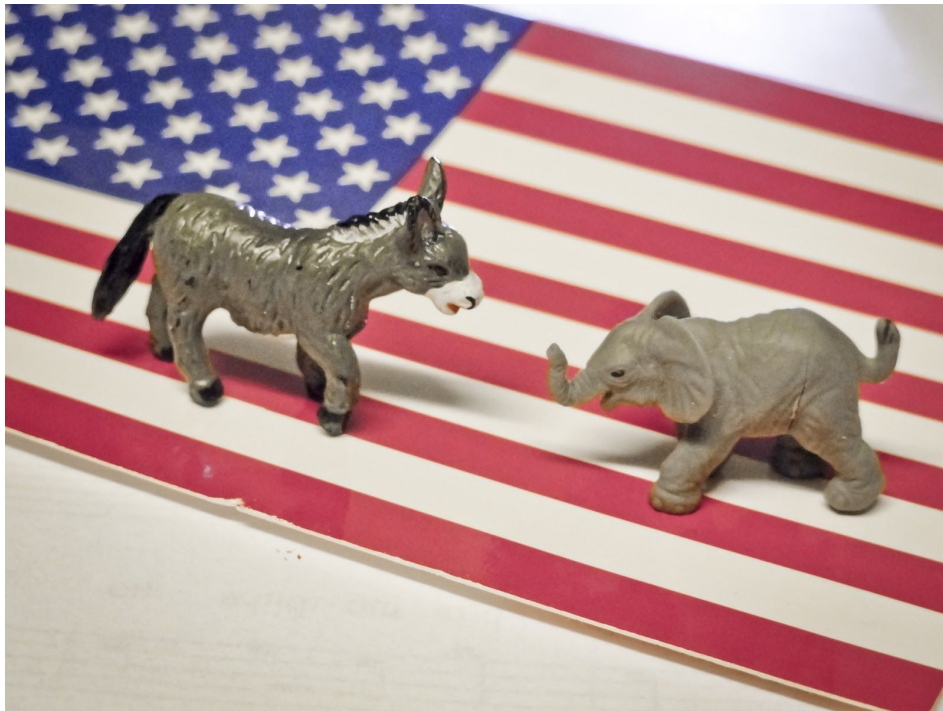


# Analyzing 2016 Presidential Election Results by Region in San Bernardino County

By Alyssa Keehan

Geog 176A

UCSB Fall 2020



## **Section 1: Project Summary**

For my final project, I used census block group data and election data to analyze the 2016 Presidential election results in San Bernardino County. My main motive for doing this project was trying to find a defining spatial difference between different residents in my area. It's always so daunting looking at election maps shown on TV because they rarely show large scale areas. I wanted to showcase my findings in such a way that it wasn't too hard for someone with little geographic knowledge, but also familiar with the San Bernardino County region to interpret. It is important for the media to create maps that spatially make sense. Since a majority of the people are getting their news from a specific source, it is that news source's job to provide the most accurate information. As you will see in the following technical report and conclusion, I have done my best to take as many different factors in as possible to produce the least biased graphs and plots I can. The majority of my exploration used a lot of statistical analyses, querying and plotting. I calculated majority numbers (i.e. majority party votes, majority sex, majority race, etc.) using binary classification and select by attribute to highlight those specific regions. After running my querying and plotting my results, I determined possible relationships between different demographics and their majority Presidential vote.

## **Section 2: Technical Report**

### **Problem**

For my research project and presentation, I performed spatial analysis on the San Bernardino County Region by comparing the relationship between different demographics and the majority presidential vote in several blocks throughout the area. San Bernardino County is California's largest with an area of over 20,000 square miles and a population of over 2 million. With a big area and population, San Bernardino is one of the most diverse regions in all of the state, which made it a good area to work with for this project.

Since this area was really diverse, it made it a perfect region to do my analysis on. This was an important project not only for me but my community because many counties and regions are ignored in the media when dealing with election results. Sometimes it is always good to zoom in and look at specific regions and their spatial heterogeneity within the area of interest. For this project, I got rid of the population that voted for any other party since there was no significant evidence that a certain group favored any third party candidate more.

