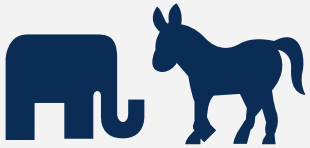
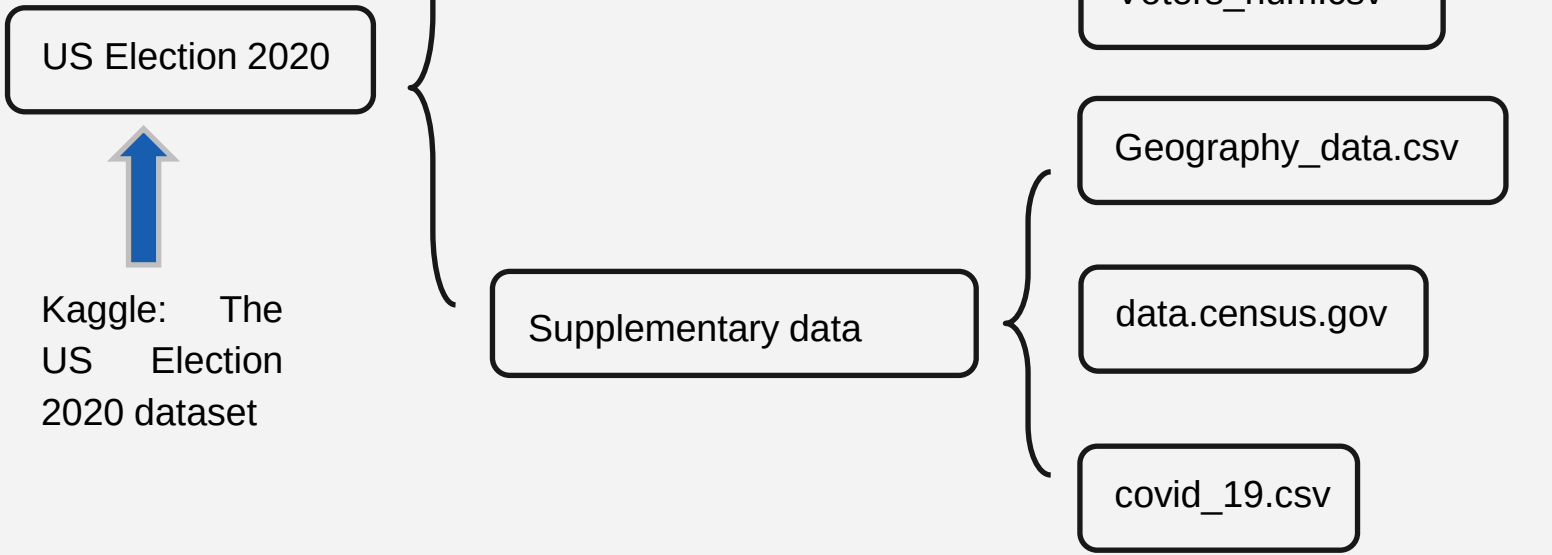


US Election 2020

Part one

Introduction



Kaggle: The US Election 2020 dataset

Background and Motivation

The election significantly influences both domestic and international policies

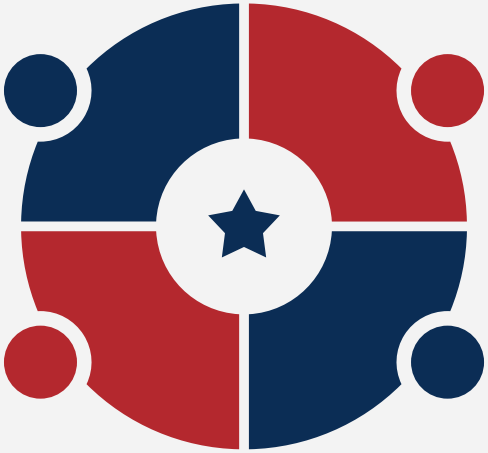
Many people are passionate about political analysis. We wold like to help them uncover key insights and correlations in the election.

serve as a valuable reference for the future election

US Election 2020

Expected Findings

Voting difference at state and county level

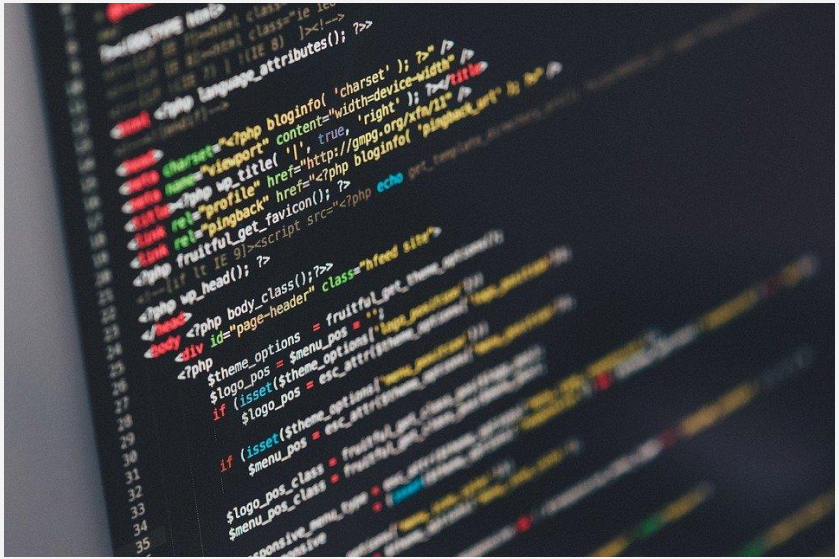


Voting difference at swing states and ‘dominant’ states

Correlation between COVID-19 and voting results

Why Trump lost the election in 2020

Data Cleaning



Remove data after the voting day (2020-11-03)

↓

Delete unnecessary columns

↓

Add additional data columns (e.g. geopraphy)

