A Data-Driven Analysis of Youth Turnout Impact on U.S. Presidential Elections

Claire Zhu

ABSTRACT

Youth voters between 18 and 29 years of age have consistently had the lowest turnout among all age groups in U.S. presidential elections.

This research was designed to study whether the number of votes from this age group would make a difference on the election results, had young voters participated more actively. An original method was devised to do so by simulating higher-than-actual youth voter turnout using various voting scenarios – combinations of voting rates and voter choices derived from past elections. The findings are:

- When 18- to 29-year olds participate in the election as actively as the entire voting population, they are definitely a formidable force in influencing the presidential election results.
- If either party energizes those young voters to participate and earns their votes with the margin it saw in its past wins in key states, the party stands a much better chance to win the election.

KEYWORDS

Youth voter turnout; Presidential elections; United States; Battleground states; Voting rate; Quantitative analysis

I. INTRODUCTION

A. Background and Objective

As the 2020 election is now behind us, a new 4-year presidential election cycle starts. There are many factors that will sway the voters and affect the next election's results: politics, economy, society, technology, healthcare, and others. These factors have played important roles in candidates' and political parties' campaign strategies in modern era elections.

In the end, however, it all comes down to the voters.

With the Census Bureau data of the presidential elections from 1980 to 2016[I-1], Diagram I-A-1 below

illustrates that younger Americans, i.e. 18- to 29-year olds, always had much lower voting rates when compared to older age groups. Although, according to estimates[I-2] by the Center for Information and Research on Civic Learning & Engagement (CIRCLE), youth voter turnout in the 2020 election was about 10% higher than that of the 2016 presidential election, this age group was still the least participating and hence a much less influential population on the nation's political stage.

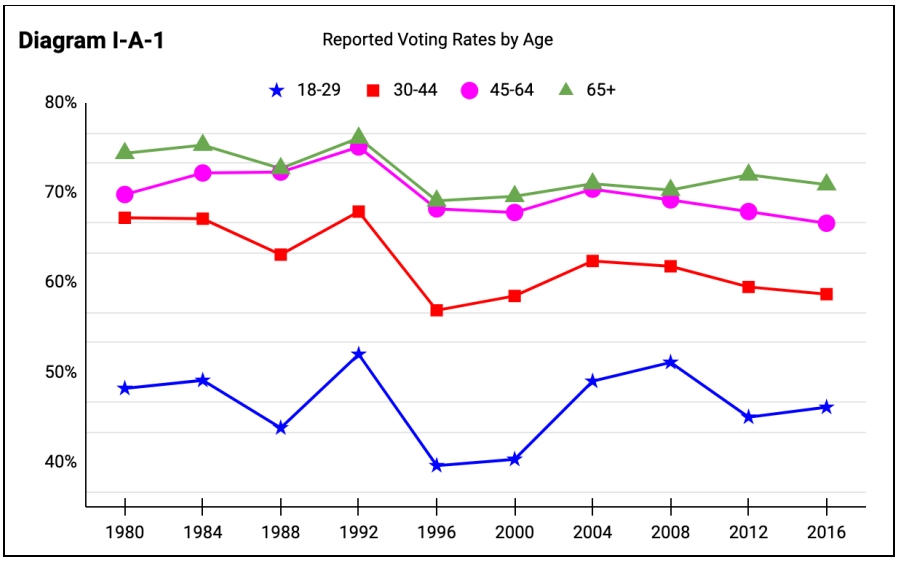


Diagram I-A-1 Reported Voting Rates by Age

Common beliefs attribute the situation to inadequate civic education, complex registration and voting, and political apathy, among others. Scholars have also discussed various theories about the cause of the stubbornly low youth turnout, including, for examples:

- Young people are habitual nonvoters and develop into habitual voters, according to Professor Eric Plutzer's article, "Becoming a Habitual Voter: Inertia, Resources, and Growth in Young Adulthood"[I-3].
- While not lacking interest in politics and motivations, young voters need the skills related to self-regulation to overcome internal and external barriers to vote, detailed in the book "Making Young Voters: Converting Civic Attitudes into Civic Action" by Professor John Holbein and Professor D. Sunshine Hillygus[I-4].
- The mutual neglect between politicians and young adults widens "the age bias in electoral

participation", and changes in media habits have led to the situation in which young people are far less likely to be exposed to news about public affairs, stated in the book, "Is Voting for Young People", by Professor Martin P. Wattenberg[I-5].

In recent decades, the 18- to 29-year-old age group has had 50 million or so eligible voters. While that accounts for about 20% of all legal voters, several factors need to be considered when trying to understand the role of youth in presidential elections:

- Are the number of votes from this group enough to make a significant impact on election results if they choose to participate more actively, and,
- If so, how do we better work with these young people, starting before they become eligible to vote, in order to improve their turnout?

This research was designed to study and answer the first question in particular. The objective was to determine the effects the young voters would have on one of the nation's most important political decisions. (A separate paper, "A Study for Improving Youth Voter Participation", details the research focused on the second question above.)

B. Methodology and Approach

This is a data-driven research with quantitative analysis of the hypothetical impact from youth voters on every presidential election since 1980. There has been an accumulation of data across various voter demographic characteristics related to the presidential elections in the last half century. With that data, this project used a scenario-based method to conduct the quantitative analysis (this method was entirely original and, to the researcher's knowledge, unused by any other research). The details of the method are as follows:

- Obtaining key metrics from data of presidential elections since 1976, for every of the 50 states and the District of Columbia, including:
  - General voting rate, i.e. total votes divided by voting age population, in each election
  - Highest, second highest, and median general voting rate among the last twelve elections

  - Percentage of votes received by each party's candidate, i.e. votes received by a candidate divided by all votes cast, in each election
  - Highest and second highest percentage of votes received by each party's candidate in the last twelve elections
  - Youth voting rate, i.e. votes from 18- to 29-year olds divided by their population, in each election
- For every presidential election from 1980 to 2020:
  - For each state, applying certain voting rate metrics, based on various higher-than-actual youth turnout scenarios, to the then 18- to 29-year-old population data to calculate the hypothetical additional youth votes
  - With the hypothetical additional votes, using certain metrics of percentage of votes received by each party's candidate, based on various youth voting decision scenarios, to calculate how many of these votes each of the candidates would get
  - Adding each candidate's additional youth votes to their actual votes from the election to determine how it would change the results for each state, as well as the results of the overall election when all states' electoral votes are added up
- Reviewing the results of all eleven elections modeled with the above steps together, upon which the research derives certain findings and then conclusions.

This method allowed the project to reach its research goal by more definitively demonstrating the significance of the young voters, by looking at how additional youth votes alone could change known election results, had the young people voted more.

C. Value and Use

Although the research focused on the presidential elections, the conclusions regarding the impact of the young voters are generally applicable to midterm elections as well as elections at state and local levels. Based on the outcome of this research, candidates, political parties, and civic groups can make an

informed decision on devoting their efforts and resources to engaging this group of voters in order to achieve their goals in elections.

II. QUANTITATIVE ANALYSIS OF YOUTH TURNOUT IMPACT

A. Data Gathering

A significant amount of time was spent on gathering reliable and multifaceted presidential election and population data spanning from 1976 to the present. While many organizations and websites have provided such data, it was difficult to arrive at a dataset that is complete, original, current, reliable, and consistently organized. After reviewing many sources with these data accessible to the public, a large volume of data was retrieved from a list of trustworthy and authoritative organizations and institutions, including:

- United States Census Bureau
- Federal Election Commission
- United States House of Representatives
- MIT Election Data and Science Lab
- Atlas of US Presidential Elections
- Kaiser Family Foundation
- Federal Register
- United States Election Project

The next major effort was to validate data accuracy before detailed analysis could be performed. Some of the validation activities were:

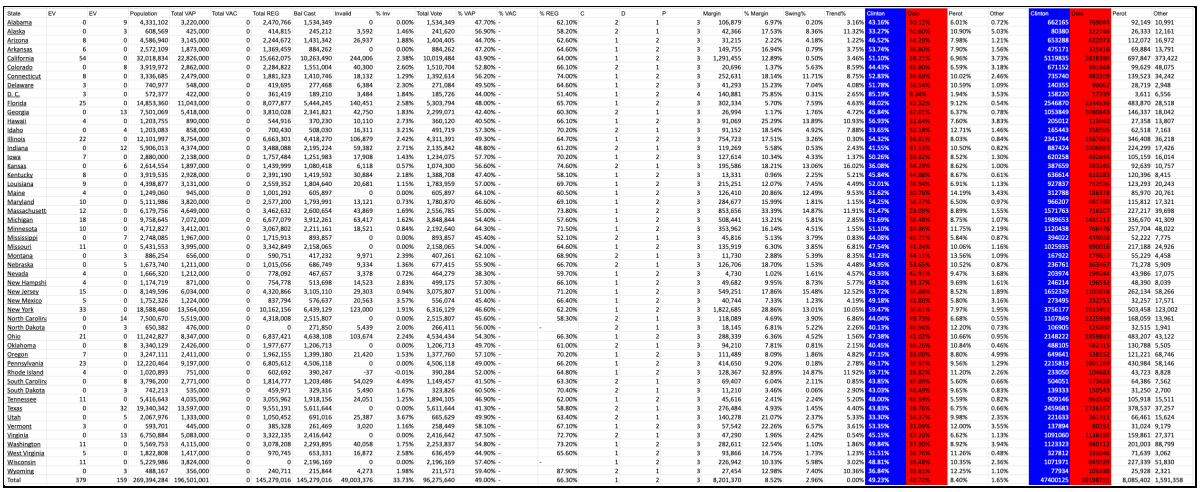
- Repeating data retrieval processes to make sure there was no clerical error
- Cross-checking data elements that are available from multiple sources
- Verifying data that was calculated by the source to ensure correctness