## **Data Sets**

The two datasets I used for this project were both found in the election tab on the California Statewide Database website. The first dataset I used was the 2010 census block group shapefiles determined by political precinct. [https://statewidedatabase.org/d10/g16\\_geo\\_conv.html](https://statewidedatabase.org/d10/g16_geo_conv.html)

This file produced the shapes of each census block, each with different areas but similar populations. The other dataset I used was the 2016 general election results dataset for San Bernardino County.

<https://statewidedatabase.org/d10/g16.html>

For both of these datasets, I made sure to use the same type of Precinct (SRPEC) so I had a common primary/foreign key to join with. The first dataset was a shapefile which contained vector polygon data in the shape of each block group. My second dataset was a database file which contained all of the attributes I looked for in this project. It contained the racial, sex, and population election data for all block groups contained in the county.

## **Analysis**

I took several steps during my research process to get to my analysis. For the most part, I wanted to perform binary classification on each region using the different attributes and their values. Since all of these attributes had a value representing the total number of people of that certain population (i.e. men/women, Asian, etc.) I was easily able to perform my classification. First, I had to figure out a way to split the data for each demographic depending on their majority vote per block group. What I ended up

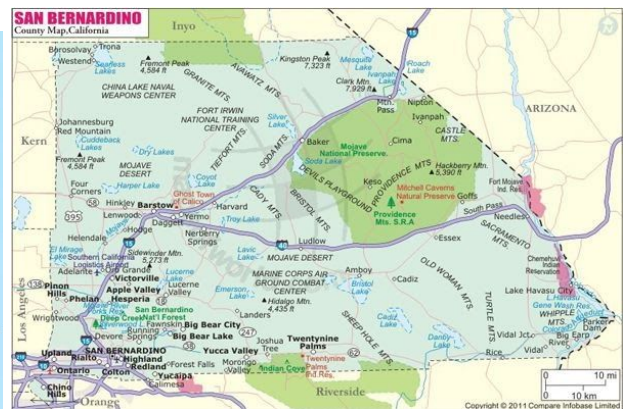
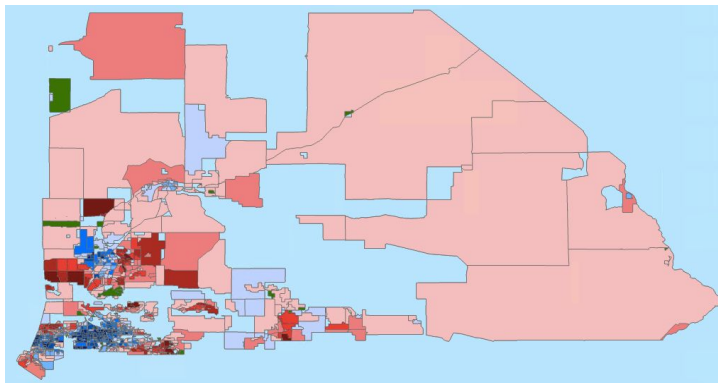
doing required an “outside of the box” process to classify different regions as majority Republican voters and majority Democratic voters.

The first layer I worked with was just the general population and the majority vote for each block. I first created a new attribute called `election_res` and it basically subtracted the number of democratic voters by the number of Republican voters in a block. After doing this, I was able to split the data between the positive values (majority Democratic) and the negative values (majority Republican) into different feature classes. Then I was able to map out my results using graduated colors and natural jenks classification for each of the majority party blocks. The natural jenks classification aided in showing which regions had a stronger favoring of a certain candidate.

