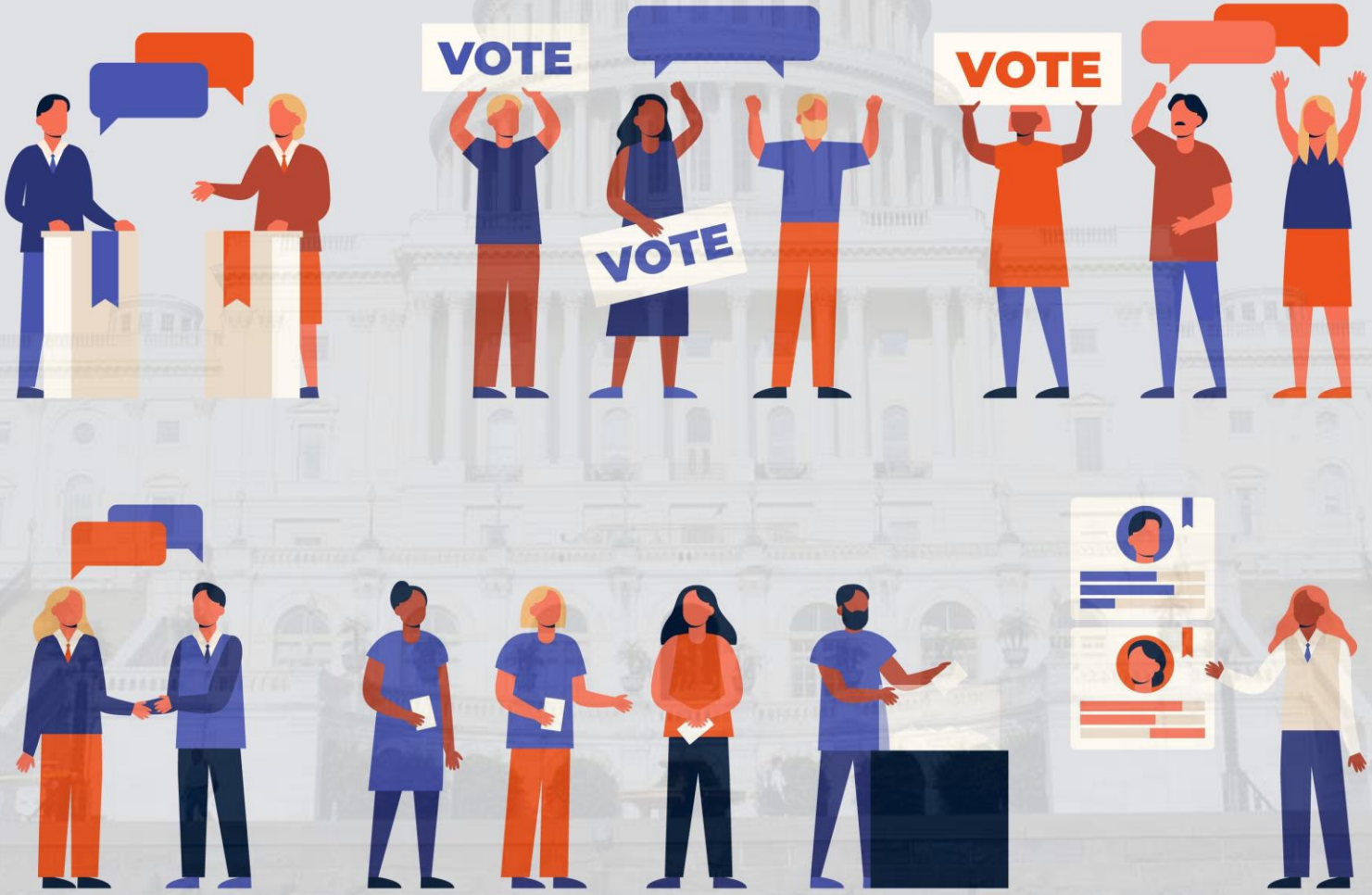
US Presidential Elections 2000 – 2020

A Data analysis of USA states votes in the 21th century



Breno Batista Caetano
IBM capstone project
2022

Outline

- Executive Summary
- Introduction
- Methodology
- Results:
  - Visualization – Charts and Maps
  - Dashboard
- Conclusion

Executive Summary

Methodologies

- **Data collection and wrangling:**
  - python, folium, dash, SQL
- **Data exploration:**
  - Python, SQL
- **Data visualization:**
  - Python-folium and plotly, dash, SQL-Power Bi
- **Machine learning:**
  - sckit-learn

Main Objective of the analysis:

Find how USA states changed their voting habit in the 21 century.

Secondary objectives of the analysis:

1- Find which states in the USA tend to vote for the Democrats

2- Find which states in the USA tend to vote for the Republicans

3- Find if and which states populations of USA changed their voting in the last 20 years.

Introduction

Background and context:

- This Presentation has been created for both general public and strategical political analysts that want to understand the voting habits of States in the USA in the 21 century.
- With only 2 major parties, the USA voting system is very rigid. Most of the States in USA have already set in which party they will vote, leaving only a handful of States open to change.
- The aim of this analysis is to show which States vote Republic, which vote Democratic and which are susceptible to change. It also aims to find party trends in the States open to change.

Main Objective of the analysis:

Find how USA states changed their voting habit in the 21 century.

Secondary objectives of the analysis:

1- Find which states in the USA tend to vote for the Democrats

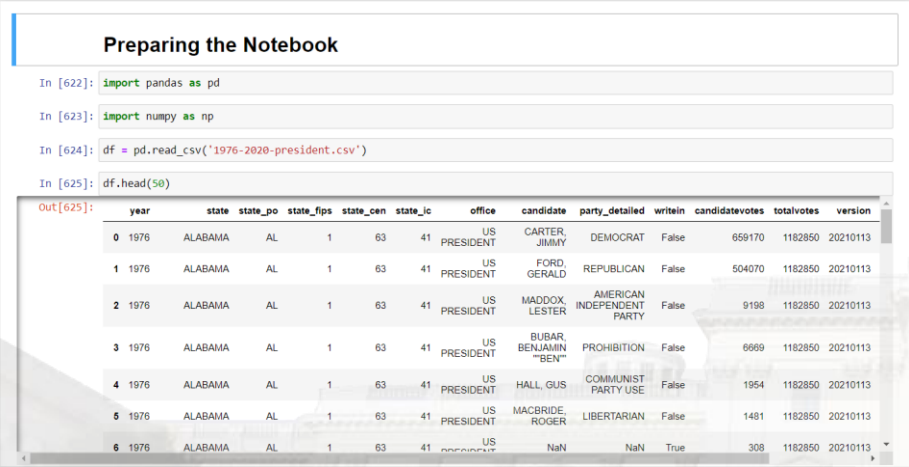
2- Find which states in the USA tend to vote for the Republicans

3- Find if and which states populations of USA changed their voting in the last 20 years.

