

Baystate Banner: LatinX Republican Support (Spring 2021)	
Contact	<p>Senior Editor: Yawu Miller yawu@bannerpub.com</p> <p>Project Manager: Lingyan Jiang, lingyan@bu.edu</p> <p>Staff Lead: Steve Backman, sbackma1@bu.edu</p> <p>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)</p>
Organization	Baystate Banner
Organization Description	The Bay State Banner is an African American-owned newsweekly 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 2014-2020 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 sub groups within the LatinX population to identify the cause. The main goal is to conclude which LatinX voters changed their votes and which voters stayed consistent.
Data Sets	2014 & 2018 Mass. Demographic dataset+Election Results Scatterpoints of the shift in Democrats and Republican votes for Presidential Bar Data for the Government Demographic

Suggested Steps	<ul style="list-style-type: none"> • Step one: Collect government election results & demographic data of majority LatinX towns in Mass. • Step two: Use pandas to read through csv files and convert to data frames. Only keep key attributes for demographic data (Tract, 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). Created a map of the Massachusetts counties using the tracts and geopanda • Step three: Relate city election data to city demographic data (either via each city's precincts). Analyze different kinds of how did their support change from 2014 to 2018 • Step four: Complete analysis on correlations between how these towns voted • Step five: Complete data visualizations by using the election results from each major city to highlight them on the geopanda map.
Strategic questions	<ol style="list-style-type: none"> 1. How has support for Trump shifted across the LatinX population 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 changes in their votes and which groups remained the same?
Additional Information	<p>Tools & Methods</p> <p><u>Data pre-processing:</u> Pandas to convert csv files into data frames, NumPy for processing and organizing the pre-processed data</p> <p><u>Data Visualization:</u> Matplotlib, Seaborn, Tableau for all kinds of interactive visualizations, Geopanda, Microsoft Excel</p> <p>Weekly Meeting Schedule: Wednesdays 11am - 12:30pm EST</p>

Collect and pre-process secondary batch of data:

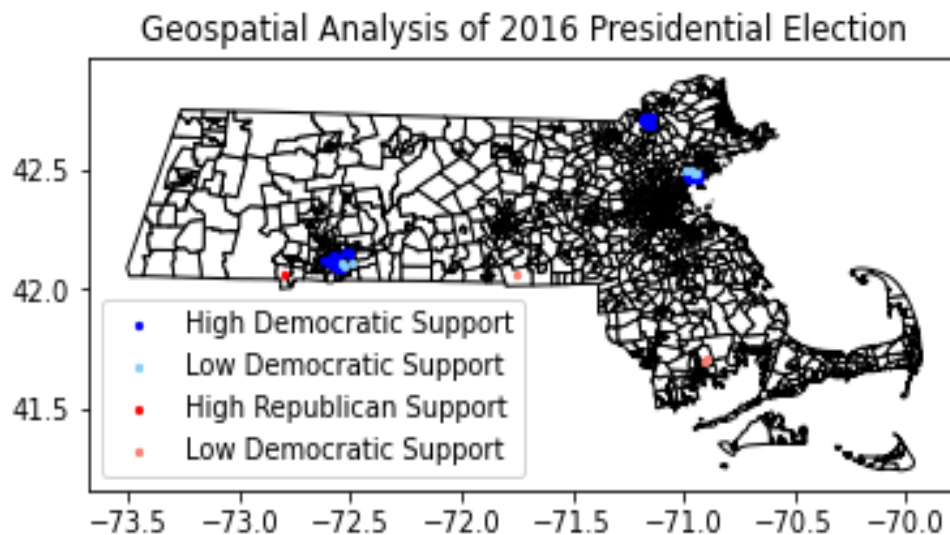
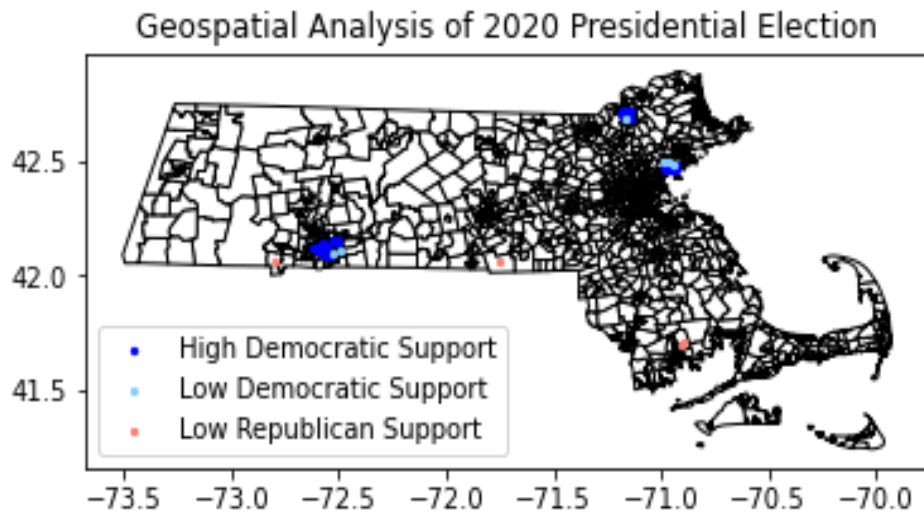
To utilize the precinct geospatial data for further analysis first we utilized the census 2010 shapefiles which include the geospatial shapes of each city and precinct of Massachusetts. We then used geopandas to convert the shapefile into a geopandas dataframe. Using the entirety of Massachusetts geospatial data, a preliminary boundary plot can be added to represent the state of Massachusetts. The geopandas data frame was then converted to transform the geometries to a new coordinate reference system by 'to_crs'.