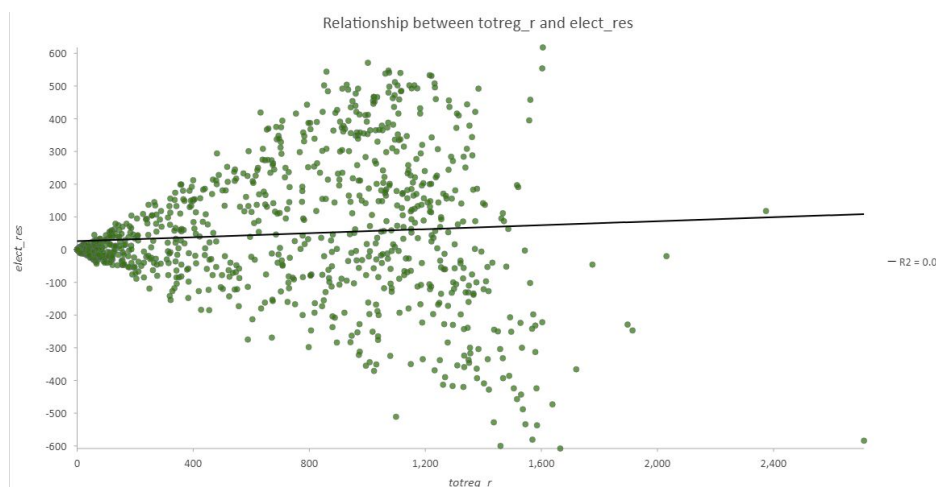
Using the same system to create binary classification, I calculated the block groups with the majority female population and those with a majority male population. Using select by attribute, I exported my features again by choosing the regions with majority female and majority democratic voters, then majority female and majority republican and so on and so forth with the majority male population. I was able to come up with a map and scatterplot graphical representation to show a possible relationship between sex and presidential vote in each block. For the map, I made the darker shade represent a majority male population and the lighter shades a majority female population.

For my last data layer, I looked at the Asian and Hispanic population and their Presidential vote. I did a similar process for this data layer like with the female and male data layer, but instead of both attributes on the same map (like I did with sex), I had a separate map for the Asian population and another for the Hispanic population. For both attributes (Hispanic and Asian results), I used natural jenks classification to show the extent of the majority vote in each region and produced scatterplots to show a relationship between the specific race and their majority vote for each block group.

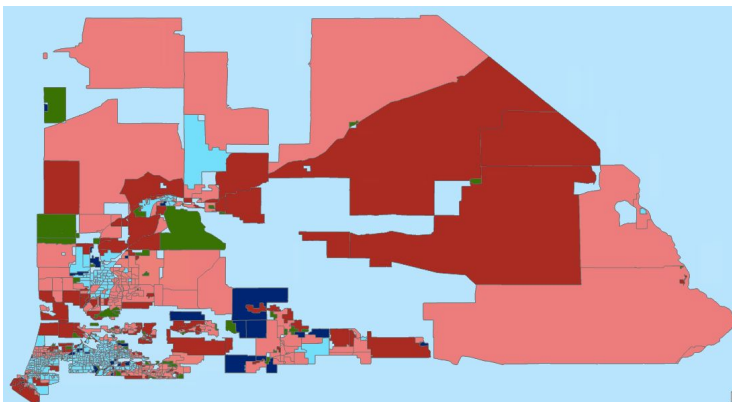
## Results



Above is the map result of the general population and the majority presidential election vote for each block group. As you can see, although the larger regions show a more majority Republican vote, the more densely populated cities, especially towards the southwest region of the map, show a large favoring towards the 2016 Democratic Presidential Candidate.

