BEST CENSUS TRACT FOR OPENING A NEW SHOP IN PHILADELPHIA

CPLN 503 Modeling Geographic Objects

Term Project: Data Interpretation | Zizhao He

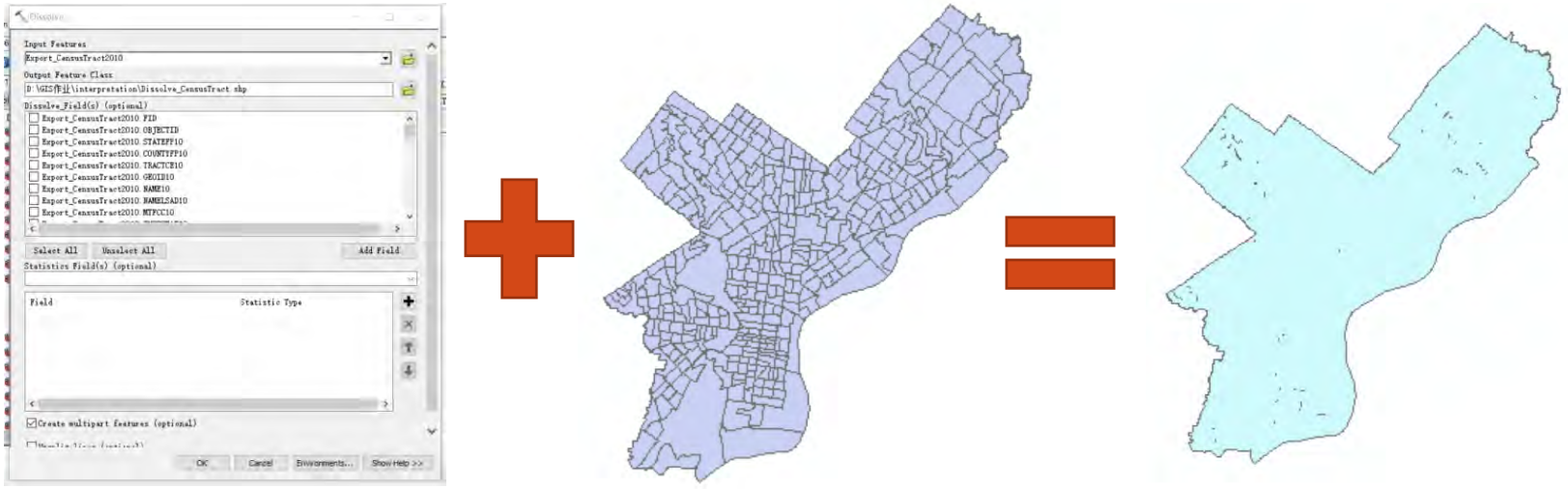
Date: 11/21/2018

INTRODUCTION

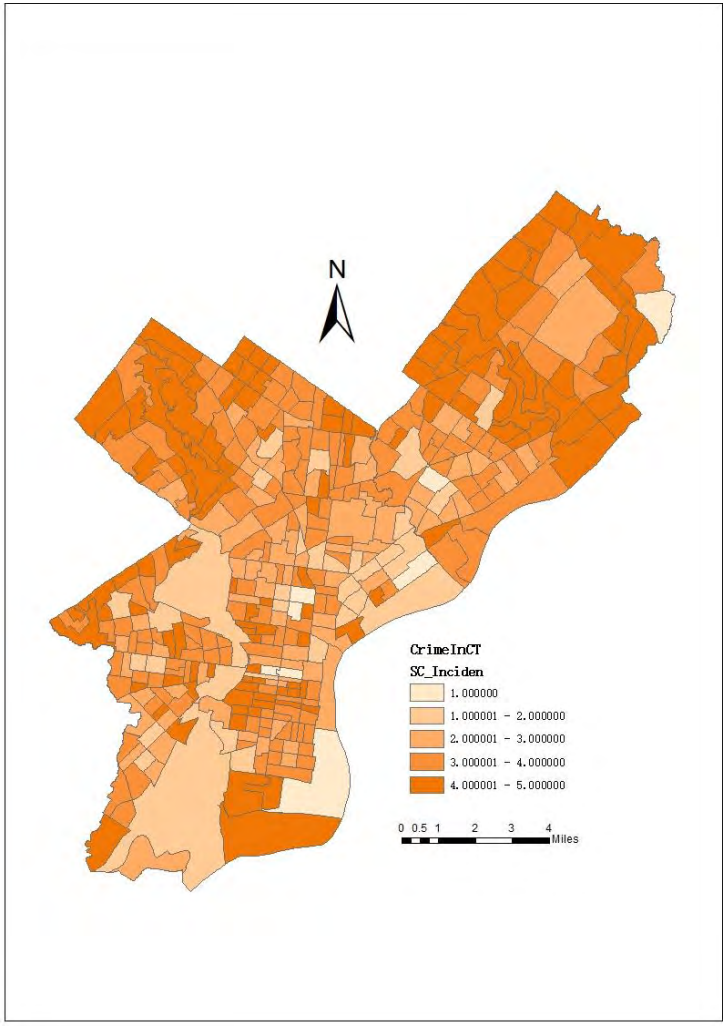
For the term project, I'm interested in evaluating the contexts of all census tracts in Philadelphia to find the best census tract for opening a new shop. In the first part of the term project (Data Preparation), I applied data acquisition skills to get the data of 6 topics, applied data creation skills to transform the raw data into the form which could be used directly, and used data conversion skills to visualize the data and get an intuitive feeling about the outcome of the project (the geographic location of the best census tract). In this part of the term project (Data Interpretation), I will appropriately apply data interpretation skills like