We then processed the dataframe to only keep necessary attributes, including the geospatial shapes, city names, and other geospatial information. Then we merge our processed demographic and voting results data with the geospatial data to a pandas dataframe. We then sort our dataframes into separate categories from those that have 15% or higher than 15% majority per party into respective data frames creating 4 data frames per voting year. By separating the voting data, we are able to plot the differences onto the map easier and color them.

Once all the geospatial information is merged together with the voting and demographic data, and then it is separated into their respective categories, the function .points_from_xy are able to convert the latitude and longitude of each tract into a geopandas shape. We then convert the pandas dataframes to geopandas dataframes by utilizing our new geopandas shapes attribute. Now we are able to plot each of the geopandas dataframes onto the original Massachusetts boundary plot. This geospatial data helps us better understand the differences in location of the various LatinX populations and the political support per party across the state of Massachusetts.

Some issues while processing this data was merging geospatial information with a pandas dataframe. By merging the geopandas information with a pandas dataframe, it is unable to recognize the appropriate geometry attribute which are made of 'Polygon' shapes. There does not exist a function to convert this object type to a geopandas object, so we can process the pandas dataframe. Further investigation is needed to understand how to do this. This step is crucial in displaying more results across Massachusetts since we have every precinct mapped and its respective Republican or Democratic Support results.

Shifts in the Presidential Election Support Across Massachusetts (2016 - 2020, 6 cities):



Geospatial Analysis:

From our geospatial analysis, there was a shift in Republican support from the 2016 presidential election to the 2020 presidential election. There was no 'High Republican Support'* present in the 2020 presidential election data for our key six cities. The cities that changed from 2016 to 2020 were from our control group, which implies that the cities which usually support Republican candidates in elections from past data, had changed their support in the 2020 presidential election. We plan to refine the maps to make it easier to read and add more tracts to show a comparison.

*The spread of High to Low support for both parties represents an arbitrary decision of a majority higher than 15% of the results was considered 'High Support' and any majority below 15% of the results was considered 'Low Support'

