

The background features a stylized American flag with red and white stripes on the left and blue with white stars on the right. The stripes are curved and flow across the top and bottom of the page. The stars are arranged in a pattern that follows the curve of the stripes.

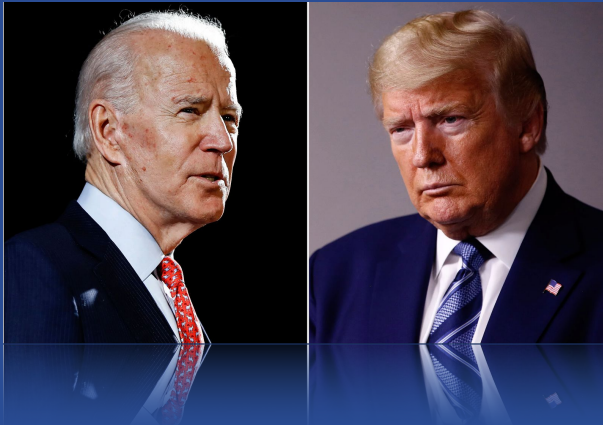
UNITED STATES PRESIDENTIAL ELECTION PREDICTION MODEL & SWING STATE BEHAVIOR STUDY

Saloni Patel & Mason Chen

Advisors: Charles Chen and Patrick Giuliano

SITUATION

- Historic 2020 Election in unprecedented times



OBJECTIVES

- Identify key swing states
- Create a prediction model based on the influence of COVID-19 and the economy
- Validate the prediction model with actual election results

CREATING THE PREDICTION MODEL

- Identify Swing State's using Swing State Index

$$(2016\% - 2012\%) * (2020\% - 2016\%) * 10,000$$

- Derive 2016-2012 Composite Win Margin for swing states

$$[(2016\% * 4) + (2012\%)] / 5$$

- Twice the weight for 2016 due to time domain
- Another twice the weight for 2016 for Trump (2016 President) over Biden (2012 Vice President)
- 2016/2012 Ratio = 4:1

State	2020 Votes	2016 Result	2012 Result	2016-2012 Composite
Oregon	7	-10.98%	-12.08%	-11.20%
New Mexico	5	-8.22%	-10.15%	-8.61%
Colorado	9	-4.91%	-5.36%	-5.00%
Virginia	13	-5.32%	-3.88%	-5.03%
Neveda	6	-2.42%	-6.68%	-3.27%
Minnesota	10	-1.52%	-7.69%	-2.75%
Michigan	16	0.23%	-9.50%	-1.72%
New Hampshire	4	-0.37%	-5.58%	-1.41%
Wisconsin	10	0.77%	-6.94%	-0.77%
Pennsylvania	20	0.72%	-5.38%	-0.50%
Florida	29	1.20%	-0.88%	0.78%
North Carolina	15	3.66%	2.04%	3.34%
Arizona	11	3.50%	9.03%	4.61%
Ohio	18	8.13%	-2.98%	5.91%
Iowa	6	9.41%	-5.81%	6.37%
Georgia	16	5.09%	7.82%	5.64%

CREATING THE PREDICTION MODEL

- Identify factors unique to 2020
 - COVID-19 and economy
- Apply Z-Standardize Transformation to avoid sampling mean and variance bias:
 - Infected cases
 - Deaths
 - Unemployment increase
- Derive Z-COVID Index

$(Z\text{-Infected}) + (Z\text{-Deaths}) + (Z\text{-Unemployment}) / 3$

- Calculate 2020 Predicted Result

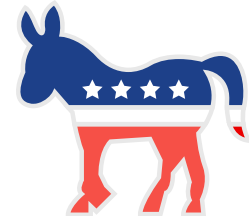
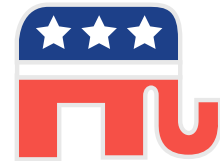
$(2016\text{-}2012 \text{ Composite Win Margin}) - [(Z\text{-COVID}) * 0.0518]$

States	2020 Predicted Result
<u>Iowa</u>	5.9%
<u>Ohio</u>	7.6%
<u>Georgia</u>	-0.2%
<u>North Carolina</u>	7.0%
<u>Arizona</u>	8.4%
<u>Florida</u>	3.4%
<u>Wisconsin</u>	2.6%
<u>Pennsylvania</u>	-8.2%
<u>Michigan</u>	-10.7%
<u>New Hampshire</u>	-1.7%
<u>Minnesota</u>	1.6%
<u>Nevada</u>	-6.1%
<u>Colorado</u>	-4.7%
<u>Virginia</u>	-3.5%
<u>New Mexico</u>	-3.9%