Measurement, Superimposition, Aggregation, Selection, Buffering, Joining and Regression to analyze and interpret the data I collected and transformed before and assign scores to each census tract based on their performance on each of the topics. Finally, I will find the best census tract in Philadelphia to open a new shop based on the final score (sum of the scores of each topic).

DISSOLVE

Step 1. Because I may want to show some of my outcomes on the map of Philadelphia, I used ArcToolbox → Data Management Tools → Generalization → Dissolve to create a layer of Philadelphia map with only the boundary lines of the county.



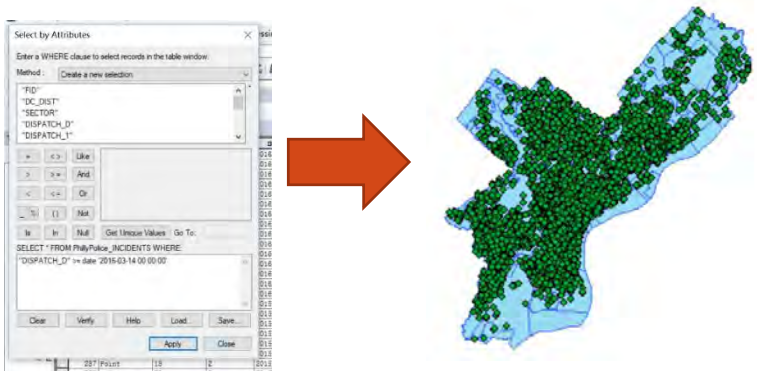
LAYER 1: CRIME



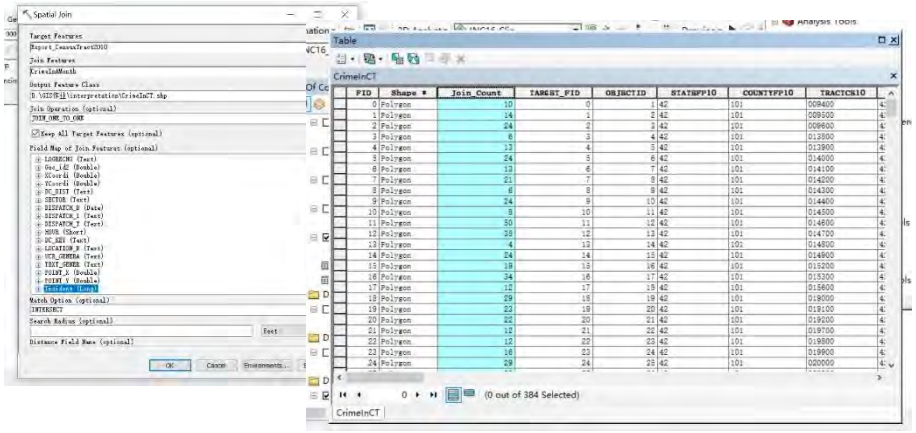
In this section, I will evaluate the crime data by scoring the census tracts based on the number of crime incidents happened in them in the recent one month. The fewer the crimes are, the better to open a new shop.