↓

Merge datasets and remove other candidates

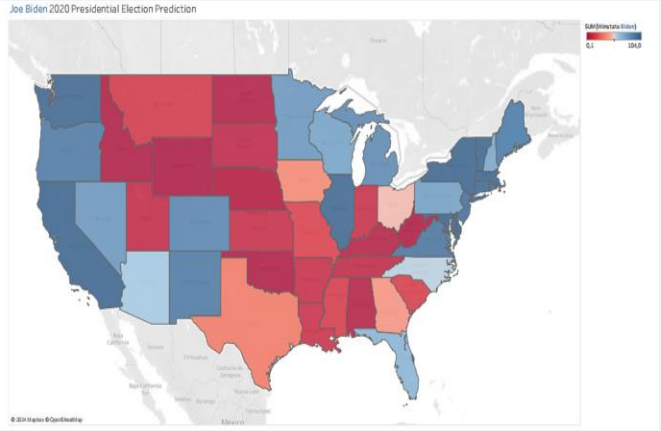
02

Part two

Explore Data Analytics （Steps of analysis and Main findings）

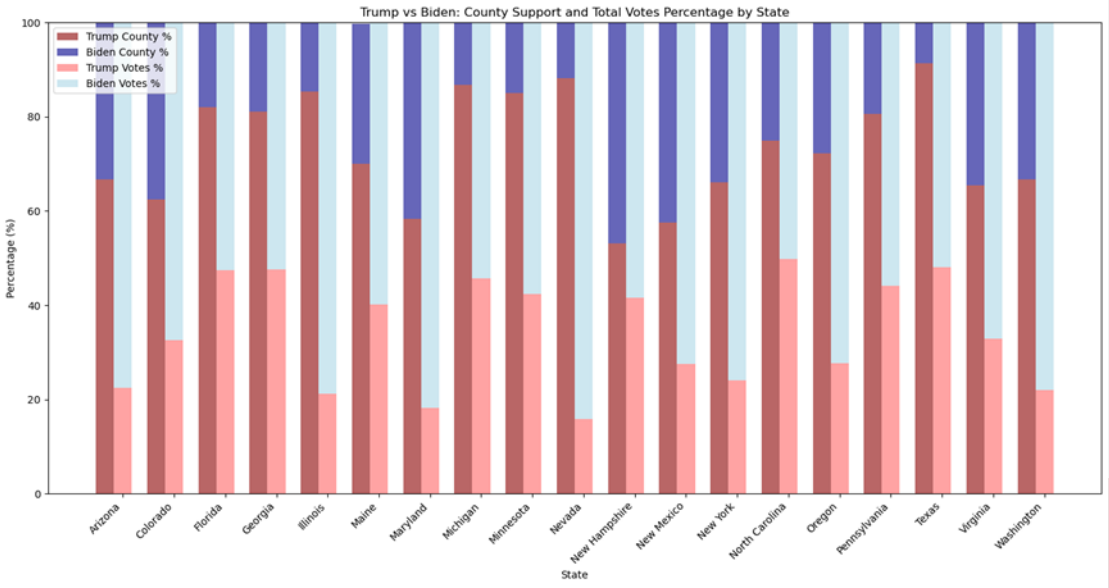
Voting Difference at State Level

- The Democratic Party maintained a strong advantage in Northeast, West Coast, Midwest Cities, and Certain Southern Cities.
- While the Republican Party Kept a powerful advantage in South, Great Plains, Mountain West, and Rural Areas Nationwide.



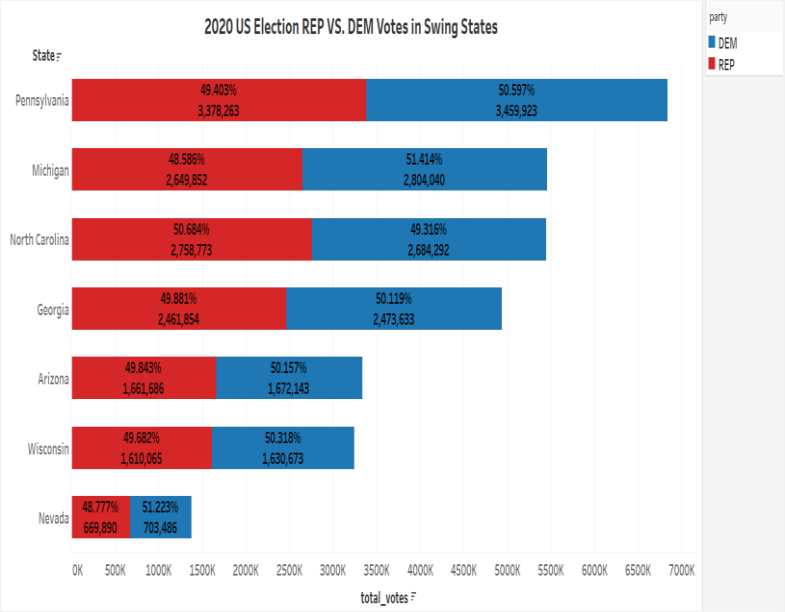
Voting Difference at County Level

There were 19 states where Donald Trump won more counties but lost the total vote to Joe Biden visualized by this stacked bar chart.



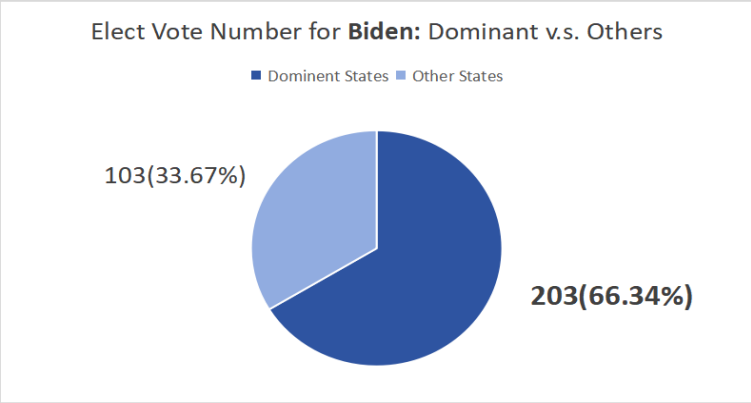
Voting Difference at Swing States

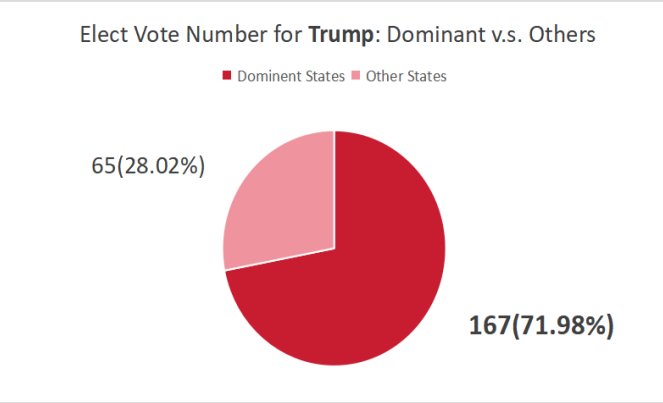
- The vote shares of both parties in the 7 swing states are all relatively small
- North Carolina tending to favour the Republicans, while Democrats win the other six states



