Ronald Chum
IDS 452
Case Study: Data Analytics Outcomes

Presidential Election

Introduction

From 1972 to 1988, the presidential elections were often one-sided. Republicans have a track record of dominating elections in those years, with their candidates frequently receiving more than 75% of the electoral votes cast across all 50 states. In recent years, from 2004 to 2020, the presidential election has started to change. The entire presidential election was no longer won by a single party in a landslide. In modern society, the odds of winning the election for both the Democratic party and the Republican party are about even (based on the research findings). A few states dedicated their votes to a specific party, while others swung back and forth between the Democratic and Republican parties during the state election.

Research Overview

In this research, the states' college degree education rate (in percentage) was selected as a model to explain election outcomes. All 50 states (plus DC and a special district) were categorized into 3 different tiers. Tier 1 states are considered to be swing states that do not have a preference for a political party and do not consistently vote for a specific political party. Tier 2 states are considered to be stable states; they do not have the likelihood of swinging, but they do not have a consistent voting pattern or political party preference. Tier 3 states are stable states that have a history of track records of dedicating their votes to a specific political party. After classifying the states into different tiers, tier 1 and tier 3 will be the main focus in this research due to time constraints. Tier 1 and tier 3 should produce enough evidence to project the next presidential election in 2024. Ideally, all 50 states should be tested. Each state's election outcome should be predicted, and the electoral votes will be concluded to project the next election's winning party.

Methodology

A central tendency is used to provide a visualization and informational value of the collected data. The mean, median, and mode were used to ensure that the data was usable. Based on the information provided by the data, the following terms were identified in the research: mean, median, mode, standard deviation, minimum value, maximum value, bin width, level of significance, sample size, confidence value, upper confidence, and lower confidence. Linear regression was also used to compare calculated values among selected states.

Data Integrity

This data was collected from various sources – mostly U.S. government sites such as the census bureau, election commissions, and departments of state and commerce. Since the dataset only contains education data related to 2020 and 2022, 2012 and 2016 education data were obtained from the National Center for Education Statistics and added to the dataset to predict the outcome under a larger sample size. The data set is more balanced as it includes education data from 2012, 2016, and 2020,

as well as election results from 2012 to 2020 (by state). These education data have minimal bias towards the political parties because they only focus on the education rate in each state, and the agency that operates under the U.S. Department of Education specializes in collecting, analyzing, and publishing statistics related to nation-wide education.

Hypothesis

The hypothesis in this research is that the increase in the education percentage of the population tends to result in the Democrats winning the state election in general.

The null hypothesis

H_0: An increase in the percentage of the (college degree) educated population will NOT increase the likelihood of the Democratic Party winning (at the state level).

The alternative hypothesis

H_a: An increase in the percentage of the (college degree) educated population will increase the likelihood of the Democratic Party winning (at the state level).

First, states are categorized into 3 tiers. Tier 3 states are sorted based on their historical track records of voting outcomes (at state level elections). If a state has not changed its voting in the past nine elections (state level), it is determined to be a tier 3, as it is a very stable state that dedicates its vote to a specific party. Tier 3 states include 20 states (plus the DC special district) (see attached Tables 1 and 2, Chart 1 and 2). Tier 1 states are those that have a history of swinging and have recently swung between 2016 and 2020. California and Texas are exceptions; they did not swing recently, but their electoral votes can change the entire election outcome at the federal level by themselves. If they were to swing, the impact should not be neglected. The remaining 21 states will fall into Tier 2 as their likelihood of swinging is between Tier 3 and Tier 1.

Tier 3 Observation

The data has revealed an interesting result after categorizing states into tiers. Based on 2020 values, 20 states (plus the DC special district) revealed a 9.77% difference between states that vote for the Democratic Party and states that vote for the Republican Party. The blue states have an average educated population (education rate) of 37.83%, while the red states have an average educated population (education rate) of 28.06%. In Tier 3, the calculated upper and lower confidence for blue and red states do not overlap each other. This represented that they use this as a tool to predict the election of a specific state if it falls within those confidence bands.