LatinX Cities: Lynn, Lawrence, Springfield

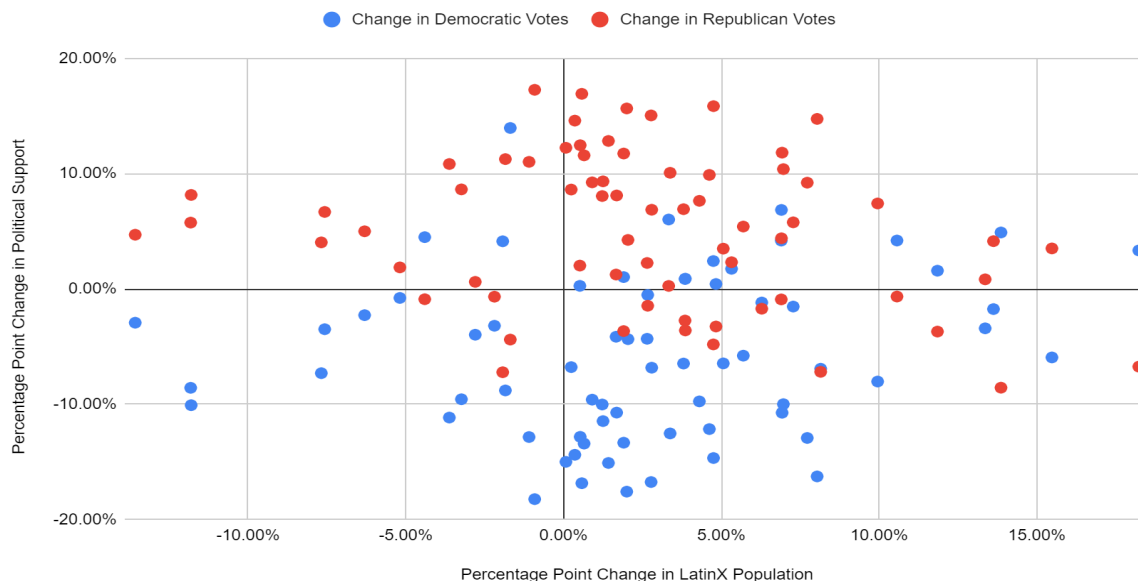
Control Group Cities(Non-LatinX): Acushnet, Douglas, Southwick

Strategic Question 1:

How has support for Trump shifted across the LatinX population from 2016 to 2020? Was there a significant shift?

Graph 1: Political Support (Presidential Election) vs. LatinX Population Changes

Change in Presidential Political Support vs Change in LatinX population (2016 - 2020)

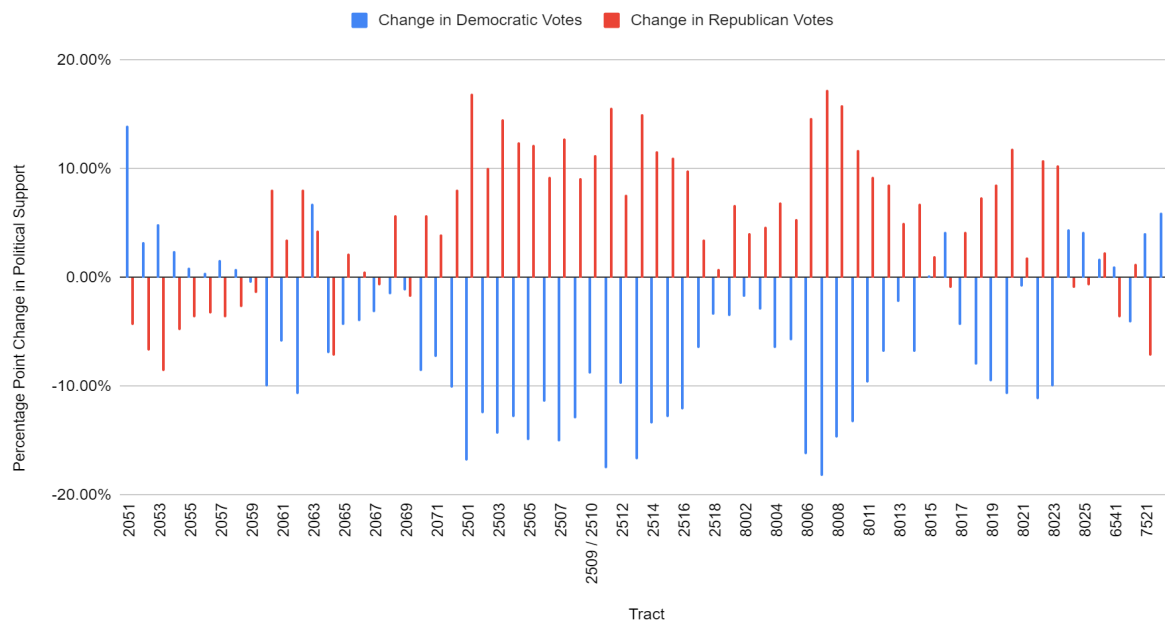


It appears that regardless of the direction of change in LatinX population, tracts within the 6 cities have seen a greater percentage point increase in votes for Trump than votes for the democratic candidate. For the change in democratic support, it has generally been a 5% - 15% percentage point decrease.

