Important PA Election Features

To determine the factors that most closely relate to voting factors, the following steps will be taken

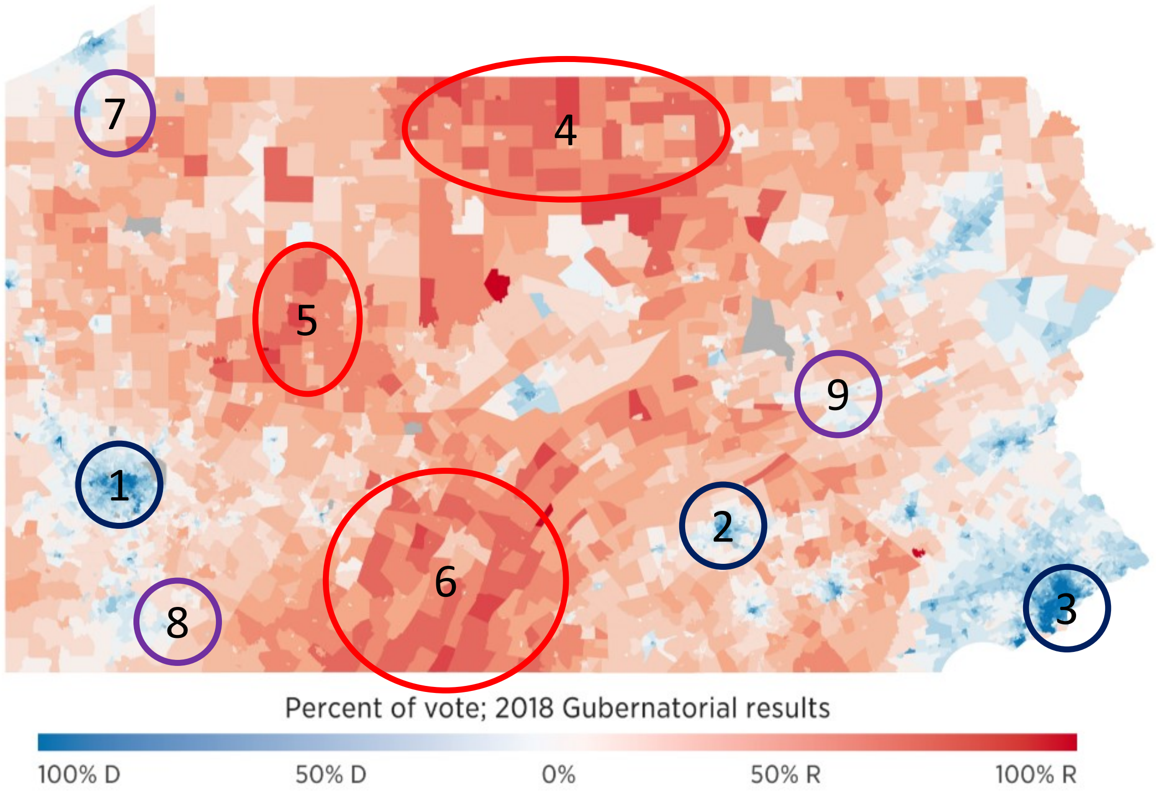
1. Preparation
2. Data Extraction
3. Data Normalization
4. Multivariate Comparisons

1. Data Preparation

First, we need to set the geoprocessing environments within our map. To do so, we set the processing extent and cell size as to those that are the same as traffic. We also set traffic as a mask.

Next, utilize the map provided to find signal / comparisons areas between voting results and variables. The following signals are chosen:

- 3 Blue (Heavily Democratic)
- 3 Red (Heavily Republican)
- 3 Purple (Largely Split)



2. Data Extractions

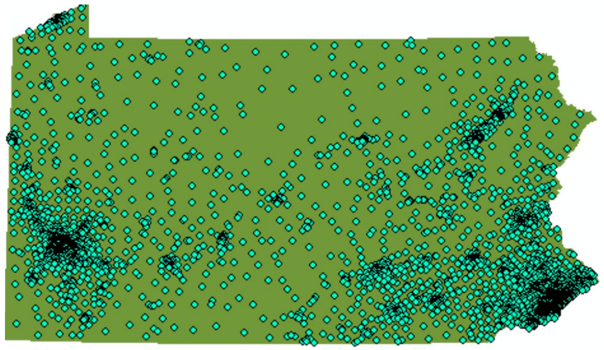
As we only have data collected at points, we need to be able to generalize them to create layers representing the entire region Pennsylvania. The following provide a summary of the operations that will be conducted for each variable.

	Variable	Source	Type	Operation	Additional Parameters
a	Population	CensusPoints	Density	Kernel Density	-
b	% White Population	CensusPoints	Interpolation	Natural Neighbors	
c	% Older than 65	CensusPoints	Interpolation	Natural Neighbors	
d	Birth Rate	CountyPoints	Interpolation	Spline	-
e	Number of people / dentist	CountyPoints	Interpolation	Spline	-
f	Income	PostalPoints	Interpolation	Natural Neighbors	
g	Golf Course Location	GolfCourses	Density	Kernel Density	Search Radius: 3 Area Units: miles^2

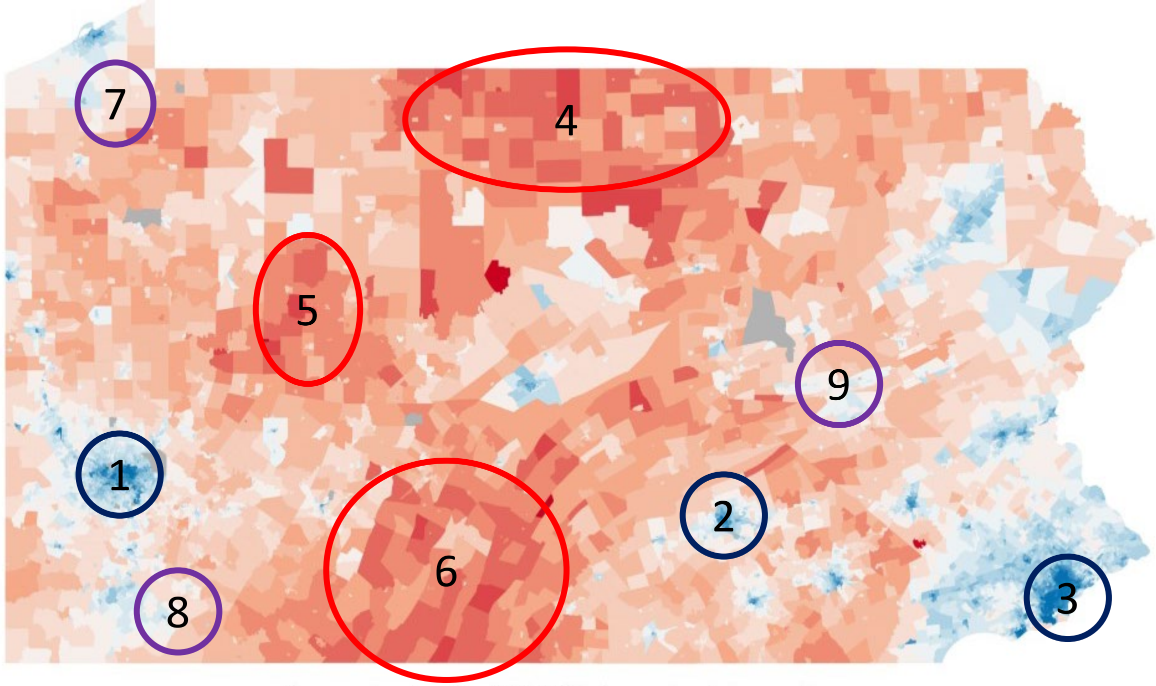
Afterwards, each data will be compared to the original voter graph to see if similar patterns occur. A stretched symbology will be used to display each variable, as a stretched symbology was used for the voter map.

a: Population

Source: CensusPoints
Type: Density
Operation: Kernel Density



CensusPoints



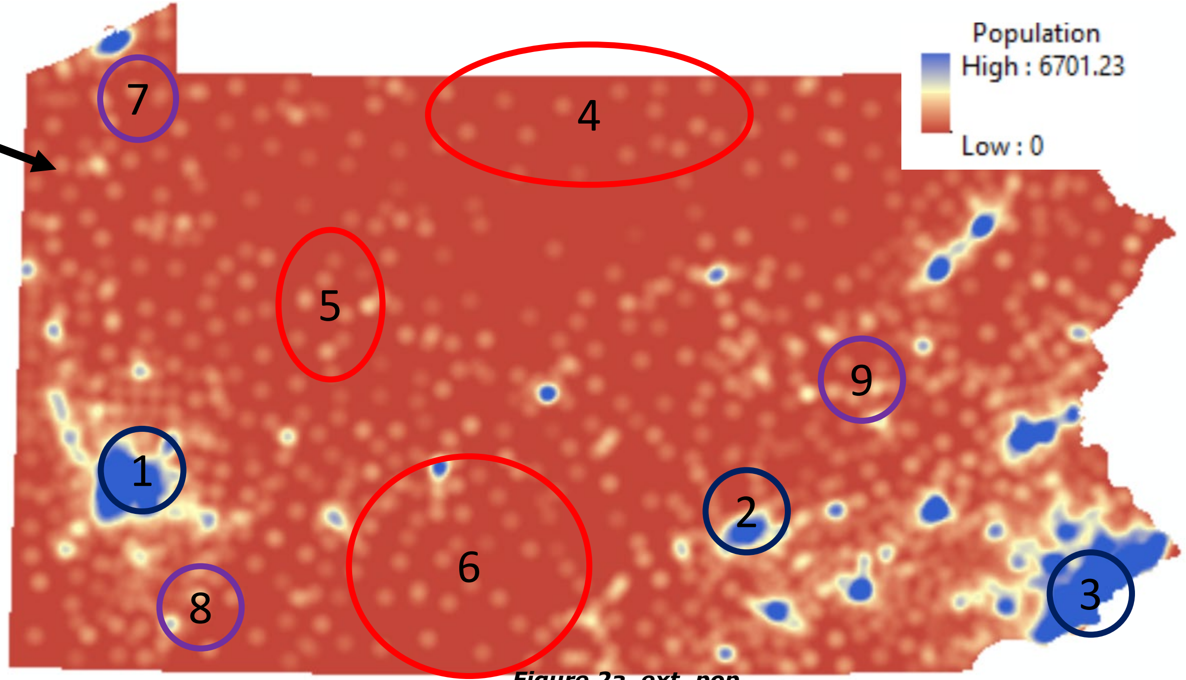


Figure 2a. ext_pop

Description: density of population in Pennsylvania
Source: Kernel Density("Population")

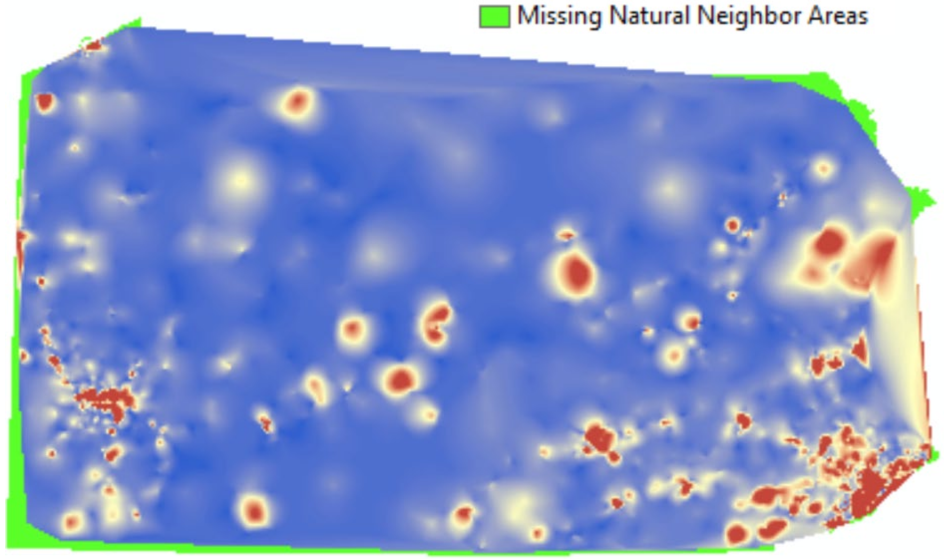
Population is a good indicator of voting preferences. Of the 3 democratic clusters, all three have high population density. Of the republic clusters, all three have low population density. Of the 3 neutral clusters 7 has a population density, while 8 and 9 have low-medium population density. Regardless, there seems to be a positive correlation between population density and democratic voting and a negative correlation between population density and republican voting.