As for the increase in education rates (Δ education rate), the average change for blue states is 2.05% and the average change for red states is 1.51%. based on the difference between 2012 and 2020. The evidence is not significant enough to conclude that the Δ education rate will result in the win of the Democratic party in a state level election. So, based on Tier 3's evidence, it is inconclusive. However, it is reliable enough to determine the possible election outcome based on the education rate (but not the change).

Tier 1 Observation

As for Tier 1 states (see Table 3 and 4, Chart 3 and 4), the average change (∆ education rate) for blue states is 1.76% and the average change for red states is 1.82%. based on the difference between 2012 and 2020. This is even weaker than the tier 3 evidence. The evidence does not support the conclusion that the D education rate will result in the win of the Democratic party in a state level election. So, based on Tier 1's evidence, it is also inconclusive.

Because neither Tier 3 nor Tier 1 states show that the ∆ education rate has any significant level of evidence, there is no correlation that an increase in education rate can result in a Democratic party victory.

Regression Analysis

4 states are selected as a representative of the tiers that they are in; they are selected based on having the closest value to their tier's median and mean. MN and SD are selected to represent tier 3, and PA and FL are selected to represent tier 1. All 4 states have a weak correlation R2 coefficient. The regression line for all 4 is around 0.2-0.3, which is too weak to support the alternative hypothesis. Therefore, the null hypothesis is accepted (failed to reject), and the alternative hypothesis is rejected in this research.

Even though no correlation was found between voting preference and D education rate, it is proven to be true that a high education rate (rate above 30%) tends to vote for the Democratic party, and a low education rate (rate below 30%) tends to vote for the Republican party. Data analytics does not necessarily show the answers to the research questions or hypotheses, but it did reveal some other useful information that is related to the presidential election outcome.

Based on the findings listed above, CA has an average education rate of 32% and TX has an average education rate of 28%. This means that CA tends to vote for the Democratic party, and TX tends to vote for the Republican party. CA should remain as a blue state as its education will not decrease unless ancillary factors are included in this research. TX should change to a blue state at some point in the future as its education rate is increasing by 2% per year.

Conclusion

Since the evidence was not strong enough to support the alternative hypothesis, it was rejected, and a null hypothesis was accepted (failed to reject) in this research. However, some other information was revealed during the research.

Research has shown that states with a high education rate (college degree) of around 31-60% have the tendency to vote for the Democratic party, and states with a low education rate (college degree) of around 20–29% have the tendency to vote for the Republican party. These variables have shown moderate correlation in the research. The research has also shown that states with 29%-31% tend to swing between both political parties. As more and more states have a high education rate, more and more states will eventually arrive at a 30% education rate or above. Therefore, the entire research concludes that the Democratic party is likely to win the upcoming election in 2024 based on the projection, which uses the 2022 education data.

There are ancillary environmental factors that could affect the research, such as party candidates' popularity, scandals, or domestic or foreign policies. This factor was

not considered in the research so as to reduce the complexity and the time spent on research. There is a possibility that some ancillary factors might change the variables that we have tested in the research. Some might even be more important than the ones that were under consideration. The ancillary factors could potentially increase or reduce the significance of the variables in the dataset, making it more or less influential.

Since the election result is actually 50 elections happening simultaneously across 50 states, the winner can be determined by adding all the electoral votes based on the projected 2024 education rate.

Tier 3 States	% OF POPULATION Education	% OF POPULATION Education	% OF POPULATION Education	% OF POPULATION Education	Average of 4 years	2020 State election outcome	State Tier	Change in education rate 2012 - 2020
Row Labels	2012	2016	2020	2022				
ARIZONA	27.5	28.65	28	30	28.5375	D	3	0.5
D.C.	53.7	57.29	57	60	56.9975	D	3	3.3
HAWAII	30.1	33.04	32	33	32.035	D	3	1.9
MASSACHUSETTS	39.5	42.79	42	44	42.0725	D	3	2.5
MINNESOTA	32.7	34.45	35	36	34.5375	D	3	2.3
RHODE ISLAND	31.2	34.43	33	34	33.1575	D	3	1.8
Total	214.7	230.65	227	237	227.3375		Total	12.3
Average of the listed states	35.7833333	38.4416667	37.8333333	39.5	37.8895833			

