Debugging the 2020 Election Predictions

Team 8: Fahim Talukder, Greg Fields, Srinivas Kakade, Jacob Pollard, and Zhehui Li

Motivations and Objective

- We analyzed a variety of trends in the 2020 election results and in the polling leading up to the election.
- In both 2016 and 2020, polling systematically underestimated former President Trump.
  - This was most pronounced in the Upper Midwest region.
  - Does it also depend on demographic?
  - How does polling methodology affect these mistakes?
- Political campaigns craft policy and plan outreach campaigns based on polling. Accurate polling can dictate outcomes of elections.

Data Collection

- We used the official election results for both the presidential and senate races.
- To get demographic information, we used exit polling data from Edison Research.
  - Their raw data is proprietary; we assembled the results we needed from reporting by the New York Times and CNN.
- For pre-election polling, there are a large number of polling firms.
  - Many different methodologies
  - Widely varying quality
  - Different polling questions and demographic breakdowns

Data Collection

- We conducted a brief survey of results from many different firms and chose demographics we found most interesting.
  - Male, female, white, black, Hispanic/Latino, Republican, Democrat, Independent, 65+ years old, white with a bachelor's degree, white without a bachelor's degree
- Chose three pollsters who collected data on these demographics, all rated A+ by fivethirtyeight.com
  - Monmouth University, Siena College, Marist College
- All three conduct polls by phone, we wanted to examine the impact of this
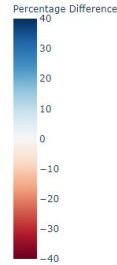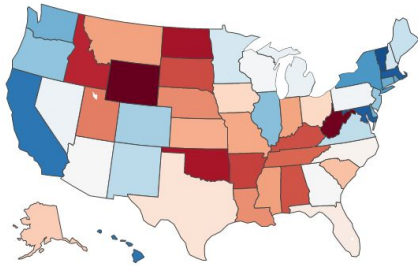 - Emerson College, A- rated and collects ~25% of data by text, ~25% by online survey

		Joe Biden and Kamala Harris, the Democrats Row %	Donald Trump and Mike Pence, the Republicans Row %	Other Row %	Undecided Row %
AZ Likely Voters		48%	48%	3%	1%
Intensity of Support	Strongly support	45%	55%	0%	0%
	Somewhat support	69%	31%	0%	0%
Already Voted		56%	42%	2%	0%
Yet to Vote		26%	66%	3%	4%
Vote Plan	In person on Election Day	20%	75%	1%	4%
	Before Election Day	54%	43%	3%	0%
Party Identification	Democrat	94%	6%	0%	0%
	Republican	7%	91%	2%	0%
	Independent	57%	35%	5%	2%
Party Identification	Strong Democrats	94%	6%	0%	0%
	Soft Democrats	93%	5%	2%	0%
	Just Independents	57%	28%	8%	7%
	Soft Republicans	15%	78%	6%	0%
	Strong Republicans	2%	97%	0%	0%
Party ID and Gender	Democrat men	93%	7%	0%	0%
	Democrat women	95%	5%	0%	0%
	Republican men	7%	90%	3%	0%
	Republican women	6%	93%	0%	0%
	Independent men	55%	35%	8%	2%
	Independent women	61%	35%	2%	2%
Region	Maricopa County (Phoenix)	51%	45%	3%	1%
	Pima County (Tucson)	54%	44%	2%	0%
	Rest of State	39%	58%	2%	2%
Race/Ethnicity	White	49%	49%	2%	0%
	Latino	49%	45%	5%	2%
	Other	51%	45%	1%	3%
Household Income	Less than $50,000	46%	49%	3%	2%
	$50,000 or more	51%	46%	2%	1%
Education	Not college graduate	43%	53%	2%	2%
	College graduate	56%	41%	3%	0%
Race and Education	White - Not College Graduate	43%	55%	2%	0%
	White - College Graduate	55%	43%	2%	0%
Gender - Race - Education	Men - White - Not College Graduate	42%	55%	3%	1%
	Men - White - College Graduate	52%	43%	4%	1%
	Women - White - Not College Graduate	44%	55%	1%	0%
	Women - White - College Graduate	57%	42%	0%	0%
Age	Under 45	52%	41%	5%	2%
	45 or older	46%	52%	1%	1%
Age	65 or older	46%	52%	1%	1%
Generation	Gen Z/Millennials (18-39)	54%	37%	7%	2%