VALIDATING THE PREDICTION MODEL

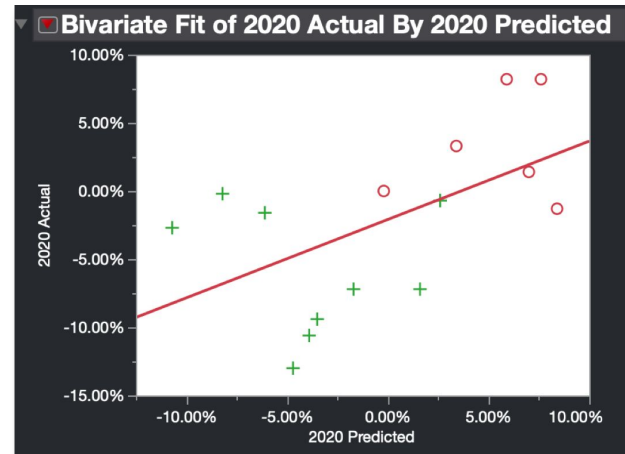
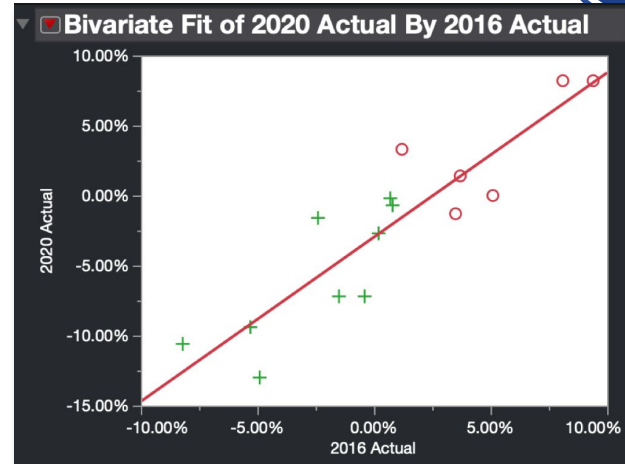
- Validate whether the factors COVID-19 and unemployment were able to predict election result
- Media (before the election) had predicted Trump will lose 3-5% of votes from 2016
- Compare the predicted results with the 2020 actual using regression test and paired t-test
- Conduct a 1-proportion hypothesis test to test predicted results accuracy

States	Electoral Votes	2012 (Actual)	2016 (Actual)	2020 (Actual)	0%	-0.5%	-1%	-2%	-3%	-4%
Iowa	6	-5.8%	9.4%	8.2%	5.9%	5.4%	4.9%	3.9%	2.9%	1.9%
Ohio	18	-3.0%	8.1%	8.2%	7.6%	7.1%	6.6%	5.6%	4.6%	3.6%
Georgia	16	7.8%	5.1%	-0.3%	-0.2%	-0.7%	-1.2%	-2.2%	-3.2%	-4.2%
North Carolina	15	2.0%	3.7%	1.3%	7.0%	6.5%	6%	5%	4%	3%
Arizona	11	9.0%	3.5%	-0.3%	8.4%	7.9%	7.4%	6.4%	5.4%	4.4%
Florida	29	-0.9%	1.2%	3.3%	3.4%	2.9%	2.4%	1.4%	0.4%	-0.6%
Wisconsin	10	-6.9%	0.8%	-0.7%	2.6%	2.1%	1.6%	0.6%	-0.4%	-1.4%
Pennsylvania	20	-5.4%	0.7%	-1.0%	-8.2%	-8.7%	-9.2%	-10.2%	-11.2%	-12.2%
Michigan	16	-9.5%	0.2%	-2.7%	-10.7%	-11.2%	-11.7%	-12.7%	-13.7%	-14.7%
New Hampshire	4	-5.6%	-0.4%	-7.2%	-1.7%	-2.2%	-2.7%	-3.7%	-4.7%	-5.7%
Minnesota	10	-7.7%	-1.5%	-7.2%	1.6%	1.1%	0.6%	-0.4%	-1.4%	-2.4%
Nevada	6	-6.7%	-2.4%	-1.6%	-6.1%	-6.6%	-7.1%	-8.1%	-9.1%	-10.1%
Colorado	9	-5.4%	-4.9%	-13.0%	-4.7%	-5.2%	-5.7%	-6.7%	-7.7%	-8.7%
Virginia	13	-3.9%	-5.3%	-9.4%	-3.5%	-4.0%	-4.5%	-5.5%	-6.5%	-7.5%
New Mexico	5	-10.2%	-8.2%	-10.6%	-3.9%	-4.4%	-4.9%	-5.9%	-6.9%	-7.9%