Based on 2020 values	
Mean	37.8333333
Median	34
Mode	#N/A
Variance	91.1388889
Standard Deviation	10.4578519
Minimum value	28
Maximum value	57
Bin Width	4.14285714
Level of Significance	0.1
Sample size	6
Confidence value	7.02253835
Upper confidence	44.8558717
Lower confidence	30.810795

Table 1: Tier 3 states that dedicated its votes to Democratic party

Tier 3 States	% OF POPULATION Education	% OF POPULATION Education	% OF POPULATION Education	% OF POPULATION Education	Average of 4 years	2020		Change in education rate 2012 - 2020
Row Labels	2012	2016	2020	2022		State election outcome	State Tier	
ALABAMA	23.4	24.72	25	26	24.78	R	3	1.6
ALASKA	28.9	29.3	29	30	29.3	R	3	0.1
IDAHO	25.8	27.29	27	28	27.0225	R	3	1.2
INDIANA	23.3	25.72	26	27	25.505	R	3	2.7
KANSAS	30.2	33.02	32	33	32.055	R	3	1.8
MISSISSIPPI	20.6	21.9	21	22	21.375	R	3	0.4
MONTANA	28.5	31.32	31	32	30.705	R	3	2.5
NEBRASKA	29.7	32.34	31	32	31.26	R	3	1.3
NORTH CAROLINA	27.3	30.47	30	31	29.6925	R	3	2.7
NORTH DAKOTA	28.8	29.16	29	30	29.24	R	3	0.2
OKLAHOMA	23.7	25.54	25	26	25.06	R	3	1.3
SOUTH CAROLINA	25.3	27.5	27	28	26.95	R	3	1.7
SOUTH DAKOTA	26.7	29.63	28	29	28.3325	R	3	1.3
UTAH	30.7	32.37	33	34	32.5175	R	3	2.3
WYOMING	25.4	26.53	27	27	26.4825	R	3	1.6
Total	398.3	426.81	421	435	420.2775		Total	22.7
Average of the listed states	26.5533333	28.454	28.0666667	29	28.0185			

Based on 2020 values	
Mean	28.0666667
Median	28
Mode	27
Variance	9.26222222
Standard Deviation	3.15020785
Minimum value	21
Maximum value	33
Bin Width	1.71428571
Level of Significance	0.1
Sample size	15
Confidence value	1.33789132
Upper confidence	29.404558
Lower confidence	26.7287753