Methodology

1- Data collection and wrangling

The data was collected via kaegle and was wrangled and cleaned using python:

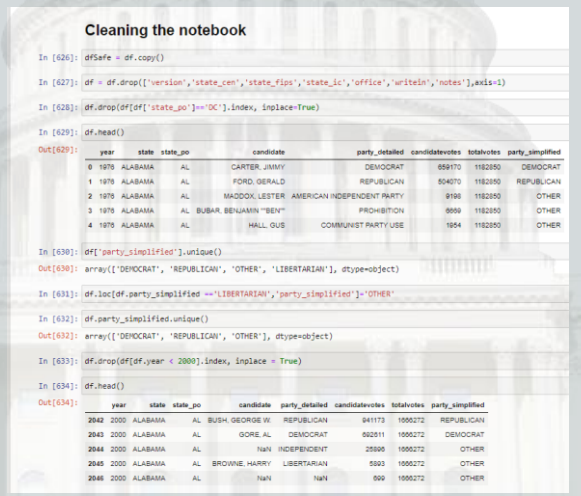


Preparing the Notebook

```
In [622]: import pandas as pd
In [623]: import numpy as np
In [624]: df = pd.read_csv('1976-2020-president.csv')
In [625]: df.head(50)
```

Out[625]:

	year	state	state_po	state_fips	state_cen	state_ic	office	candidate	party_detailed	writein	candidatevotes	totalvotes	version
0	1976	ALABAMA	AL	1	63	41	US PRESIDENT	CARTER, JIMMY	DEMOCRAT	False	659170	1182850	20210113
1	1976	ALABAMA	AL	1	63	41	US PRESIDENT	FORD, GERALD	REPUBLICAN	False	504070	1182850	20210113
2	1976	ALABAMA	AL	1	63	41	US PRESIDENT	MADDOX, LESTER	AMERICAN INDEPENDENT PARTY	False	9198	1182850	20210113
3	1976	ALABAMA	AL	1	63	41	US PRESIDENT	BUBAR, BENJAMIN ""BEN""	PROHIBITION	False	6669	1182850	20210113
4	1976	ALABAMA	AL	1	63	41	US PRESIDENT	HALL, GUS	COMMUNIST PARTY USE	False	1954	1182850	20210113
5	1976	ALABAMA	AL	1	63	41	US PRESIDENT	MACBRIDE, ROGER	LIBERTARIAN	False	1481	1182850	20210113
6	1976	ALABAMA	AL	1	63	41	US PRESIDENT	NaN	NaN	True	308	1182850	20210113



Cleaning the notebook

```
In [626]: dfSafe = df.copy()
In [627]: df = df.drop(['version','state_cen','state_fips','state_ic','office','writein','notes'],axis=1)
In [628]: df.drop(df[df['state_po']=='DC'].index, inplace=True)
In [629]: df.head()
```

Out[629]:

	year	state	state_po	candidate	party_detailed	candidatevotes	totalvotes	party_simplified
0	1976	ALABAMA	AL	CARTER, JIMMY	DEMOCRAT	659170	1182850	DEMOCRAT
1	1976	ALABAMA	AL	FORD, GERALD	REPUBLICAN	504070	1182850	REPUBLICAN
2	1976	ALABAMA	AL	MADDOX, LESTER	AMERICAN INDEPENDENT PARTY	9198	1182850	OTHER
3	1976	ALABAMA	AL	BUBAR, BENJAMIN ""BEN""	PROHIBITION	6669	1182850	OTHER
4	1976	ALABAMA	AL	HALL, GUS	COMMUNIST PARTY USE	1954	1182850	OTHER

```
In [630]: df['party_simplified'].unique()
```

Out[630]: array(['DEMOCRAT', 'REPUBLICAN', 'OTHER', 'LIBERTARIAN'], dtype=object)

```
In [631]: df.loc[df.party_simplified =='LIBERTARIAN','party_simplified']='OTHER'
In [632]: df.party_simplified.unique()
```

Out[632]: array(['DEMOCRAT', 'REPUBLICAN', 'OTHER'], dtype=object)

```
In [633]: df.drop(df[df.year < 2000].index, inplace = True)
In [634]: df.head()
```

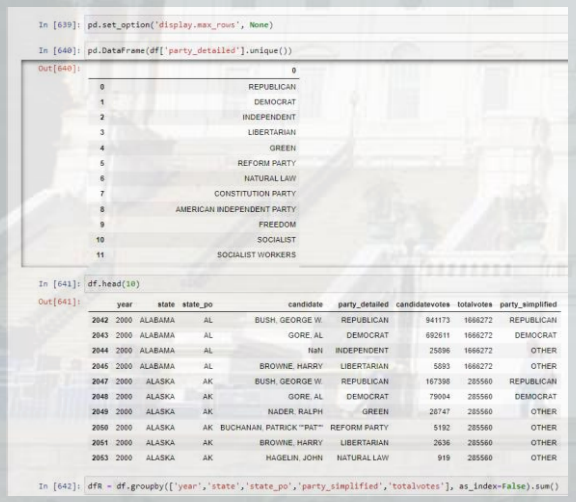
Out[634]:

	year	state	state_po	candidate	party_detailed	candidatevotes	totalvotes	party_simplified
2042	2000	ALABAMA	AL	BUSH, GEORGE W.	REPUBLICAN	941173	1666272	REPUBLICAN
2043	2000	ALABAMA	AL	GORE, AL	DEMOCRAT	692611	1666272	DEMOCRAT
2044	2000	ALABAMA	AL	NaN	INDEPENDENT	25896	1666272	OTHER
2045	2000	ALABAMA	AL	BROWNE, HARRY	LIBERTARIAN	5893	1666272	OTHER
2046	2000	ALABAMA	AL	NaN	NaN	699	1666272	OTHER

```
In [635]: df.tail()
```