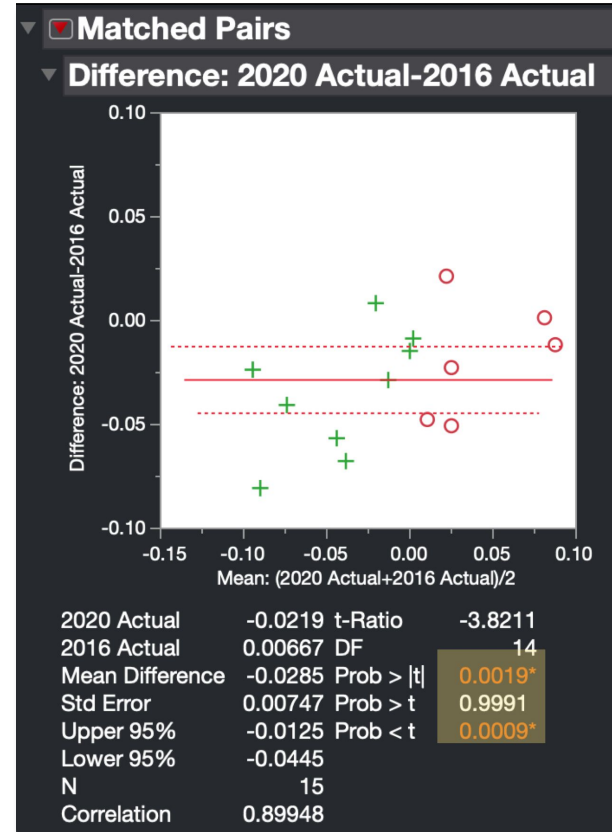
REGRESSION TEST

- 2012 actual vs. 2020 actual
- 2016 actual vs. 2020 actual (<.001)
 - R2: 0.8
 - Slope: 1.17
- 2020 predicted vs. 2020 actual (<.05)
 - R2: 0.3
 - Slope: 0.58
- 2016 election results are the most correlated with the 2020 election results, despite having declared a different winner
 - R2 higher, slope closer to 1



PAIRED T-TEST

- Why Paired T-Test?
 - Election results were compared within the same state at different times
- 2012 actual vs. 2020 actual
- 2016 actual vs. 2020 actual (<.001)
- 2020 predicted vs. 2020 actual
- 2012 actual and 2020 predicted are similar to the 2020 actual
 - Both results had declared same party as the winner
 - Shows that correlation between 2020 predicted and 2020 actual was not strong enough to find difference in paired t-test



1-PROPORTION HYPOTHESIS TEST

- Unlike regression test and paired t-test, this test tests the results based on the states, not the election result percent in each state
- 2020 predicted vs. 2020 actual
- Used to check whether model is 90% accurate
- Alt. Hypothesis - Model accuracy is more than 90% accurate
- Result: Failed to reject null hypothesis at 95% confidence

▼ **Choose Type of Alternative Hypothesis**

Population proportion is not equal to hypothesized proportion (two-tailed)