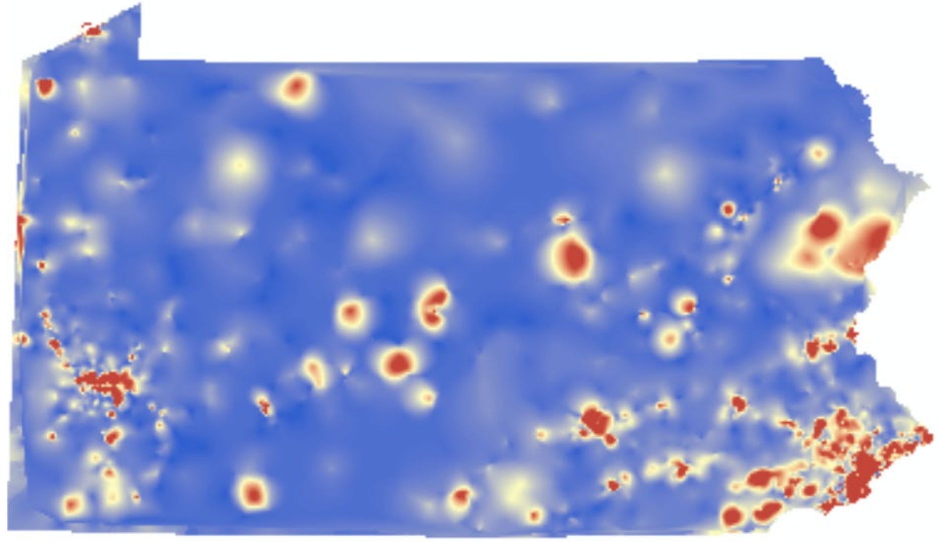
2. Data Extractions

For % white, % Old, and Income, Natural Neighbors was selected due to the irregular spatial distribution of the census and postal points. This is to reduce bias towards areas further from data points. However, Natural Neighbors, unlike other interpolation methods, even with setting geoprocessing environments, does not interpolate outside of its point boundaries. Thus, the natural neighbor layer has some missing values. To fill these missing values, IDW was calculated for its variable as well. Then, the two layers were combined into a single layer, with IDW values being used to fill only where natural neighbor values were missing.

Example: Con(IsNull("nat_white"), "idw_white", "nat_white")

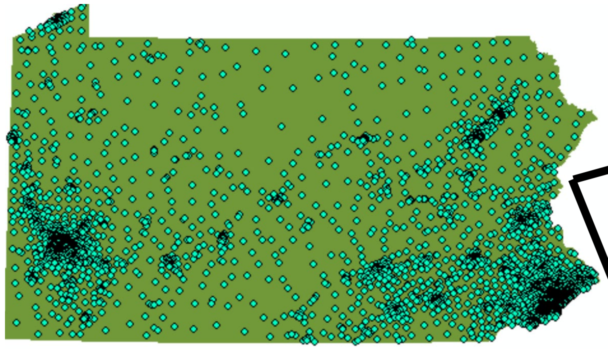
This results in the majority (~98%) using the natural neighbor distribution and only ~2% of boundary, outlying pixels using IDW. This solves both Natural Neighbor's lack of values in boundaries and IDW's spatial bias. Alternatively, Euclidian Allocation could have also been used by setting pixels missing Natural Neighbor values to the closest natural neighbor value



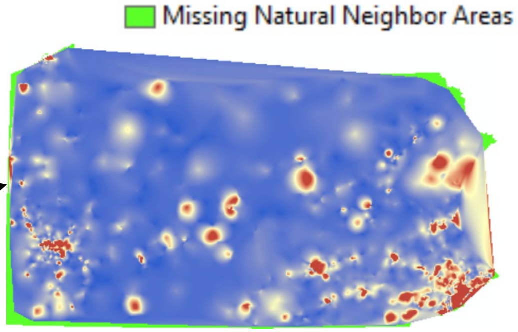


b: % white

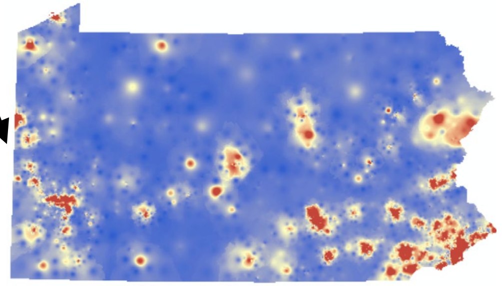
Source: CensusPoints
Type: Interpolation
Operation: Natural Neighbors (with IDW filling boundary)



CensusPoints

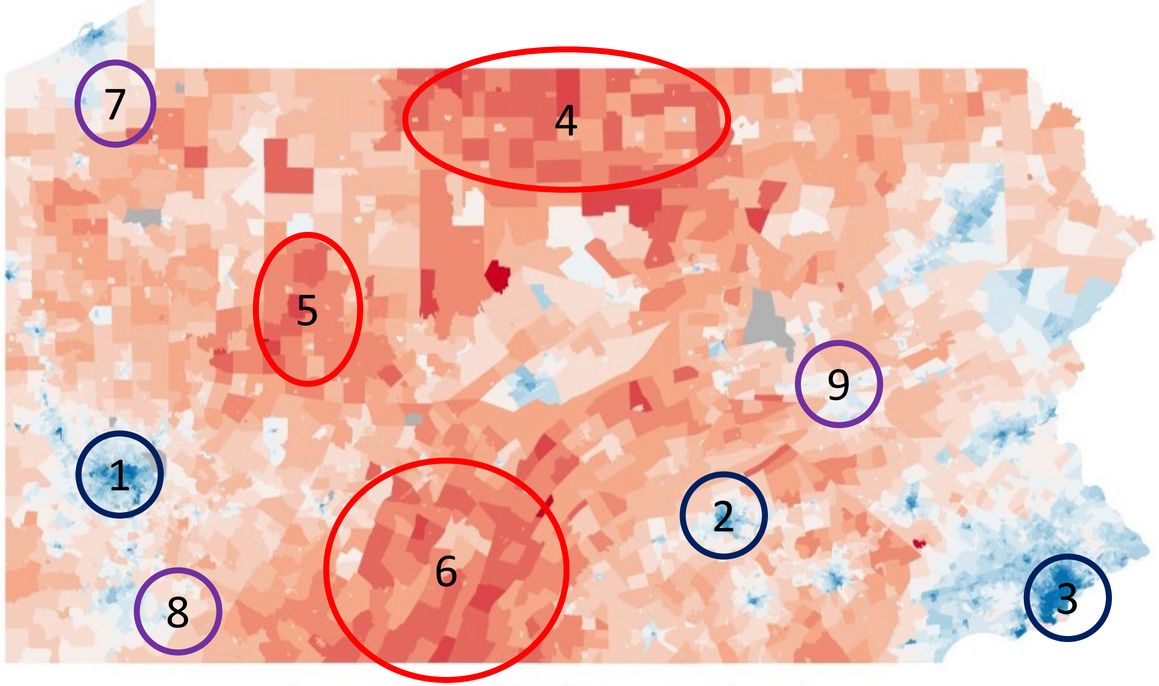


Natural Neighbors: nat_white



IDW: idw_white

% White is a good indicator of voting preferences. Of the 3 democratic clusters, all three have low white percentages. Of the republic clusters, all three have high white percentages. Of the 3 neutral clusters 7 and 8 have high-medium % white density, while 9 has a mixed % white. Regardless, there seems to be a negative correlation between % white and Democrat and positive correlation between % white and Republican.



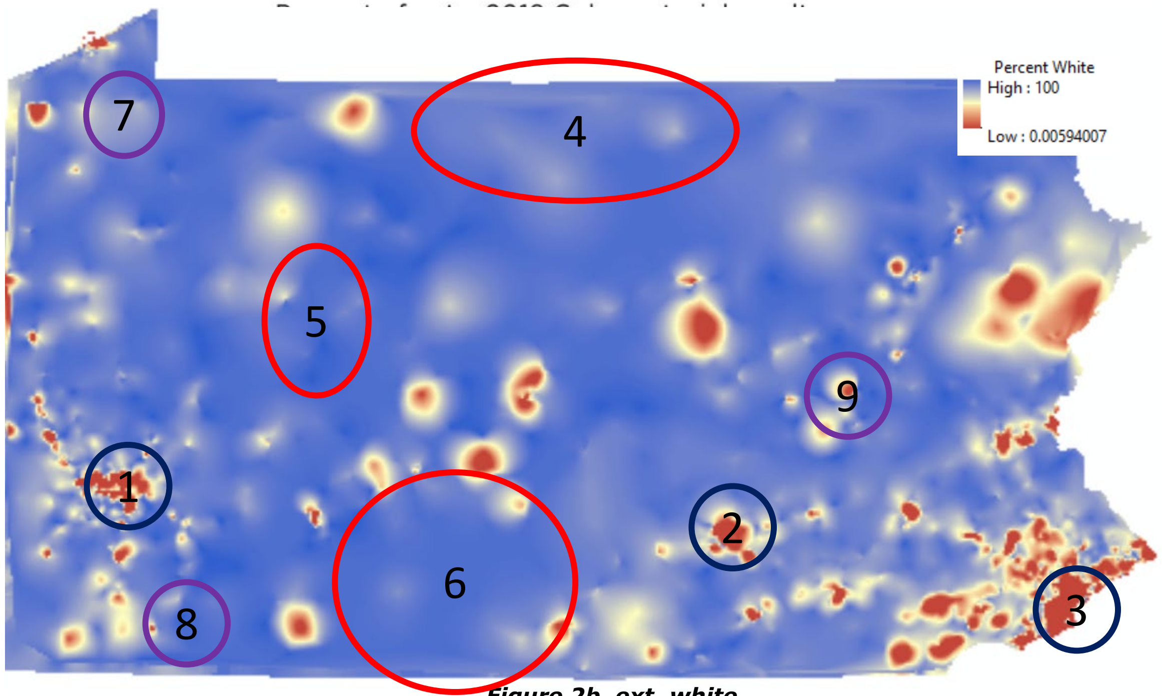
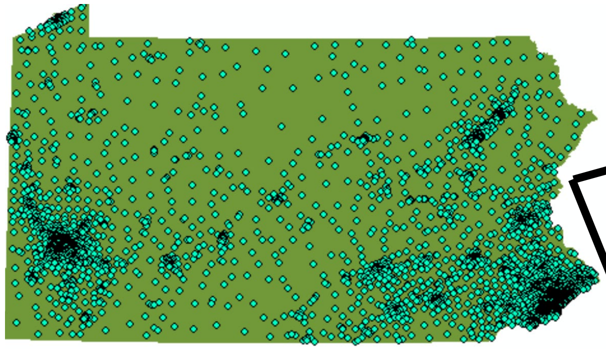