Out[635]:

	year	state	state_po	candidate	party_detailed	candidatevotes	totalvotes	party_simplified
4282	2020	WYOMING	WY	JORGENSEN, JO	LIBERTARIAN	5768	278503	OTHER
4283	2020	WYOMING	WY	PIERCE, BROCK	INDEPENDENT	2208	278503	OTHER
4284	2020	WYOMING	WY	NaN	NaN	1739	278503	OTHER
4285	2020	WYOMING	WY	OVERVOTES	NaN	279	278503	OTHER
4286	2020	WYOMING	WY	UNDERVOTES	NaN	1459	278503	OTHER



```
In [639]: pd.set_option('display.max_rows', None)
In [640]: pd.DataFrame(df['party_detailed'].unique())
```

Out[640]:

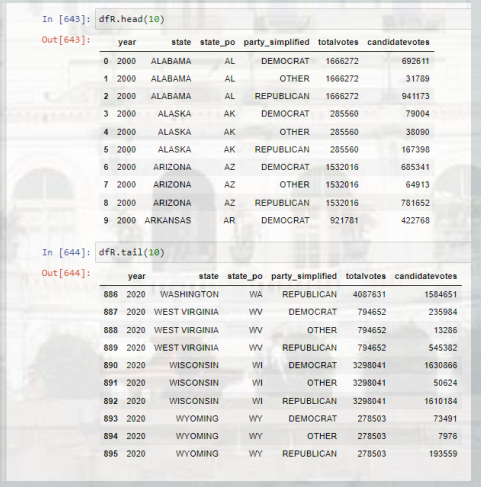
	0
0	REPUBLICAN
1	DEMOCRAT
2	INDEPENDENT
3	LIBERTARIAN
4	GREEN
5	REFORM PARTY
6	NATURAL LAW
7	CONSTITUTION PARTY
8	AMERICAN INDEPENDENT PARTY
9	FREEDOM
10	SOCIALIST
11	SOCIALIST WORKERS

```
In [641]: df.head(10)
```

Out[641]:

	year	state	state_po	candidate	party_detailed	candidatevotes	totalvotes	party_simplified
2042	2000	ALABAMA	AL	BUSH, GEORGE W.	REPUBLICAN	941173	1666272	REPUBLICAN
2043	2000	ALABAMA	AL	GORE, AL	DEMOCRAT	692611	1666272	DEMOCRAT
2044	2000	ALABAMA	AL	NaN	INDEPENDENT	25896	1666272	OTHER
2045	2000	ALABAMA	AL	BROWNE, HARRY	LIBERTARIAN	5893	1666272	OTHER
2047	2000	ALASKA	AK	BUSH, GEORGE W.	REPUBLICAN	167398	285560	REPUBLICAN
2048	2000	ALASKA	AK	GORE, AL	DEMOCRAT	79004	285560	DEMOCRAT
2049	2000	ALASKA	AK	NADER, RALPH	GREEN	28747	285560	OTHER
2050	2000	ALASKA	AK	BUCHANAN, PATRICK ""PAT""	REFORM PARTY	5192	285560	OTHER
2051	2000	ALASKA	AK	BROWNE, HARRY	LIBERTARIAN	2636	285560	OTHER
2053	2000	ALASKA	AK	HAGELIN, JOHN	NATURAL LAW	919	285560	OTHER

```
In [642]: dfR = df.groupby(['year','state','state_po','party_simplified','totalvotes'], as_index=False).sum()
```



```
In [643]: dfR.head(10)
```

Out[643]:

	year	state	state_po	party_simplified	totalvotes	candidatevotes
0	2000	ALABAMA	AL	DEMOCRAT	1666272	692611
1	2000	ALABAMA	AL	OTHER	1666272	31789
2	2000	ALABAMA	AL	REPUBLICAN	1666272	941173
3	2000	ALASKA	AK	DEMOCRAT	285560	79004
4	2000	ALASKA	AK	OTHER	285560	38090
5	2000	ALASKA	AK	REPUBLICAN	285560	167398
6	2000	ARIZONA	AZ	DEMOCRAT	1532016	685341
7	2000	ARIZONA	AZ	OTHER	1532016	64913
8	2000	ARIZONA	AZ	REPUBLICAN	1532016	781652
9	2000	ARKANSAS	AR	DEMOCRAT	921781	422768

```
In [644]: dfR.tail(10)
```

Out[644]:

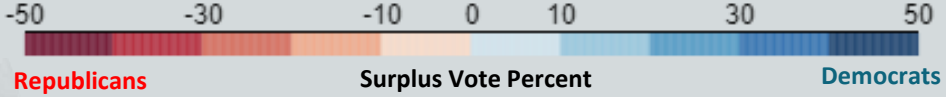
	year	state	state_po	party_simplified	totalvotes	candidatevotes
886	2020	WASHINGTON	WA	REPUBLICAN	4087631	1584651
887	2020	WEST VIRGINIA	WV	DEMOCRAT	794652	235984
888	2020	WEST VIRGINIA	WV	OTHER	794652	13286
889	2020	WEST VIRGINIA	WV	REPUBLICAN	794652	545382
890	2020	WISCONSIN	WI	DEMOCRAT	3298041	1630866
891	2020	WISCONSIN	WI	OTHER	3298041	50624
892	2020	WISCONSIN	WI	REPUBLICAN	3298041	1610184
893	2020	WYOMING	WY	DEMOCRAT	278503	73491
894	2020	WYOMING	WY	OTHER	278503	7976
895	2020	WYOMING	WY	REPUBLICAN	278503	193559