With this effort, the following key data elements, among others, were reliably obtained:

- Population

- Voting age population
- Votes cast
- Democratic and Republican candidates
- Votes received by Democratic and Republican candidates respectively
- Electoral vote count
- Electoral votes received by Democratic and Republican candidates respectively
- Population of 18- to 29-year-old

Each of the data elements above was obtained for every of the last twelve elections and every of the 50 states and the District of Columbia. In total, there were over 16,000 data points making up the input datasets for the next step study. These datasets were then used to calculate the key metrics needed for the scenario-based analysis. Chart II-A-1 below is an illustration of a subset of the 1996 election data used in the research:



State	EV	EV	Population	Total VAP	Total VAC	Total REG	Bal Cast	Invalid	% Inv	Total Vote	% VAP	% VAC	% REG	C	D	P	Margin	% Margin	Swing%	Trend%	Clinton	Dole	Perot	Other	Clinton	Dole	Perot	Other
Alabama	0	9	4,331,102	3,220,000	0	2,470,766	1,534,349	0	0.00%	1,534,349	47.70%	-	62.10%	2	1	3	106,879	6.97%	0.20%	3.16%	43.10%	50.12%	6.01%	0.72%	662165	769984	92,149	10,991
Alaska	0	3	608,569	425,000	0	414,815	245,212	3,592	1.46%	241,620	56.90%	-	58.20%	2	1	3	42,366	17.53%	8.36%	11.32%	33.27%	50.80%	10.90%	5.03%	80380	122746	26,333	12,161
Arizona	8	0	4,586,940	3,145,000	0	2,244,672	1,431,342	26,937	1.88%	1,404,405	44.70%	-	62.60%	1	2	3	31,215	2.22%	4.18%	1.22%	46.52%	44.29%	7.98%	1.21%	653288	622093	112,072	16,972
Arkansas	6	0	2,572,109	1,873,000	0	1,369,459	884,262	0	0.00%	884,262	47.20%	-	64.60%	1	2	3	149,755	16.94%	0.79%	3.75%	53.74%	36.80%	7.90%	1.56%	475172	325416	69,884	13,791
California	54	0	32,018,834	22,826,000	0	15,662,075	10,263,490	244,006	2.38%	10,019,484	43.90%	-	64.00%	1	2	3	1,291,455	12.89%	0.50%	3.46%	51.10%	38.21%	6.96%	3.73%	5119835	3828380	697,847	373,422
Colorado	8	0	3,919,972	2,862,000	0	2,284,822	1,551,004	40,300	2.60%	1,510,704	52.80%	-	66.10%	2	1	3	20,696	1.37%	5.63%	8.59%	44.43%	45.80%	6.59%	3.18%	671152	691848	99,629	48,075
Connecticut	8	0	3,336,685	2,479,000	0	1,881,323	1,410,746	18,131	1.29%	1,392,614	56.20%	-	74.00%	1	2	3	252,631	18.14%	11.71%	9.75%	52.83%	34.69%	10.02%	2.46%	735740	483205	139,523	34,242
Delaware	3	0	740,977	548,000	0	419,695	277,468	6,384	2.30%	271,084	49.50%	-	64.60%	1	2	3	41,293	15.23%	7.04%	4.08%	51.78%	36.54%	10.59%	1.09%	140355	99061	28,719	2,948
D. C.	3	0	572,377	422,000	0	361,419	189,210	3,484	1.84%	185,726	44.00%	-	51.40%	1	2	4	140,881	75.85%	0.31%	2.65%	85.10%	9.54%	1.94%	3.53%	158220	17729	3,611	6,556
Florida	25	0	14,853,360	11,043,000	0	8,077,877	5,444,245	140,451	2.58%	5,303,794	48.00%	-	65.70%	1	2	3	302,334	5.70%	7.59%	6.63%	48.02%	42.32%	9.12%	0.54%	2546870	2244536	483,870	28,518
Georgia	0	13	7,501,069	5,418,000	0	3,810,028	2,341,821	42,750	1.83%	2,299,071	42.40%	-	60.30%	2	1	3	26,994	1.17%	1.76%	4.72%	45.84%	47.01%	6.37%	0.78%	1053849	1080843	146,337	18,042
Hawaii	4	0	1,203,755	890,000	0	544,916	370,230	10,110	2.73%	360,120	40.50%	-	66.10%	1	2	3	91,069	25.29%	13.89%	10.93%	56.19%	31.64%	7.60%	3.83%	205012	115364	27,358	13,807
Idaho	4	0	1,203,083	858,000	0	700,430	508,030	16,311	3.21%	491,719	57.30%	-	70.20%	2	1	3	91,152	18.54%	4.92%	7.88%	33.65%	52.18%	12.71%	1.46%	165443	256595	62,518	7,163
Illinois	22	0	12,101,997	8,754,000	0	6,663,301	4,418,270	106,879	2.42%	4,311,391	49.30%	-	64.70%	1	2	3	754,723	17.51%	3.26%	0.30%	54.32%	38.81%	8.03%	0.84%	2341744	1672021	346,408	36,218
Indiana	0	12	5,906,013	4,374,000	0	3,488,088	2,195,224	59,382	2.71%	2,135,842	48.80%	-	61.20%	2	1	3	159,269	5.58%	0.33%	2.43%	41.35%	47.13%	10.50%	0.82%	887424	1006693	224,299	17,426
Iowa	7	0	2,880,000	2,138,000	0	1,757,484	1,251,983	17,908	1.43%	1,234,075	57.70%	-	70.20%	1	2	3	127,614	10.34%	4.33%	1.37%	50.26%	39.92%	8.52%	1.30%	620258	492644	105,159	16,014
Kansas	0	6	2,614,554	1,897,000	0	1,439,999	1,080,418	6,118	0.57%	1,074,300	56.60%	-	74.60%	2	1	3	195,586	18.21%	13.06%	16.02%	36.08%	54.29%	8.62%	1.00%	387659	583245	92,639	10,757
Kentucky	8	0	3,919,535	2,928,000	0	2,391,190	1,419,592	30,884	2.18%	1,388,708	47.40%	-	58.10%	1	2	3	13,331	0.96%	2.25%	5.21%	45.84%	44.88%	8.67%	0.61%	636514	623183	120,396	8,415
Louisiana	9	0	4,398,877	3,131,000	0	2,559,352	1,804,640	20,681	1.15%	1,783,959	57.00%	-	69.70%	1	2	3	215,251	12.07%	7.45%	4.49%	52.01%	39.94%	6.91%	1.13%	927837	712586	123,293	20,243
Maine	4	0	1,249,060	945,000	0	1,001,292	605,897	0	0.00%	605,897	64.10%	-	60.50%	1	2	3	126,410	20.86%	12.49%	9.59%	51.62%	30.76%	14.19%	3.43%	312788	186228	85,970	20,761
Maryland	10	0	5,111,986	3,820,000	0	2,577,200	1,793,991	13,121	0.73%	1,780,870	46.60%	-	69.10%	1	2	3	284,677	15.99%	1.81%	1.15%	54.25%	38.27%	6.50%	0.97%	966207	681936	115,812	17,321
Massachusett	12	0	6,179,756	4,649,000	0	3,462,632	2,600,654	43,869	1.69%	2,556,785	55.00%	-	73.80%	1	2	3	853,656	33.39%	14.87%	11.91%	61.47%	28.09%	8.89%	1.55%	1571763	718107	227,217	39,698
Michigan	18	0	9,758,645	7,072,000	0	6,677,079	3,912,261	63,417	1.62%	3,848,844	54.40%	-	57.60%	1	2	3	508,441	13.21%	5.81%	2.85%	51.69%	38.48%	8.75%	1.07%	1989653	1481212	336,670	41,309
Minnesota	10	0	4,712,827	3,412,001	0	3,067,802	2,211,161	18,521	0.84%	2,192,640	64.30%	-	71.50%	1	2	3	353,962	16.14%	4.51%	1.55%	51.10%	34.96%	11.75%	2.19%	1120438	766439	257,704	48,022
Mississippi	0	7	2,748,085	1,967,000	0	1,715,913	893,857	0	0.00%	893,857	45.40%	-	52.10%	2	1	3	45,816	5.13%	3.79%	0.89%	44.08%	49.21%	5.84%	0.87%	394022	439838	52,222	7,775
Missouri	11	0	5,431,553	3,995,000	0	3,342,849	2,158,065	0	0.00%	2,158,065	54.00%	-	64.60%	1	2	3	135,919	6.30%	3.85%	6.81%	47.54%	41.24%	10.06%	1.16%	1025935	890016	217,188	24,926
Montana	0	3	886,254	656,000	0	590,751	417,232	9,971	2.39%	407,261	62.10%	-	68.90%	2	1	3	11,730	2.88%	5.39%	8.35%	41.23%	44.11%	13.56%	1.09%	167922	179702	55,229	4,458
Nebraska	0	5	1,673,740	1,211,000	0	1,015,056	686,749	9,334	1.36%	677,415	55.90%	-	66.70%	2	1	3	126,706	18.70%	1.53%	4.48%	34.95%	53.65%	10.52%	0.87%	236761	363467	71,278	5,909
Nevada	4	0	1,666,320	1,212,000	0	778,092	467,657	3,378	0.72%	464,279	38.30%	-	59.70%	1	2	3	4,730	1.02%	1.61%	4.57%	43.93%	42.91%	9.47%	3.68%	203974	199244	43,986	17,075
New Hampshi	4	0	1,174,719	871,000	0	754,778	513,698	14,523	2.83%	499,175	57.30%	-	66.10%	1	2	3	49,682	9.95%	8.73%	5.77%	48.32%	39.37%	9.69%	1.61%	246214	196532	48,390	8,039
New Jersey	15	0	8,149,596	6,034,000	0	4,320,866	3,105,110	29,303	0.94%	3,075,807	51.00%	-	71.20%	1	2	3	549,251	17.86%	15.48%	12.52%	53.72%	35.46%	8.52%	1.89%	1652329	1090394	262,134	58,266
New Mexico	5	0	1,752,326	1,224,000	0	837,794	576,637	20,563	3.57%	556,074	45.40%	-	66.40%	1	2	3	40,744	7.33%	1.23%	4.19%	48.18%	41.80%	5.80%	3.16%	273495	232751	32,257	17,571
New York	33	0	18,588,460	13,564,000	0	10,162,156	6,439,129	123,000	1.91%	6,316,129	46.60%	-	62.20%	1	2	3	1,822,685	28.86%	13.01%	10.05%	59.47%	30.61%	7.97%	1.95%	3756177	1933401	503,458	123,002
North Carolin	14	0	7,300,670	5,519,000	0	4,318,008	2,515,807	0	0.00%	2,515,807	45.60%	-	58.30%	2	1	3	118,089	4.69%	3.90%	6.86%	44.04%	48.73%	6.68%	0.55%	1107849	1225938	168,059	13,961
North Dakota	0	3	650,382	476,000	0	0	271,850	5,439	2.00%	266,411	56.00%	-	-	2	1	3	18,145	6.81%	5.22%	2.26%	40.13%	46.94%	12.20%	0.73%	106905	125050	32,515	1,941
Ohio	21	0	11,242,827	8,347,000	0	6,837,421	4,638,108	103,674	2.24%	4,534,434	54.30%	-	66.30%	1	2	3	288,339	6.36%	4.52%	1.56%	47.38%	41.02%	10.66%	0.95%	2148222	1859883	483,207	43,122
Oklahoma	8	0	3,340,129	2,426,000	0	1,977,677	1,206,713	0	0.00%	1,206,713	49.70%	-	61.00%	2	1	3	94,210	7.81%	0.81%	2.15%	40.45%	48.26%	10.84%	0.46%	488105	582315	130,788	5,505
Oregon	7	0	3,247,111	2,411,000	0	1,962,155	1,399,180	21,420	1.53%	1,377,760	57.10%	-	70.20%	1	2	3	111,489	8.09%	1.86%	4.82%	47.15%	39.06%	8.80%	4.99%	649641	538102	121,221	68,746
Pennsylvania	23	0	12,220,464	9,197,000	0	6,805,612	4,506,118	0	0.00%	4,506,118	49.00%	-	66.20%	1	2	3	414,650	9.20%	0.18%	2.78%	49.17%	39.97%	9.56%	1.29%	2215819	1801169	430,984	58,146
Rhode Island	4	0	1,020,893	751,000	0	602,692	390,247	-37	-0.01%	390,284	52.00%	-	64.80%	1	2	3	128,367	32.89%	14.87%	11.92%	59.71%	26.82%	11.20%	2.26%	233050	104601	43,723	8,828
South Carolin	8	0	3,796,200	2,771,000	0	1,814,777	1,203,486	54,029	4.49%	1,149,457	41.50%	-	63.30%	2	1	3	69,407	6.04%	2.11%	0.85%	43.85%	49.89%	5.60%	0.66%	504051	573458	64,386	7,562
South Dakota	3	0	742,213	535,000	0	459,971	329,316	5,490	1.67%	323,826	60.50%	-	70.40%	2	1	3	11,210	3.46%	0.06%	2.90%	43.09%	46.49%	9.65%	0.83%	139333	150343	31,250	2,700
Tennessee	11	0	5,416,643	4,035,000	0	3,055,962	1,918,156	24,051	1.25%	1,894,105	46.90%	-	62.00%	1	2	3	45,616	2.41%	2.24%	5.20%	48.02%	45.59%	5.59%	0.82%	909146	863530	105,918	15,511
Texas	0	32	19,340,342	13,597,000	0	9,551,191	5,611,644	0	0.00%	5,611,644	41.30%	-	58.80%	2	1	3	276,484	4.93%	1.45%	4.40%	43.83%	48.76%	6.75%	0.66%	2459683	2736107	378,537	37,257
Utah	0	5	2,067,976	1,333,000	0	1,050,452	691,016	25,387	3.67%	665,629	49.90%	-	63.40%	2	1	3	140,278	21.07%	2.37%	5.33%	33.30%	54.37%	9.98%	2.35%	221633	361911	66,461	15,624
Vermont	3	0	593,701	445,000	0	385,328	261,469	3,020	1.16%	258,449	58.10%	-	67.10%	1	2	3	57,542	22.26%	6.57%	3.61%	53.35%	31.09%	12.00%	3.55%	137894	80352	31,024	9,179
Virginia	0	13	6,750,884	5,083,000	0	3,322,135	2,416,642	0	0.00%	2,416,642	47.50%	-	72.70%	2	1	3	47,290	1.96%	2.42%	0.54%	45.15%	47.10%	6.62%	1.13%	1091060	1138356	159,861	27,371
Washington	11	0	5,569,753	4,115,000	0	3,078,208	2,293,895	40,058	1.75%	2,253,837	54.80%	-	73.20%	1	2	3	282,611	12.54%	1.10%	1.86%	49.84%	37.30%	8.92%	3.94%	1123323	840712	201,003	88,799
West Virginia	5	0	1,822,808	1,417,000	0	970,745	653,331	16,872	2.58%	636,459	44.90%	-	65.60%	1	2	3	93,866	14.75%	1.73%	1.23%	51.51%	36.76%	11.26%	0.48%	327812	233946	71,639	3,062
Wisconsin	11	0	5,229,986	3,824,000	0	0	2,196,169	0	0.00%	2,196,169	57.40%	-	-	1	2	3	226,942	10.33%	5.98%	3.02%	48.81%	38.48%	10.35%	2.36%	1071971	845029	227,339	51,830
Wyoming	0	3	488,167	356,000	0	240,711	215,844	4,273	1.98%	211,571	59.40%	-	87.90%	2	1	3	27,454	12.98%	7.40%	10.36%	36.84%	49.81%	12.25%	1.10%	77934	105388	25,928	2,321
Total	379	159	269,394,284	196,501,001	0	145,279,016	145,279,016	49,003,376	33.73%	96,275,640	49.00%	-	66.30%	1	2	3	8,201,370	8.52%	2.96%	0.00%	49.23%	40.72%	8.40%	1.65%	47400125	39197830	8,085,402	1,591,358