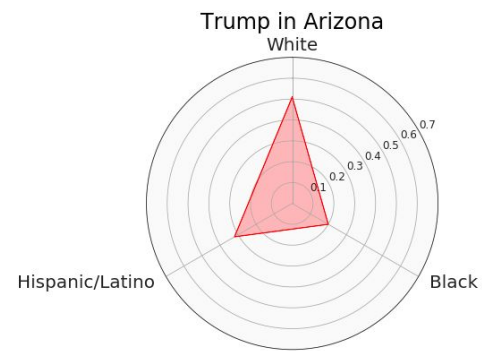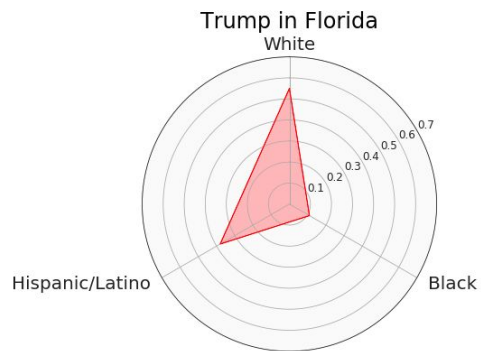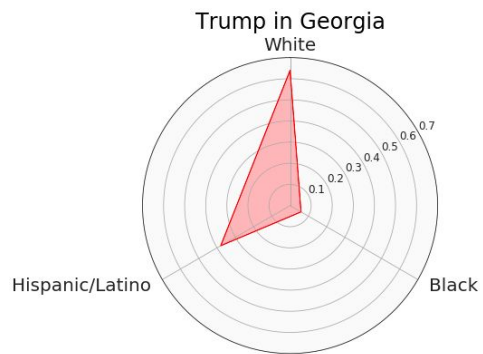|	State	Male	Female	White	Black	Hispanic/Latino	Asian	White + BA/BS	White + No BA/BS	Republican	Democrat	Independent	65+	Candidate	Pollster
0	Alabama	0.280	0.400	0.230	0.840	0.370	0.460	0.355	0.170	0.030	0.950	0.450	0.300	Biden	SurveyMonkey
1	Alabama	0.670	0.540	0.730	0.070	0.510	0.470	0.605	0.785	0.940	0.020	0.270	0.660	Trump	SurveyMonkey
2	Arizona	0.507	0.453	0.451	0.636	0.127	0.545	NaN	NaN	0.061	0.868	0.567	0.488	Biden	Emerson
3	Arizona	0.450	0.520	0.490	NaN	0.490	NaN	0.550	0.430	0.070	0.940	0.570	0.460	Biden	Marist
4	Arizona	0.480	0.400	0.510	NaN	NaN	NaN	0.480	0.540	0.900	0.050	0.330	0.420	Biden	Monmouth
5	Arizona	0.510	0.560	0.430	NaN	0.660	NaN	0.540	0.360	0.050	0.960	0.520	0.470	Biden	Siena
6	Arizona	0.400	0.550	0.450	0.730	0.590	0.510	0.380	0.385	0.050	0.950	0.490	0.470	Biden	SurveyMonkey
7	Arizona	0.412	0.499	0.507	0.333	0.300	0.364	NaN	NaN	0.911	0.122	0.304	0.507	Trump	Emerson
8	Arizona	0.490	0.460	0.490	NaN	0.450	NaN	0.410	0.530	0.910	0.060	0.350	0.520	Trump	Marist
9	Arizona	0.430	0.560	0.420	NaN	NaN	NaN	0.470	0.400	0.090	0.930	0.540	0.560	Trump	Monmouth
10	Arizona	0.490	0.380	0.500	NaN	0.260	NaN	0.390	0.580	0.910	0.020	0.340	0.480	Trump	Siena
11	Arizona	0.550	0.400	0.510	0.200	0.320	0.410	0.540	0.575	0.920	0.010	0.250	0.490	Trump	SurveyMonkey
12	California	0.540	0.660	0.580	0.800	0.640	0.610	0.695	0.470	0.070	0.960	0.500	0.610	Biden	SurveyMonkey
13	California	0.400	0.280	0.370	0.140	0.270	0.310	0.265	0.475	0.890	0.020	0.210	0.340	Trump	SurveyMonkey
14	Colorado	0.470	0.610	0.540	0.690	0.630	0.630	0.655	0.410	0.040	0.970	0.480	0.530	Biden	SurveyMonkey
15	Colorado	0.470	0.340	0.410	0.230	0.300	0.310	0.305	0.525	0.930	0.010	0.250	0.420	Trump	SurveyMonkey
16	Florida	0.414	0.586	0.421	0.750	0.746	0.600	NaN	NaN	0.150	0.880	0.461	0.531	Biden	Emerson
17	Florida	0.440	0.570	0.450	0.840	0.460	NaN	0.580	0.360	0.090	0.950	0.550	0.530	Biden	Marist
18	Florida	0.540	0.370	0.550	NaN	NaN	NaN	0.390	0.630	0.910	0.030	0.370	0.510	Biden	Monmouth
19	Florida	0.450	0.500	0.400	0.780	0.550	NaN	0.490	0.350	0.080	0.910	0.470	0.480	Biden	Siena