Voting Difference at ‘dominat’ States

- Those states with more than 15% Diff-Total ratio were defined as dominant states.(non-swing states)

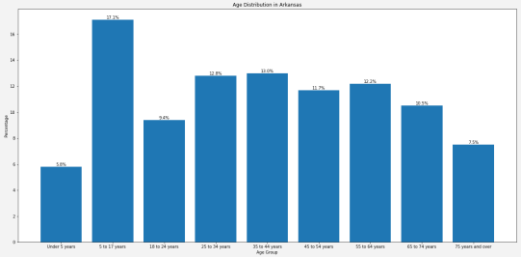


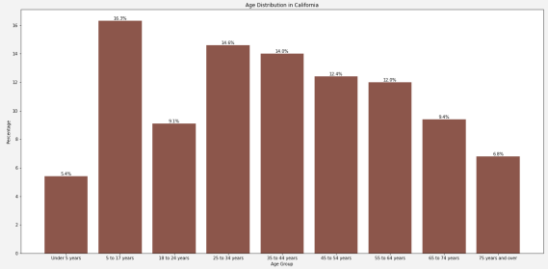


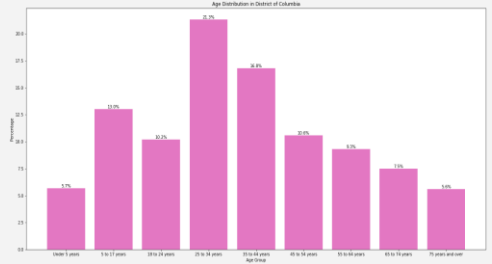
It was found that Trump seemed to have more faithful followers compared with Biden.

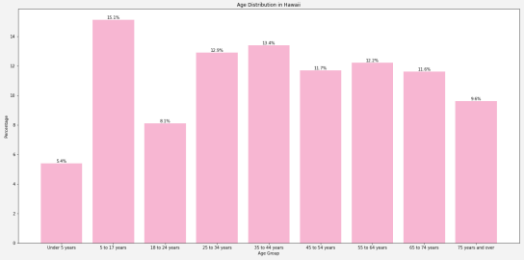
Looking into These ‘Dominant’ States

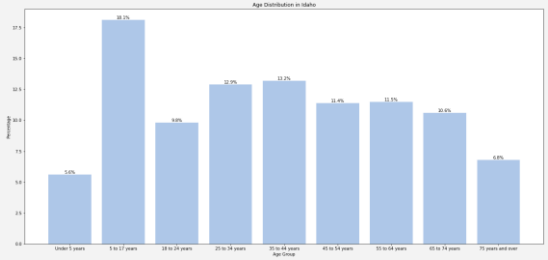
Age

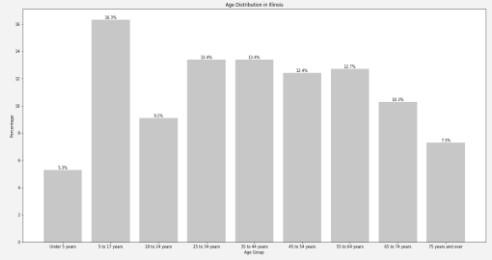






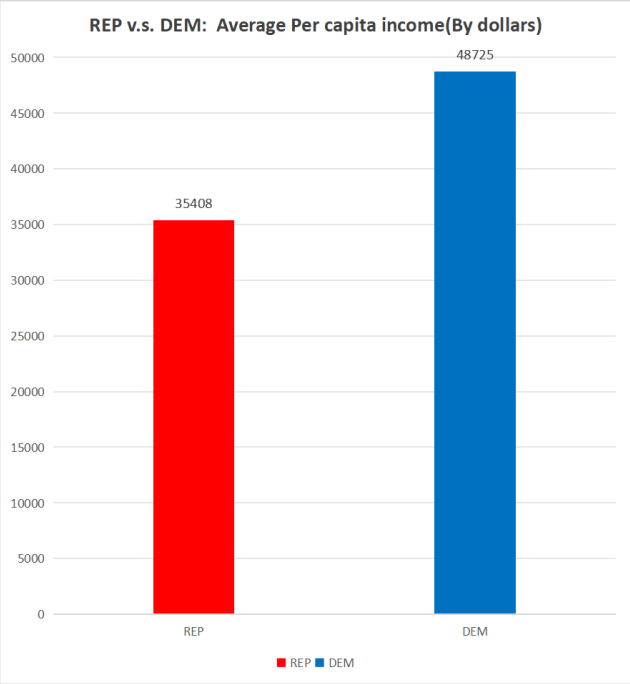






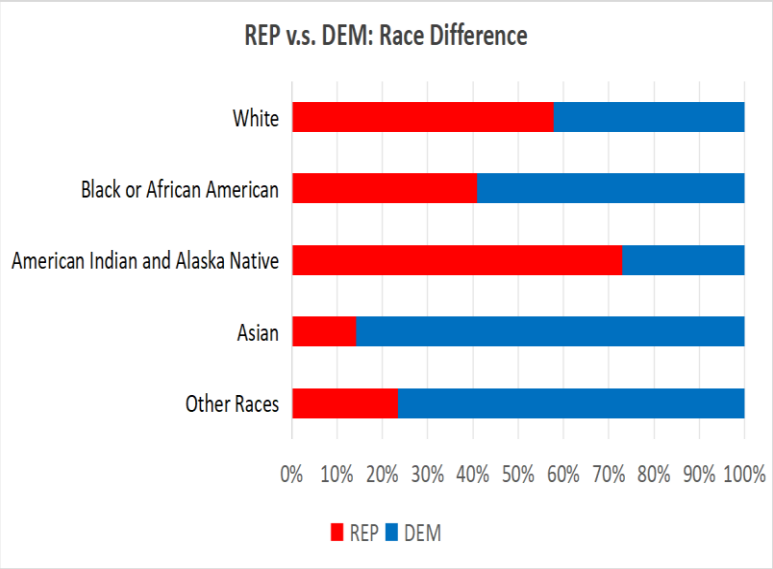
There were only slight differences in age distribution between different states.