2- Visualization

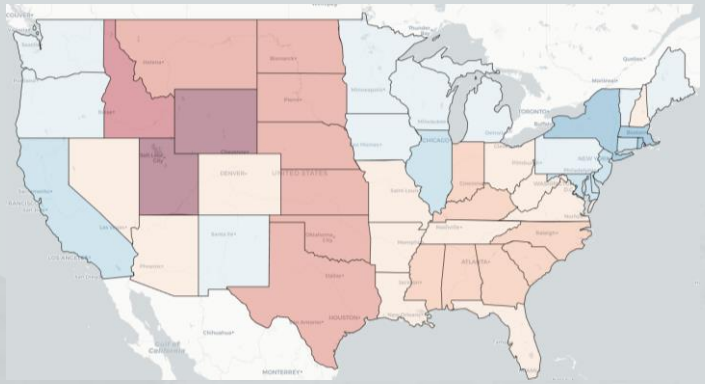
Elections map of the last 20 years

The following maps shows the intensity of United States votes of the last 20 elections using the Surplus Vote Percnet*

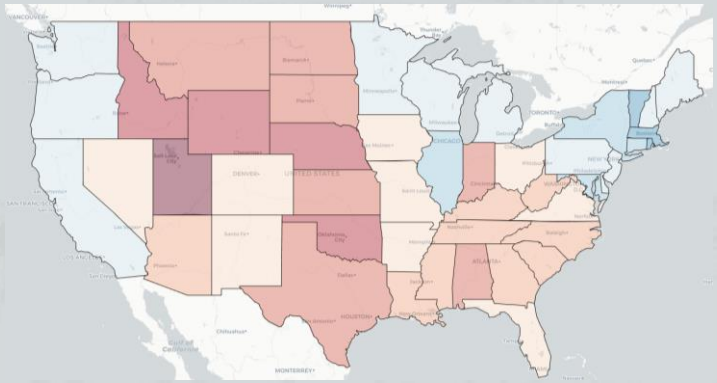
Surplus Vote Percent: shows how much more was the voting percentage of the winning state party. The republicans are represented by the the negative percentages (red), while the democrats are represented by the positive percentages (blue)