Figure 2b. ext_white

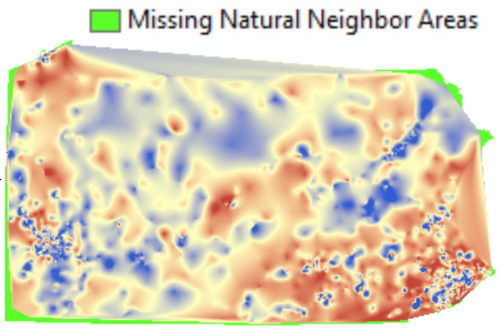
Description: natural neighbor interpolation of white in Pennsylvania (with Nat Neighbor missing values using IDW

c: % old

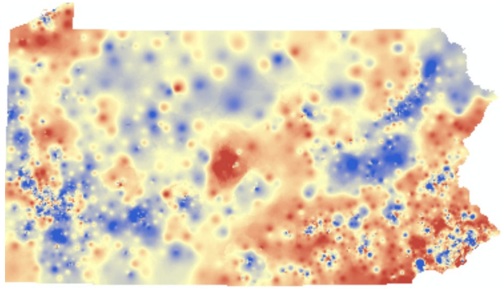
Source: CensusPoints
Type: Interpolation
Operation: Natural Neighbors (with IDW filling boundary)



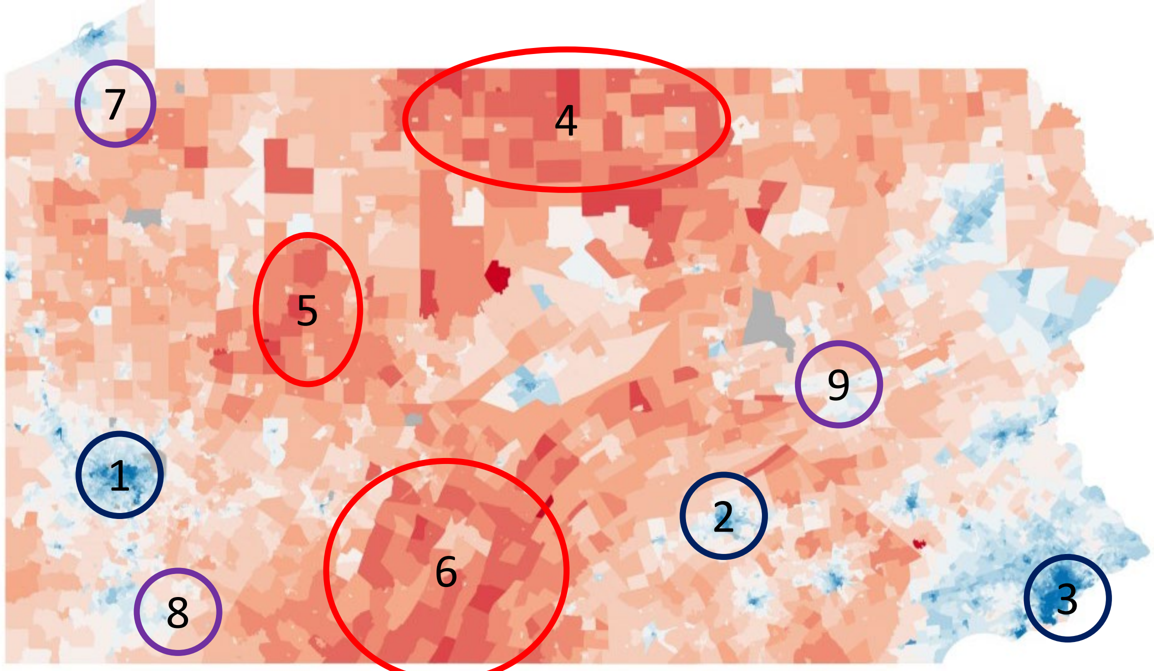
CensusPoints



Natural Neighbors: nat_old



IDW: idw_old



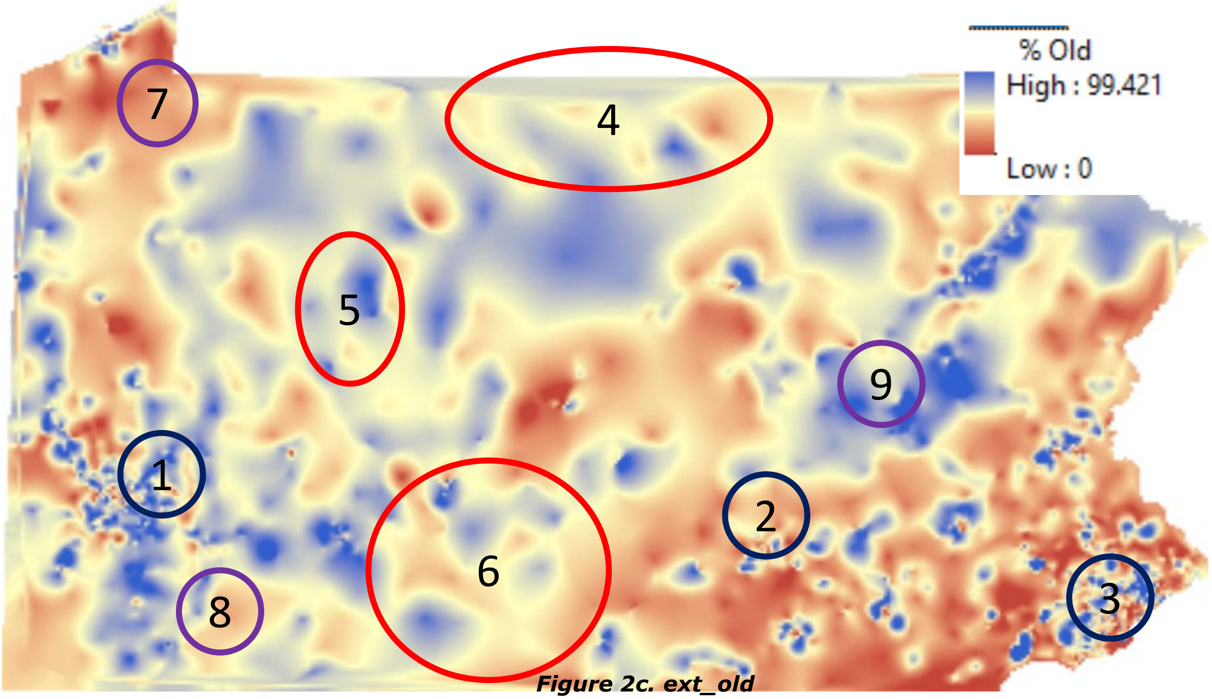


Figure 2c. ext_old

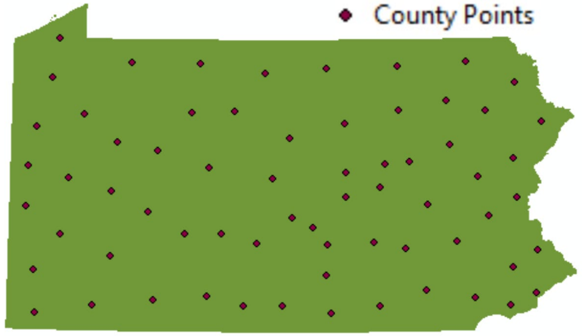
Description: natural neighbor interpolation of ld in Pennsylvania (with Nat Neighbor missing values using IDW

% Old is a poor indicator of voting preferences. Of the republic clusters, all three have higher rates of older populations, with some low rate clusters showing up. Of the 3 neutral and democratic cluster, no apparent pattern seems to exist, since 2 and 7 has mostly young, 3 and 8 is mixed, and 1 and 9 is mostly old.

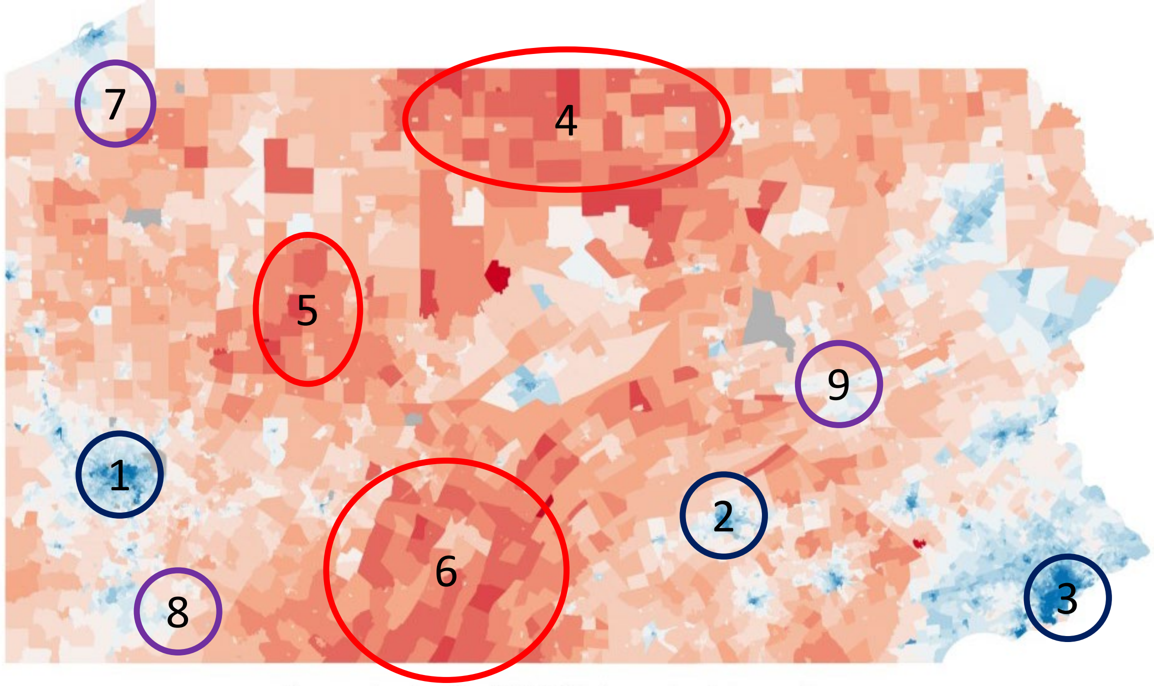
d: Birth Rate

Source: County Points
Type: Interpolation
Operation: Spline.

Spline was chosen since the original data points, county points, are displayed over a wide region and we want to interpolate over a vast region without biasing points closer to county points.