Income



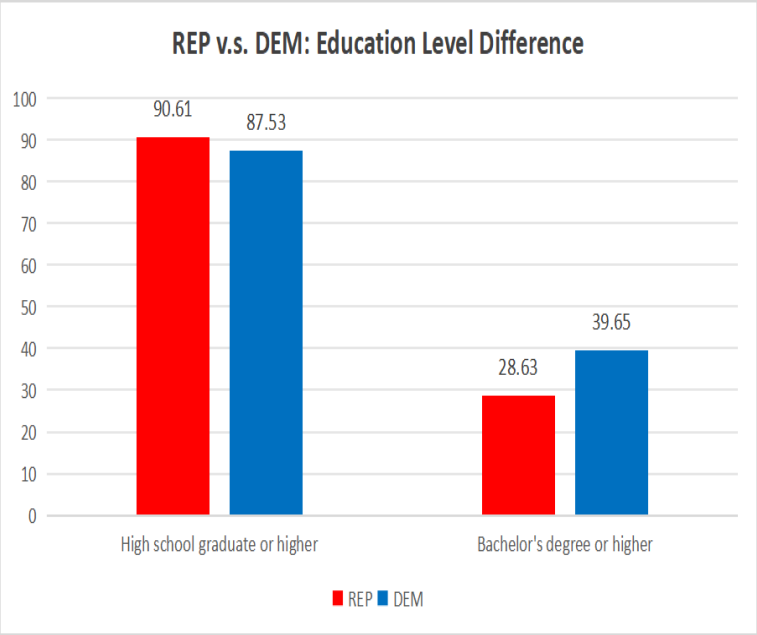
Red states had a lower per capita income compared to blue states.

Race



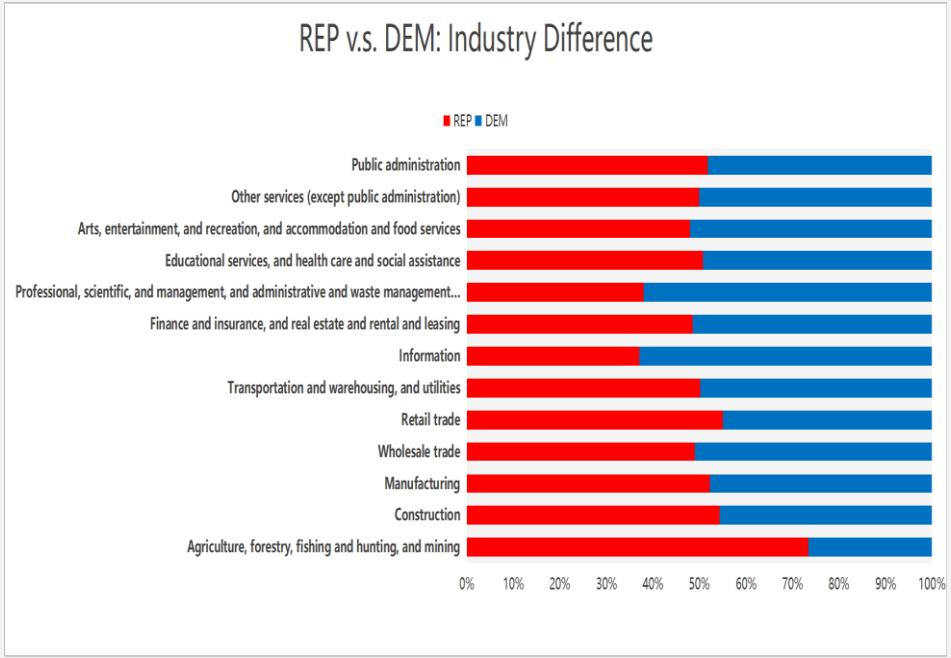
- The proportion of white people,as well as American indian and Alaska Native in red states were higher than in blue states, while Black or African American and Asian were higher in blue states, showing the race difference between red and blue states.

Education Level



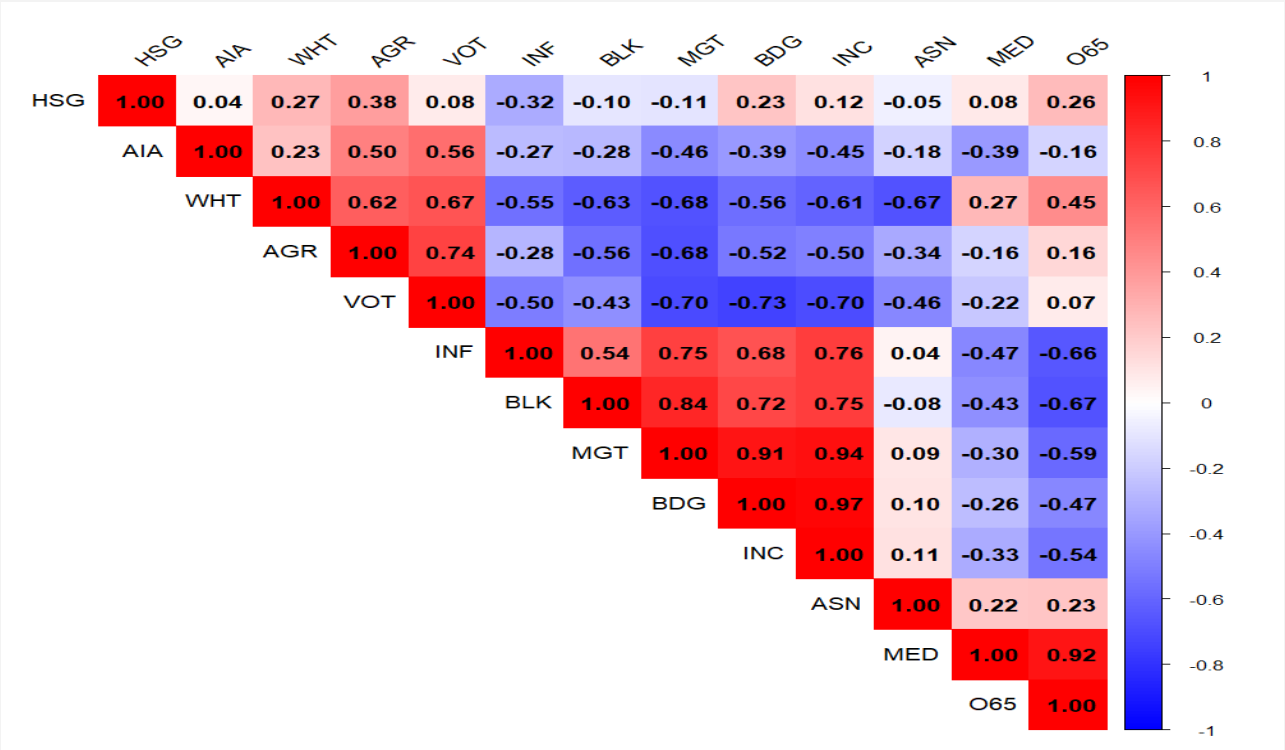
- Higher education level accounted for a higher proportion(39.65%) in blue states(28.63% in red states)