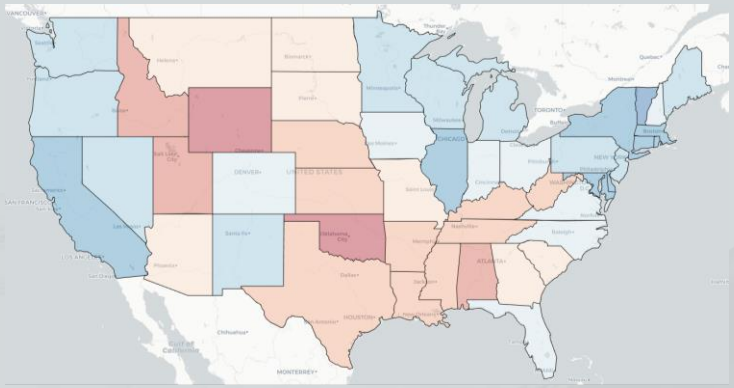
2000 Elections



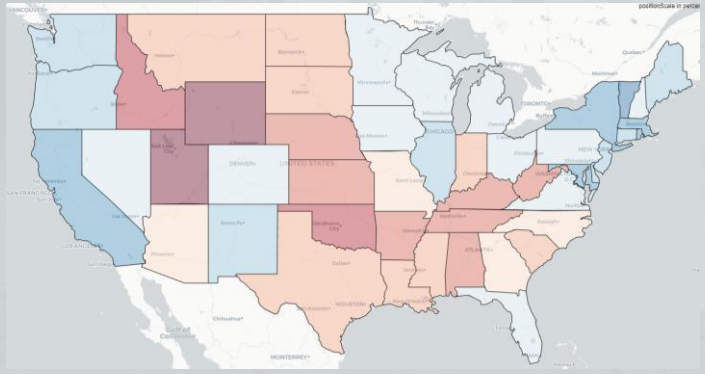
2004 Elections



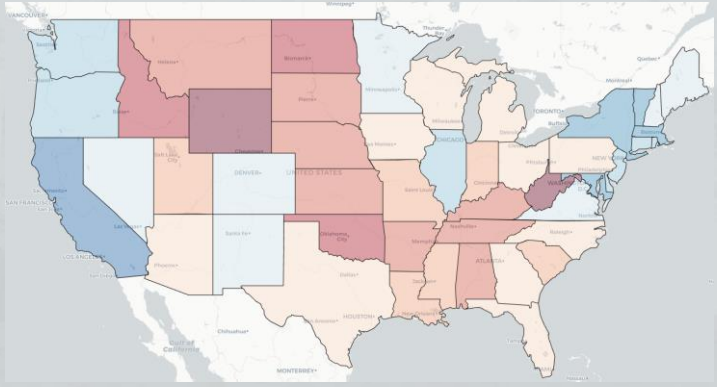
2008 Elections



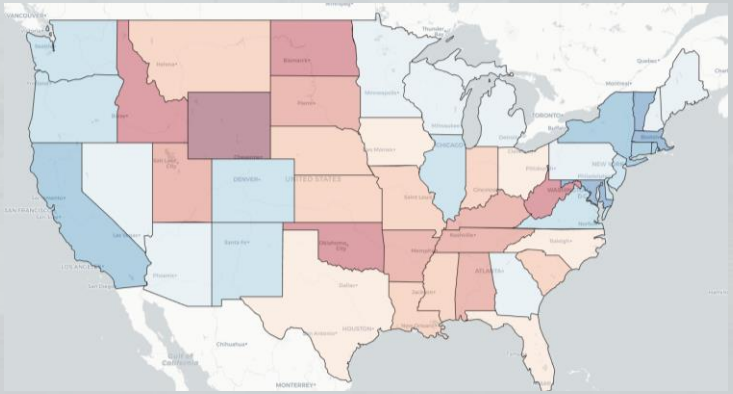
2012 Elections



2016 Elections



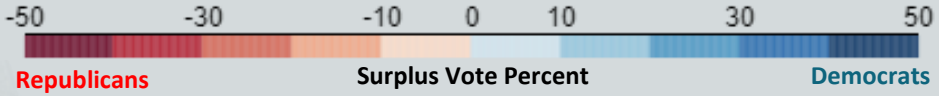
2020 Elections

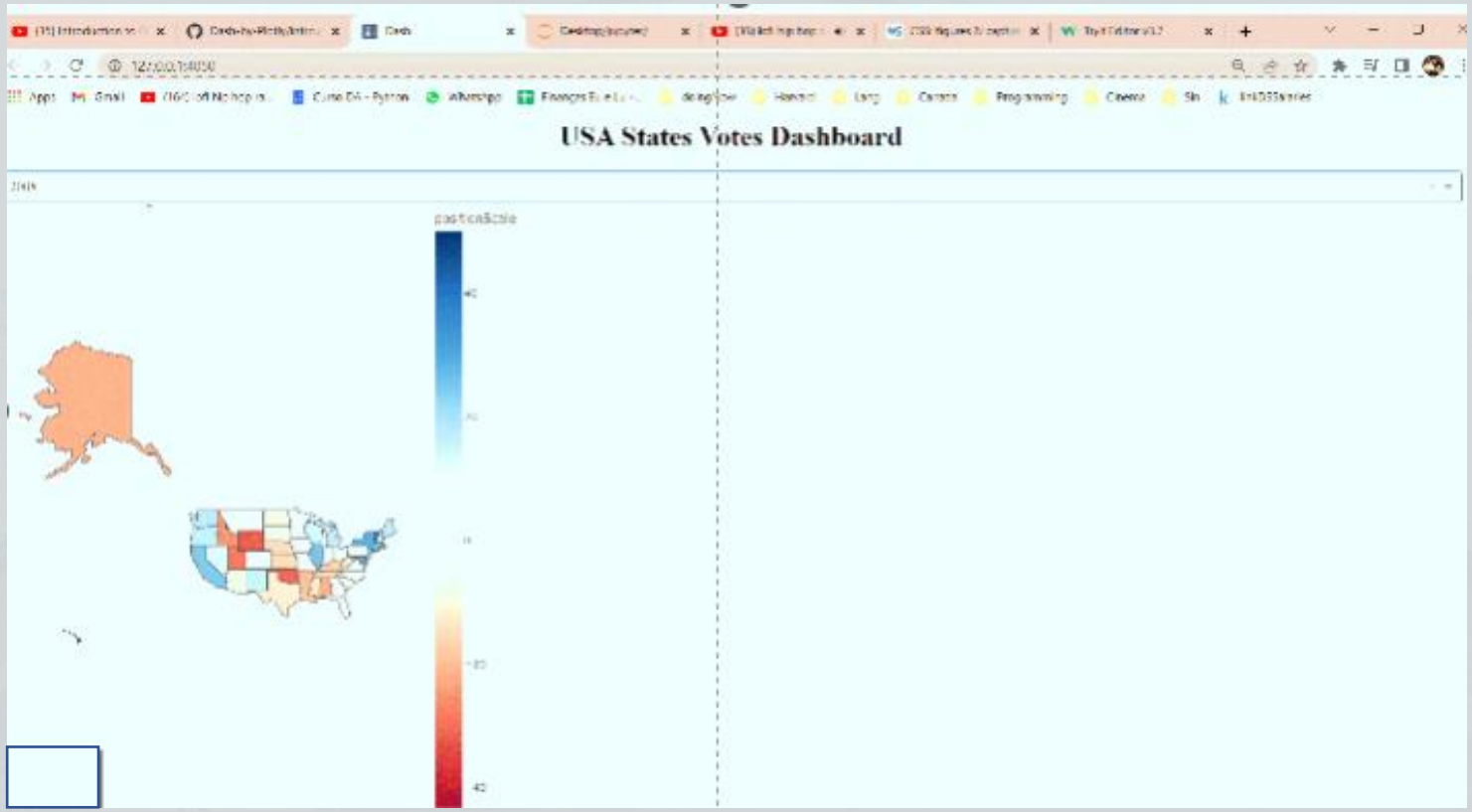


Elections map of the last 20 years With Plotly Dash

The following interactive map shows the intensity of United States votes of the last 20 elections using the Surplus Vote Percent*.
It also shows the total, Republican and Democrats per state in United States during the period of (2000 – 2020).

Surplus Vote Percent: shows how much more was the voting percentage of the winning state party. The republicans are represented by the the negative percentages (red), while the democrats are represented by the positive percentages (blue)