(Each red dot represents the percentage point change in Republican support for a given percentage point change in LatinX population and each blue dot represents the percentage point change in Democratic support for the same change in LatinX population, both of which pertaining to a single tract)

Graph 2: Presidential Support per Tract

Change in Presidential Political Support per Tract (2016 - 2020)

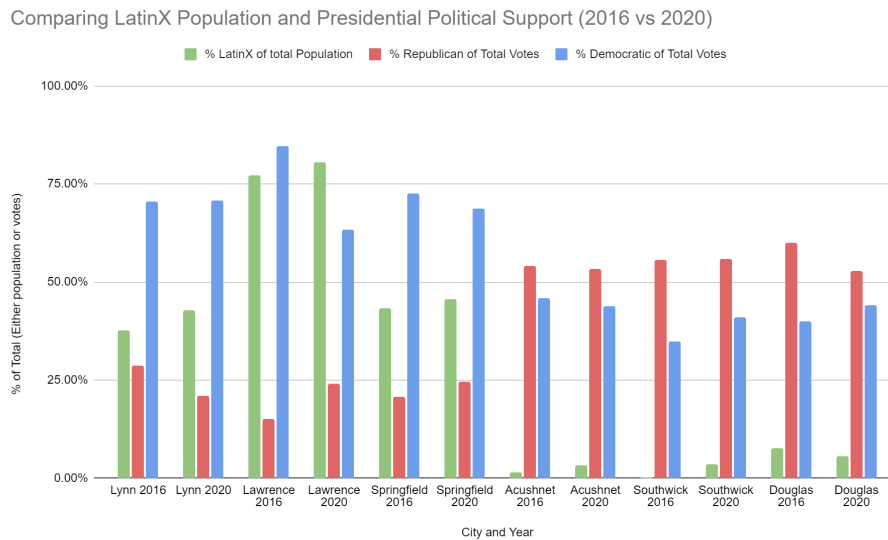


Control Group Tracts(Non-LatinX): Acushnet(6541, 6542),Douglas(7521),Southwick(8131)

LatinX Tracts: Lynn(2051-2072), Lawrence(2501-2518),Springfield(8001-8026)

As shown in Graph 2, Lawrence and Springfield, predominantly LatinX populations, experienced increases in Republican support and decreases in Democratic support from 2016-2020 and Lynn experienced a decrease in Republican support and a large increase in democratic support. The control groups did not experience significant changes in political party support from 2016-2020. This leads us to believe that we should analyze whether or not there is significant variable change explained by changes in the LatinX population from 2016-2020. Further investigation is needed, however, it is not possible to conclude that the changes in LatinX population in these tracts can explain the changes in voting pattern from 2016-2020.

Graph 3: LatinX Population and Political Party Support Presidential Election (2016 vs. 2020)



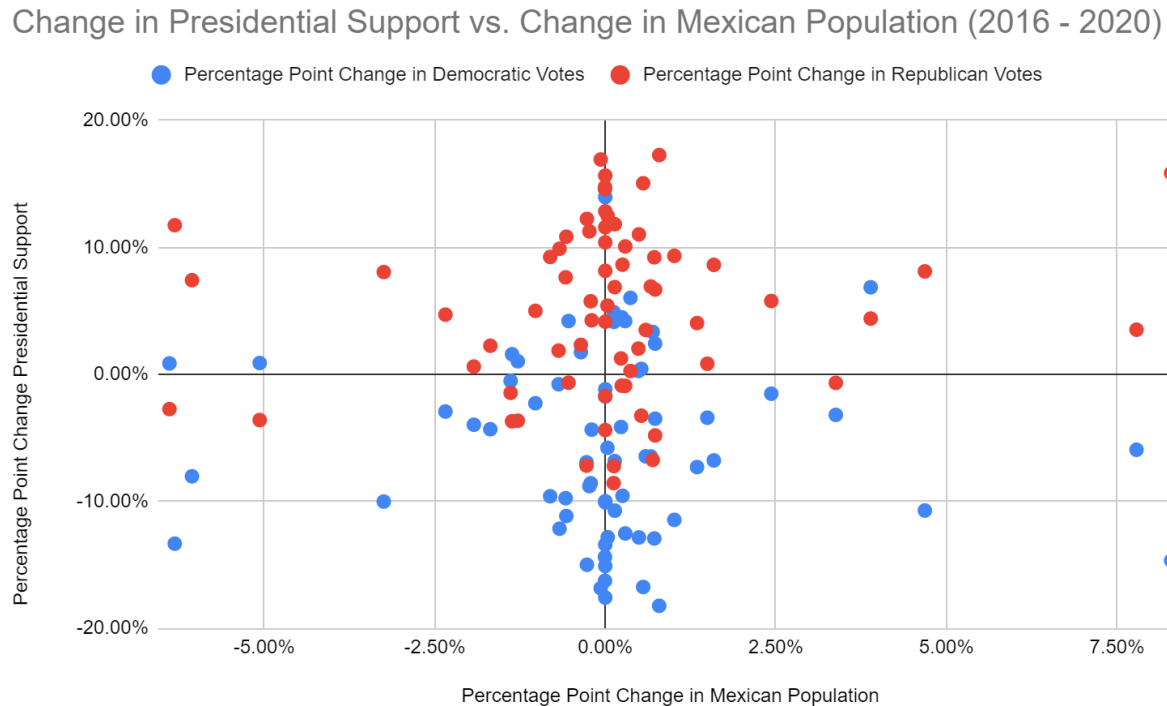
LatinX Cities: Lynn, Lawrence, Springfield