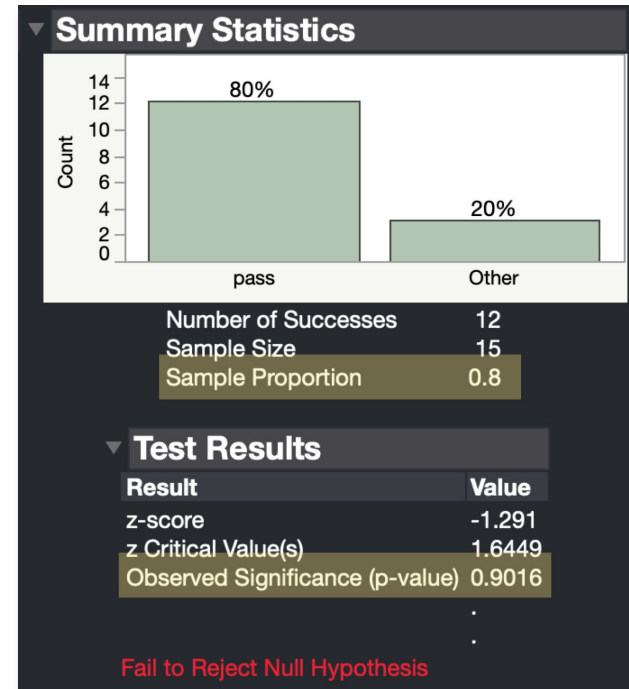
Population proportion is less than hypothesized proportion (one-tailed)

Population proportion is greater than hypothesized proportion (one-tailed)

▼ **Test Inputs**

Hypothesized Proportion

Significance Level (alpha)



CONCLUSIONS

- Regression Test showed significant relationship between 2016 actual vs. 2020 actual and 2020 predicted vs. 2020 actual
- Paired T-Test showed significant differences between 2016 actual vs. 2020 actual
- 1-Proportion Hypothesis Test failed to reject null hypothesis
- 3 states were not predicted correctly: Arizona, Wisconsin, and Minnesota
- COVID-19 and Unemployment rates ratio may not have applied for all states

	2012 actual	2016 actual	2020 predicted
Regression Test	not significant	<u>significant</u>	<u>significant</u>
Paired T-Test	not significant	<u>significant</u>	not significant
1-Proportion Test	N/A	N/A	not significant

QUESTIONS

- Which events and factors influenced these swing states to vote the way they did?
- How much more or less did voters care about COVID-19 than the economy?
- Can we use statistical tools to link political events with voting patterns?

GOALS

- Study 15 key swing states' voting patterns by linking statistical clustering methods to political events
- Adjust the Z-Ratio with new ratios to better understand the importance of COVID-19 and the economy in voting behavior

SWING STATE INDEX

- Swing State Index

$$(2016\% - 2012\%) - (2020\% - 2016\%) * 10,000$$

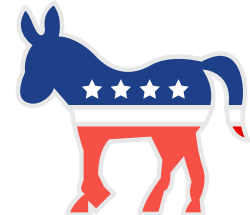
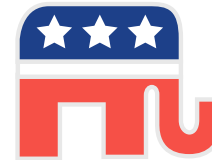
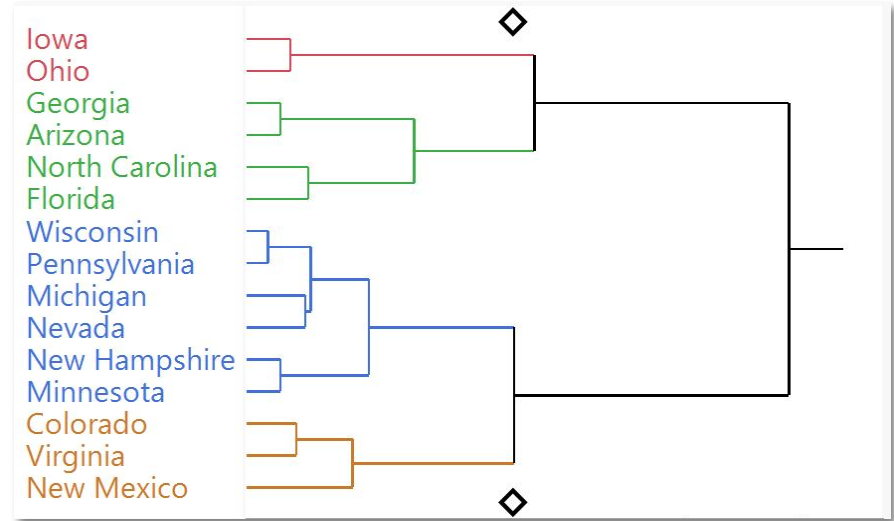
- Larger the magnitude and more negative the swing index = more “swing”