Chart II-A-1 A Subset of the 1996 Presidential Election Data

B. Scenario Development

The purpose of this research was to study the impact of voting by the 18- to 29-year-old age group, had they had a more active participation in the presidential elections. The research designed and used a scenario-based approach, or a what-if approach, to tackle the task at hand.

Four scenarios were developed, each representing a hypothetical situation of higher-than-actual youth

turnout and their voting decisions. Every such scenario was constructed by using a combination of the options of the variables detailed below:

- The level of youth participation in terms of 18- to 29-year olds voting rate at state level, with the following options:
  - Youth voting rate is the same as the highest voting rate of the state's entire population since 1976.
  - Youth voting rate is the same as the 2nd highest voting rate of the state's entire population since 1976.
  - Youth voting rate is the same as the median voting rate of the state's entire population since 1976.
- The portion of the hypothetical additional youth votes received by a party in a state, with the following options:
  - The portion of the hypothetical additional youth votes received by a party is calculated with the highest percentage of votes the party ever received since 1976 in that state.
  - The portion of the hypothetical additional youth votes received by a party is calculated with the 2nd highest percentage of votes the party ever received since 1976 in that state.
- The states used to model the impact of hypothetical additional youth votes on their election results, with the following options:
  - The battleground states or swing states only
  - The states that were not carried by a candidate who lost the election
  - All states
- The party that would be favored in voting decision among the hypothetical additional youth voters, with the following options:
  - The party that lost the election
  - The party that won the election

- Both parties

Chart II-B-1 below illustrates the combinations of these options, as well as how they make up the four scenarios and their variations.