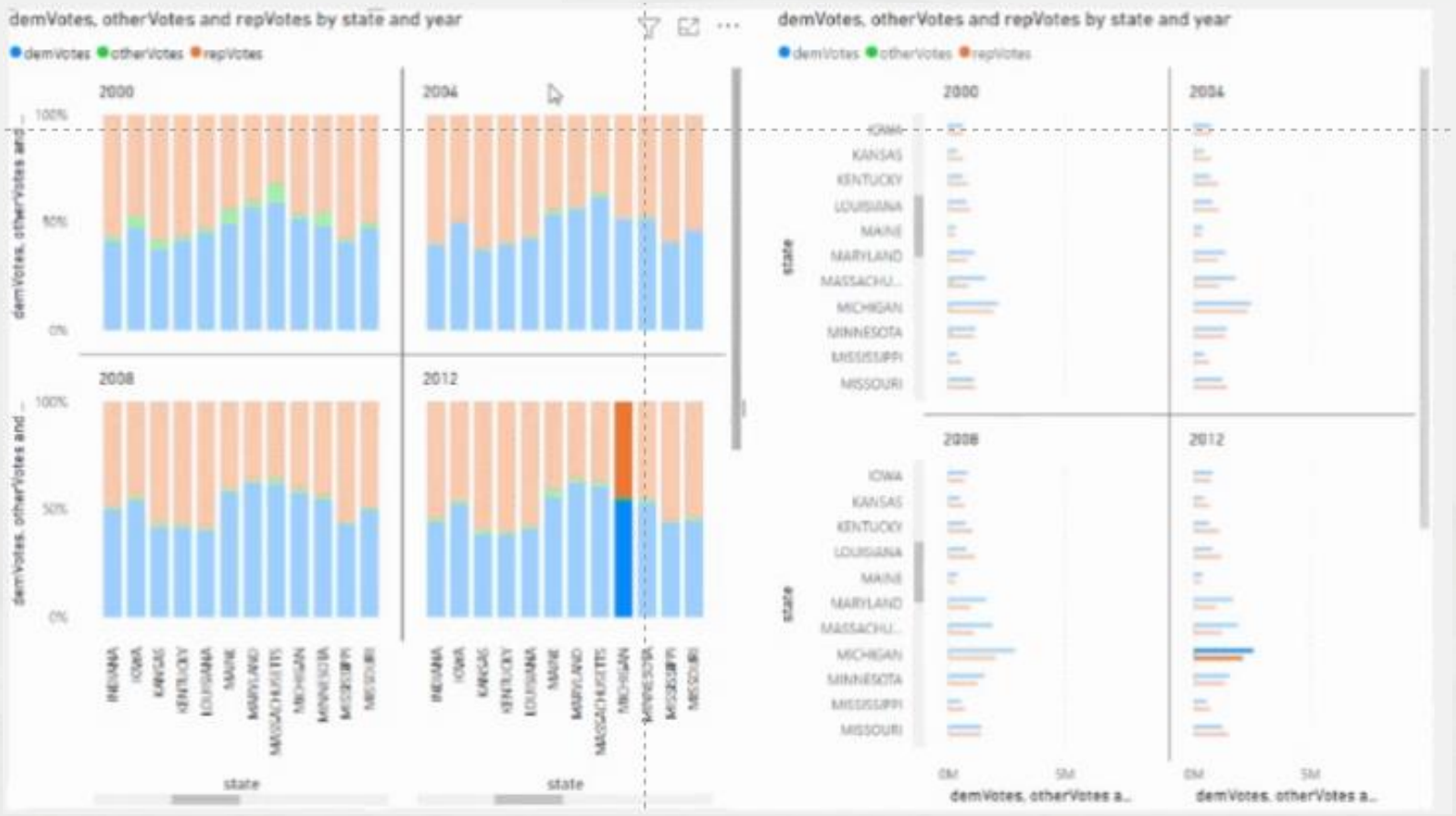


Elections graph of the last 20 years with SQL and Power Bi

The graphs (ploted in power bi) shows the elections results in each state for the 2 major parties (Democrat and Republican) for the last 20 years.

Republicans are represented in red, democrats in blue and other parties (green, libertarian, etc.) are represented in green.

- **Republican**
- **Democrat**
- **Other parties**

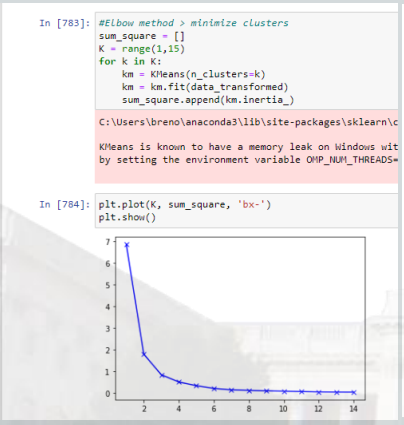


3- Machine Learning

Using machine learning – kcluster in sckit-learn – we were able to classify the states in 3 clusters (using the elbow technique): states that vote republican (red), states that vote democrat (blue) and states in between (white), meaning that they can vote wither republican or democrat.

Elbow technique:

```
In [783]: #Elbow method > minimize clusters
sum_square = []
K = range(1,15)
for k in K:
    km = KMeans(n_clusters=k)
    km = km.fit(data_transformed)
    sum_square.append(km.inertia_)

C:\Users\breno\anaconda3\lib\site-packages\sklearn\c

KMeans is known to have a memory leak on Windows wit
by setting the environment variable OMP_NUM_THREADS=

In [784]: plt.plot(K, sum_square, 'bx-')
plt.show()
```

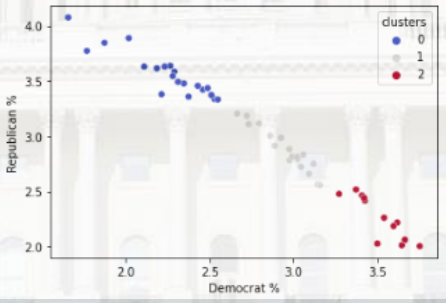


Kmeans cluster (sckit-learn) using the ideal numbers of clusters in the elbow technique:

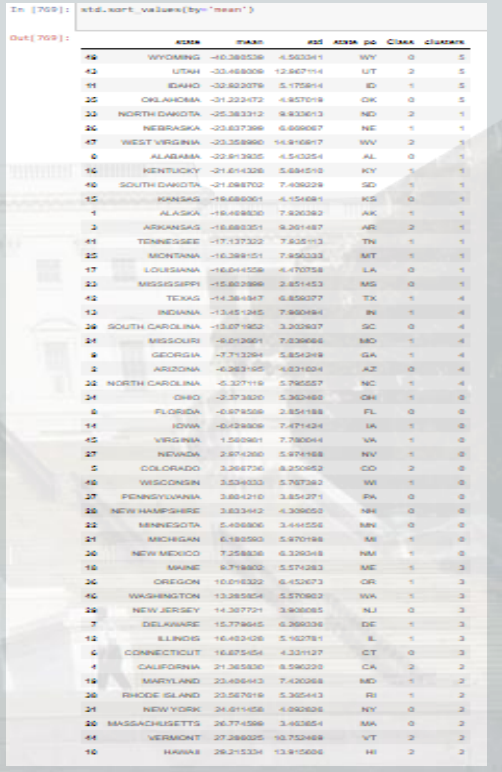
```
In [785]: kmeans = cluster.KMeans(n_clusters=3, init='k-means++')
kmeans = kmeans.fit(mcDF[['d%','r%']])

In [786]: kmeans.cluster_centers_

Out[786]: array([[2.24361139, 3.56519533],
                 [3.52001849, 2.25562644],
                 [2.95616523, 2.87131004]])

In [787]: std['clusters'] = kmeans.labels_

In [788]: std['Democrat %'] = mcDF['d%']

In [789]: std['Republican %'] = mcDF['r%']

In [790]: std.head()

Out[790]:
```

	state	mean	std	state_po	Class	clusters	Democrat %	Republican %
0	ALABAMA	-22.913935	4.543254	AL	0	0	2.264383	3.639219
1	ALASKA	-19.469830	7.926392	AK	1	0	2.212081	3.380270
2	ARIZONA	-5.263195	4.031024	AZ	0	2	2.733288	3.109080
3	ARKANSAS	-18.880351	9.261487	AR	2	0	2.345837	3.478658
4	CALIFORNIA	21.365830	8.596220	CA	2	1	3.542172	2.260223

```
In [792]: import seaborn as sns

In [793]: sns.scatterplot(x='Democrat %',y='Republican %',hue ='clusters',data=std, palette='coolwarm')

Out[793]: <AxesSubplot:xlabel='Democrat %', ylabel='Republican %'>
```