Results of the senate and presidential elections

Presidential Election

Senate Election

Relative performance of senate and presidential candidates

Relative Performance of Democratic Presidential and Senate Candidate

Relative Performance of Republican Presidential and Senate Candidate

Ethnicity

Trump in Georgia

Trump in Florida

Trump in Arizona
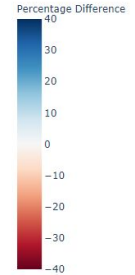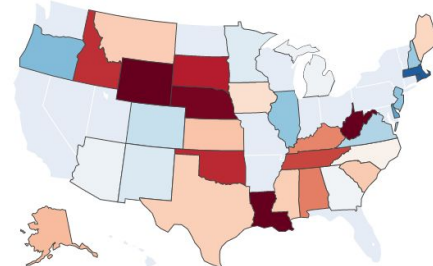
Hispanic/Latino

Biden won Arizona by .3% where Latinos make up 23.6% of the voting population. He was the first Democrat to win there since 1996.

Shift in pre/post-election poll data among Hispanic/Latino voters for Biden
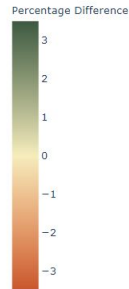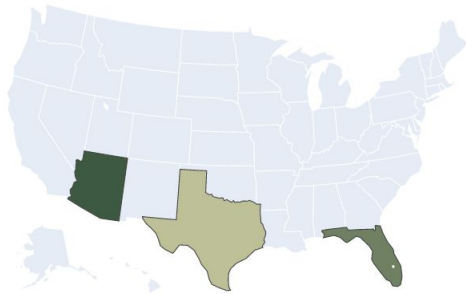
Shift in pre/post-election poll data among Hispanic/Latino voters for Trump
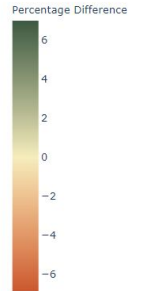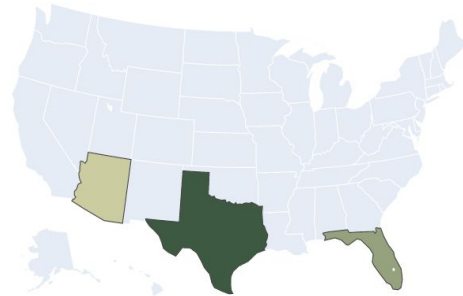
Black