Scenario	Variation	Level of Participation - Voting Rate			Portion of Votes Received		States			Party	
		Highest	2nd Highest	Median	Highest	2nd Highest	Battleground	Lost	All	Lost	Won
1	a	X			X		X			X	
	b	X				X	X			X	
	c		X		X		X			X	
	d		X			X	X			X	
	e			X	X		X			X	
	f			X		X	X			X	
2	a	X			X		X				X
	b	X				X	X				X
	c		X		X		X				X
	d		X			X	X				X
	e			X	X		X				X
	f			X		X	X				X
3	a	X			X			X		X	
	b	X				X		X		X	
	c		X		X			X		X	
	d		X			X		X		X	
4	a	X			X				X	X	X
	b	X				X			X	X	X
	c		X		X				X	X	X
	d		X			X			X	X	X

Chart II-B-1 Scenarios and Scenario Variations

Described in details below are the what-if situations represented by the four scenarios that were used to analyze the impact on the past presidential election results:

Scenario 1:

- For the elections that the Democratic Party lost, adding the additional youth votes, resulting from their turnout being the same as the highest, the 2nd highest, or the median level turnout of the entire population the state has ever seen, into the actual votes of all battleground states, with the Democratic Party gaining a portion of these votes at a magnitude the same as the most or the 2nd most votes in percentage the Democratic Party has ever had in that state
- For the elections that the Republican Party lost, adding the additional youth votes, resulting from their turnout being the same as the highest, the 2nd highest, or the median level turnout of the entire population the state has ever seen, into the actual votes of all battleground states, with the Republican Party gaining a portion of these votes at a magnitude the same as the most or the 2nd most votes in percentage the Republican Party has ever had in that state

Scenario 2:

- For the elections that the Democratic Party won, adding the additional youth votes, resulting from their turnout being the same as the highest, the 2nd highest, or the median level turnout of the entire population the state has ever seen, into the actual votes of all battleground states, with the Democratic Party gaining a portion of these votes at a magnitude the same as the most or the 2nd most votes in percentage the Democratic Party has ever had in that state
- For the elections that the Republican Party won, adding the additional youth votes, resulting from their turnout being the same as the highest, the 2nd highest, or the median level turnout of the entire population the state has ever seen, into the actual votes of all battleground states, with the Republican Party gaining a portion of these votes at a magnitude the same as the most or the 2nd most votes in percentage the Republican Party has ever had in that state

Scenario 3:

For each election a party lost, adding the additional youth votes, resulting from their turnout being the same as the highest or the 2nd highest turnout of the entire population the state has ever seen, into the actual votes of all the states the party didn't carry, with the party gaining a portion of these votes at a magnitude the same as the most or the 2nd most votes in percentage the party has ever had in that state

Scenario 4:

- In every election, for each state that the Democratic Party didn't carry, adding the additional youth votes, resulting from their turnout being the same as the highest or the 2nd highest turnout of the entire population the state has ever seen, into the actual votes of that state, with the Democratic Party gaining a portion of these votes at a magnitude the same as the most or the 2nd most votes in percentage the Democratic Party has ever had in that state
- In every election, for each state that the Republican Party didn't carry, adding the additional youth votes, resulting from their turnout being the same as the highest or the 2nd highest turnout of the entire population the state has ever seen, into the actual votes of that state, with the Republican

Party gaining a portion of these votes at a magnitude the same as the most or the 2nd most votes in percentage the Republican Party has ever had in that state

Every such scenario was repeatedly played out for every presidential election since 1980, and then a collective assessment was made on all these modeled election results. The objective was to understand how many of these elections would see the results changed when each of the hypothetical scenarios was applied, which would then support the conclusion about the potential impact the youth could levy on the nation's biggest political event when they participate actively.

C. Data Transformation

The raw data from multiple sources were transformed into datasets that were ready for analysis and extrapolation. The data transformation process is visualized with a flow diagram, Diagram II-C-1, below.

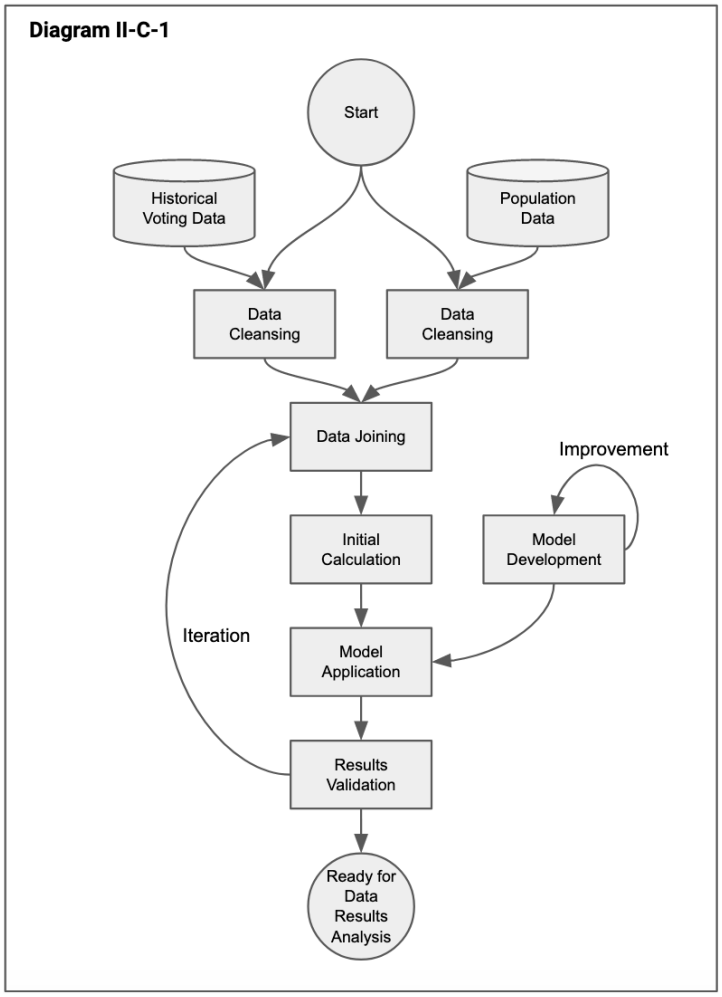


Diagram II-C-1 Data Transformation Flow

The following are the activities that were included in the transformation process:

- Data cleansing: to ensure the accuracy and completeness of source data being used
- Data joining: to join different data points based on certain criteria and common keys
- Initial calculation: to derive key metrics to be used for what-if scenarios
- Model development: to establish data processing models that represent the what-if scenarios, and to verify and improve the models in multiple iterations
- Model application: to run data through the models
- Validation and iteration: to review the data results for accuracy and completeness, and to iterate as necessary

While this research focused specifically on the modern era presidential elections, i.e. the twelve elections from 1976 to 2020, the models were built with reusability and scalability in mind. A reusable and scalable model in this case allowed:

- Processing data from multiple elections from before 1976, providing an opportunity to compare and contrast the difference of key metrics between modern day elections and those from almost half a century ago
- Plugging in additional data from new elections when they become available, enriching the dataset to produce more reliable results for analysis and opportunity for ongoing study
- Simulating the impact of young voters on past elections with other hypothetical or survey data based youth voting patterns

Such capabilities of the models let this research easily and efficiently include and update 2020 presidential election details as soon as they became available, without needing to rework from the beginning.

D. Data Results Analysis

Analysis of the data results made sense of the data after the transformation step. The processed data presented information that not only enabled the research to meet its fundamental objective – determining

the effects the young voters could have on the presidential elections – but also demonstrated the level of youth turnout that would be needed to have such effects on the election results.

The following activities were involved in data results analysis, in order to identify useful information and form conclusions:

- Ascertain how data results contribute to the research objective, through comparative study among scenarios, over-the-time trend review on multiple factors, probability and statistical analysis, etc.
- Decipher and interpret the meaning of the data, to determine if the results supported or refuted the hypothesis that young voters of 18- to 29-year-old can impact the election results in a meaningful way
- Assess the consistency, or lack thereof, of the data results and their meaning across all data points used in this research

Results analysis went through several rounds, as the successive iterations saw improved selection of data, further clarity on the information derived from the data results, and more definitive conclusions.

The outcome of this step brought forward the key findings of the research discussed later.

E. Other Considerations

There were items taken note of throughout the initiation, definition, execution, and conclusion phases of the research. Those items reflect the areas of imperfection that were given specific considerations, concerning data accuracy, exception process, trade-off, etc. Details are available in Supplementary Information II.

While each of these items alone had minimum to no impact on the overall findings of the research, the project still took additional steps to verify the collective effect of those items. This was done using synthetically enlarged data variances caused by those items in both directions and rerunning the data models, which yielded no change to the data results until the variances were unreasonably larger than realistically possible, thus proving that the collective effect of the various imperfections had no impact on the findings either.

F. Key Findings

1. If they participate more actively, youth voters have the potential to make a significant difference in presidential elections.

Diagram II-F-1 below illustrates the difference between the population of 18- to 29-year-old and the popular vote margin in each election. Other than Reagan's landslide win in 1984, most of the popular vote margins are much smaller than the youth population. Similarly, at state level, vote count differences are also considerably smaller than the state's youth population.