Table 2: Tier 3 states that dedicated its votes to Republican party

Visualization of Table 1 and Table 2 Below

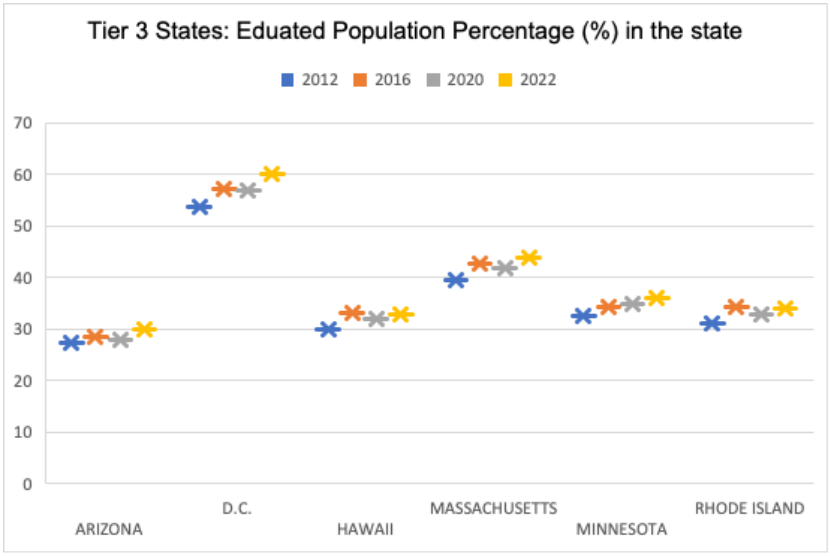

Chart 1: Tier 3 states that dedicated its votes to Democratic party

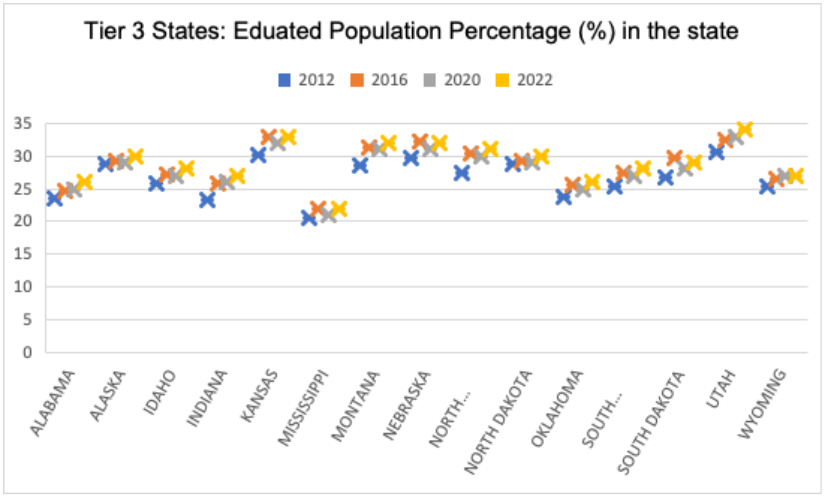

Chart 2: Tier 3 states that dedicated its votes to Democratic party

Tier 1 States	% OF POPULATION Education	% OF POPULATION Education	% OF POPULATION Education	% OF POPULATION Education	Average of 4 years	2020		Change in education rate 2012 - 2020	Based on 2020 values	
Row Labels	2012	2016	2020	2022		State election outcome	State Tier		Mean	30
CALIFORNIA	31	32.97	33	34	32.7425	D	1	2	Median	30
GEORGIA	28.3	30.57	30	31	29.9675	D	1	1.7	Mode	30
MICHIGAN	26.3	28.19	28	29	27.8725	D	1	1.7	Variance	2.8
PENNSYLVANIA	28.1	30.84	30	31	29.985	D	1	1.9	Standard Deviation	1.870828693
WISCONSIN	27.5	29.83	29	30	29.0825	D	1	1.5	Minimum value	28
									Maximum value	33
Total	141.2	152.4	150	155	149.65			8.8	Bin Width	0.714285714
Average of the listed states	28.24	30.48	30	31	29.93				Level of Significance	0.1
									Sample size	5
									Confidence value	1.376183279
									Upper confidence	31.37618328
									Lower confidence	28.62381672

Table 3: Tier 1 states swing but voted for Democratic party in 2020

Tier 1 States	% OF POPULATION Education	% OF POPULATION Education	% OF POPULATION Education	% OF POPULATION Education	Average of 4 years	2020		Change in education rate 2012 - 2020	Based on 2020 values	
Row Labels	2012	2016	2020	2022		State election outcome	State Tier		Mean	28.25
FLORIDA	26.8	28.66	29	30	28.615	R	1	2.2	Median	28.5
IOWA	26.8	28.95	28	29	28.1875	R	1	1.2	Mode	29
OHIO	25.2	27.76	27	28	26.99	R	1	1.8	Variance	0.6875
TEXAS	26.9	28.92	29	30	28.705	R	1	2.1	Standard Deviation	0.957427108
									Minimum value	27
Total	105.7	114.29	113	117	112.4975			7.3	Maximum value	113
Average of the listed states	26.425	28.5725	28.25	29.25	28.12438				Bin Width	12.28571429
									Level of Significance	0.1
									Sample size	6
									Confidence value	0.642920615
									Upper confidence	28.89292061
									Lower confidence	27.60707939