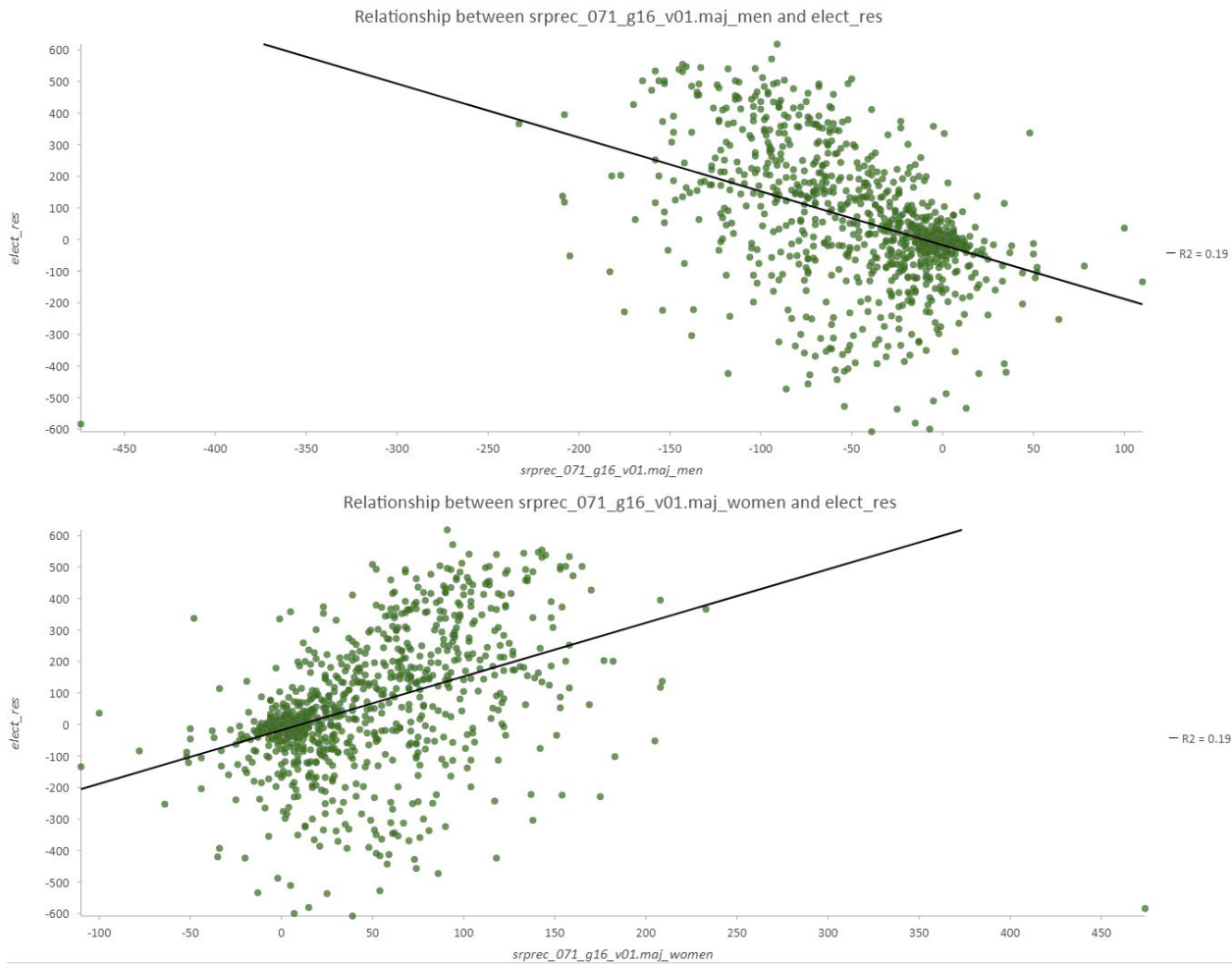


Above is the scatterplot comparing the total population of people who voted to the majority election result per block group. As you can see, there is really no linear relationship between the two variables. However, there is more variation along the election\_res axis when the number of people registered gets higher. This shows that with a bigger pool of voters, we're gonna see a bigger variety amongst the populations. When looking at the y-axis, we should think of a positive value being majority democratic and a negative value being majority republican vote.