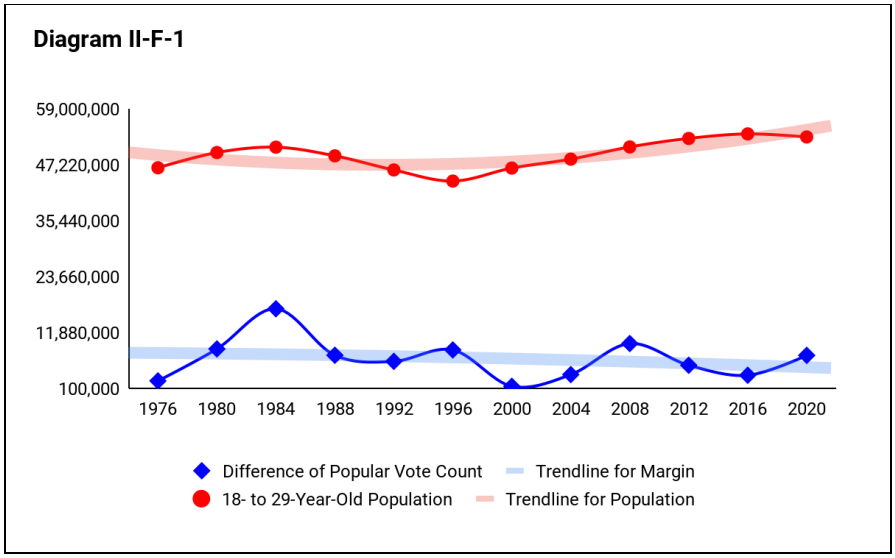


Diagram II-F-1 Youth Population Vs. Popular Vote Margin

Also in the diagram, the trend lines indicate that the margin of victory has been on a downward trajectory since the 1990s, while the population of 18- to 29-year-old has been upticking. This is a clear indication that the possibility of youth voters changing the overall election results is increasing, as long as they participate in the elections actively.

While the popular vote differences between the two parties have been decreasing over the years, the voting age population and the actual presidential votes cast have both been increasing, as illustrated in Diagram II-F-2 below.

The fact that these metrics are moving in opposite directions proves that the elections are becoming more competitive. In the paper, "Reinterpreting the 2016 Election and Presidential Election

Competition"[II-1], using the Presidential Competition Index, the study scored all 48 presidential elections from 1828 to 2016 to define the list of top ten most competitive elections. Three of the elections from the scope of this research, i.e. 1992, 2000 and 2016, made the list, also proving that the races for the White House are becoming more closely-contested in recent decades.

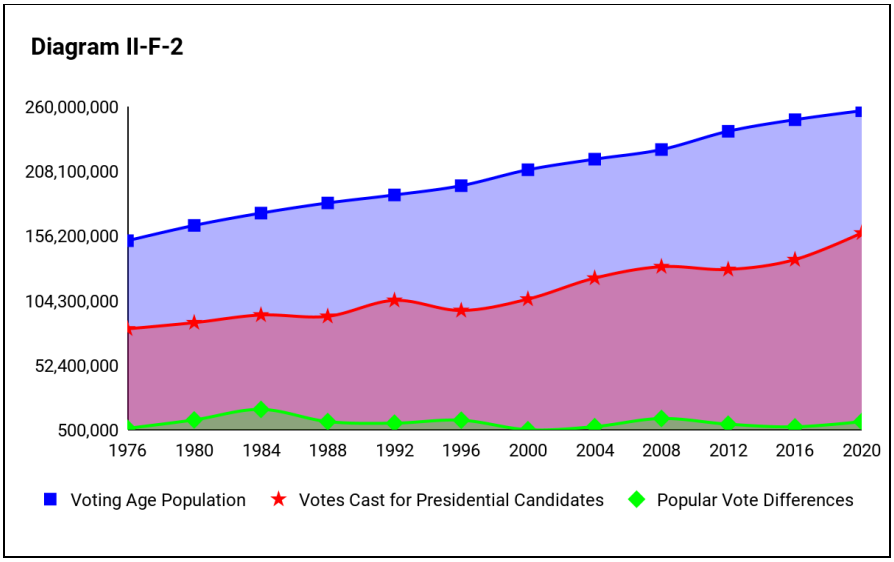


Diagram II-F-2 Voting Age Population Vs. Votes Cast Vs. Popular Vote Margin

In increasingly competitive elections, a smaller vote difference decides the election results. With youth being the least active voting group, they have the votes with the most potential to overcome the margin and make a difference to the election results.

However, because of the Electoral College system, the popular vote at the country level does not determine the winner of the election. Knowing that, the research focused on studying youth voting at state level to assess the impact on overall election outcomes. Diagram II-F-3 below illustrates the parties, blue Democratic or red Republican, that won each state in each presidential election from 1976 to 2020.

While there are a handful of states that consistently voted for the same party over the years, most of the states had varied results in different elections. To identify the states that are more likely to have different outcomes, the research used the election data from 1976 to 2020 to define the list of battleground states, or swing states. They are Colorado, Florida, Iowa, Michigan, Minnesota, Nevada, New Hampshire, North Carolina, Ohio, Pennsylvania, Virginia, and Wisconsin.

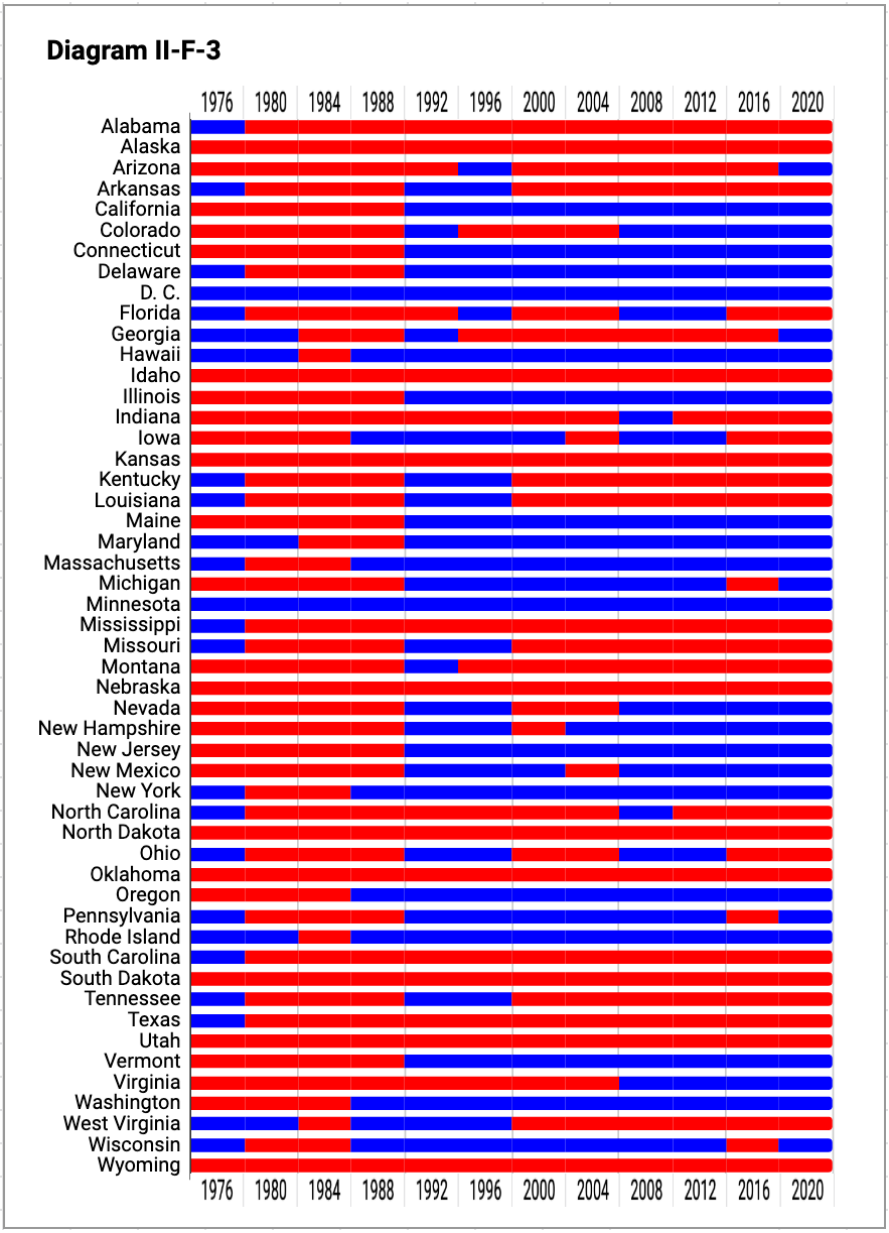


Diagram II-F-3 Presidential Election Results for the 50 States and the District of Columbia from 1976 - 2020

Battleground states have changed over time. These states made the list of this research for meeting one or more of the following criteria:

- They gave the winning candidates the least amount of vote count margins, based on the results over the years.
- They were carried by each of the two parties in close to half of the last twelve elections.
- They had back-to-back-to-back changes of presidential election winners in the last decade.

The following two diagrams demonstrate why the battleground states are critical in the elections.