CountyPoints



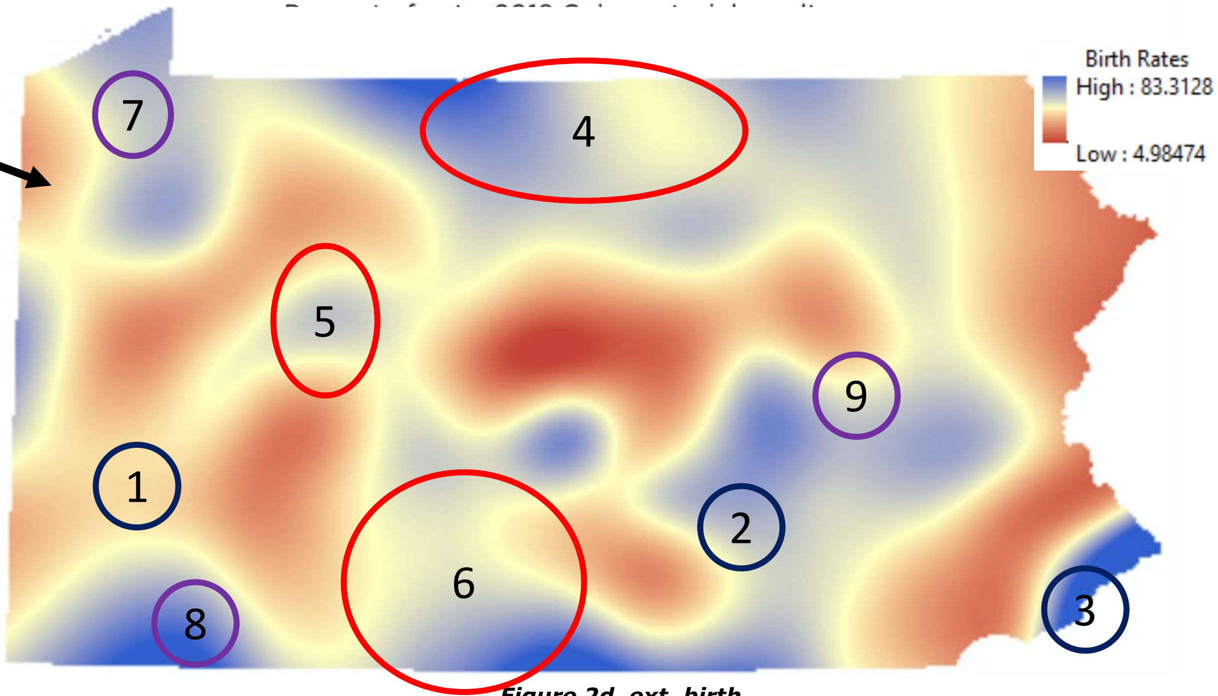


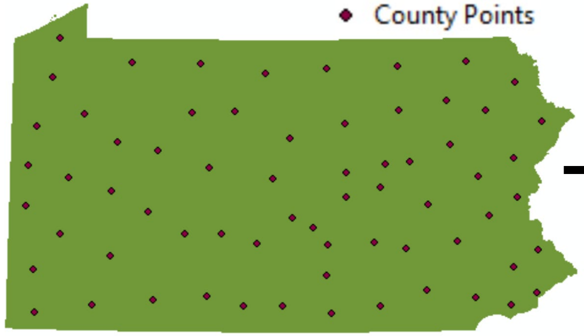
Figure 2d. ext_birth
Description: interpolated birth rates in Pennsylvania
Source: Spline("CountyPoints"), Z value field: BirthRate

Birth rate is a poor indicator of voting preference. Of the 3 democratic clusters, 1 is has low birth rate, 2 has moderate, and 3 has high. Since all three clusters are different, birth rate would fail to identify democratic clusters. Of the republic clusters, 4,5,6 all have moderate/high rates. Of the 3 neutral clusters, 7 and 9 have moderate/high rates and 8 has higher rate. With both republican and neutral clusters having moderate/high birth rates, it would be difficult to differentiate between the two and create a gradient correlating birth rate to leaning democratic/republican.

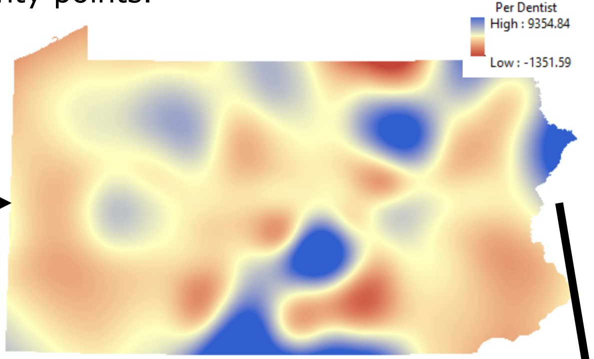
e: Dentists

Source: County Points
Type: Interpolation
Operation: Spline.

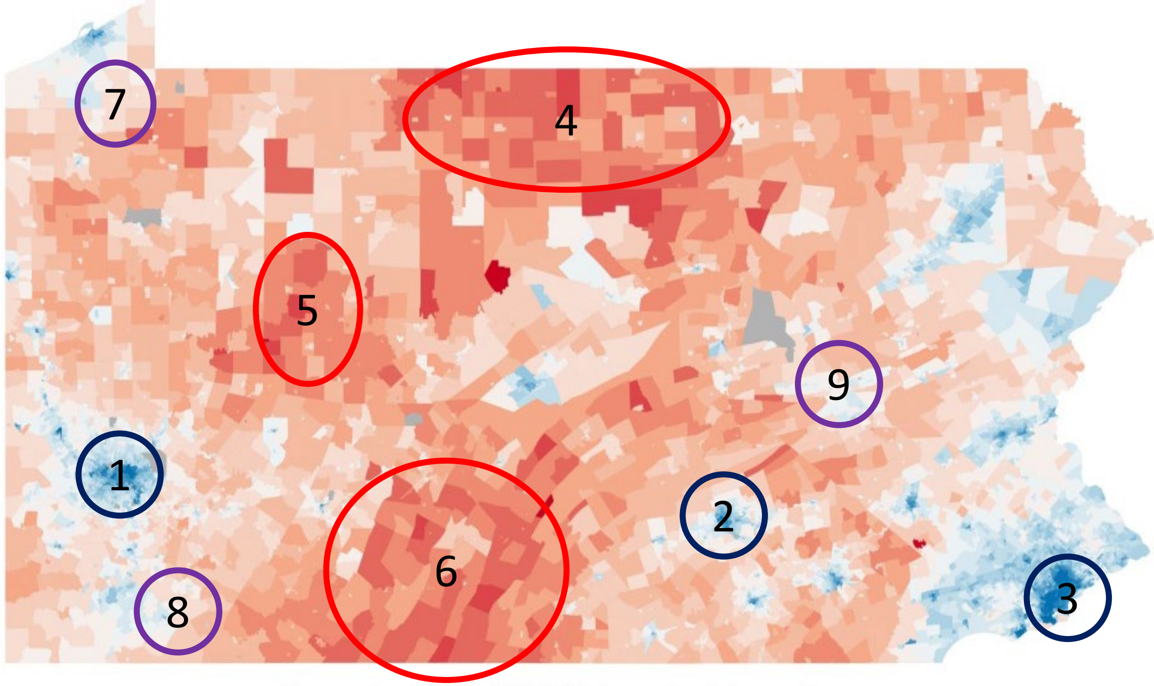
Spline was chosen since the original data points, county points, are displayed over a wide region and we want to interpolate over a vast region without biasing points closer to county points.



CountyPoints



dist_dentist
Description: interpolated per dentists rates Pennsylvania
Source: Spline("CountyPoints"), Z value field: PerDentist



It does not make sense to have negative dentist per person, so after using Spline interpolation, I set the per dentist rate to 0 for all values lower than 0.

Per dentist is a poor indicator of voting preference. Of the democratic clusters, all three clusters have low dentist rates. Of the republic clusters, 4 has a high-medium per dentist, 5 has a low-medium, and 6 has both extremely high and extremely low. Of the 3 neutral clusters, all 7 and 8 have medium low rates while 9 has medium high rates. There seems to be no positive or negative correlation between party voting and per dentist rates..

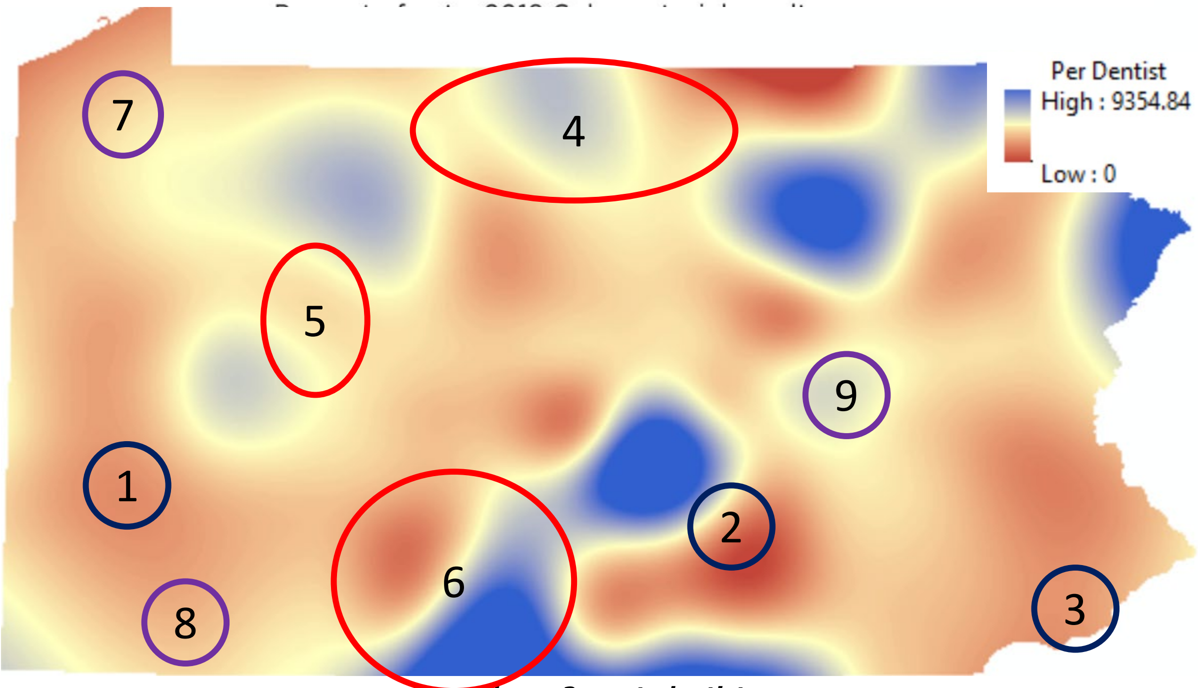
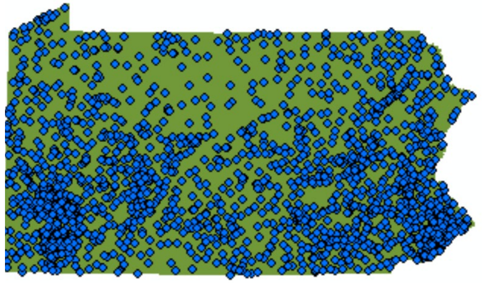


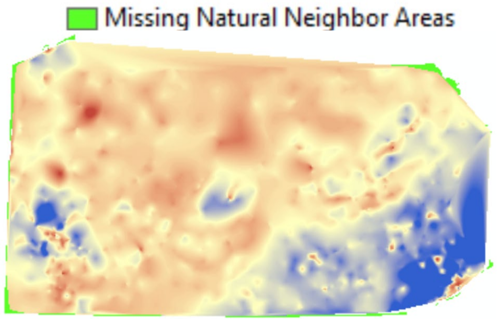
Figure 2e. ext_dentist
Description: interpolated dentists Pennsylvania, minimum of 0
Source: Raste Calculator("dist_dentist" < 0, 0, "dist_dentist")

f: income

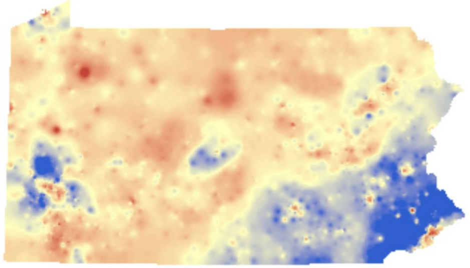
Source: PostalPoints
Type: Interpolation
Operation: Natural Neighbors (with IDW filling boundary)