Table 4: Tier 1 states swing but voted for Republican party in 2020

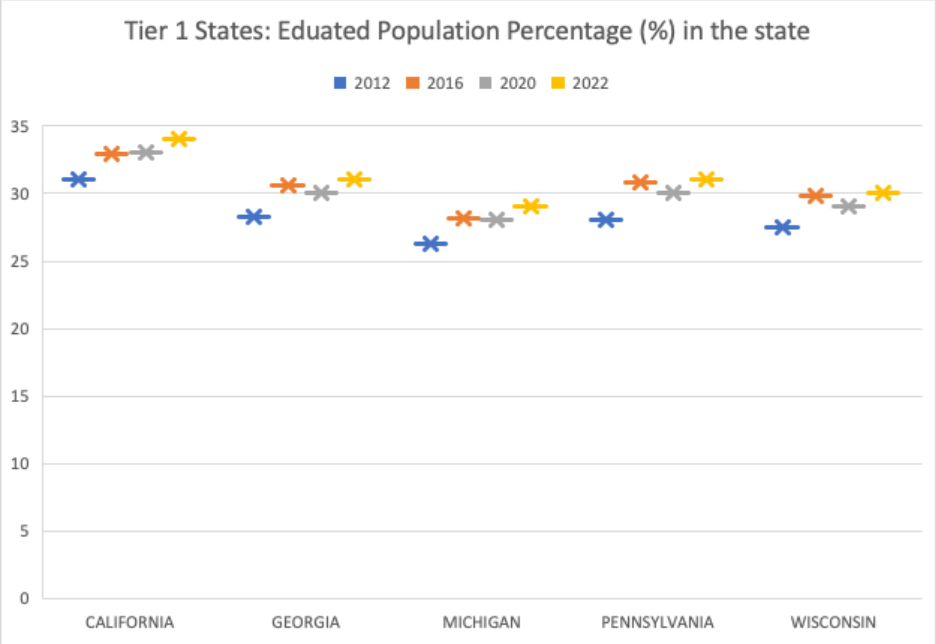

Chart 3: Tier 1 states swing but voted for Democratic party in 2020

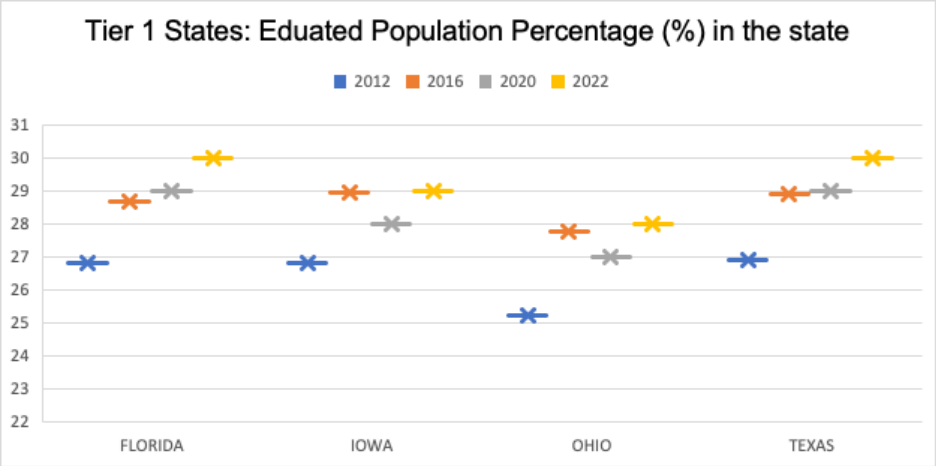

Chart 4: Tier 1 states swing but voted for Republican party in 2020

Regression analysis on Selected States
(2 states are selected to represent the Tiers that they are in; they selected based on having the closest value to its Tier's Median, and Mean)

Figure 5