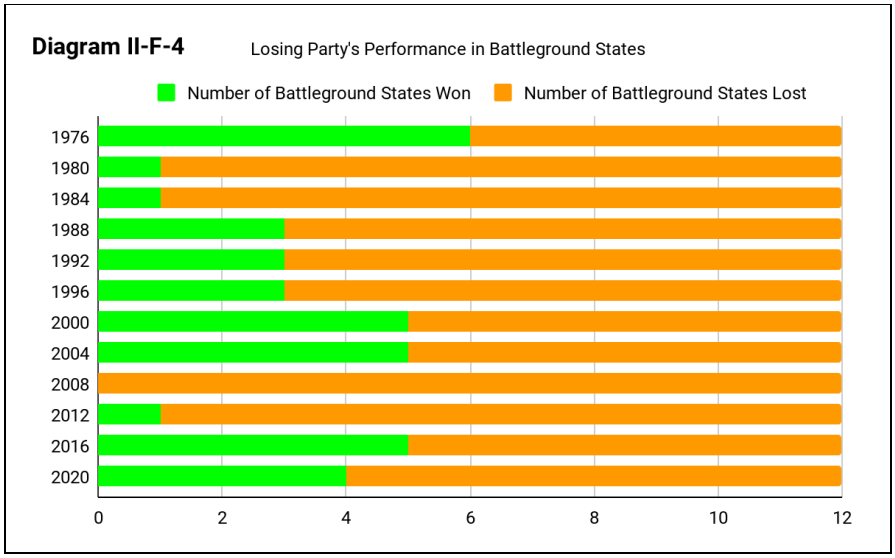


Diagram II-F-4 Battleground State Wins and Losses by the Party that Lost the Election

Diagram II-F-4 shows that, in each election except 1976, the losing party always lost more battleground states than it won.

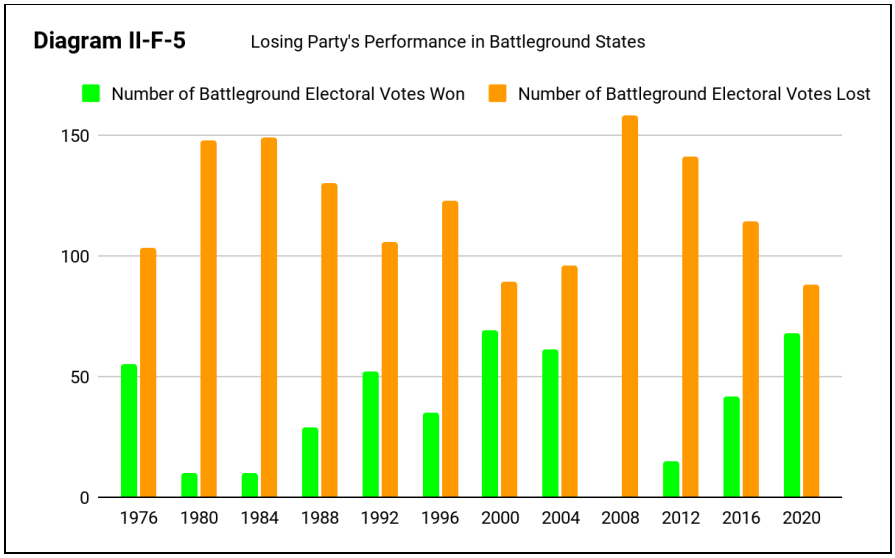


Diagram II-F-5 Battleground State Electoral Votes Wins and Losses by the Party that Lost the Election

Diagram II-F-5 is more telling, showing that the losing party in every election lost significantly more electoral votes among the battleground states than it won.

The importance of battleground states goes beyond swaying the results of the presidential election. For example, the paper, “Battleground States and Voter Participation in U.S. Presidential Elections: An

Empirical Test"[II-2], provided empirical support for the hypothesis that "the greater the degree to which a given state is a battleground state, the greater the expected benefits from voting in that state and hence the greater the voter turnout in that state".

The following two findings focus on youth voter turnout impact in battleground states.

2. In battleground states, the youth voting population can meaningfully influence the election results with their votes, especially in tight races.

This finding was drawn from the data results of Scenario 1.

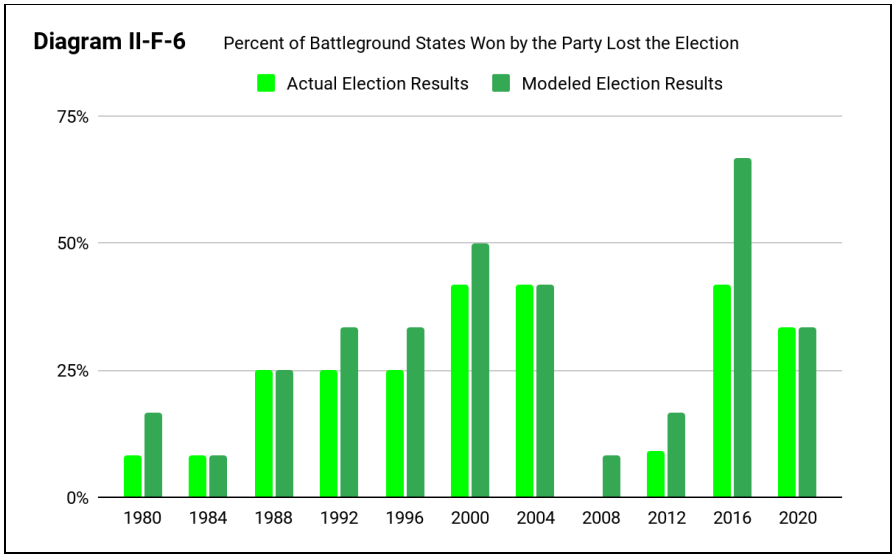


Diagram II-F-6 Actual Vs. Modeled Results - Percent of Battleground States Won by the Party that Lost the Election

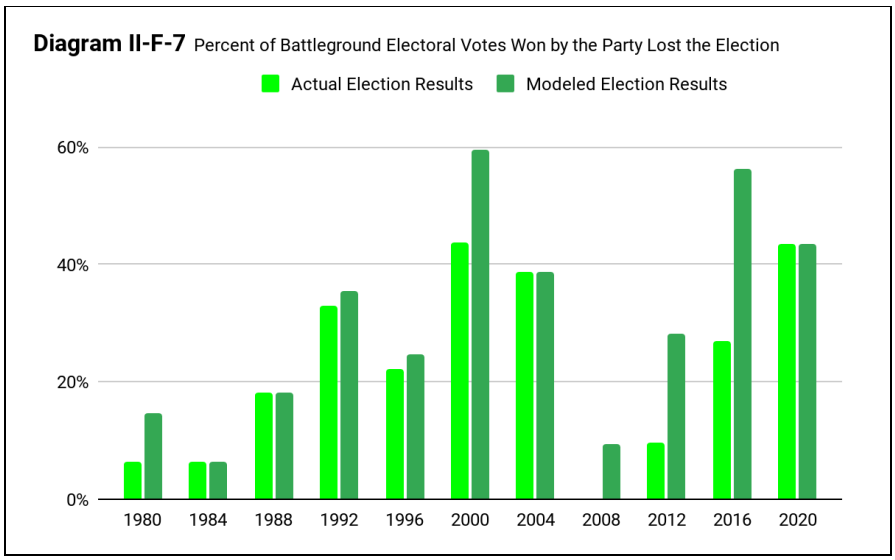


Diagram II-F-7 Actual Vs. Modeled Results - Percent of Battleground State Electoral Votes Won by the Party that Lost the Election

As shown in both Diagram II-F-6 and II-F-7, the hypothetical additional young voters' participation and their candidate choices notably changed the election results in these battleground states. Based on the modeled results, seven of the eleven elections since 1980 saw that the losing party carried more swing states than they actually did in the elections. Consequently, the losing party won more electoral votes in the model than in the actual elections.

Furthermore, Scenario 1 data also shows that the changed results in battleground states affected the outcomes of the presidential elections: those won with comfortable margins became tight races, and those already tight races saw winners changed. The diagram below illustrates the two elections that were most significantly impacted in Scenario 1:

- the 2000 election, one of the most competitive elections in recent history, and
- the 2016 election, the only election since 1888, other than 2000, that the popular vote winner didn't win the election.

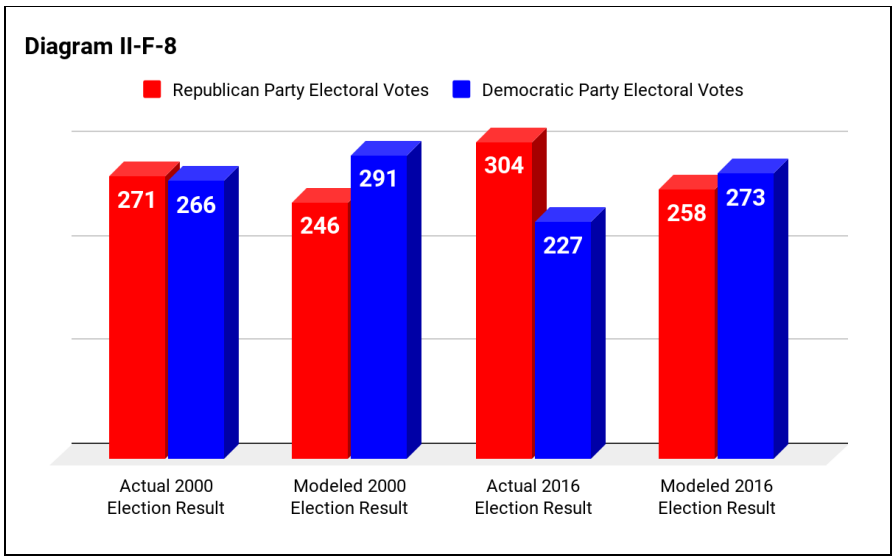


Diagram II-F-8 Actual Vs. Modeled Results - The 2000 and 2016 Presidential Elections

Based on the modeled results, in the 2000 election, Republicans won 25 less electoral votes in these battleground states, and in the 2016 election, 46 less, which were enough to flip the results of both elections. All it needed for that to happen was the youth voting rate being the same as the highest general voting rate these battleground states have ever had, and the Democratic Party gaining a portion of these

youth votes at a magnitude the same as the most votes, percentage-wise, the party has ever had in those battleground states.