1.1 CRIME

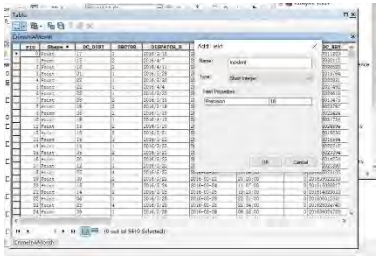
Step 1. Use table option → **select by attributes** to get the crime incidents data in most recent month of the crime dataset.

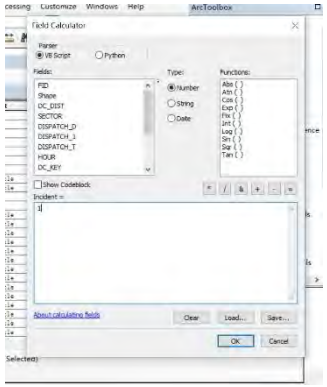


Step 3. Use ArcToolbox → Analysis Tools → Overlay → **Spatial Join (Intersect) and use "sum" as the merge rule** with the census tract layer as Target Features and the crime layer as the Join Features. After the spatial join, I get total number of crimes in each census tracts.

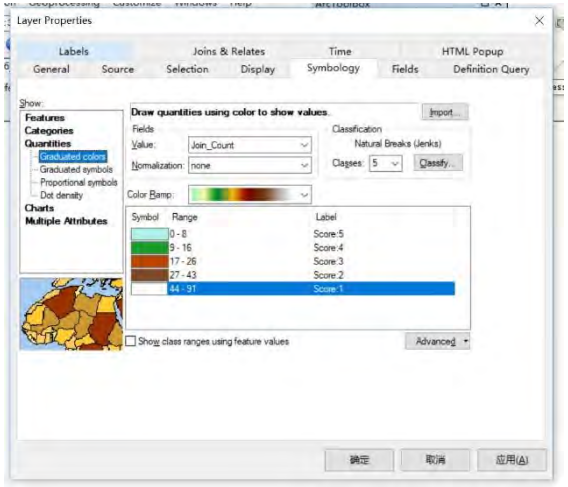


Step 2. Add a new field in the attribute table of the new crime shapefile. Then, right click on the new field → **Field Calculator** to give the same number to all locations of reported crimes which is "1".