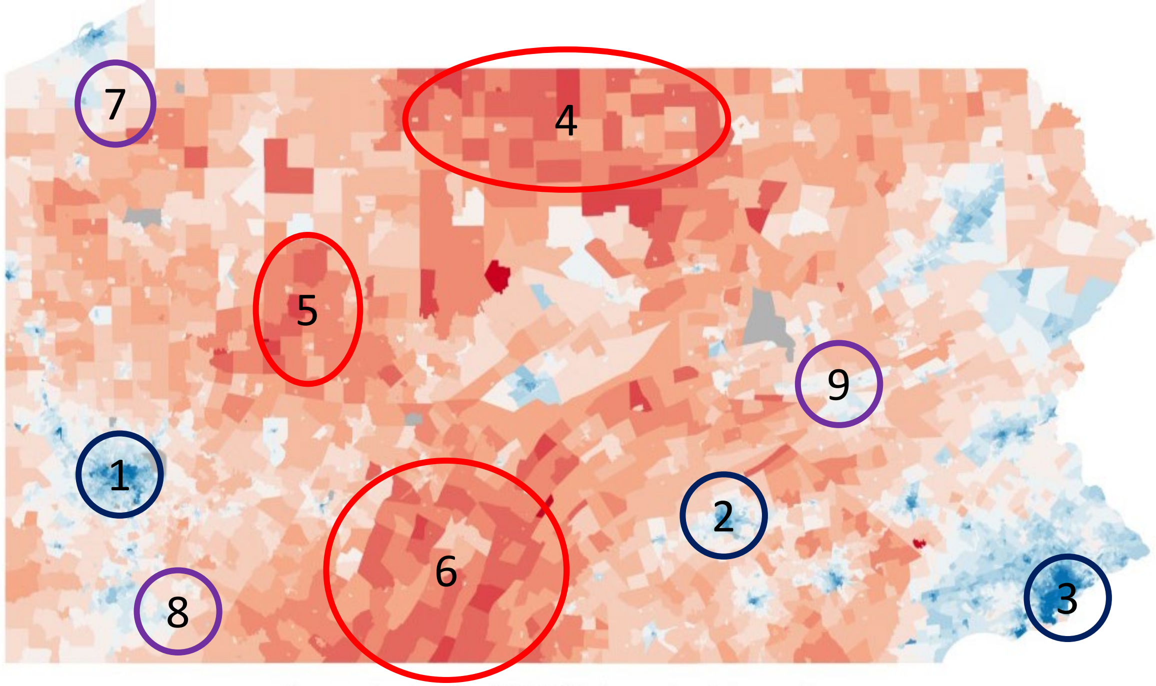
PostalPoints



Natural Neighbors: nat_income



IDW: idw_income



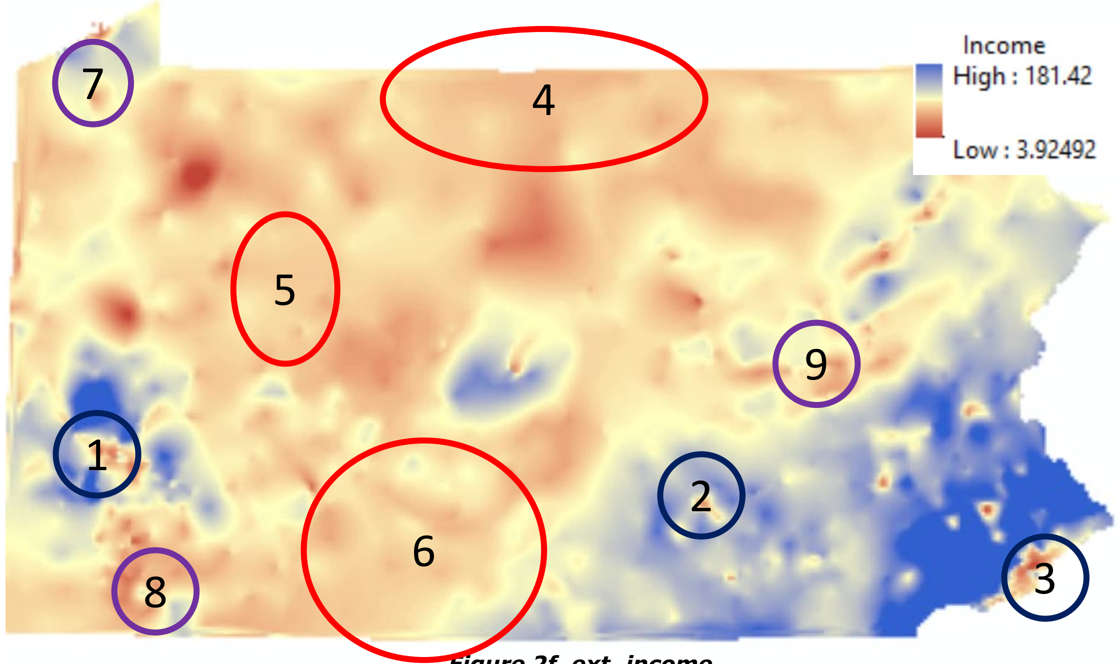


Figure 2f. ext_income

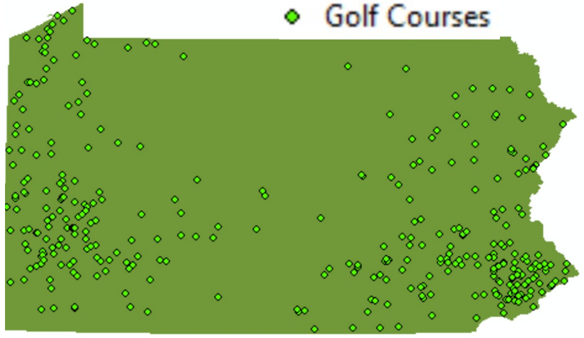
Description: natural neighbor interpolation of income in Pennsylvania (with Nat Neighbor missing values using IDW

Income is an ok indicator of voting preferences. Of the 3 democratic clusters, all three have a low income cluster surrounded by a high income. That is, areas that strongly vote democratic are very poor, but areas that moderately vote democrat are very rich. Of the republic clusters, all three have low income. Of the 3 neutral clusters, all three have mixed income patches that span multiple levels. For each voting preference cluster, income patterns are apparent, but overlap with one another.

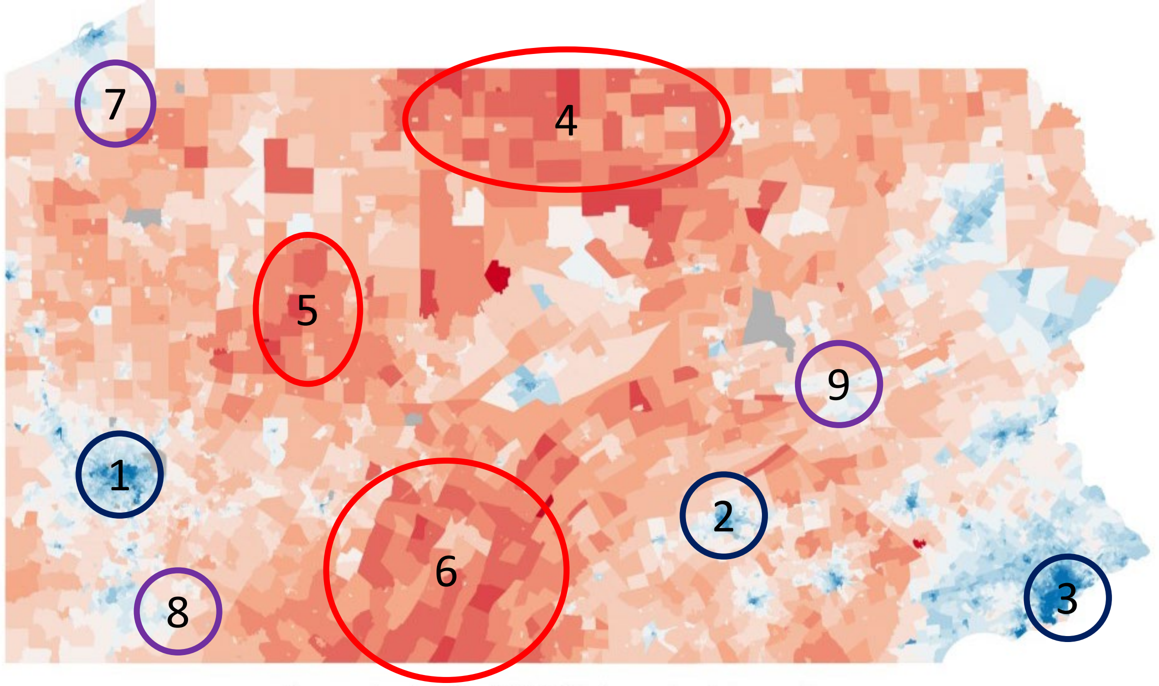
g: Golf Courses

Source: GolfCourses
Type: Density
Operation: Kernel Density
Additional Parameters: Search Radius (30000), Units (m^2)

A search radius of 30000 m^2 is approximately 20 miles. I am assuming individuals are not willing to drive more than 20 miles to play golf



GolfCourses



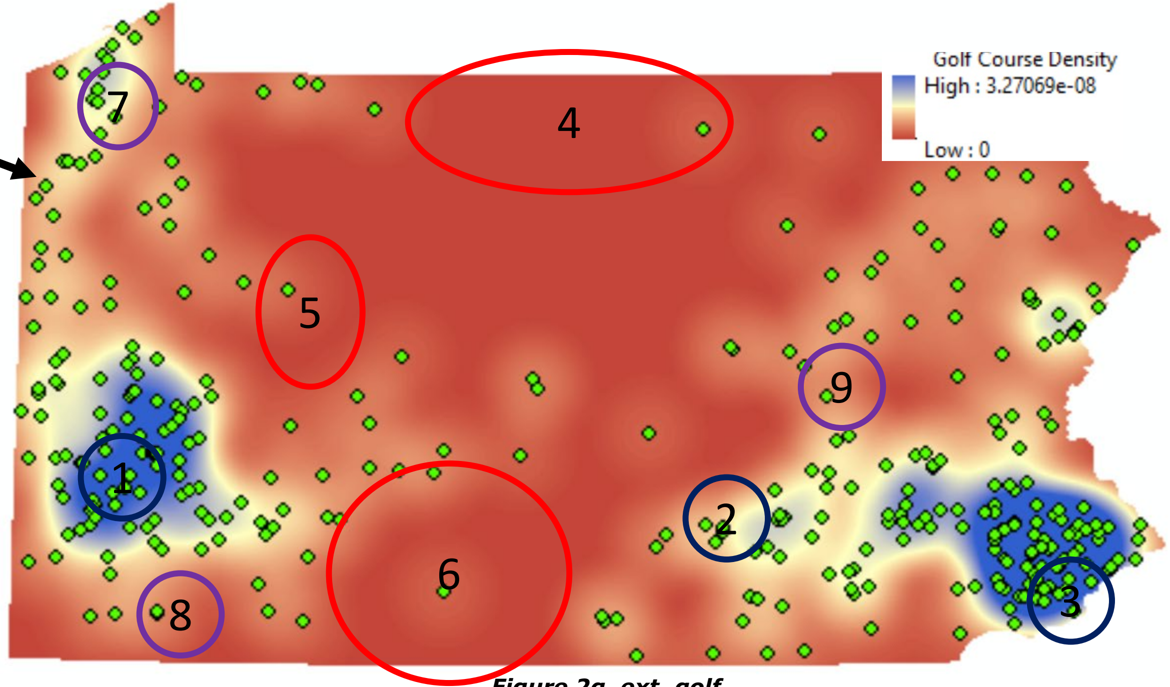


Figure 2g. ext_golf

Description: density of golf courses in Pennsylvania
Source: Kernel Density("GolfCourses"), Search Radius = 30000, Units = m^2

Golf Course Density is a good indicator of voting preferences. Of the 3 democratic clusters, all three have high golf course density. Of the republic clusters, all three have low golf course density. Of the 3 neutral clusters 7 has a moderate golf course density, while 8 and 9 have low golf course density. Regardless, there seems to be a positive correlation between golf course density and democratic voting and a negative correlation between golf course density and republican voting.