3. Motivating youth in battleground states and earning their votes are of essential importance for both winning and losing parties.

Scenario 2 modeled the what-if situation in which the winning party earns more votes from the additional 18- to 29-year-old voters. In this model, the winner gained slightly more electoral votes from battleground states in five of the eleven elections, while keeping the same already large margins in the other six elections. Based on the results from both Scenario 1 and Scenario 2, it is evident that the amount of untapped youth votes in battleground states could either flip enough electoral votes for the trailing candidate, or solidify a convincing win for the leading candidate.

Moreover, when comparing the modeled results of Scenario 1 and 2, it shows that these additional youth votes would provide a bigger boost to the losing party than they would to the winning party in terms of electoral vote gain, as shown in below diagram.

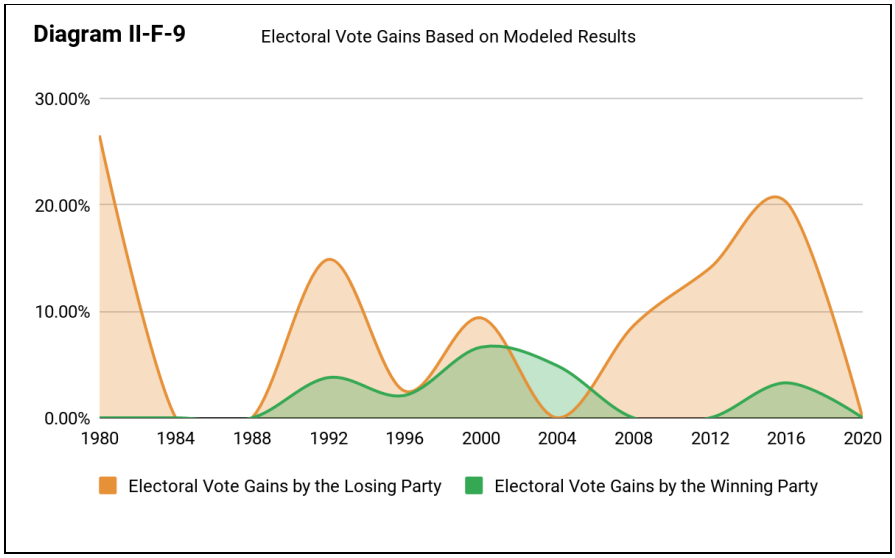


Diagram II-F-9 Modeled Electoral Vote Gains by the Losing Party Vs. the Winning Party

4. In the states that the opposing party is likely to win, improving youth turnout and being favored by the untapped voters will benefit the trailing party in a substantial way.

Diagram II-F-10 below illustrates the winners and the electoral vote differences of both the actual

results and Scenario 3 modeled results in each election.

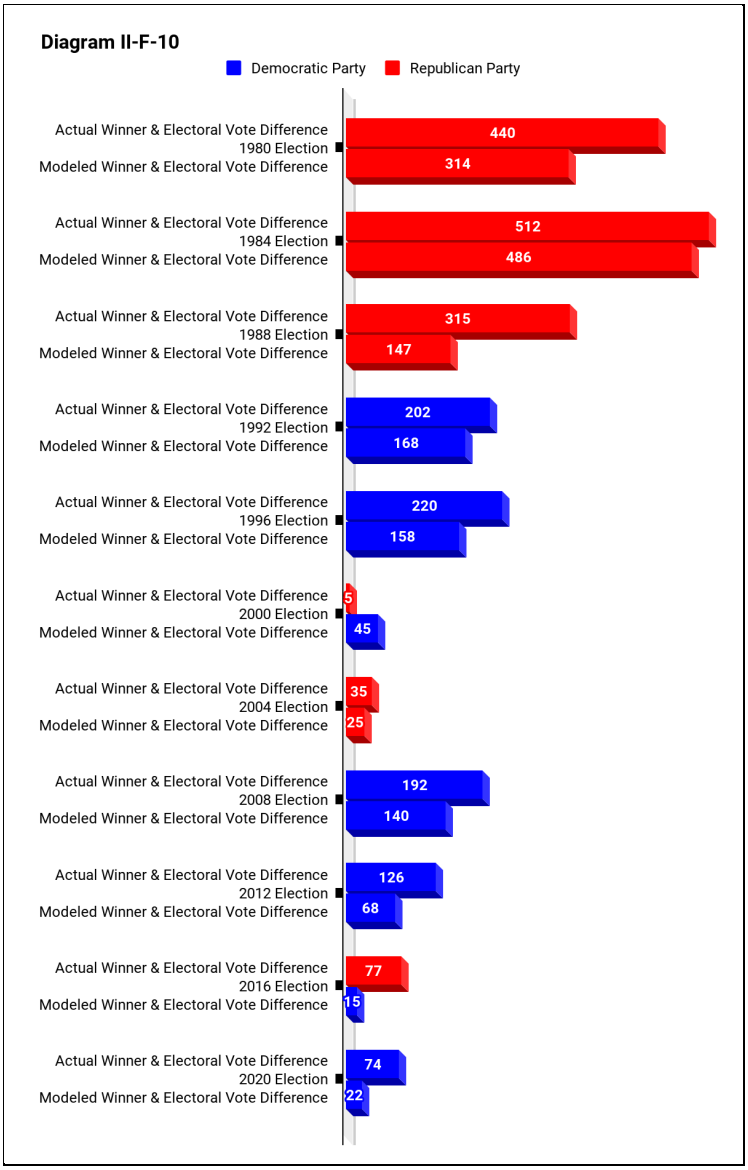


Diagram II-F-10 Actual Vs. Modeled Results - The Presidential Election Winner and Electoral Vote Difference

Here are several key observations from the diagram:

- For all eleven elections from 1980, the modeled results showed reductions in electoral vote deficit for the losing party. Some of the reductions were big enough to make the winning party's comfortable victories into very competitive races.
- Two out of the eleven elections, 2000 and 2016, saw different winners, the same as the modeled result from Scenario 1.

- Even Reagan's 1984 decisive victory lost some electoral votes in this scenario. Although Reagan had a big 18% margin of popular vote at the country level, his wins in some states, e.g. Massachusetts, were smaller and were flipped with hypothetical youth votes in the modeled results.

It can be concluded that had more youth voters come out to vote and voted more for the losing party in the states it lost, the losing party would have a better chance of winning these presidential elections.

Built on the outcome of Scenario 3, Scenario 4 continued to model the situation that both the winning and losing parties campaigned for and gained additional youth votes in the states they each lost in the elections. The diagram below illustrates electoral vote split between the two parties for both the actual election results and modeled results of Scenario 4, side by side, in the last eleven elections.

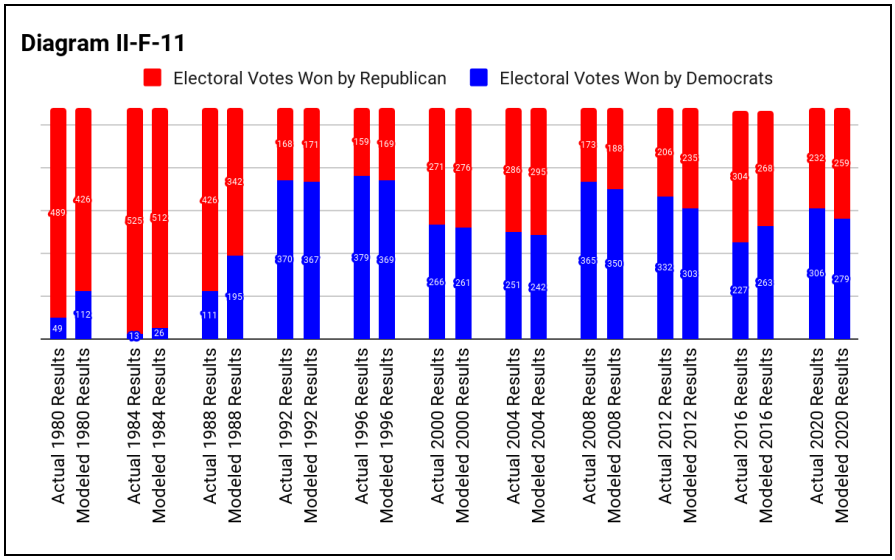


Diagram II-F-11 Actual Vs. Modeled Results - Electoral Votes Won by the Republican Party and by the Democratic Party

The diagram shows that although there were slight differences in each party's electoral vote count between the actual and the modeled results, none of the election outcome was changed and the level of competitiveness was mostly unchanged as well.

It proves that, had the party that won the election energized additional youth to vote for it in the states it lost, the party would hold onto its overall advantages and victories, even if the opposing party had managed to flip some of the states.

The conclusion drawn from the analysis of Scenario 3 and Scenario 4 is that both the leading and the

trailing party can benefit substantially from winning the untapped youth voters in opposing party's territories.