Control Group Cities(Non-LatinX): Acushnet, Douglas, Southwick

There are no obvious patterns displayed from the changing LatinX population from 2016 to 2020 and the Political Party each respective city had majority support. From graph 2 we can gather that from these six cities, higher LatinX populations tend to vote more for Democratic Candidates from 2016 to 2020. Cities with lower LatinX populations tend to vote for Republican Candidates from 2016 to 2020. This data conflicts with the more specialized analysis we had used in Graph 2, as we look more closely at each tract of the city it had shown that the LatinX population tracts would vote more for Republican candidates from 2016 to 2020. One limiting factor when analyzing this data is that we are only looking at the three cities in Massachusetts with the highest LatinX populations which represent high population totals and comparing them to three cities with very low LatinX populations and much smaller total population sizes. In order to get a more accurate representation of the changes in support and population, collecting a larger sample size would be greatly beneficial.

Strategic Question 2: What is the breakdown of LatinX sub-groups in their support for Democratic vs. Republican candidates?

Graph 4: Presidential Support v. Mexican Population Change (2016-2020)

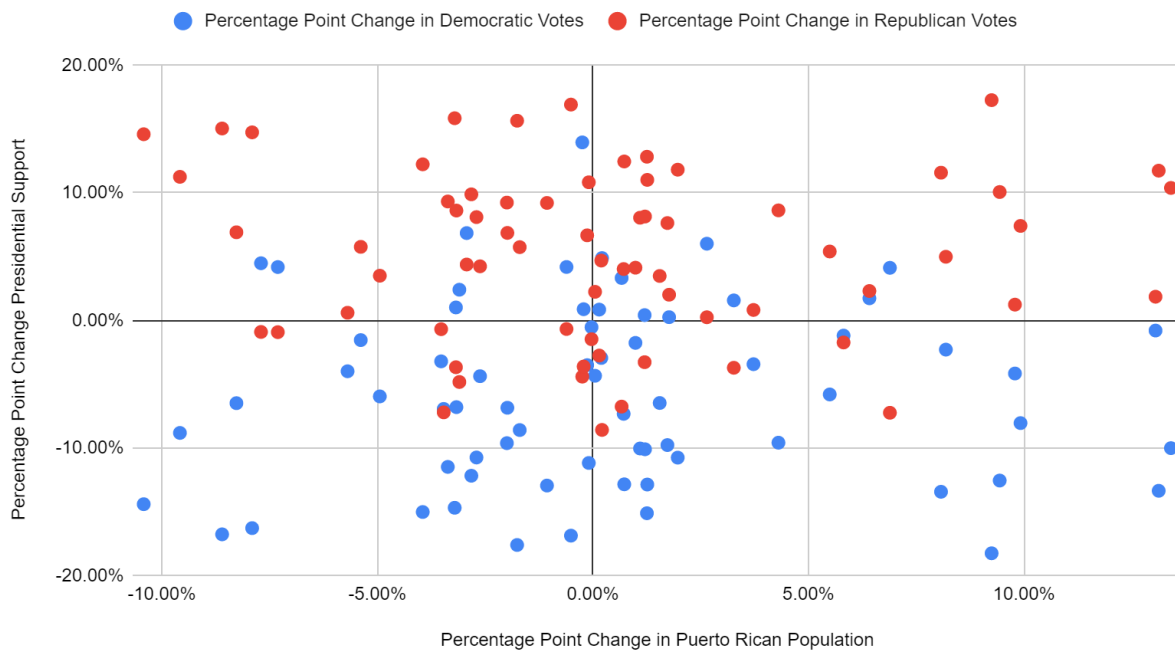


Note: Each dot on the scatter plot represents a tract and the respective change in republican vote or democratic vote for their Mexican population.

In graph 4, it is observed that there are very minimal population changes for the Mexican LatinX subgroup. When we compare these changes from 2016 to 2020, there is no apparent pattern present. The scatter plots show that the data is centered mostly around 0% change in Mexican population and there was an increase in support for Republican support from 2016 to 2020 for these respective tracts.

Graph 5: Presidential Support v. Puerto Rican Population Change (2016-2020)

Change in Presidential Support vs. Change in Puerto Rican Population (2016 - 2020)

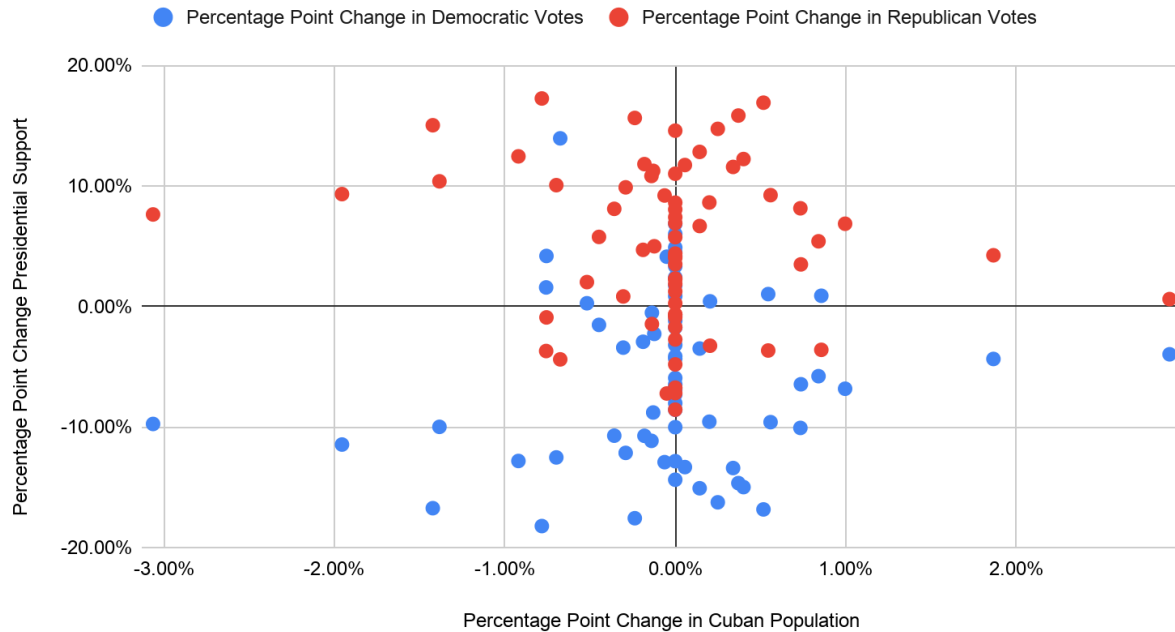


Note: Each dot on the scatter plot represents a tract and the respective change in republican vote or democratic vote for their Puerto Rican population.

Based on the graph, it seems that regardless of the direction of change in the Puerto Rican population, the percentage point increase in Republican support has generally been higher than the percentage point increase in Democratic support. There does not seem correlation between shifts in the Puerto Rican population and support for a specific political candidate.