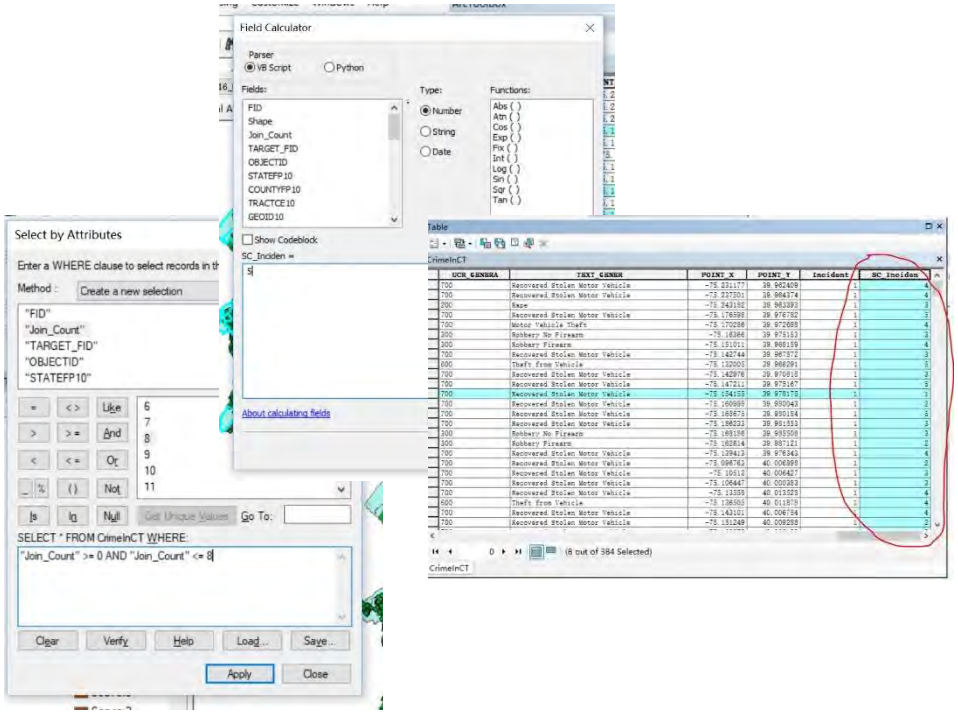


Step 4. Properties → Quantities → Color → Joint_Count for Value with 5 classes. Set the standard and range (1-5) of the scores according to the amount of incidents in the census tracts.

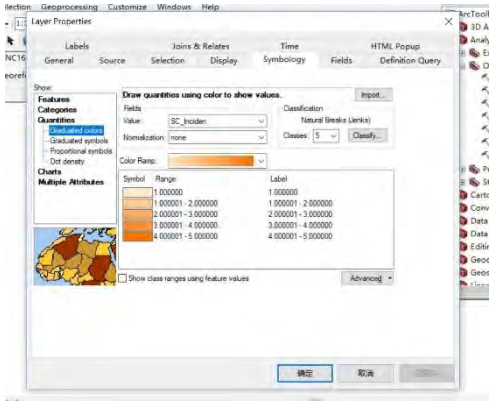


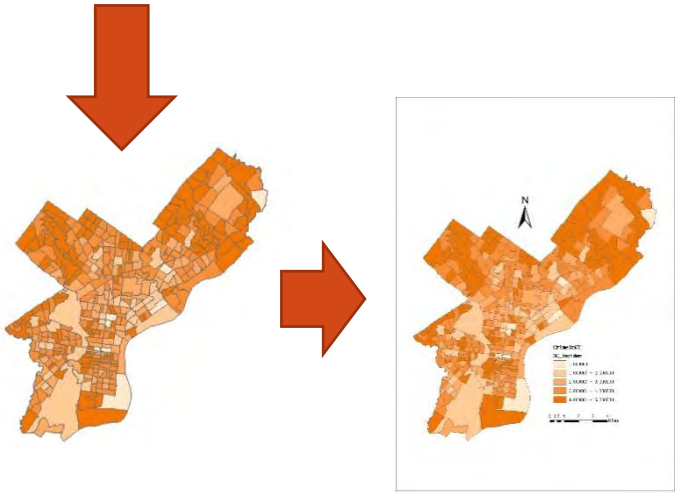
1.2 CRIME

Step 5. Add a **new field** Sc_Incident for the score of the crime incidents. Then, using **select by attributes** to distinguish the 5 classes of scoring. For the selecting features, use field calculator to give them scores (Fewer crime, higher score).