Industry Structure



- Blue states had more employers in Information and Professional, scientific, and management, and administrative and waste management services industry, while red states had more employment ratio in Agriculture, forestry, fishing and hunting, and mining ratio. Comparatively, differences in other industry were smaller.

Heatmap - voters



Heatmap - voters

Correlation	Parametres	Factors	Correlation Coefficient(With VOT)
Positively Correlated	WHT	Race	0.62
	AIA	Race	0.56
	AGR	Industry	0.62
Negatively Correlated	INC	Income	-0.7
	BDG	Education	-0.73
	ASN	Race	-0.46
	BLK	Race	-0.43
	MGT	Industry	-0.7
	INF	Industry	-0.5
Weakly Correlated	MED	Age	-0.22
	O65	Age	0.07
	HSG	Education	0.08

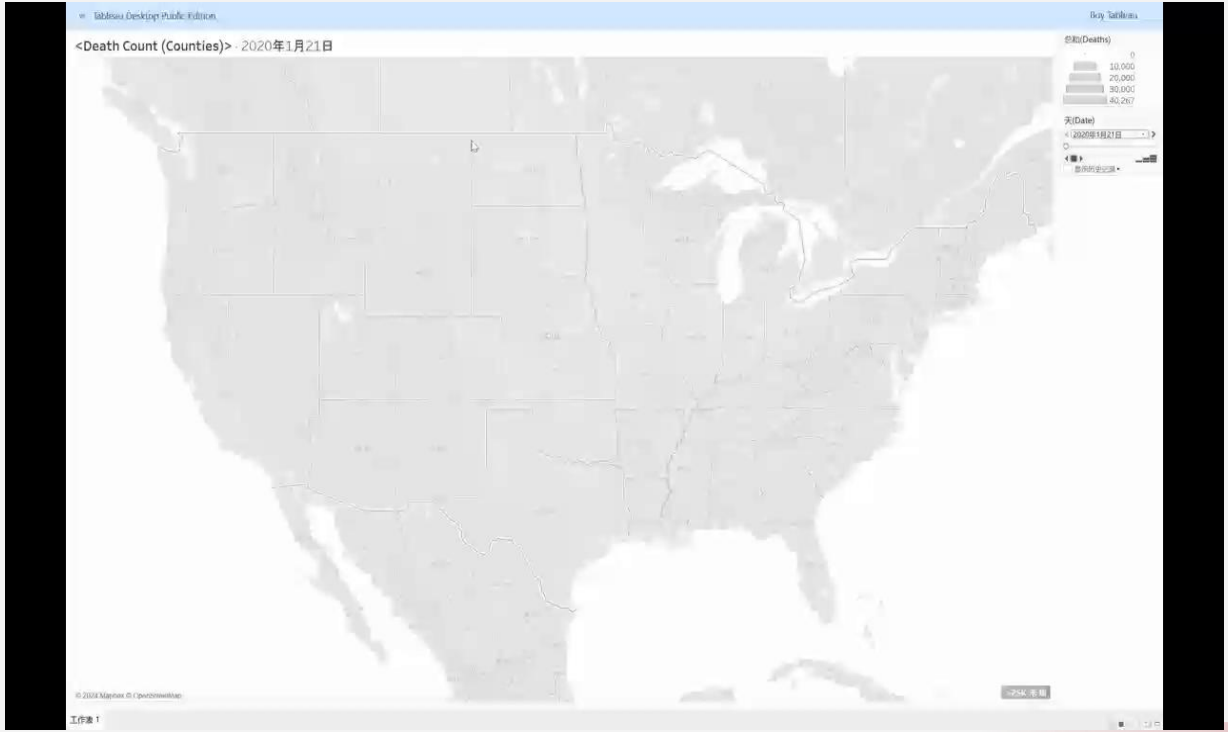
Positively Correlated: Race (White & American.Indian.and.Alaska.Native), Industry (Agriculture)

Negatively Correlated: Income, Education (Bachelor), Race (Asian & Black), Industry (Information & management)

Weekly Correlated: Age, Education(High school)