Graph 6: Presidential Support v. Cuban Population Change (2016-2020)

Change in Presidential Support vs. Change in Cuban Population (2016 - 2020)

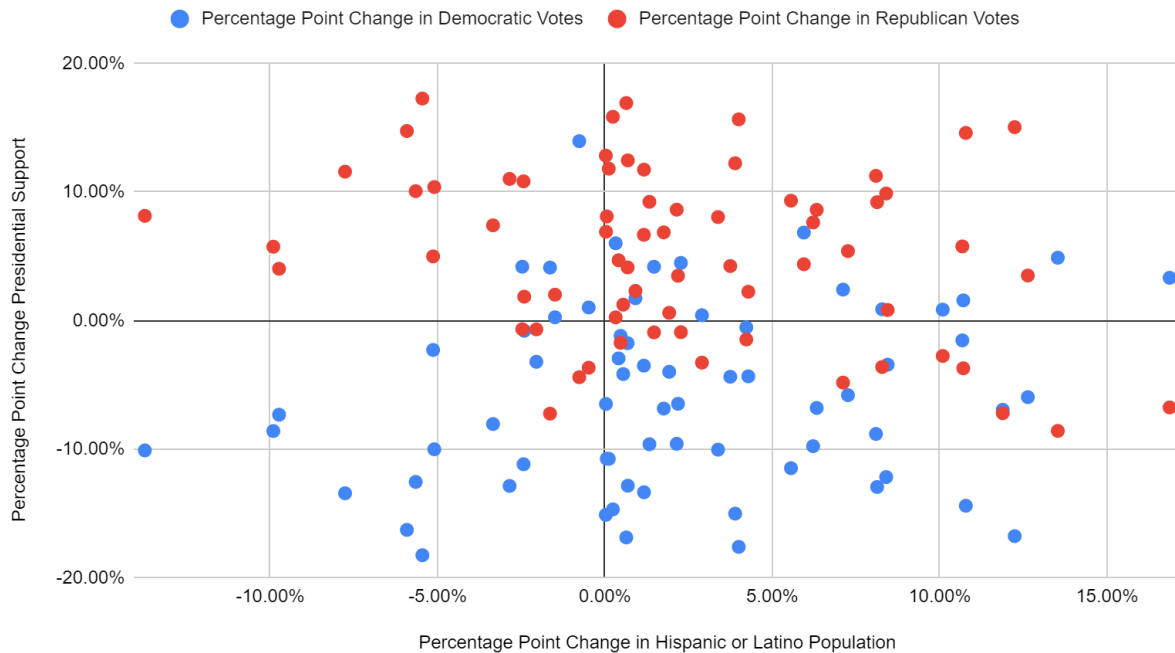


Note: Each dot on the scatter plot represents a tract and the respective change in republican vote or democratic vote for their Cuban population.

It does not appear that there is a correlation between the percentage point changes in the Cuban population and the percentage point changes in presidential support. This is supported by how shifts in the Cuban population are clustered around 0.00% despite a general decrease in Democratic support and increase in Republican support.

Graph 7: Presidential Support v. Other LatinX Population Change (2016-2020)

Change in Presidential Support vs. Change in Other Latino Population (2016 - 2020)



Note: Each dot on the scatter plot represents a tract and the respective change in republican vote or democratic vote for their Other LatinX subgroup populations.

There is not a correlation between the percentage point change of the other Latino population and percentage point change of election results. The percentage point change of the other latino shows some of the points are around the range of 0.00% and 5.00% for both slides but there is more of a decrease in percentage point for democrats support while the republican has increased in percentage point change.

Potential Limitations and Risks:

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. The conclusion on voting pattern changes may not be concrete due to 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. If possible, we also plan to analyze the voting turnout changes: LatinX vs. Non LatinX.
3. There is a possibility of finding no correlation between the LatinX demographic and the election results for both presidential and government elections.

Shifts in the Project Scope:

Since the election data does not go into detail about the demographic breakdown of votes, we cannot infer a direct cause and effect relationship between shifts in demographic data and shifts in election results. Thus, we decided to refine our project scope to encompass both the shifts in the overall LatinX population and shifts within each sub group and make inferences solely based on the data's correlations. However, finding a correlation between shifts in LatinX demographics and shifts in election results does not necessarily mean that LatinX voter preferences shifted in any way because correlation does not imply causation.

Our project focus is now to analyze whether or not there is a correlation between changes in the LatinX population and its demographics and the election results. For the data on the 6 cities that we are currently analyzing, the amount of data is not sufficient to form a substantiated inference. However, we plan to increase our sample size and generate data for more cities in Massachusetts in order to help us better infer a correlation between demographic changes and election results.