Variables: Summary

	Variable	Democrat	Republican	Assessment Potential	Max Value
a	Population	Positive	Negative	Good	6701.23
b	% White Population	Negative	Positive	Good	100
c	% Older than 65	No Correlation	Positive	Poor	100
d	Birth Rate	No Correlation	Positive	Poor	83.3128
e	Number of people / dentist	Negative	No Correlation	Poor	9354.84
f	Income	Negative Center, Positive Outer	Negative	Ok	180.337
g	Golf Course Density	Positive	Negative	Good	3.27 ^ 10 -8

It seems as if population, % white, and golf courses are the three variables that are the best at independently predicting voter preference. However, income is also seem to be a variable we should consider.

Thus, we will see of the four, which three, in combination, is the best for predicting. We will run ISO Clusters with the following combinations

- Combination 1: a, b, g: population, white, golf course
- Combination 2: a, b, f: population, white, income
- Combination 3: a, g, f: population, golf, income
- Combination 4: b, g, f: white, golf, income

Step 3: Data Normalization

We want to run classifications with our raster layers to see which combination of three variables is best at predicting voter preference. Before we do so, we need want to normalize the variables ranges so that each variable holds equal weight when we run classifications. Within our classification, we want to have a scale from 0 to 1, with 0 being most republican and 1 being most democrat.

Case 1: For population, income, and golf courses, since these variables are positively correlated with democrat and negative with republican, we will simply normalize the range by dividing each pixel's value by the max value of the variable. We will apply the following function: value / (max+1).

Case 2: For % white, since this variable is negatively correlated with democrat and positive with republicans. We will apply the following function: (100 – value)/100. We are essentially counting non-white populations.

Both formula will preserve the proportions of values, but have 1 representing close to democrat and 0 as close to republicans.

Step 3: Data Normalization

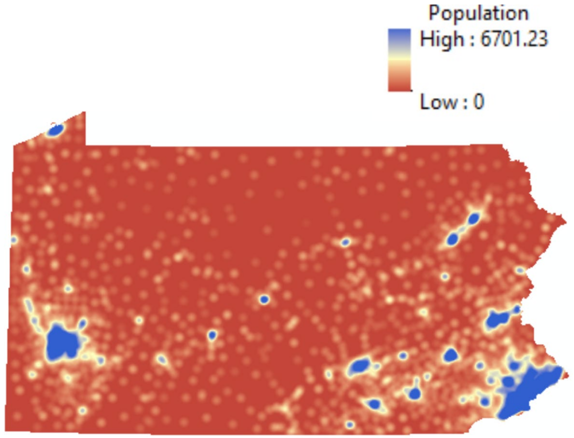


Figure 2a. ext_pop

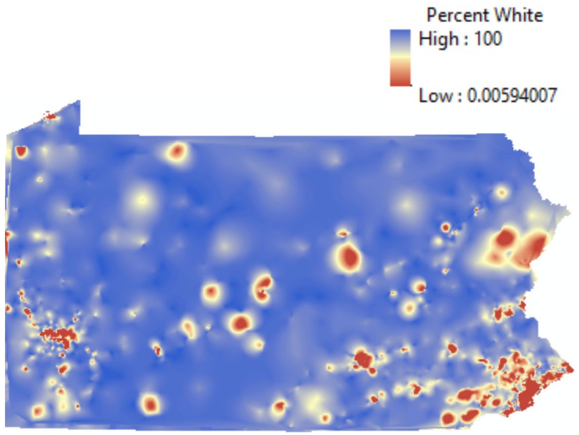


Figure 2b. ext_white

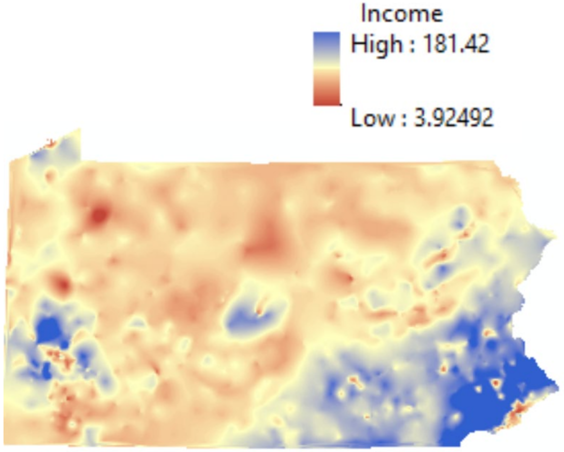


Figure 2f. ext_income

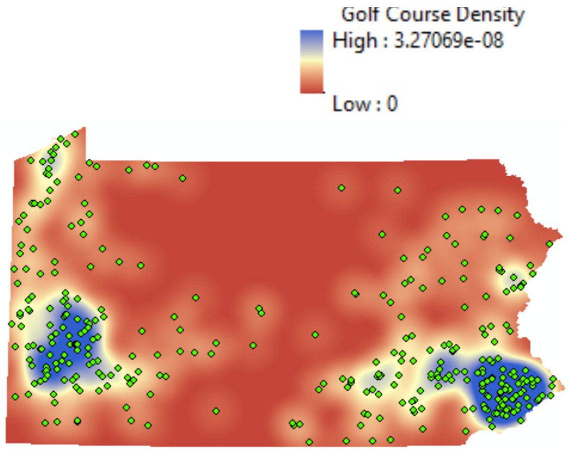


Figure 2g. ext_golf

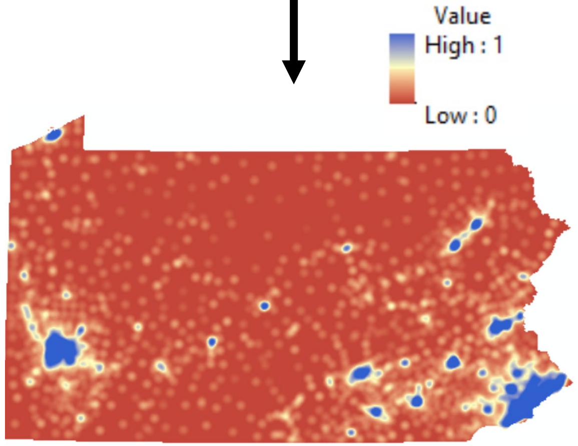


Figure 3a. norm_pop
Description: population scaled to 1
Source: Raster Calculator
("ext_pop" / 6702)

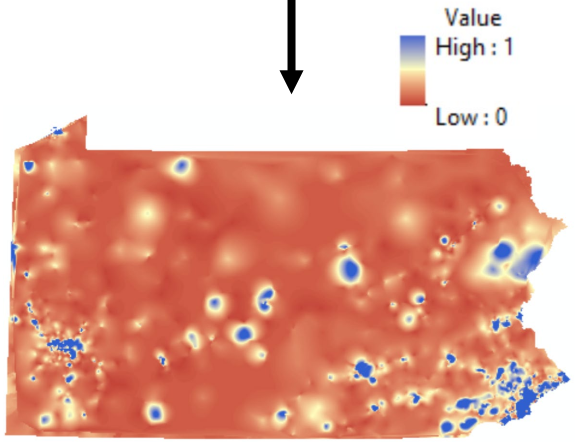


Figure 3a. norm_white
Description: white scaled to 1
Source: Raster Calculator (100
- "ext_white" / 100)

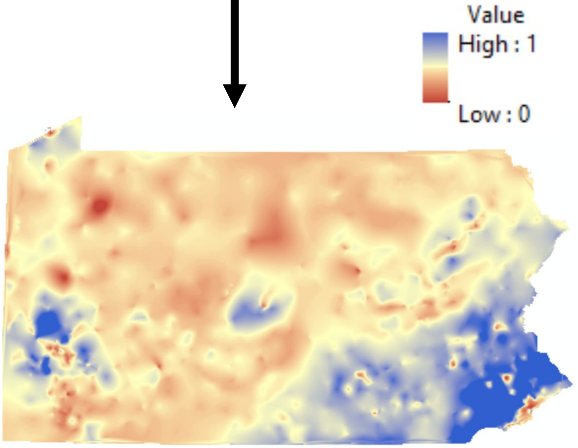


Figure 3f. norm_income
Description: income scaled to 1
Source: Raster Calculator
("ext_income" / 181)

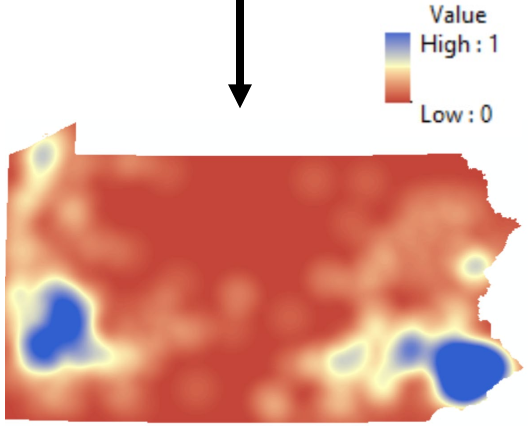
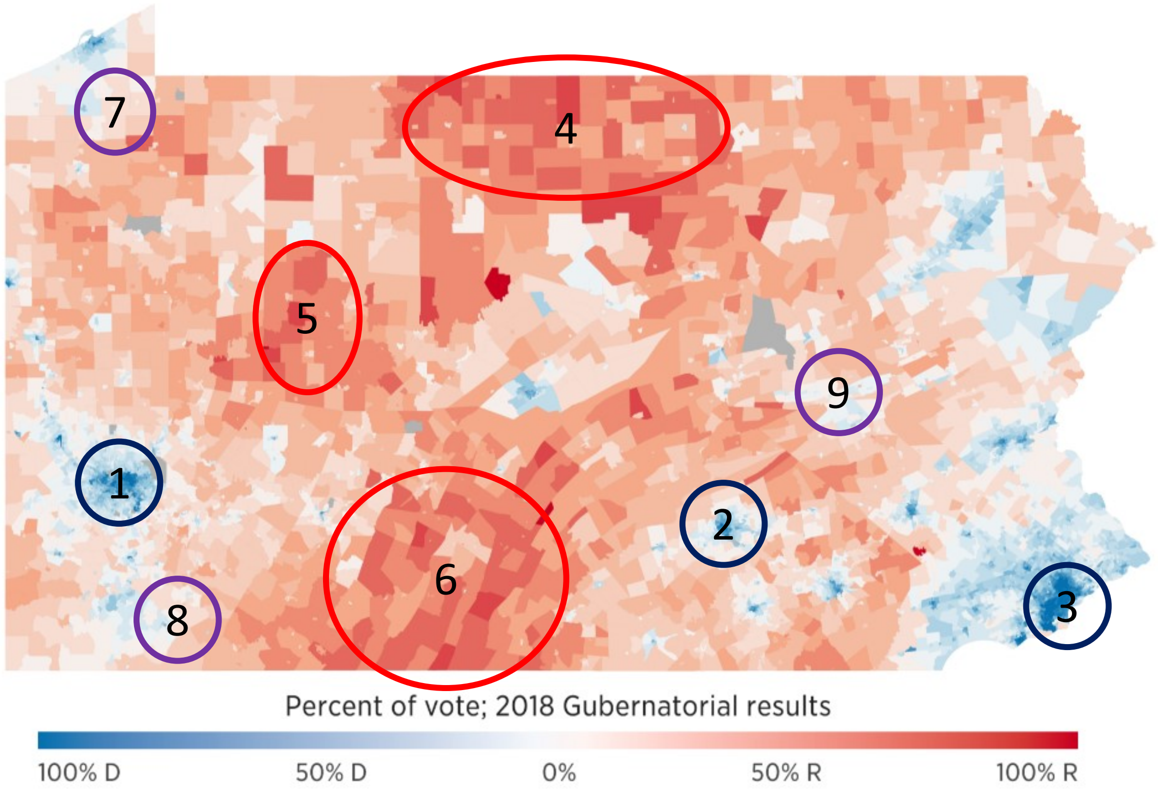