States divided in their cluster



Republican states: Wyoming, Utah, Idaho, Oklahoma, North Dakota, Nebraska, West Virginia, Alabama, Kentucky, South Dakota, Kansas, Alaska, Arkansas, Tennessee, Montana, Louisiana, Mississippi, Texas, Indiana, South Carolina

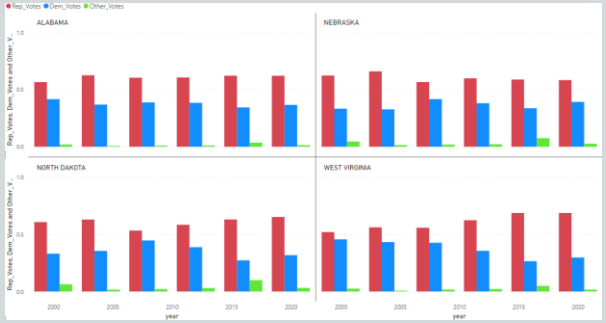
Democrat states: Washington, New Jersey, Delaware, Illinois, Connecticut, California, Maryland, Rhode Island, New York, Massachusetts, Vermont, Hawaii

Neutral states: Missouri, Georgia, Arizona, North Carolina, Ohio, Florida, Iowa, Virginia, Nevada, Colorado, Wisconsin, Pennsylvania, New Hampshire, Minnesota, Michigan, New Mexico, Maine, Oregon

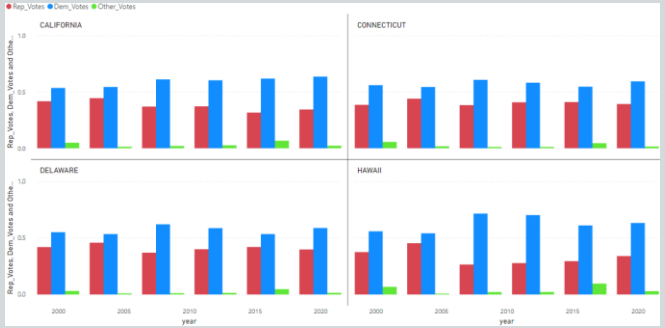
Results

Elections in the 3 groups (2000 – 2020):

Republican States

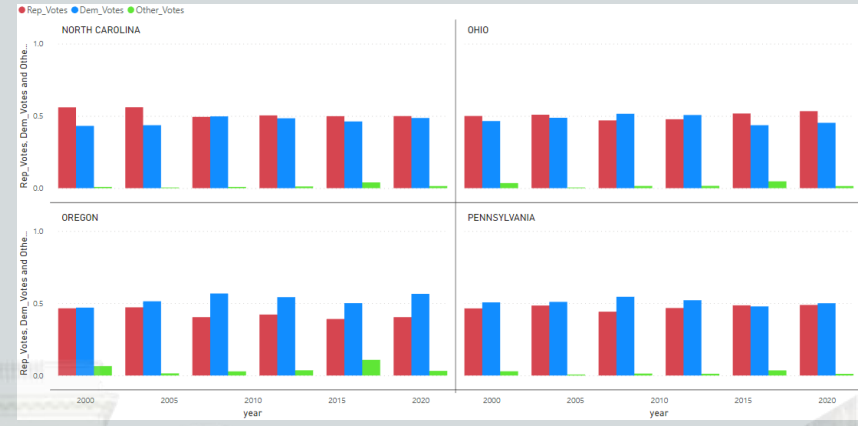


Democrat States



Elections in the 3 groups (2000 – 2020):

Neutral States



3- States that changed their way of voting in the last 20 years.

As seen in the last graphs, both states in the parties cluster (Democratic and Republican) did not change their way of voting in the last 20 years. In the Republican cluster the Republican party won every single election during the whole period and the same goes to the Democratic cluster, that the Democratic Party won every election for the last 20 years.

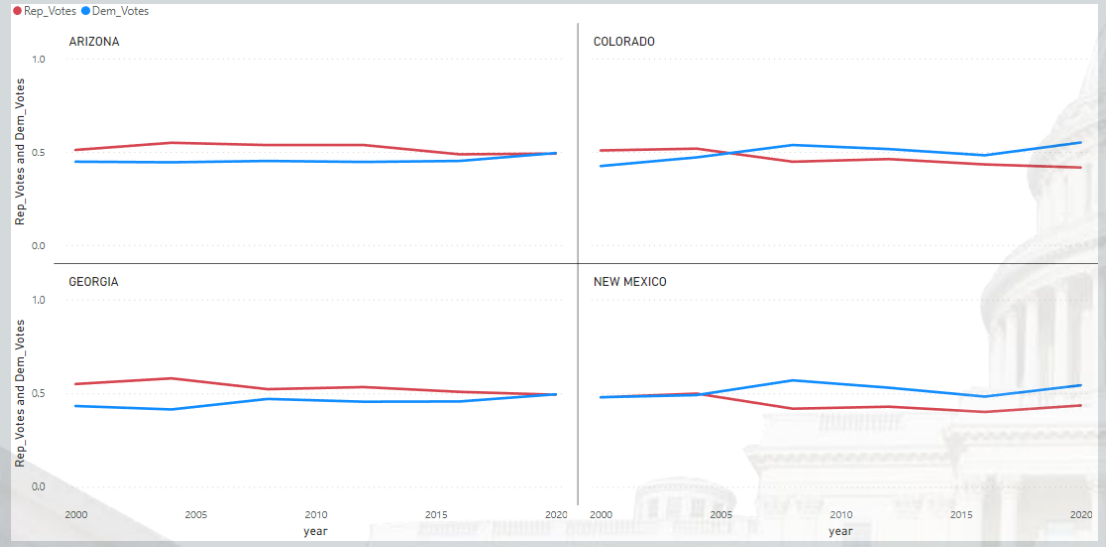
That situation changes in the neutral cluster, that some states, but not all, changed their way of voting.