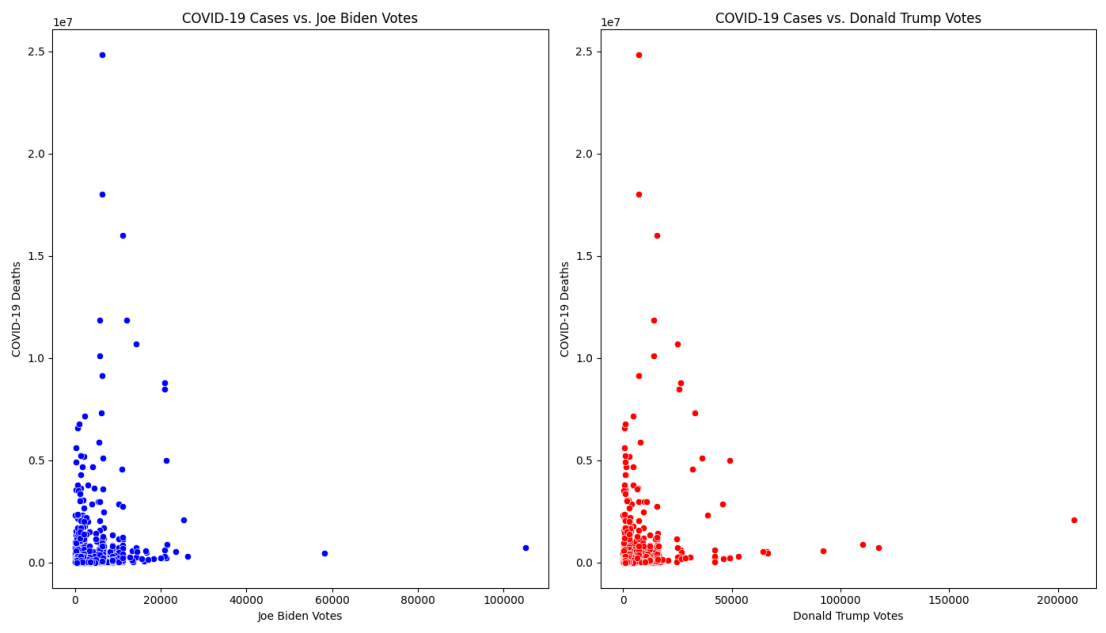
Did Covid-19 Influence Voters' Behavior

COVID-19 Pandemic



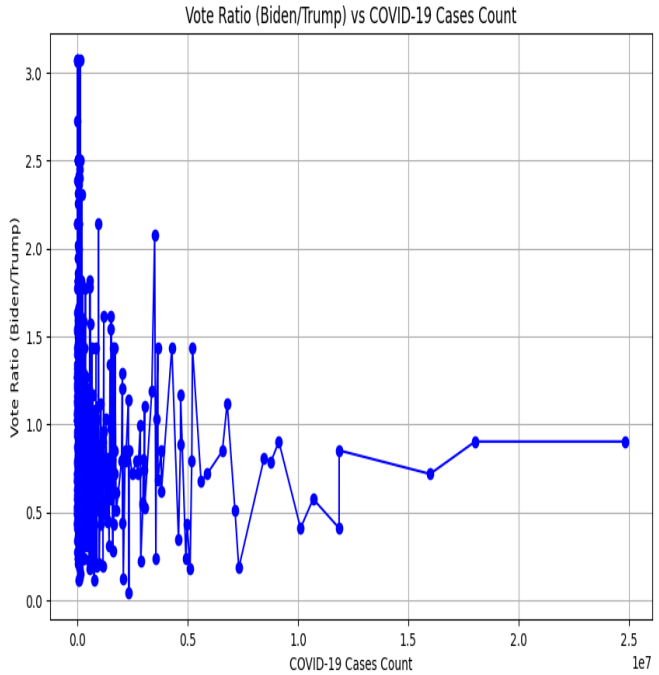
Distribution of covid19 deaths across the US

COVID-19 Pandemic



- In the left plot, the X-axis showed Biden's vote count in each county, and the Y-axis shows COVID-19 cases.
- In the right plot, the X-axis showed Trump's vote count in each county.

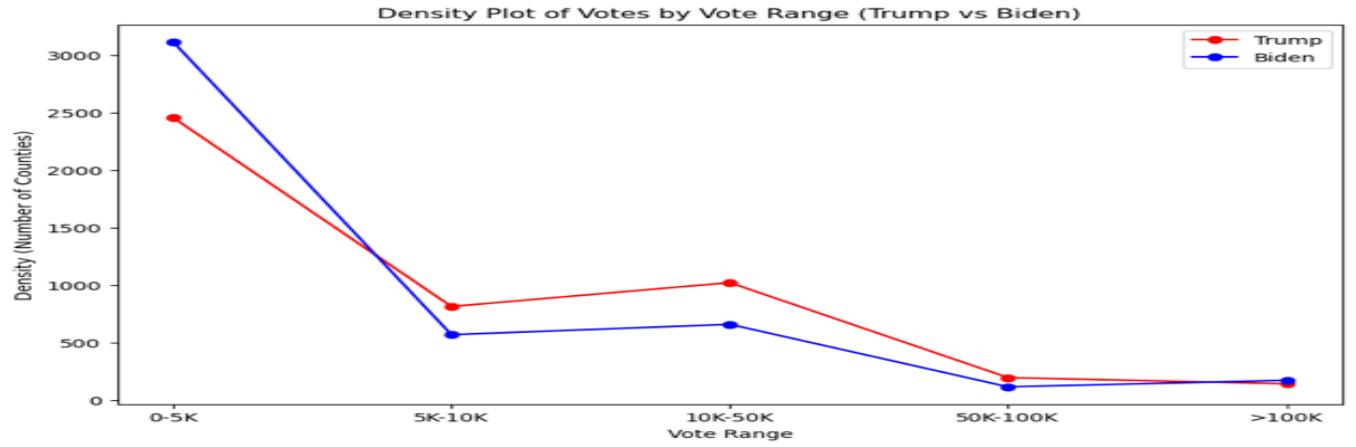
COVID-19 Pandemic



- The Y-axis showed the Biden-to-Trump vote ratio, where values greater than 1 meant Biden received more votes.
- The X-axis showed the number of COVID-19 cases (in millions). Most counties had relatively low COVID-19 cases, clustering to the left.
- The distribution suggested a negative trend: counties with more COVID-19 cases generally had a lower Biden/Trump vote ratio.