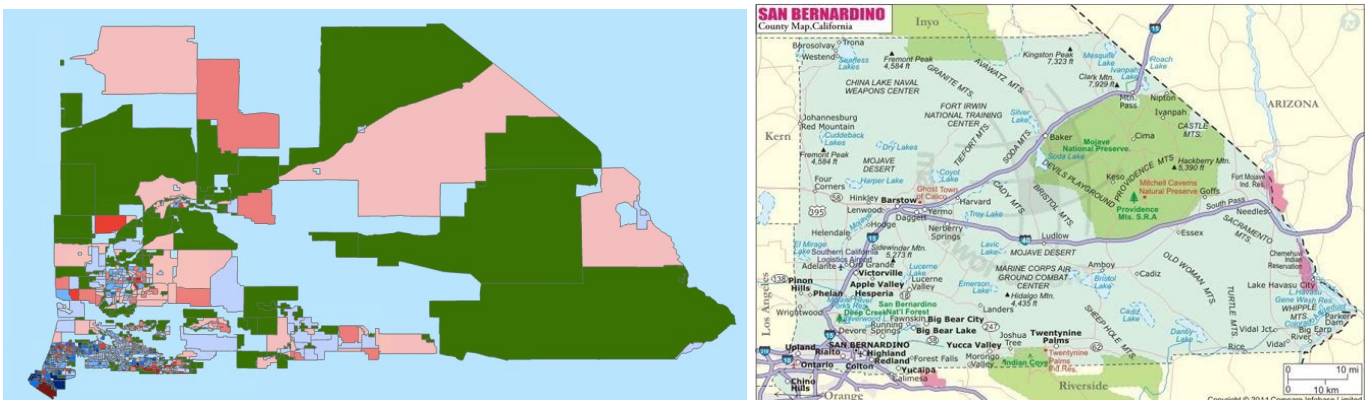


Above is the map result using the majority sex in each block group and their majority Presidential Election vote. As explained in my analysis, the lighter shades representing a majority women appear more frequently than majority male populations. While there are very few men compared to women who vote a majority Democrat, the distribution between majority men and women for the majority republican vote is pretty much the same.

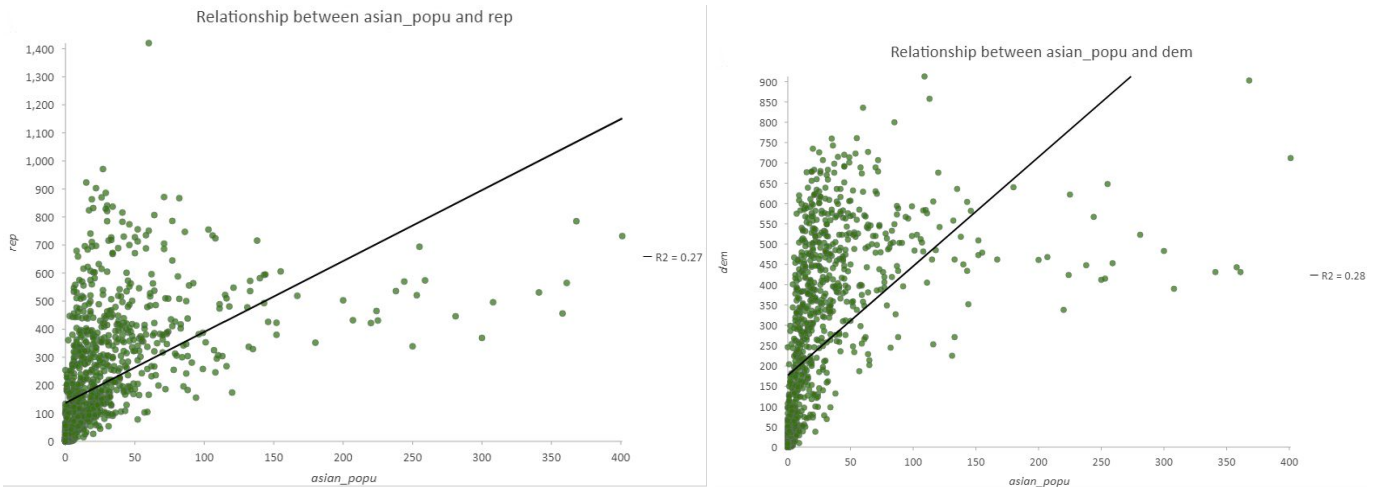




Above are the graphs representing the majority sex (one for majority women and the other for majority men) and their relationship with the election results. When looking at the first graph, there seems to be a negative relationship between the population of males and the election result meaning the more men in a population, the more likely that population will vote more republican. Conversely, with a positively linear relationship, the more women in a population, the more likely that region will have a majority Democratic vote.