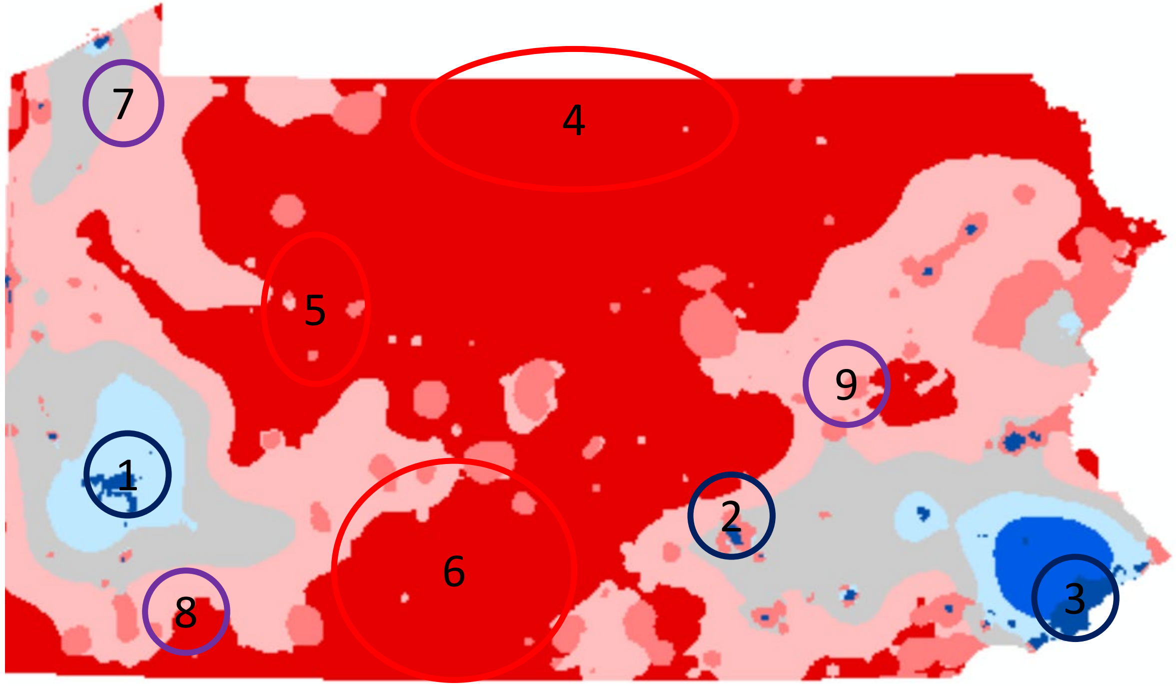
Figure 3g. norm_golf
Description: golf scaled to 1
Source: Raster Calculator
("ext_golf" * 100000000 / 3.28)

Combination 1:pop, white, golf

- Moderate Republican
- Strong Republican
- Slight Republican
- Neutral
- Slight Democrat
- Strong Democrat
- Moderate Democrat

7 classes from isocluseters are created. Hopefully, they can find areas that are strongly democratic, medium democratic, slightly democratic, neutral, slightly republican, medium republican, strongly republican



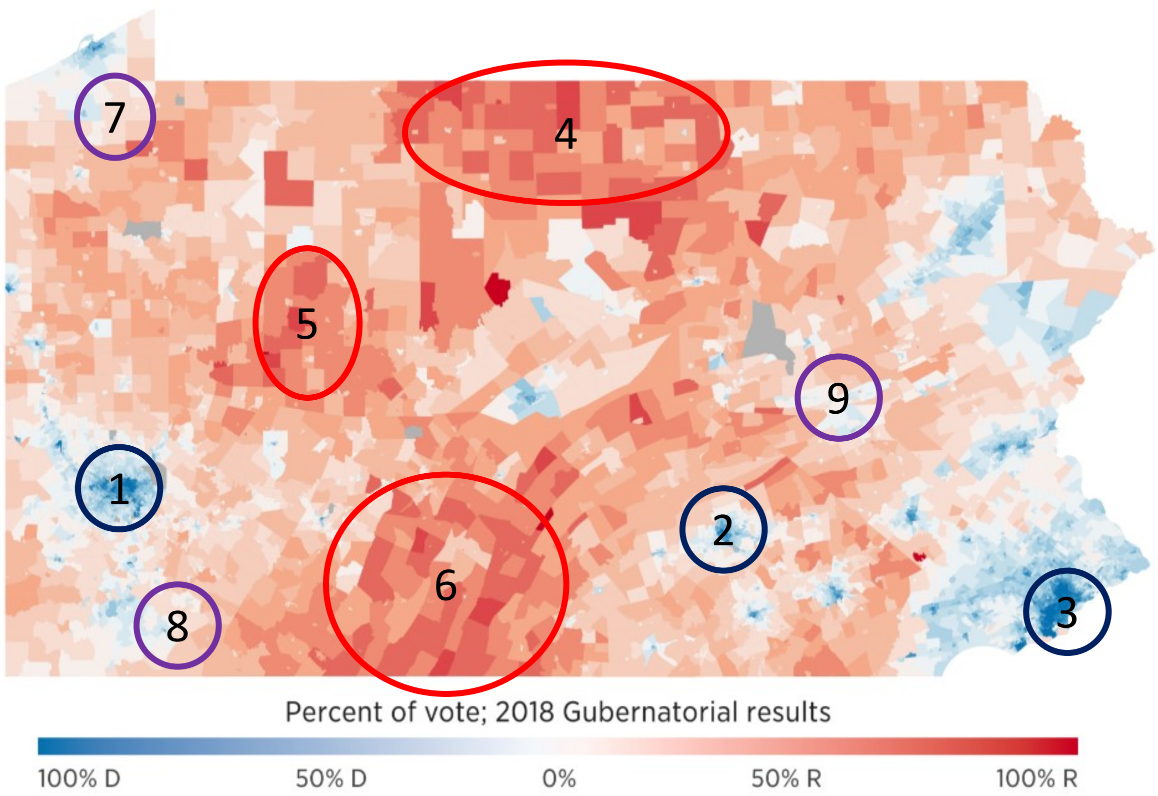


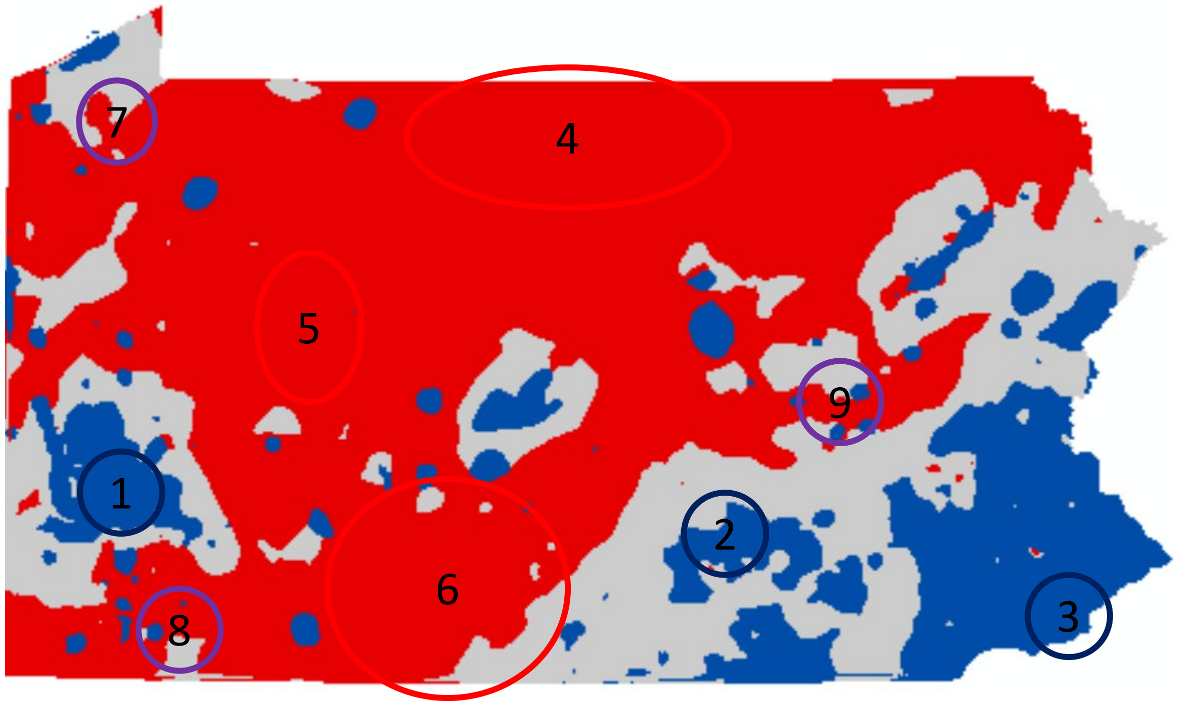
Source: Iso Cluster Unsupervised, Input Raster("norm_pop", "norm_white", "norm_golf"), Classes = 7)

This combination is able to accurately identify strong republican clusters (4, 5, 6). For Democratic clusters, it was able to strongly identify 1 and 3, but seems to have underestimated the democratic cluster for 2. For neutral clusters, 7 seems to be predicted accurately, 9 seems to tilt a bit more republican, and 8 seems to tilt too republican. Thus, this combination is good, but seems to tilt towards republican a bit too much.

Combination 2:pop, white, income

7 classes from isocluseters are created. Hopefully, they can find areas that are strongly democratic, medium democratic, slightly democratic, neutral, slightly republican, medium republican, strongly republican