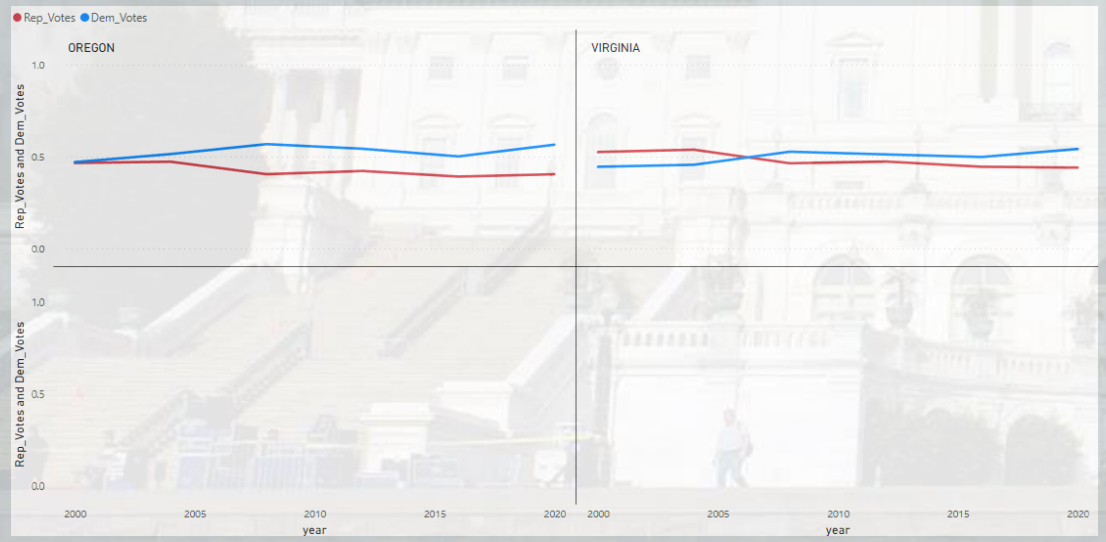
Looking into the graphs we can divide the states in 2 groups, that is, states that changed their way of voting, that means, States that we can clearly see a trend in their voting way and States that we can not, namely, States that might change the winning party but that not showcase a trend in their voting habit.

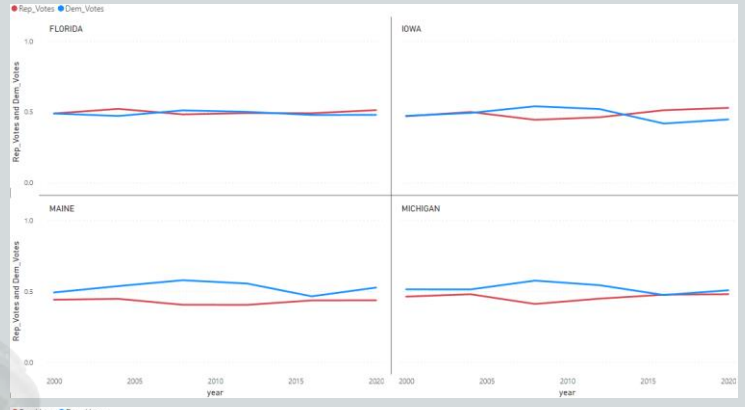
States that presented a trend in their way of voting:

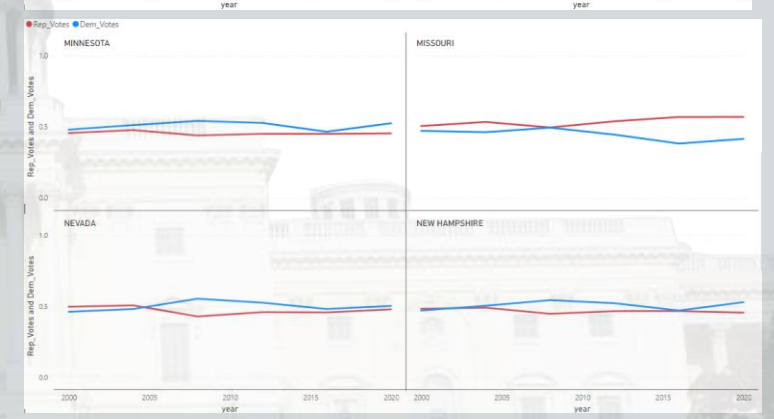
Neutral States

States that did not presented a trend in their way of voting:









Conclusion

Conclusions:

Of the 50 States in USA:

- **20 have the Republican Party as their major party:** Wyoming, Utah, Idaho, Oklahoma, North Dakota, Nebraska, West Virginia, Alabama, Kentucky, South Dakota, Kansas, Alaska, Arkansas, Tennessee, Montana, Louisiana, Mississippi, Texas, Indiana, South Carolina.
- **12 have the Democratic party as their major party:** Washington, New Jersey, Delaware, Illinois, Connecticut, California, Maryland, Rhode Island, New York, Massachusetts, Vermont, Hawaii
- **18 are Neutral, namely, do not have a clear major party:** Missouri, Georgia, Arizona, North Carolina, Ohio, Florida, Iowa, Virginia, Nevada, Colorado, Wisconsin, Pennsylvania, New Hampshire, Minnesota, Michigan, New Mexico, Maine, Oregon
- **Of the 18 Neutral States, 6** (Arizona, Colorado, Georgia, New Mexico, Oregon, Virginia) **presented the same voting trend for the last 20 years, expressly going away from the Republican Party and into the Democratic party.**