States	Electoral Votes	2012 (Actual)	2016 (Actual)	2020 (Actual)	Swing Index
Iowa	6	-5.8%	9.4%	8.2%	-1.1
Ohio	18	-3.0%	8.1%	8.2%	0.14
Georgia	16	7.8%	5.1%	-0.3%	2.31
North Carolina	15	2.0%	3.7%	1.3%	-0.55
Arizona	11	9.0%	3.5%	-0.3%	2.92
Florida	29	-0.9%	1.2%	3.3%	1.27
Wisconsin	10	-6.9%	0.8%	-0.7%	-1.13
Pennsylvania	20	-5.4%	0.7%	-1.0%	-1.12
Michigan	16	-9.5%	0.2%	-2.7%	-4.56
New Hampshire	4	-5.6%	-0.4%	-7.2%	-1.42
Minnesota	10	-7.7%	-1.5%	-7.2%	-3.50
Nevada	6	-6.7%	-2.4%	-1.6%	0.21
Colorado	9	-5.4%	-4.9%	-13.0%	-0.33
Virginia	13	-3.9%	-5.3%	-9.4%	0.76
New Mexico	5	-10.2%	-8.2%	-10.6%	-0.23



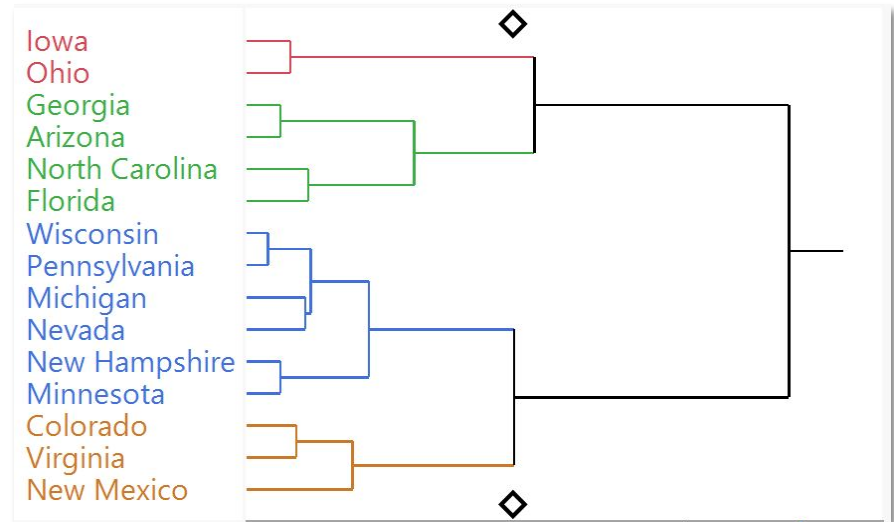
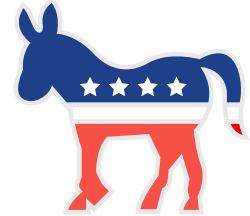
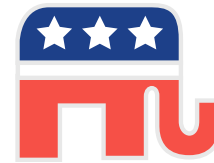
HIERARCHICAL CLUSTERING

- Swing State Index reveals voting patterns from past 3 elections, it does not describe the key events
- Hierarchical clustering grouped the swing states into 4 clusters based on their Swing State Index result
- “Bottom-up” approach where every state has its own cluster and then is merged with other states, moving it up the hierarchy
 - Iowa and Ohio in red in same cluster



FOUR CLUSTERS

- **First cluster:** Iowa & Ohio
 - Voted “blue” in 2012, “red” in 2016 and 2020
- **Second Cluster:** Georgia, Arizona, North Carolina, & Florida
 - Becoming “bluer” or “redder”, except North Carolina
- **Third Cluster:** Wisconsin, Pennsylvania, Michigan, Nevada, New Hampshire, & Minnesota
 - Most inconsistent swing states (all with negative swing index, except Nevada)
- **Fourth Cluster:** Colorado, Virginia, & New Mexico
 - Relatively “blue”