Similarly strong, non-uniform shifts among Black voters

Shift in pre/post-election poll data among Black voters for Biden
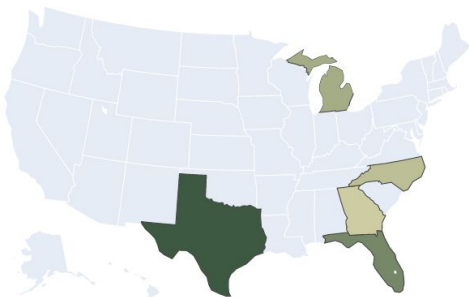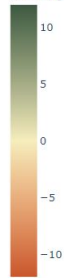
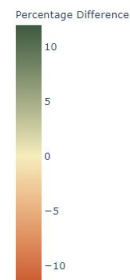Shift in pre/post-election poll data among Black voters for Trump
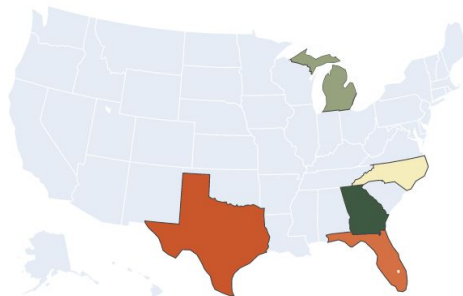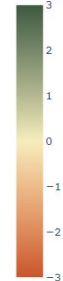
Age and education
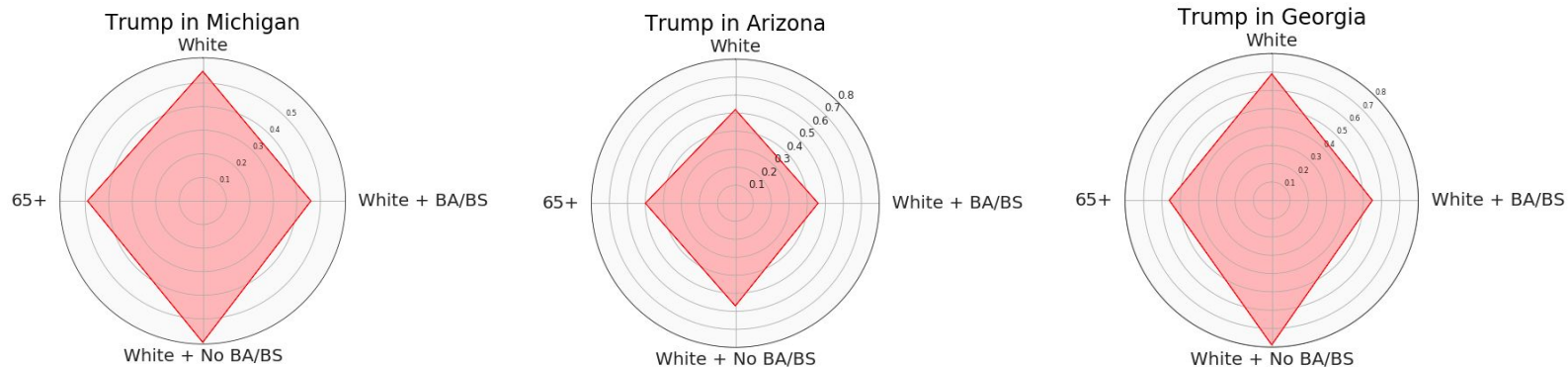
Whites with bachelor's degree

Widespread shift in Trump's favor, particularly throughout much of the midwest and southeast.

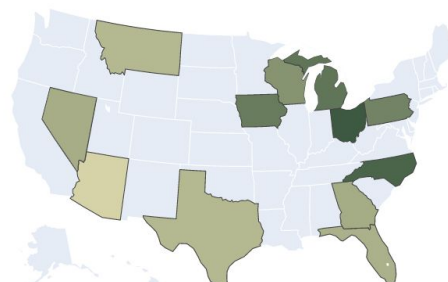Shift in pre/post-election poll data among White + BA/BS voters for Biden
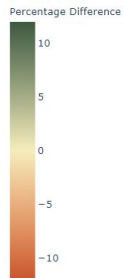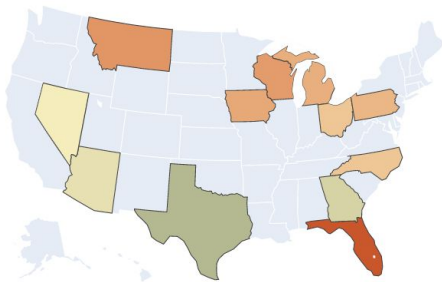
Shift in pre/post-election poll data among White + BA/BS voters for Trump
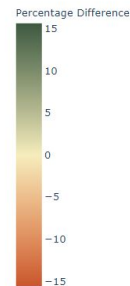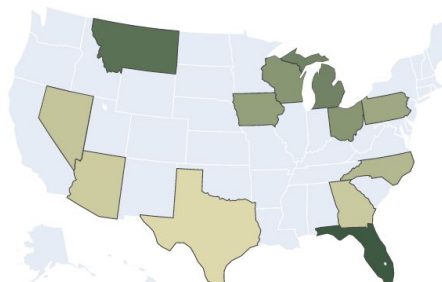
Whites without bachelor's degree

These mirror the previous results, despite the two populations having different voting tendencies

