Baystate Banner: LatinX Republican Support (Spring 2021)	
Contact	**Senior Editor: Yawu Miller** yawu@bannerpub.com **Project Manager: Lingyan Jiang, lingyan@bu.edu** **Staff Lead: Steve Backman, sbackma1@bu.edu** **Team Members:** Ngozi Omatu (**nomatu@bu.edu**), **Song Xie (sxie2@bu.edu), Matan Ziegel (mziegel@bu.edu), Gil Lotzky (glotzky@bu.edu), Anna Xie (annaxyw@bu.edu)**
Organization	Baystate Banner
Organization Description	The Bay State Banner is an African American owned news weekly that reports on the political, economic, social, and cultural issues that are of interest to African American and English speaking Latinos in Boston and throughout New England.
Project Type	Data Science
Project Description	The client would like to understand the components of support for Republicans over the time period of 2015-2020 including Presidential elections and the Governor's races. The goal of the project is to find whether or not there is a significant difference in the voting pattern of the LatinX community from 2016 to 2020. We will be collecting data from cities with a majority LatinX population and non-LatinX populations. We will then compare the two populations to see whether there is a significant difference in voting patterns between the LatinX community and the control group(which has voted Republican consistently in the past). If there is a significant difference between the communities, we will analyze the various changes from 2016 to 2020 in those communities. The main goal is to conclude which LatinX voters changed their votes and which voters stayed consistent.
Data Sets	● PD43 MA state portal ○ 2016 and 2020 Presidential races ○ 2014 and 2018 MA Governor (General Elections) ● Census data (Demographics) ○ 6 cities Census data 2019 ○ Lawrence: 2016 and 2019 ○ Lynn: 2016 and 2019 ○ Springfield: 2016 and 2019

	○ Southwick: 2016 and 2019 ○ Acushnet: 2016 and 2019 ○ Douglas: 2016 and 2019 Optional Support data ● LatinX Origin Demographics (2014)
Suggested Steps	**Step one:** Collect city voter & demographic data of majority LatinX towns (~80%) ● Lawrence ● Lynn ● Springfield and compare with representative mostly white towns very likely to vote Republican as control groups ● Southwick ● Acushnet ● Douglas About the data: ● Voter data includes total ballots cast, registered voters, eligible voters, and demographics for that city. ● Election races will include 2016 and 2020 Presidential races as well as 2014 and 2018 MA Governor-general elections. **Step two:** Use pandas to read through csv files and convert to data frames. Only keep key attributes for demographic data (zip code, estimated total pop, LatinX sub-group pop, and voting age pop (18+)) and election data (city zip codes, city names, total votes per candidate, total votes) **Step three:** Relate city election data to city demographic data (either via each city's zip codes or precincts). Analyze different kinds of LatinX voters supporting Trump - Mexican, Dominican, Cuban American. Where is their support concentrated? How did their support change from 2016 to 2020 **Step three:** Examine the difference between LatinX people registered as Democrat, Republican, Independent in these different geographic and elections. **Step four:** Complete analysis on correlations between how these towns to vote, do they move with each other in different elections> **Step five:** Complete data visualizations based to present which neighborhoods, cities, and districts have higher levels of support for Donald Trump.

Strategic questions	1. How has support for Trump shifted across LatinX sub-groups from 2016 to 2020? Was there a significant shift? 2. What is the breakdown of LatinX sub-groups in their support for Democratic vs. Republican candidates? 3. Which LatinX groups exhibited changed their votes and which groups remained the same?
Additional Information	**Tools & Methods** *Data pre-processing:* Pandas to convert csv files into data frames, NumPy for processing and organizing the pre-processed data *Data Visualization:* Matplotlib, Seaborn, Tableau for all kinds of interactive visualizations **Weekly Meeting Schedule:** **Wednesdays 11am - 12:30pm EST**

Potential Risks and Limitations:

1. Lack of connection between city election results and city demographic data, i.e. unable to break down city election results by demographics as we are only given each city's aggregate election results
2. Difficult to apply clustering methods as we are only given aggregate data for demographics and election results for each city (lack of large dataset)
3. The conclusion on voting pattern changes may not be concrete due to the other factors such as changes in voter turnout or each demographic's population which is not indicative of a change in preference for a political candidate.