CLUSTERING JOIN HISTORY

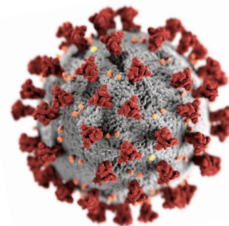
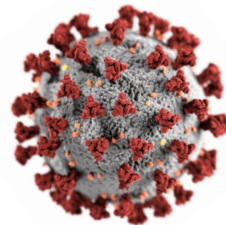
- Identifies top pairs of states (which two states most “similar”)
- First two pairs Wisconsin/Pennsylvania and Georgia/Arizona
 - Part of clusters that had states that switched from “red” to “blue” in the 2020 election
 - Wisconsin/Pennsylvania - economic, health, and environmental reasons
 - Georgia/Arizona - issues regarding civil rights and racial movements
- Study these pairs to understand which events in 2020 may have most affected their clusters’ voting patterns

Number of Clusters	Distance	Leader	Joiner
14	0.194112187	Wisconsin	Pennsylvania
13	0.302497298	Georgia	Arizona
12	0.305005698	New Hampshire	Minnesota
11	0.394441634	Iowa	Ohio
10	0.446849363	Colorado	Virginia
9	0.527686265	Michigan	Nevada
8	0.553445115	North Carolina	Florida
7	0.577730743	Wisconsin	Michigan
6	0.951416439	Colorado	New Mexico
5	1.096840607	Wisconsin	New Hampshire
4	1.505588474	Georgia	North Carolina



Z-RATIO

- Previously, we had attempted to predict the 2020 election result using a Z-Ratio
- Z-Ratio represents two key topics voters cast their ballot on: COVID-19 and the economy
- However, each state was assumed to have the same Z-Ratio
 - Twice the weight was given to COVID-19
 - Resulting ratio was 2:1 for all
- Each state has a unique situation



ADJUSTING Z-RATIO

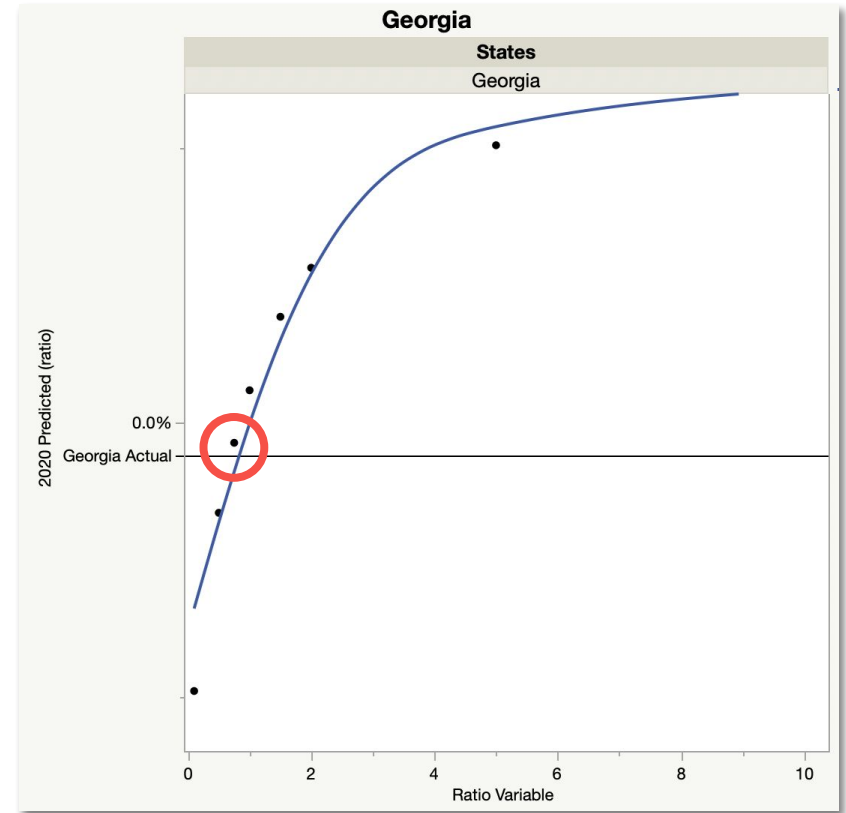
- Ratio Variable will determine the ratio for the importance of the Z-COVID and Z-Unemployment
- Once adjusted, Z-Ratio is implemented into the adjusted 2020 predicted result
- The adjusted 2020 predicted result is compared against the 2020 actual result to determine which ratio best explains the state's situation

