveyMonkey	Online			210	7.1	84%	+2.3	+2.6	D-		D+5.0

(screen grab from
FiveThirtyEight)



Possible poll inclusion strategy

- Not much can be gleaned from RCP data at a base level of analysis -- and that's part of the problem
- The bulk of the included polls trend toward the same results/range, while RCP also includes polls & pollsters that objectively appear to be outliers which display drastically different views of the race / reality than the rest of the data
- This might be the exact strategy behind their selections: aim to provide a majority of polls with similar variation + results, but also toss in outlier polls regardless of quality, reliability, etc. In other words, a methodology designed to “cover all bases” or “cover the full spread” rather than prioritizing **quality** and **utility** of data



Why this matters

- As important as accurate polling + sampling methods are, even more important is how that information is presented. Presenting casual viewers and non-experts with an overwhelming amount of information, without context and/or with an inaccurate representation of key takeaways, **does more harm than good** -- especially when it's unclear how/why certain information was included.
- I was left with more questions than answers, which is the opposite of what RCP presents itself as providing (implicitly or explicitly)
- Threatens to sow additional doubt on the usefulness of polling when all polls are presented as equal, especially when actual outcomes are different than the above information and/or encourages cherry picking pollsters whose results align best with preconceptions or desires. At its most extreme, it also casts doubt on the integrity of the system when things go awry



Next Steps

- Finish building out a Plotly Dash app to allow users to explore the data, and how including/excluding some of RCP's chosen polls would affect outcomes
- Scrape and integrate data from FiveThirtyEight, attempt to use their expanded & more transparent data to find some semblance of a coherent methodology in RCP's aggregation choices
- Time series analysis
 - Plot a moving average (likely 2-week intervals) for each individual pollster
 - Plot standard deviation lines for each pollster