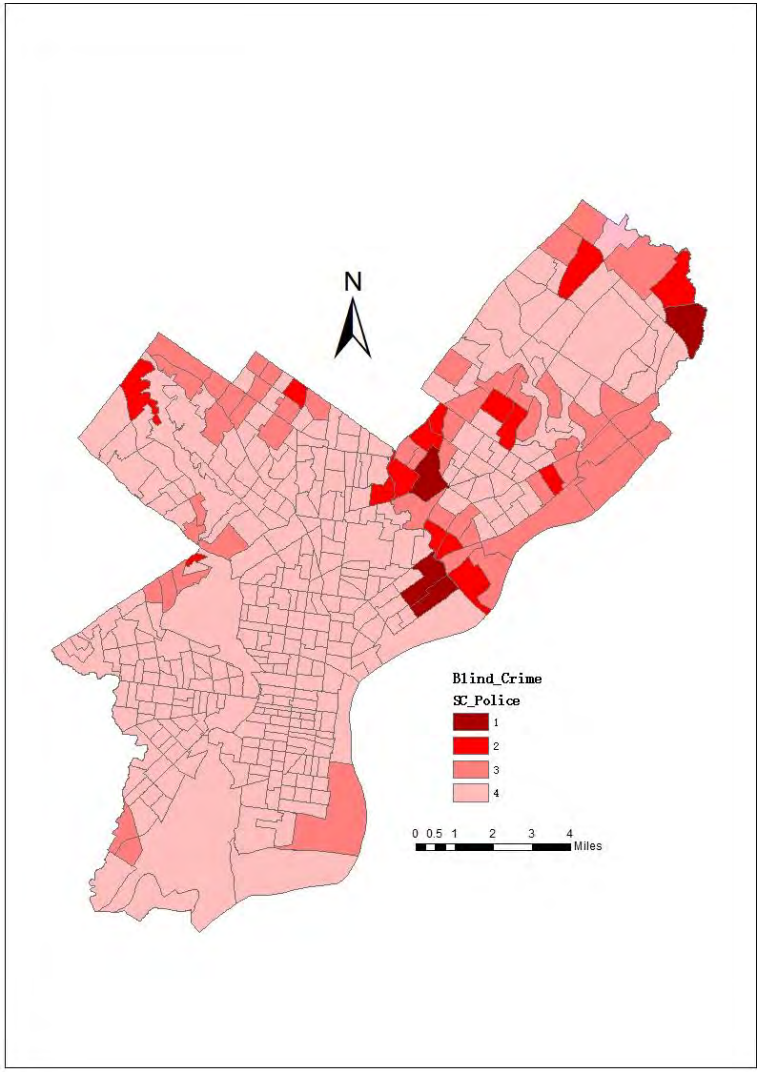


Step 6. Change **Symbology**: Right click on the layer → properties → Quantities (Graduated colors), select Sc_Incident in Value Field. Finally, choose an appropriate Color Ramp and then add map elements.





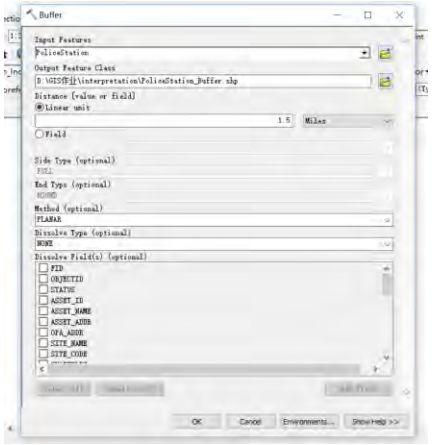
LAYER 2: POLICE STATION



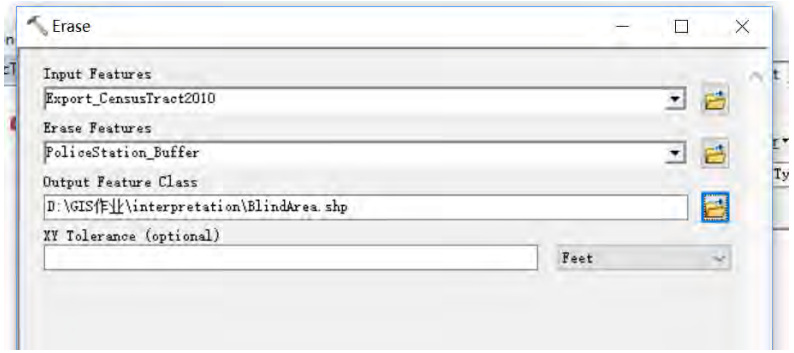
In this section, I will evaluate the police power and the safety level in and out the region controlled by police stations. According to my criteria, the regions controlled by police station are safest. In the area out of the control of police station, the fewer the crimes, the safer the census tract is.