	2020 Result	States	2016-2012 AVG	Z-Infected	Z-Death	Z-Unemployment	Z-COVID	Z-RATIO	Ratio Variable 2	2020 Predicted (ratio)	2012 Result	2016 Result
1	8.2%	Iowa	6.36%	0.47	1.27	-0.41	0.87	0.23	1	5.2%	-5.8%	9.4%
2	8.2%	Ohio	5.88%	-0.45	0.80	-0.45	0.17	-0.14	1	6.6%	-3.0%	8.1%
3	-0.3%	Georgia	5.64%	0.56	0.21	1.68	0.38	1.03	1	0.3%	7.8%	5.1%
4	1.3%	North Carolina	3.36%	-1.00	-0.85	-0.47	-0.93	-0.70	1	7.0%	2.0%	3.7%
5	-0.3%	Arizona	4.60%	-0.97	-0.68	-0.74	-0.82	-0.78	1	8.6%	9.0%	3.5%
6	3.3%	Florida	0.78%	-0.43	-0.44	-0.36	-0.43	-0.40	1	2.8%	-0.9%	1.2%
7	-0.7%	Wisconsin	-0.74%	-0.80	-1.09	-0.54	-0.94	-0.74	1	3.1%	-6.9%	0.8%
8	-1.0%	Pennsylvania	-0.52%	1.86	-0.68	1.36	0.59	0.98	1	-5.6%	-5.4%	0.7%
9	-2.7%	Michigan	-1.74%	2.24	1.21	1.50	1.72	1.61	1	-10.1%	-9.5%	0.2%
10	-7.2%	New Hampshire	-1.44%	-0.35	-1.32	0.60	-0.84	-0.12	1	-0.8%	-5.6%	-0.4%
11	-7.2%	Minnesota	-2.74%	-1.04	1.39	-0.65	0.17	-0.24	1	-1.5%	-7.7%	-1.5%
12	-1.6%	Nevada	-3.26%	-0.35	-0.56	1.37	-0.45	0.46	1	-5.6%	-6.7%	-2.4%
13	-13.0%	Colorado	-5.00%	0.71	-0.79	-1.19	-0.04	-0.62	1	-1.8%	-5.4%	-4.9%
14	-9.4%	Virginia	-5.02%	-0.06	-0.09	-0.81	-0.07	-0.44	1	-2.7%	-3.9%	-5.3%
15	-10.6%	New Mexico	-8.60%	-0.37	1.62	-0.88	0.63	-0.13	1	-7.9%	-10.2%	-8.2%
16	8.2%	Iowa	6.36%	0.47	1.27	-0.41	0.87	-0.29	0.1	7.9%	-5.8%	9.4%
17	8.2%	Ohio	5.88%	-0.45	0.80	-0.45	0.17	-0.39	0.1	7.9%	-3.0%	8.1%
18	-0.3%	Georgia	5.64%	0.56	0.21	1.68	0.38	1.56	0.1	-2.5%	7.8%	5.1%
19	1.3%	North Carolina	3.36%	-1.00	-0.85	-0.47	-0.93	-0.51	0.1	6.0%	2.0%	3.7%
20	-0.3%	Arizona	4.60%	-0.97	-0.68	-0.74	-0.82	-0.75	0.1	8.5%	9.0%	3.5%
21	3.3%	Florida	0.78%	-0.43	-0.44	-0.36	-0.43	-0.37	0.1	2.7%	-0.9%	1.2%
22	-0.7%	Wisconsin	-0.74%	-0.80	-1.09	-0.54	-0.94	-0.58	0.1	2.2%	-6.9%	0.8%
23	-1.0%	Pennsylvania	-0.52%	1.86	-0.68	1.36	0.59	1.29	0.1	-7.2%	-5.4%	0.7%
24	-2.7%	Michigan	-1.74%	2.24	1.21	1.50	1.72	1.52	0.1	-9.6%	-9.5%	0.2%
25	-7.2%	New Hampshire	-1.44%	-0.35	-1.32	0.60	-0.84	0.47	0.1	-3.9%	-5.6%	-0.4%
26	-7.2%	Minnesota	-2.74%	-1.04	1.39	-0.65	0.17	-0.58	0.1	0.2%	-7.7%	-1.5%
27	-1.6%	Nevada	-3.26%	-0.35	-0.56	1.37	-0.45	1.20	0.1	-9.5%	-6.7%	-2.4%
28	-13.0%	Colorado	-5.00%	0.71	-0.79	-1.19	-0.04	-1.09	0.1	0.6%	-5.4%	-4.9%
29	-9.4%	Virginia	-5.02%	-0.06	-0.09	-0.81	-0.07	-0.74	0.1	-1.2%	-3.9%	-5.3%
30	-10.6%	New Mexico	-8.60%	-0.37	1.62	-0.88	0.63	-0.74	0.1	-4.8%	-10.2%	-8.2%
31	8.2%	Iowa	6.36%	0.47	1.27	-0.41	0.87	0.02	0.5	6.3%	-5.8%	9.4%
32	8.2%	Ohio	5.88%	-0.45	0.80	-0.45	0.17	-0.24	0.5	7.1%	-3.0%	8.1%
33	-0.3%	Georgia	5.64%	0.56	0.21	1.68	0.38	1.25	0.5	-0.8%	7.8%	5.1%
34	1.3%	North Carolina	3.36%	-1.00	-0.85	-0.47	-0.93	-0.62	0.5	6.6%	2.0%	3.7%
35	-0.3%	Arizona	4.60%	-0.97	-0.68	-0.74	-0.82	-0.77	0.5	8.6%	9.0%	3.5%
36	3.3%	Florida	0.78%	-0.43	-0.44	-0.36	-0.43	-0.38	0.5	2.8%	-0.9%	1.2%
37	-0.7%	Wisconsin	-0.74%	-0.80	-1.09	-0.54	-0.94	-0.67	0.5	2.8%	-6.9%	0.8%
38	-1.0%	Pennsylvania	-0.52%	1.86	-0.68	1.36	0.59	1.10	0.5	-6.2%	-5.4%	0.7%
39	-2.7%	Michigan	-1.74%	2.24	1.21	1.50	1.72	1.57	0.5	-9.9%	-9.5%	0.2%
40	-7.2%	New Hampshire	-1.44%	-0.35	-1.32	0.60	-0.84	0.12	0.5	-2.1%	-5.6%	-0.4%
41	-7.2%	Minnesota	-2.74%	-1.04	1.39	-0.65	0.17	-0.38	0.5	-0.8%	-7.7%	-1.5%
42	-1.6%	Nevada	-3.26%	-0.35	-0.56	1.37	-0.45	0.76	0.5	-7.2%	-6.7%	-2.4%
43	-13.0%	Colorado	-5.00%	0.71	-0.79	-1.19	-0.04	-0.81	0.5	-0.8%	-5.4%	-4.9%
44	-9.4%	Virginia	-5.02%	-0.06	-0.09	-0.81	-0.07	-0.56	0.5	-2.1%	-3.9%	-5.3%
45	-10.6%	New Mexico	-8.60%	-0.37	1.62	-0.88	0.63	-0.38	0.5	-6.6%	-10.2%	-8.2%