Why Trump lost the election in 2020

Population

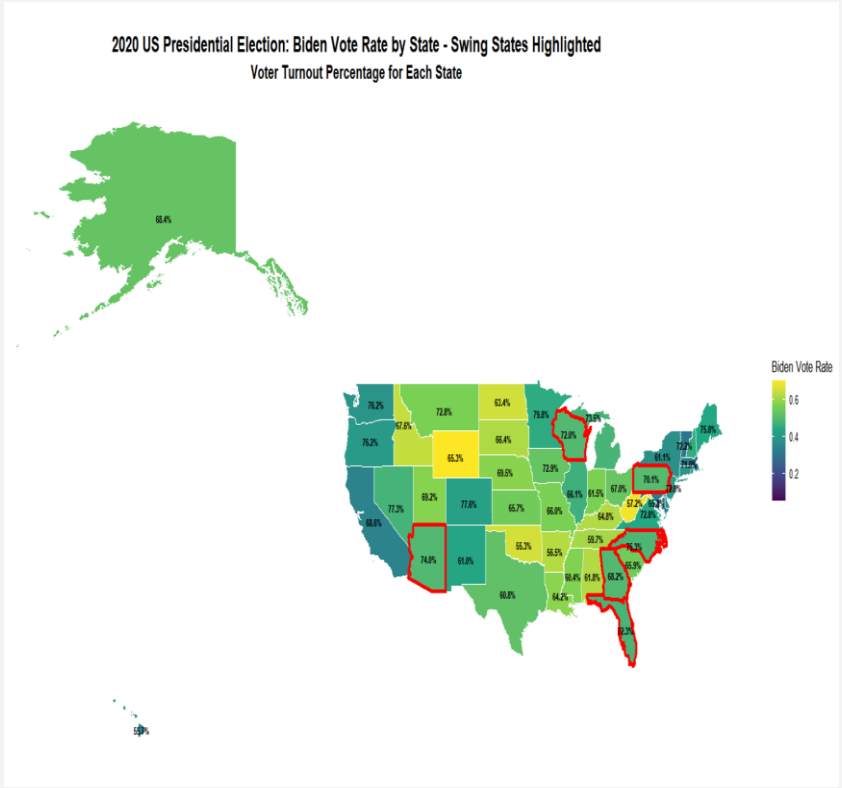


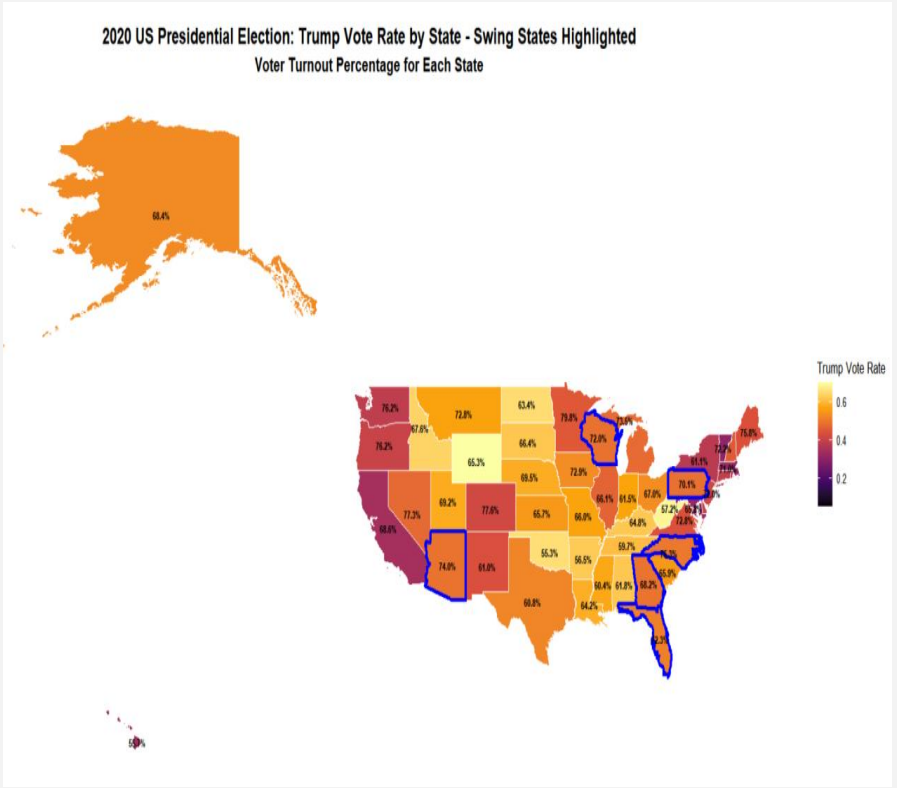
There existed a tendency that Biden received more votes in counties with more voters, especially in the range over 100K.

Turnout Rate

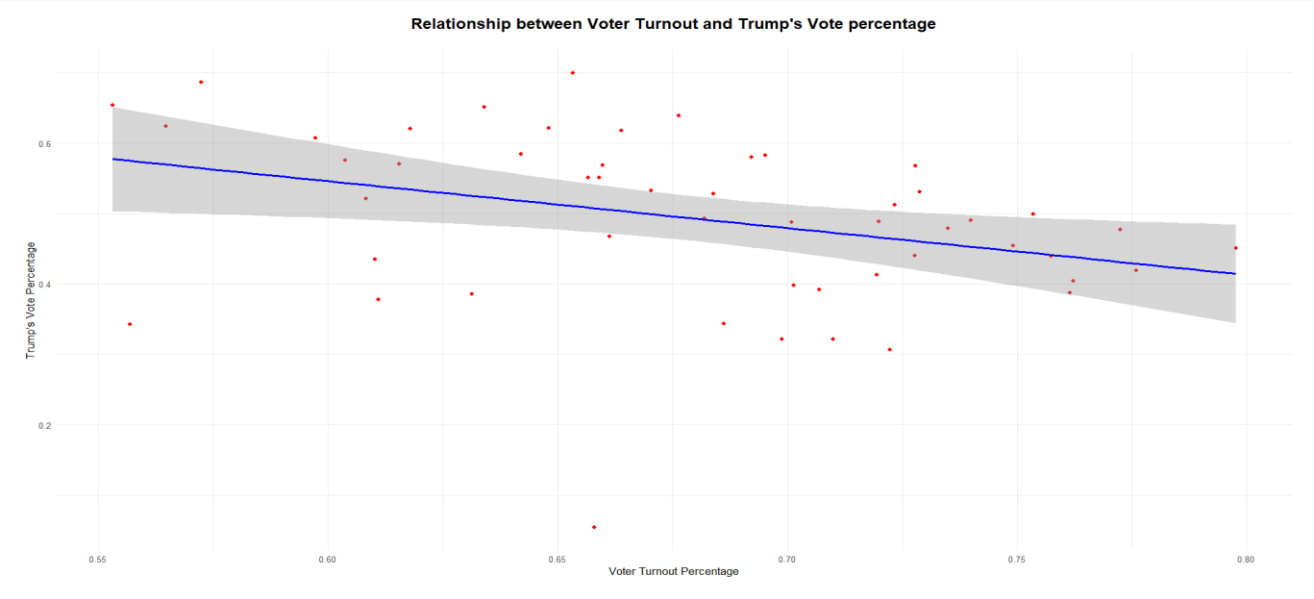
- **Turnout Rate:** Refers to the number of voters who actually vote in an election as a proportion of the total number of voters eligible to vote. It is an important indicator of voter participation and is usually expressed as a percentage.