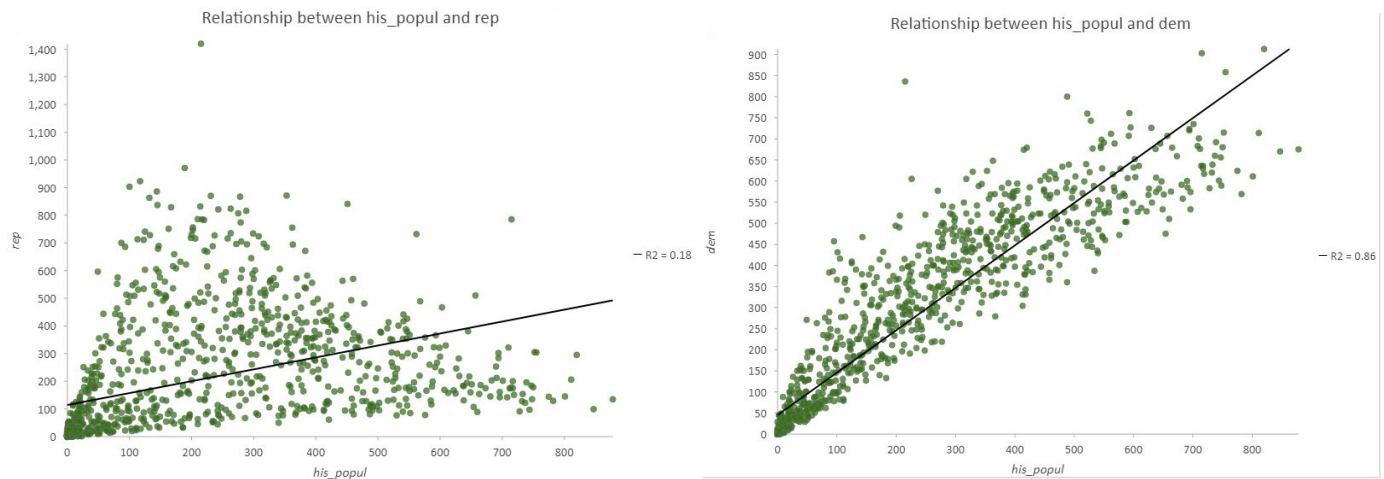
Above is the map representation of the Asian population and their majority election vote. As you can see with the neutral green value in some regions, There isn't much significant evidence of a drastic difference between political parties. Similarly with the general population map, you can see more blue in the densely populated regions while red still lingers in the larger blocks.



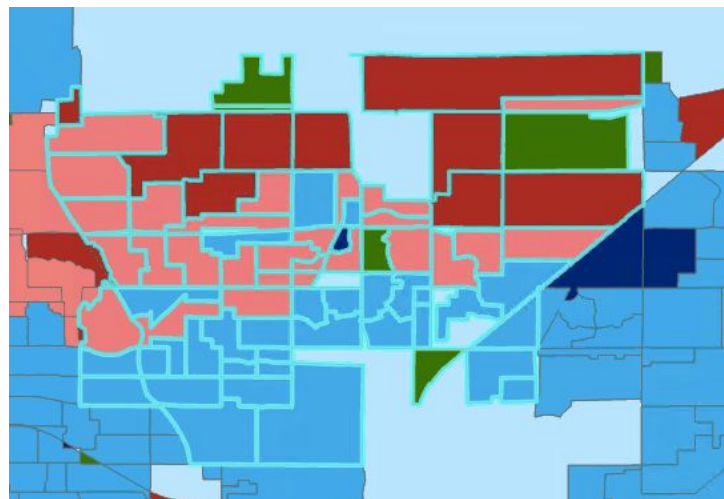
The scatter plots above represent the relationship between the Asian population and the number of democratic and republican voters in their region. Although both graphs have a positive linear relationship between the x and y variables, since they both have around the same R squared value, there is no significant difference between their results. So we say that the total Asian population is independent of the Election result. Looking at the small details of the plot, we can see that there is a large cluster towards the left hand side of both graphs. This means that a majority of blocks in our dataset have a low population of Asians.



Above is the map representation of the Hispanic population and their election results. There is a staggering difference between the Democratic and Republican voters in the Hispanic Population. There are significantly more regions in San Bernardino County that are both Hispanic and vote for the Democratic party.