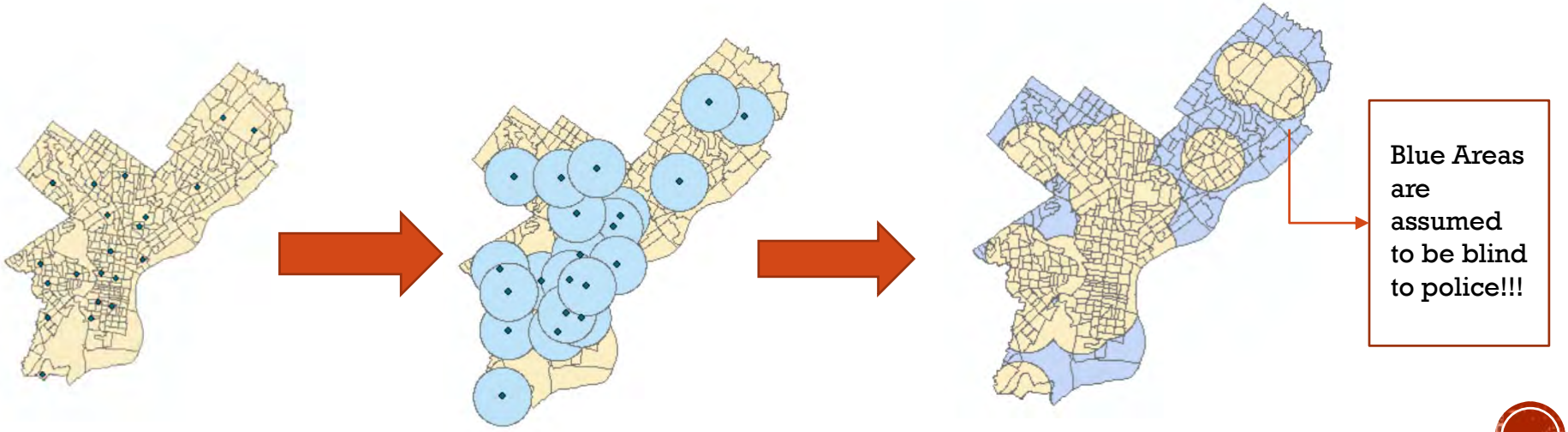
2.1 POLICE STATION

Step 1. Make a buffer around the police stations for 1.5 miles (Police Region → Safer!) by using ArcToolbox → Analysis Tools → Proximity → **Buffer**.



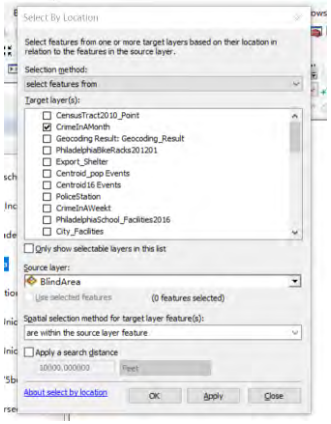
Step 2. Erase the buffer layer from the census tract layer to get the area out of the region of police stations. Using ArcToolbox → Analysis Tools → Overlay → **Erase**.

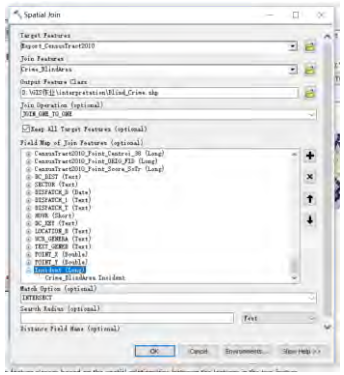




2.2 POLICE STATION

Step 3. Select the crimes happened in the area out of police control by using selection → select by location. Method is "are within the source layer feature". Then, export the selected data.





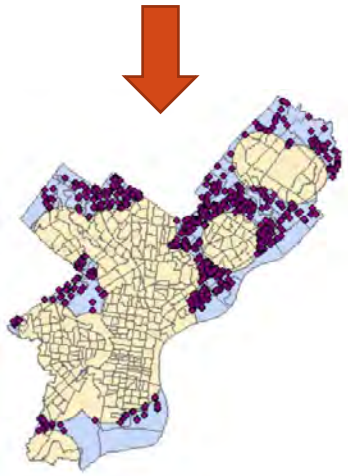
Step 4. Use ArcToolbox → Analysis Tools → Overlay → Spatial Join to spatially join the selected crime layer to the census tract layer, which provides the number of crimes in the census tract that out of common protection of police in the field Joint_Count.