	STATE_NAME	Voter_turnout
2	Alabama	61.80%
3	Alaska	68.40%
4	Arizona	74%
5	Arkansas	56.5%
6	California	68.6%
7	Colorado	77.6%
8	Connecticut	70.7%
9	Delaware	70.1%
10	District of Colu	65.8%
11	Florida	72.3%
12	Georgia	68.2%
13	Hawaii	55.7%
14	Idaho	67.6%
15	Illinois	66.1%
16	Indiana	61.5%
17	Iowa	72.9%
18	Kansas	65.7%
19	Kentucky	64.8%
20	Louisiana	64.2%
21	Maine	75.8%
22	Maryland	69.9%
23	Massachusetts	71.0%
24	Michigan	73.50%
25	Minnesota	79.8%
26	Mississippi	60.4%
27	Missouri	66.0%
28	Montana	72.8%
29	Nebraska	69.5%


2020 US Presidential Election: Biden Vote Rate by State - Swing States Highlighted
Voter Turnout Percentage for Each State
Biden Vote Rate
0.6
0.4
0.2

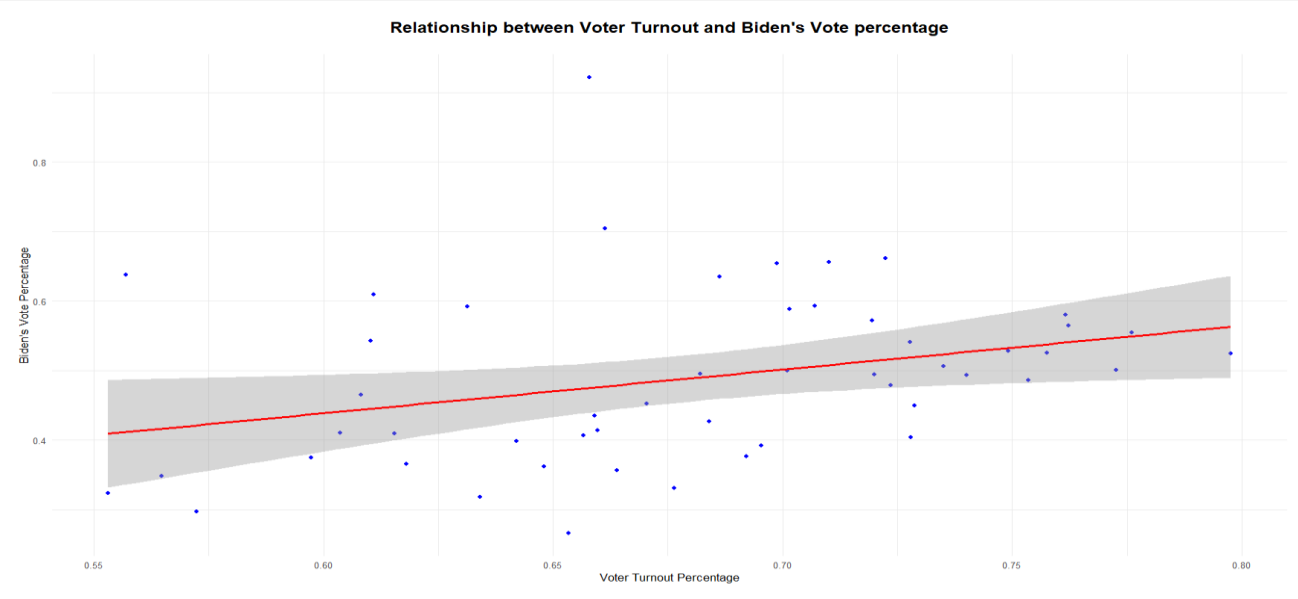

2020 US Presidential Election: Trump Vote Rate by State - Swing States Highlighted
Voter Turnout Percentage for Each State
Trump Vote Rate
0.6
0.4
0.2

Regression - Turnout Rate (Trump)



- X-axis: represented the percentage of voter turnout from approximately 0.55 (55%) to 0.80 (80%).
- Y-axis: represented the percentage of votes Biden received, from approximately 0.40 (40%) to 0.80 (80%).
- There was a negative correlation between turnout and Trump's vote share

Regression - Turnout Rate (Biden)



- There was a positive correlation between turnout and Biden's vote share
- This may reflect the link between voter participation and candidate-specific support, as well as Biden's strength in mobilizing voters to turn out.

03

Part three

Summary and Conclusion

Future Directions

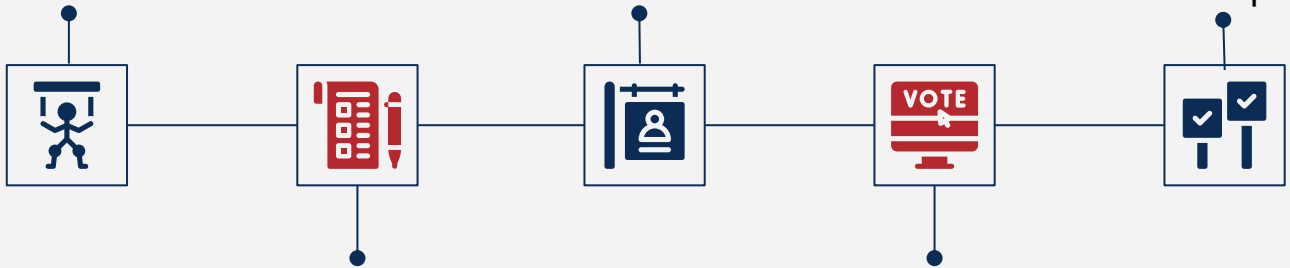
Summary and Conclusion

Found some highlights

- **Race and industry structure** were the most significant factors influencing voter behavior.
- Higher **voter turnout** could mean more voter participation, which could have a positive impact on Biden's vote percentage.
- Higher **Covid-19** cases may influence voter preference.

Presidential election data
Supplementary data

tableau dynamic graph, stacked bar chart, pie chart, scatter plot, heat map



python

tableau

Serve as a valuable reference for the future election

Future Directions

- We found in our studies that higher Covid-19 cases may influence voter preferences, but further study is needed to find out the mechanism (**mail-in ballots**).
- We found in our studies that High voter turnout appeared to be positively associated with Biden's high vote share and negatively associated with Trump's vote share. Further study is needed to find out the mechanism (**younger voters and minority communities**).
- In the future, we can compare the study with the **2024 election** to find out the difference in Trump’s tactic.

Thanks!