Above is the result of plotting the Hispanic population against the number of Democratic and Republican voters in a block region. As you can see in the left plot, although there is a positive relationship between the Hispanic population and the number of Republican voters, there isn't a large correlation. However, if we look at the right plot, there is a strong linear relationship between the Hispanic population and the Democratic voters in a region. Although there is still a small cluster in the bottom left corner of the plots, the graphs are both very different and an obvious conclusion can be made about the relationship between the Hispanic population and the number of Democratic voters in a region.



Above is a close up of my hometown, Rancho Cucamonga (selected in teal outline). This screenshot was taken from the gender graph so we can see how a majority of our population is women. However, you can see a difference in political parties the more north you go. The map coverage stops because of the Mountains, but you can see that the closer to the mountains, the more the region is majority Republican. Since I am familiar with the area, I know that in those regions, there are nicer, bigger and more expensive houses. Based on this evidence, you can see some spatial heterogeneity in terms of economic affluence in my town.

Below is the link to my folder containing my datasets

<https://drive.google.com/drive/folders/1x2xgX6tO0S8qsb4e5U8PiR1xDMrDohuy?usp=sharing>

Below is the link to my ARCGIS project

<https://drive.google.com/file/d/1COdaPyuVkDpj3P1fhVJ-9wHktYH2MYcR/view?usp=sharing>



### Section 3: Conclusion

There are many things I would like to potentially do involving the work I've already done with my project. This is a good stepping stone for more political spatial analyses in the future. One thing I would hope to add or develop more in the future regarding this project is using more data. In addition to adding more data, I would like to extend my processes to explore the 2020 Presidential Election data or possibly even look at smaller, local elections.

One thing I can add to improve the work presented above is make sure the maps are more well organized. Since my shapefiles were randomly placed in a region close to the equator instead of in the South East region of Southern California, It made it hard for my work to have a nice layout. I can improve in the visual aspect by spending more time fixing that projection problem.

One challenge I faced while working on this project was the inability for my data to join together. I wanted to compare the election results of each block group with different demographics (i.e. Median Household Income, average Education level, unemployment rate, etc.), but the data wasn't working well with joining. Since I got my election data, and some of my demographic data from different sources, it was impossible to find a proper key to link the data together.

This project was a great experience for me because I learned about more than just how to make a map and some graphs. I learned that most processes aren't always straight forwards and the best approaches are sometimes the least expected. For example, my binary classification process was outside of the box and if it were never mentioned to me, I would've never thought of it. It became a huge part of my whole project. Another thing I learned from this project is that it is always good to gain more knowledge in the domain you're working with. At first, the data was a bit confusing and I didn't know where to start. After reading the metadata and several different websites, I was sure the data I retrieved was what I wanted to show in the end.

If I had to do this project again, I wouldn't necessarily change my project, but more my workflow. I wish I got all my data from the Census website instead of the California Statewide Database. Although the Statewide Database was a lot easier to work with, the Census had the data I really wanted to work with. In addition, I would've dived in a little bit more with the analysis. Adding the data about income and education would've made an even more compelling project.

## Section 4: Career-Readiness Component

- In one sentence (25 words or less), describe this course to your friend who might be interested in taking the course.

Geog 176A is a class where you can apply what you already know about geography and develop your skills to become a better spatial analyst.

- In one sentence (25 words or less), describe how what you learned in this course could contribute to the society or community in which you are involved.

With the material I learned in this class, I can create visualizations that can accompany my explanations in future data science projects.

- In one sentence (25 words or less), summarize how what you learned in this project could apply to your future career in research or GIS industry.

This project helped me show that I have the skills to perform a compelling spatial analysis on demographic data for a large piece of land.

- If you were to interview for a GIS position, describe how would you articulate your readiness and qualifications for that position (in one or two sentences, less than 100 words)

I believe I would be a great addition to this company because I have project experience working with spatial data and ArcGIS pro. My additional data science background has allowed me to develop the skills needed to become an outstanding spatial analyst and fully prepared me for this position.