5. In addition to battleground states, higher youth voter turnout can also affect the election results of the states that are not considered strongholds.

Chart II-F-12 below shows a list of states, based on Scenario 4, that would have a different state winner with the hypothetical additional youth voters in each presidential election from 1980 to 2020.

Election Year	States	Actual State Winner	Modeled State Winner	Battleground	Electoral Votes
1980	Alabama	Republican	Democratic	No	9
	Arkansas	Republican	Democratic	No	6
	Delaware	Republican	Democratic	No	3
	Massachusetts	Republican	Democratic	No	14
	North Carolina	Republican	Democratic	Yes	13
	South Carolina	Republican	Democratic	No	8
	Tennessee	Republican	Democratic	No	10
1984	Massachusetts	Republican	Democratic	No	13
1988	California	Republican	Democratic	No	47
	Illinois	Republican	Democratic	No	24
	Maryland	Republican	Democratic	No	10
	Vermont	Republican	Democratic	No	3
1992	Georgia	Democratic	Republican	No	13
	New Hampshire	Democratic	Republican	Yes	4
	North Carolina	Republican	Democratic	Yes	14
1996	Arizona	Democratic	Republican	No	8
	Colorado	Republican	Democratic	Yes	8
	Georgia	Republican	Democratic	No	13
	Kentucky	Democratic	Republican	No	8
	Nevada	Democratic	Republican	Yes	4
	Tennessee	Democratic	Republican	No	11
2000	Florida	Republican	Democratic	Yes	25
	Iowa	Democratic	Republican	Yes	7
	New Mexico	Democratic	Republican	No	5
	Oregon	Democratic	Republican	No	7
	Wisconsin	Democratic	Republican	Yes	11
2004	New Hampshire	Democratic	Republican	Yes	4
	New Mexico	Republican	Democratic	No	5
	Wisconsin	Democratic	Republican	Yes	10
2008	Indiana	Democratic	Republican	No	11
	Missouri	Republican	Democratic	No	11
	North Carolina	Democratic	Republican	Yes	15
2012	Florida	Democratic	Republican	Yes	29
2016	Michigan	Republican	Democratic	Yes	16
	Neveda	Democratic	Republican	Yes	6
	New Hampshire	Democratic	Republican	Yes	4
	Pennsylvania	Republican	Democratic	Yes	20
	Wisconsin	Republican	Democratic	Yes	10
2020	Arizona	Democratic	Republican	No	11
	Georgia	Democratic	Republican	No	16

Chart II-F-12 States with Different Winners in Modeled Results

The following was observed from the chart:

- Among the eleven elections, based on the model, there were 40 instances that the hypothetical additional youth votes changed the state election results, and 23 of them were not in battleground

states.

- For 21 times, the modeled results showed a state flipped to the Democratic Party, and, for 19 times, to the Republican Party.

This is evident that higher young voter turnout can influence the election results in more states than those typically considered as battleground states. Elections in the last four decades have presented almost the same amount of opportunities for both parties to win over a competitive state with the votes from young voters, had they been sufficiently motivated to cast their ballots.

III. CONCLUSION

The historical data and the modeled scenarios prove that youth of 18- to 29-year-old make a truly formidable force, as they have enough votes to make a decisive difference in close races. When they show up in a presidential election with the same voting rate as other age groups, the party that wins their support stands a much better chance to win the White House for the next four years.

Updated with the newly available 2020 election results and population data, the study has solidified the foundation of the research, quantitative analysis, and findings.

As the next step, other youth voting patterns, derived from publicly available survey data of nonvoters' opinion and demographic characteristics, will be plugged into the models developed in this project to assess the effect of additional young voters on past election results. Together with the findings of this research, which was based on scenario-driven, hypothetical youth voting patterns, the combined results will be more conclusively defining youth impact on U.S. presidential elections.

ACKNOWLEDGEMENTS

I am most grateful to political scientist Robert Y. Shapiro, a professor and former chair of the Department of Political Science at Columbia University, for his enthusiastic support. Professor Shapiro reviewed multiple versions of my paper, and provided invaluable feedback on not only the research project itself but also the writing of the paper. His continued encouragement and advice resulted in important improvements in the paper, a clearer path to submission for publication, and next step research directions. As a young researcher, I couldn't be more fortunate to have had Professor Shapiro's valuable time and attention on my project.

This research relies heavily on presidential election and population data. I am grateful to those working at the organizations that collect and curate these data, and most importantly generously make these data available for others and me to analyze.

REFERENCES

[I-1] File, Thom. “Voting in America: A Look at the 2016 Presidential Election”. United States Census Bureau. 10 March 2017. https://www.census.gov/newsroom/blogs/random-samplings/2017/05/voting_in_america.html

[I-2] “Youth Voter Turnout Increased in 2020”. CIRCLE. Updated 18 November 2020. https://circle.tufts.edu/index.php/latest-research/election-week-2020#youth-voter-turnout-increased-in-2020

[I-3] Plutzer, Eric. “Becoming a Habitual Voter: Inertia, Resources, and Growth in Young Adulthood”. The American Political Science Review, Volume 96, No. 1, Pages 41–56. March 2002. https://www.jstor.org/stable/3117809

[I-4] Holbein, John B., Hillygus, D. Sunshine. “Making Young Voters: Converting Civic Attitudes into Civic Action”. Cambridge University Press. February 2020. https://doi.org/10.1017/9781108770446

[I-5] Wattenberg, Martin P. “Is Voting for Young People?” (5th Edition). Routledge. 27 April 2020. https://www.routledge.com/Is-Voting-for-Young-People/Wattenberg/p/book/9780367445522

[II-1] Rose, Douglas D. “Reinterpreting the 2016 Election and Presidential Election Competition”. ResearchGate. 20 August 2017. http://dx.doi.org/10.13140/RG.2.2.35415.78247

[II-2] Cebula, Richard J., Duquette, Christopher M., Mixon, Franklin G. “Battleground States and Voter Participation in U.S. Presidential Elections: An Empirical Test”. Applied Economics, Volume 45, 2013 - Issue 26, Pages 3795-3799. 05 November 2012. https://doi.org/10.1080/00036846.2012.727981

Supplementary Information

Here is additional information not shown in the main paper: the data, data models, modeling results in Section I, other considerations on the data and process in Section II, the study for improving youth participation in Section III and the study of key factors impacting youth turnout in Section IV.

I. THE FULL SET OF DATA, DATA MODELS, AND MODELING RESULTS

Please contact the research and author of this paper for a complete copy of data, data models, and modeling results.

II. OTHER CONSIDERATIONS ON THE DATA AND PROCESS

Considerations made in the areas concerning data accuracy, exception process, trade-offs, and etc:

- Maine and Nebraska do not award all electoral votes to the state-level winner; rather, some of the votes are awarded based on the congressional district level results. This would make it difficult to analyze as it requires the youth voting scenarios to be applied at the congressional district level. However, other than 2008 for Nebraska, 2016 for Maine, and 2020 for both of the two states, the winners at the state level actually carried all congressional districts in all other elections from 1976, hence producing a de facto winner-take-all situation. Given that, the research decided to merely mirror how electoral votes were actually awarded for its modeled calculation, instead of allocating based on the congressional district level modeled results, as this discrepancy would not impact the findings.
- As discussed above, the research used historical voting participation rate and percent of votes received to perform scenario-based analysis. Due to the fact that there were candidates that carried landslide victories in some of the elections since 1976, the research paid attention to its data use so potential distortion caused by outliers was mitigated.
- As the presidential elections are on the first Tuesday after the first Monday in the month of November, the most accurate way to count the population of 18- to 29-year-old would be using that election day as the birthday cutoff. However, the best available data is the July 1st estimates based on the then most recent decennial census data from the Census Bureau's Population Estimates Program. The variance between July and November was determined to be acceptable to the purpose of this research, as fluctuation of birth rates between July and November is not significant enough to impact the findings.
- The 2016 presidential election is the only one, among the twelve elections in this study, to have faithless voting, so 7 electoral votes were allocated to neither the Republican candidate nor the Democratic candidate. To reflect this, the 7 votes were excluded from the analysis.
- The population of the 18- to 29-year-old age group for most of the presidential election years were directly available in the data files published by the Census Bureau. Except that:
    - For 1976, the Census Bureau only published data by age groups 15-19, 20-24 and 25-29. To obtain the 18-29 population, the research added two fifths of the 15-19 population, representing

the 18-19 population, to the 20-24 and 25-29 population. It is believed that the possible fluctuation of population caused by the use of indirect data is not big enough to impact the findings.

- At the time of the data analysis in late 2020, the Census Bureau had not yet released population estimates by state and by age for 2020. Because of this, the 17-28 population from the 2019 data published by the Census Bureau was used in place of the 18-29 population when modeling and analyzing the 2020 election. It is believed that the possible fluctuation of population caused by the use of indirect data is not big enough to impact the findings.

III. THE STUDY FOR IMPROVING YOUTH PARTICIPATION

Please contact the research and author of this paper for a copy of the research paper, “A Study for Improving Youth Voter Participation”.

IV. THE STUDY OF KEY FACTORS IMPACTING YOUTH TURNOUT

Please contact the research and author of this paper for a copy of the research paper, “A Study of Key Factors Influencing Youth Voter Turnout”.