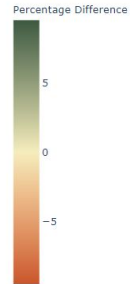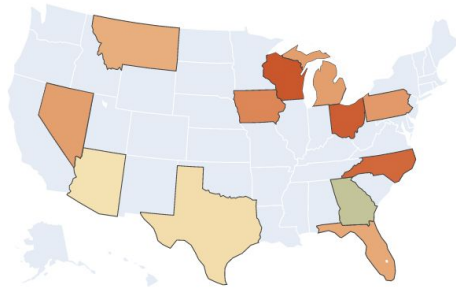Shift in pre/post-election poll data among White + No BA/BS voters for Biden
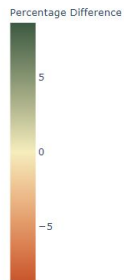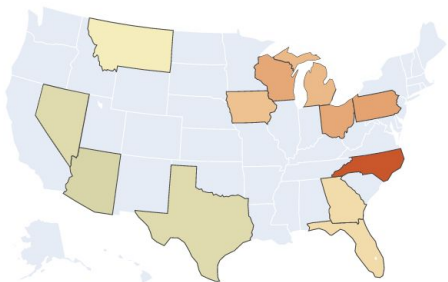
Shift in pre/post-election poll data among White + No BA/BS voters for Trump
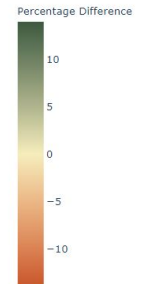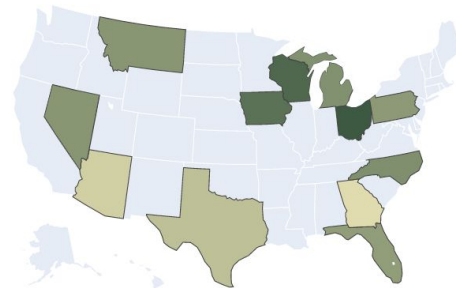
Age 65+

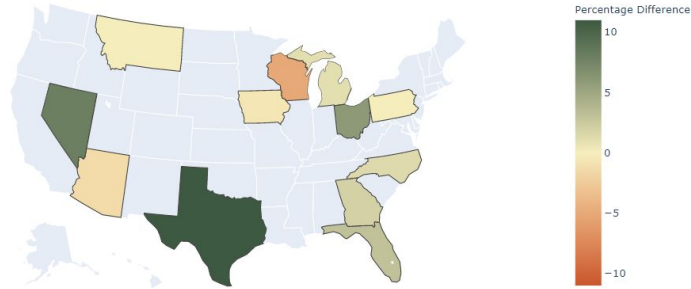Shift in pre/post-election poll data among 65+ voters for Biden

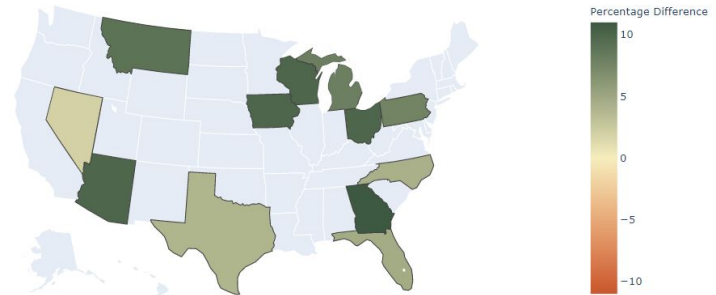Shift in pre/post-election poll data among 65+ voters for Trump

Independent voters

Independents broke more for Trump than for Biden in most of the country, particularly in the Upper Midwest.

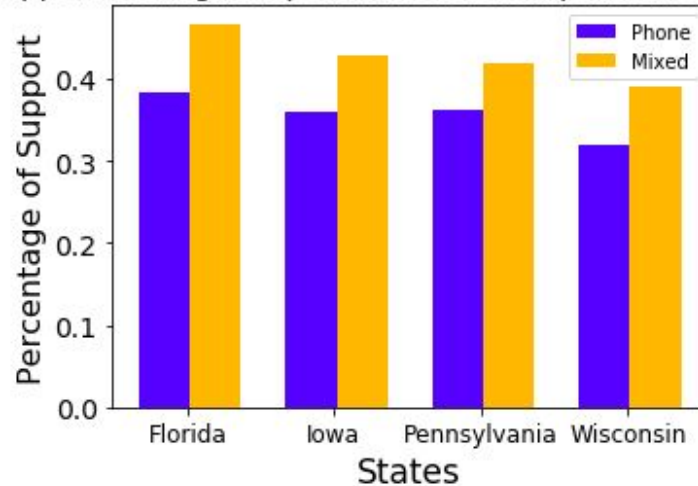Shift in pre/post-election poll data among Independent voters for Biden

Shift in pre/post-election poll data among Independent voters for Trump

Polling methodology

Polling by phone interview in blue and by a mix of phone, text, and online survey in yellow.

Support Among Independents for Trump in Various States

Future Directions

- This is a very rich problem with many dimensions that we didn't have the time or data to explore. Some of the most interesting follow-up questions include:
    - Formulating an objective score of candidate ideology and conditioning on that
    - A direct comparison of presidential and congressional candidates per district
    - Incorporating turnout data and comparing with polling sample distributions
    - Conditioning on preferred source of news and social media

Thank you!

Questions?