Column Name → Column Properties → Formula

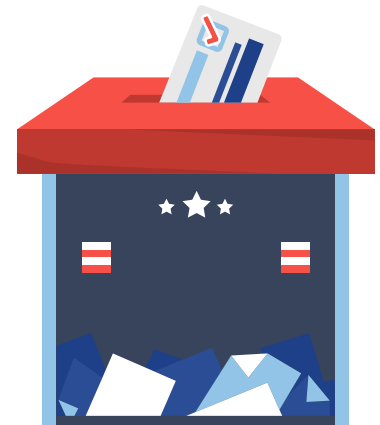
GEORGIA

- Georgia is a state that stood out in 2020
 - First state to reopen business in April
 - President Trump had an eye on Georgia's votes and attempted to overturn results several times
- Georgia's actual result: **-0.3%**
- Ratio of **3:4** was closest: **-0.2%**
- Indicates that economy was more important



CONCLUSIONS

- Using Hierarchical Clustering
 - In second cluster, Arizona and Georgia were mostly affected by issues regarding civil rights
 - In third cluster, Pennsylvania and Wisconsin voted for Biden due to economic, health, and environmental reasons.
 - Worsening COVID-19 situation, racial movements like Black Lives Matter, and increasing dissatisfaction towards Trump's policies (mostly healthcare-related) caused the switch from red to blue in a large number of swing states
- Adjusting Z-Ratio
 - In Georgia, 3:4 matched best suggesting economy was a more important to voters compared to COVID-19
 - Makes sense because Georgia was the first state to reopen business in April





THANK YOU!