General Questions:

1. How can we combine the election data with the demographic data?
    a. It will be difficult to combine data frames because we cannot correspond the zip codes from the demographic data to the precincts in the election data
    b. Can't correlate changes in demographics to changes in election results
    c. Worst case scenario, should we combine data frames based on the city?
2. Should we organize election data based on Republicans vs Non Republicans rather than Republicans vs Democrats vs Independents?
3. We can't directly answer any of the key questions until we are able to relate election data to demographic data

Key Question Answers:

1. In our non-control group cities (Springfield, Lawrence, Lynn), Springfield saw a 1.79 percentage point increase in LatinX population (relative to the city's total population) and a 5.13 percentage point increase in Trump votes(across all races). For Lawrence, there was a 3.44 percentage point increase in LatinX population(relative to the city's total population) and a 10.85 percentage point increase in Trump votes(across all races). For Lynn, there was a 4.79 percentage point increase in LatinX population(relative to the city's total population) and 2.96 percentage point increase in Trump votes(across all races).

   Listed below are the screenshots of the code for the preprocessed data:

In [1]:
```python
import pandas as pd
import numpy as np
```

In [2]:
```python
acs2016 = pd.read_csv('2016data.csv')
acs2019 = pd.read_csv('2019data.csv')
```

In [3]:
```python
acs2016
```

Out[3]:

	Zipcode	Geographic Area Name	Estimate Total population	Margin of Error!!SEX AND AGE!!Total population	Percent!!SEX AND AGE!!Total population	Percent Margin of Error!!SEX AND AGE!!Total population	Estimate!!SEX AND AGE!!Total population!!Male	Margin of Error!!SEX AND AGE!!Total population!!Male	Percent!!SEX AND AGE!!Total population!!Male	Percent Margin of Error!!SEX AND AGE!!Total population!!Male	...	Perc POPULAT
0	1077	ZCTA5 01077	9650	48	9650	(X)	4649	272	48.2	2.8	...	
1	1103	ZCTA5 01103	2430	312	2430	(X)	1217	217	50.1	5.0	...	
2	1104	ZCTA5 01104	23183	872	23183	(X)	10829	587	46.7	1.9	...	
3	1105	ZCTA5 01105	11576	814	11576	(X)	5585	458	48.2	2.8	...	
4	1107	ZCTA5 01107	11474	753	11474	(X)	5794	426	50.5	2.1	...	
		ZCTA5										

In [4]:
```python
acs2019
```

In [5]:
```python
acs2016['Zipcode'] = acs2016['Zipcode'].astype(str)
acs2019 = acs2019.rename(columns={"id":"Zipcode"})
```

In [6]:
```python
acs2019['Zipcode'] = acs2019['Zipcode'].astype(str)
```

In [7]:
```python
acs2016['Zipcode'] = acs2016['Zipcode'].str.zfill(5)
acs2019['Zipcode'] = acs2019['Zipcode'].str.zfill(5)
```

In [9]:
```python
# Estimate!!HISPANIC OR LATINO AND RACE!!Total population!!Hispanic or Latino (of any race)!!Mexican
# Estimate!!HISPANIC OR LATINO AND RACE!!Total population!!Hispanic or Latino (of any race)!!Puerto Rican
# Estimate!!HISPANIC OR LATINO AND RACE!!Total population!!Hispanic or Latino (of any race)!!Cuban
# Estimate!!HISPANIC OR LATINO AND RACE!!Total population!!Hispanic or Latino (of any race)!!Other Hispanic or Latin
# Estimate!!HISPANIC OR LATINO AND RACE!!Total population!!Not Hispanic or Latino
# Estimate!!CITIZEN, VOTING AGE POPULATION!!Citizen, 18 and over population
# Estimate Total population
# Zipcode
processed2016 = acs2016[['Zipcode','Estimate Total population','Estimate!!HISPANIC OR LATINO AND RACE!!Total populat
```

In [10]:
```python
processed2019 = acs2019[['Zipcode','Estimate Total population','Estimate!!HISPANIC OR LATINO AND RACE!!Total populat
```

In [11]:
```python
processed2016 = processed2016.rename(columns={'Estimate!!HISPANIC OR LATINO AND RACE!!Total population!!Hispanic or
```

In [12]:
```python
processed2019 = processed2019.rename(columns={'Estimate!!HISPANIC OR LATINO AND RACE!!Total population!!Hispanic or
```

In [13]:
```python
city = []
for row in processed2016['Zipcode']:
    if row == '01840' or row == '01841' or row == '01842' or row == '01843':
        city.append('Lawrence')
    elif row == '01103' or row == '01104' or row == '01105' or row == '01107' or row == '01108' or row == '01109' or
        city.append('Springfield')
    elif row == '01901' or row ==  '01902' or row == '01904' or row == '01905':
        city.append('Lynn')
    elif row == '02743':
        city.append('Acushnet')
    elif row == '01077':
        city.append('Southwick')
    elif row == '01516':
        city.append('Douglas')
    else:
        city.append('Missing City Name, Fix')
```