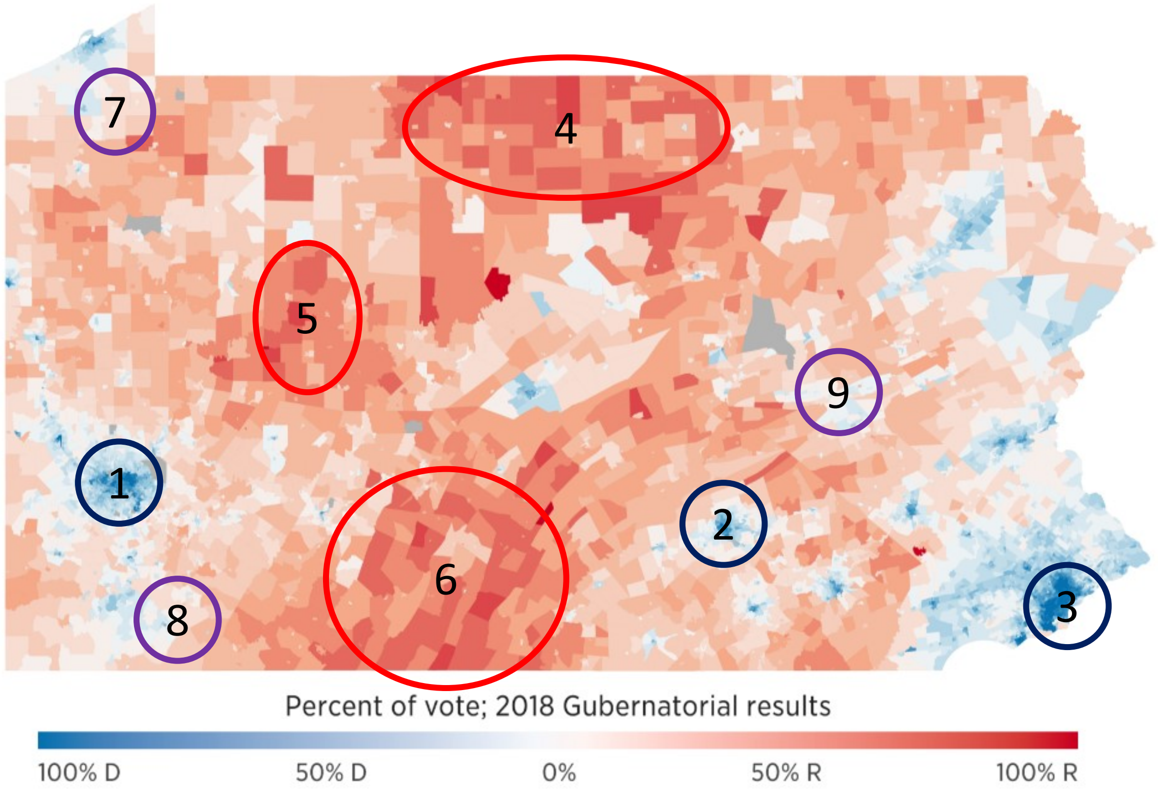


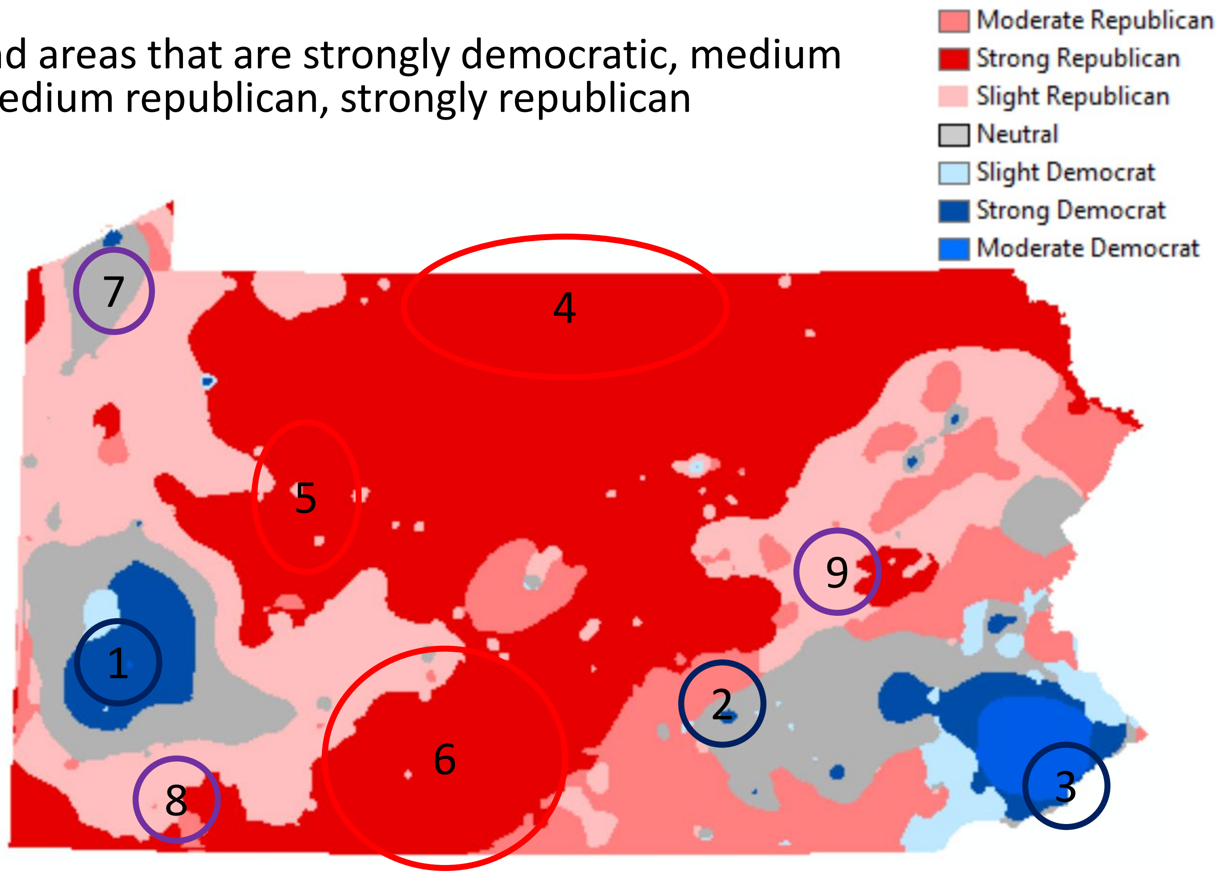
Source: Iso Cluster Unsupervised, Input Raster("norm_pop", "norm_white", "norm_income"), Classes = 7)

This combination was only able to create 3 classes of a minimum of 20 sample points. Thus, it cannot distinguish between very or slightly democratic or republican. This combination is weak and should not be used.

Combination 3: population, golf, income

7 classes from isocluseters are created. Hopefully, they can find areas that are strongly democratic, medium democratic, slightly democratic, neutral, slightly republican, medium republican, strongly republican



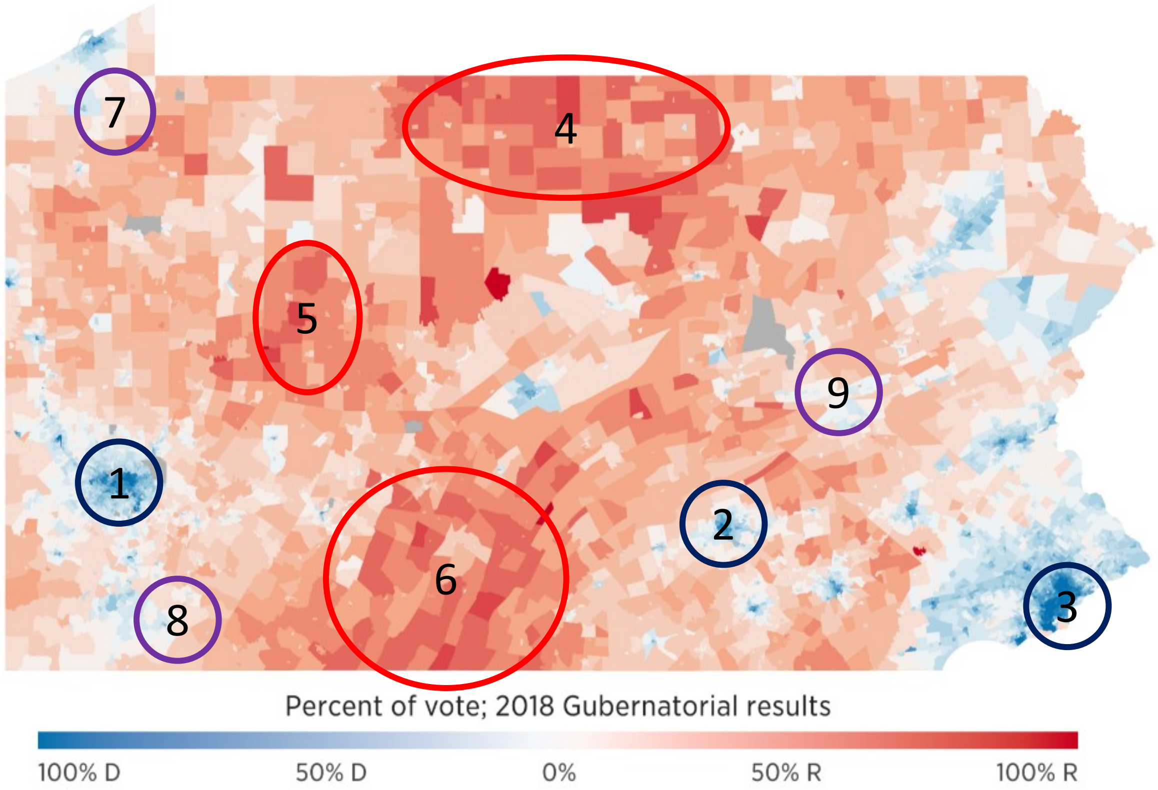


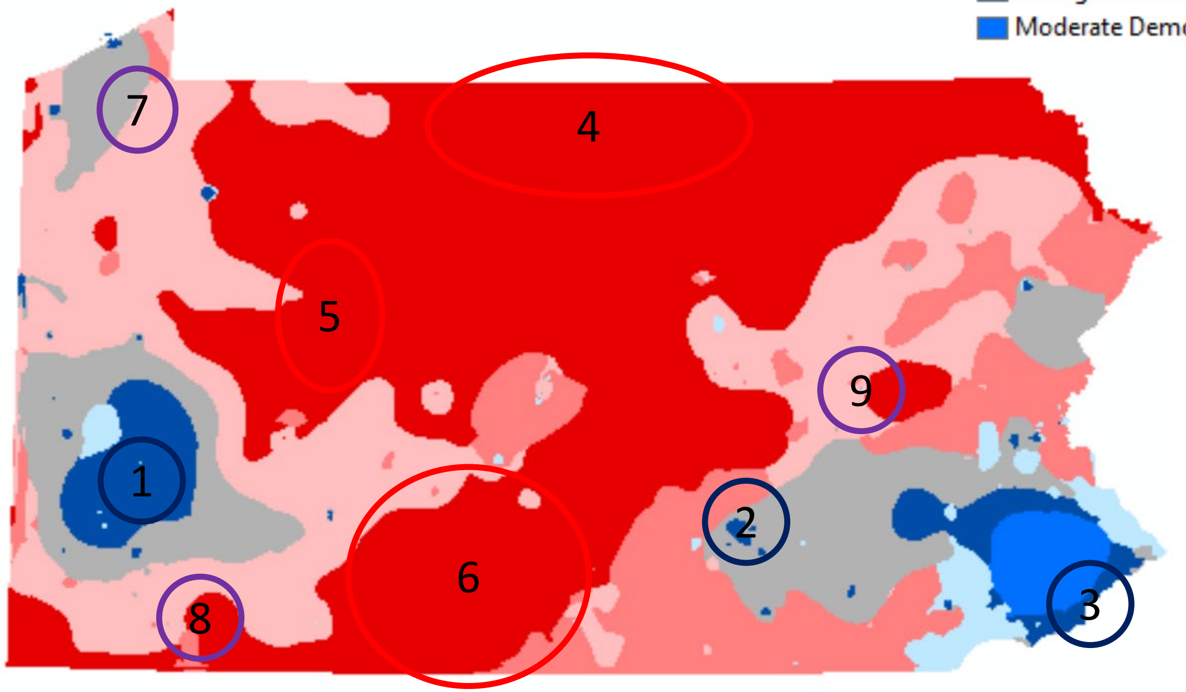
Source: Iso Cluster Unsupervised, Input Raster("norm_pop", "norm_golf", "norm_golf"), Classes = 7)

This combination is able to accurately identify strong republican clusters (4, 5, 6). For Democratic clusters, it was able to strongly identify 1 and 3, but seems to have underestimated the democratic cluster for 2. That said, while combination 1 assigned cluster 2 to have more republican, this combination assigned cluster 2 to be more neutral, which is closer to the actual voting. For neutral clusters, 7 seems to be predicted accurately, 9 seems to tilt a bit more republican, but is more neutral than combination 1, and 8 seems to tilt too republican. This combination, like combination 1, seems to tilt too republican, but it seems tilt less (have more neutral), and so is a better predictor.

Combination 4: white, golf, income

7 classes from isocluseters are created. Hopefully, they can find areas that are strongly democratic, medium democratic, slightly democratic, neutral, slightly republican, medium republican, strongly republican





Source: Iso Cluster Unsupervised, Input Raster("norm_white", "norm_white", "norm_golf"), Classes = 7)

This combination seems pretty similar to combination three, but is a bit weaker in identifying small noises. This should also not be used, as it serves the same purpose as combination 3, but is a bit weaker.

Analysis: Travel Time

The best predictors seem to be

Individually: population, % white, and golf courses are the three variables.

In Combination: population, golf, income. This suggest that perhaps % white has overlapping information with population or golf, and income has information that the other two does not have