In [15]:
```python
processed2016.insert(1,"City",city)
```

In [16]:
```python
processed2016
```

In [20]:
```python
pres2016 = pd.read_csv('2016PresidentPrecinct.csv')
pres2020 = pd.read_csv('2020PresidentPrecint.csv')
gov2018 = pd.read_csv('2018gov.csv')
```

In [21]:
```python
pres2020.head()
```

Out[21]:

	City/Town	Ward	Pct	Joseph R. Biden, Jr.	Donald J. Trump	Jo Jorgensen	Howard Hawkins	All Others	No Preference	Blanks	Total Votes Cast
0	NaN	NaN	NaN	Democratic	Republican	Libertarian	Green-Rainbow	NaN	NaN	NaN	NaN
1	Abington	-	1	1,037	764	29	6	8	0.0	16	1,860
2	Abington	-	2	960	838	23	16	4	0.0	9	1,850
3	Abington	-	3	954	848	31	8	3	0.0	9	1,853
4	Abington	-	4	1,117	904	28	7	6	0.0	26	2,088

In [22]:
```python
gov2018.head()
```

Out[22]:

	City/Town	Ward	Pct	Baker and Polito	Gonzalez and Palfrey	All Others	Blanks	Total Votes Cast
0	NaN	NaN	NaN	Republican	Democratic	NaN	NaN	NaN
1	Abington	-	1	984	311	3	45	1,343
2	Abington	-	2	967	330	2	62	1,361
3	Abington	-	3	1,047	325	6	68	1,446
4	Abington	-	4	1,247	350	1	68	1,666

In [25]:
```python
procpres2016 = pres2016[['City/Town','Pct','Clinton/ Kaine','Trump/ Pence','Total Votes Cast']]
procpres2020 = pres2020[['City/Town','Pct','Joseph R. Biden, Jr.','Donald J. Trump','Total Votes Cast']]
procgov2018 = gov2018[['City/Town','Pct','Gonzalez and Palfrey','Baker and Polito','Total Votes Cast']]
```

In [26]:
```
procpres2016.drop([0])
procpres2020.drop([0])
procgov2018.drop([0])
```

Out[26]:

	City/Town	Pct	Gonzalez and Palfrey	Baker and Polito	Total Votes Cast
1	Abington	1	311	984	1,343
2	Abington	2	330	967	1,361
3	Abington	3	325	1,047	1,446
4	Abington	4	350	1,247	1,666
5	Abington	5	370	1,167	1,596
...	...	...	...	...	...
2170	Yarmouth	4	372	1,254	1,681
2171	Yarmouth	5	366	1,085	1,486
2172	Yarmouth	6	374	1,089	1,516
2173	Yarmouth	7	433	1,802	2,307
2174	TOTALS	NaN	885,770	1,781,341	2,752,665

In [28]:
```
pres2016 = procpres2016[(procpres2016['City/Town'] == 'Lawrence') | (procpres2016['City/Town'] == 'Springfield')| (p
```

In [30]:
```
pres2020 = procpres2020[(procpres2020['City/Town'] == 'Lawrence') | (procpres2020['City/Town'] == 'Springfield')| (p
```

In [31]:
```
gov2018 = procgov2018[(procpres2016['City/Town'] == 'Lawrence') | (procgov2018['City/Town'] == 'Springfield')| (proc
```

```
<ipython-input-31-ef57abbf539a>:1: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  gov2018 = procgov2018[(procpres2016['City/Town'] == 'Lawrence') | (procgov2018['City/Town'] == 'Springfield')| (p
rocgov2018['City/Town'] == 'Lynn')|(procgov2018['City/Town'] == 'Acushnet')|(procgov2018['City/Town'] == 'Southwick
')|(procgov2018['City/Town'] == 'Douglas')]
```

In [33]:
```
pres2016 = pres2016.reset_index()
pres2016 = pres2016.drop(columns=['index'])
pres2020 = pres2020.reset_index()
pres2020 = pres2020.drop(columns=['index'])
gov2018 = gov2018.reset_index()
gov2018 = gov2018.drop(columns=['index'])
```

In [37]:
```
processed2016.to_csv(index = False)
processed2019.to_csv(index = False)
pres2016.to_csv(index = False)
pres2020.to_csv(index = False)
gov2018.to_csv(index = False)

#to_excel(r'Demo2016.xlsx', index = False)
#processed2019.to_excel(r'Demo2019.xlsx', index = False)
#pres2016.to_excel(r'Pres2016.xlsx', index = False)
#pres2020.to_excel(r'Pres2020.xlsx', index = False)
#gov2018.to_excel(r'Gov2018.xlsx', index = False)
```