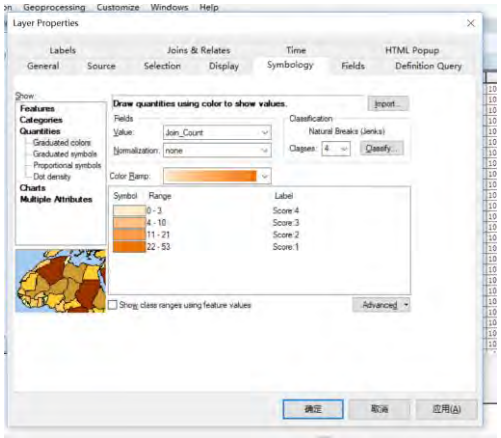
Table

Blind_Crime

FID	Shape *	Join_Count	TARGET_FID
0	Polygon	0	0
1	Polygon	0	1
2	Polygon	0	2
3	Polygon	0	3
4	Polygon	0	4
5	Polygon	0	5
6	Polygon	0	6
7	Polygon	0	7
8	Polygon	0	8
9	Polygon	0	9
10	Polygon	0	10
11	Polygon	0	11
12	Polygon	0	12
13	Polygon	0	13
14	Polygon	0	14
15	Polygon	0	15

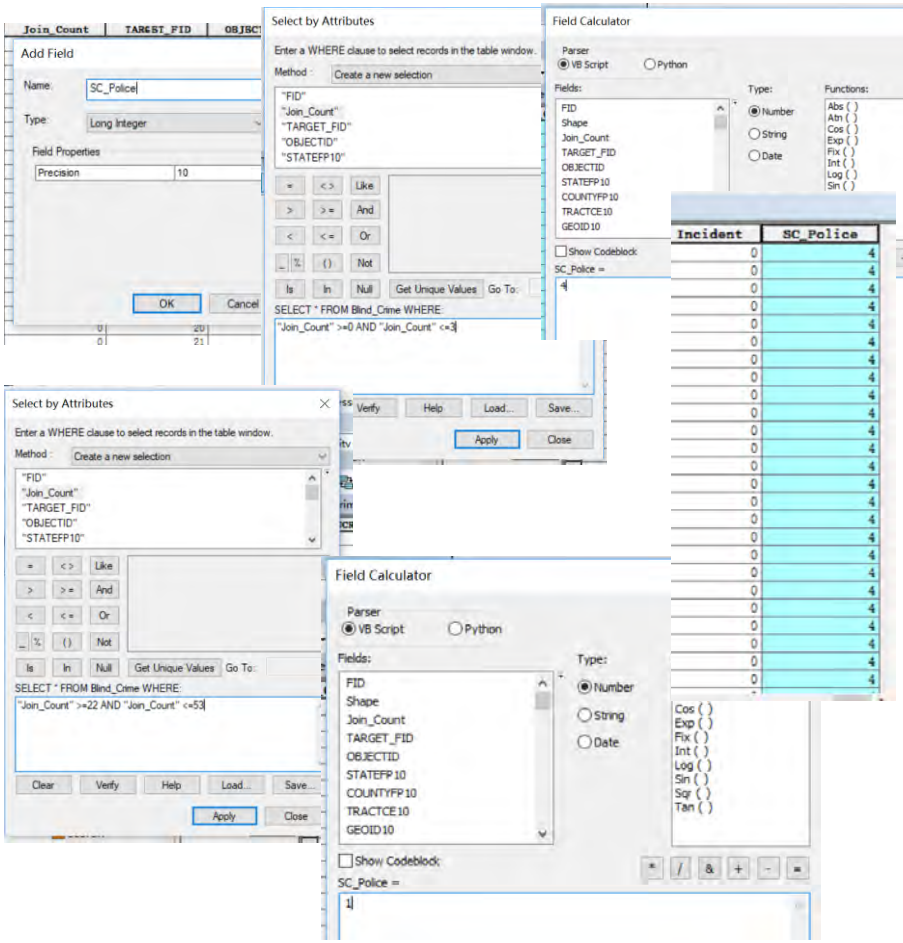
Step 5. Properties → Quantities → Color → Joint_Count for Value with 4 classes



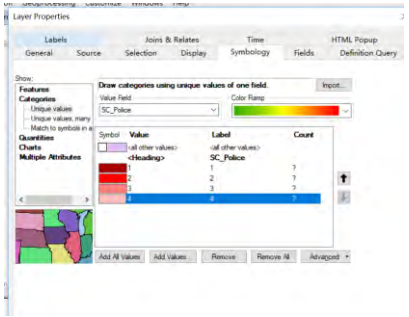


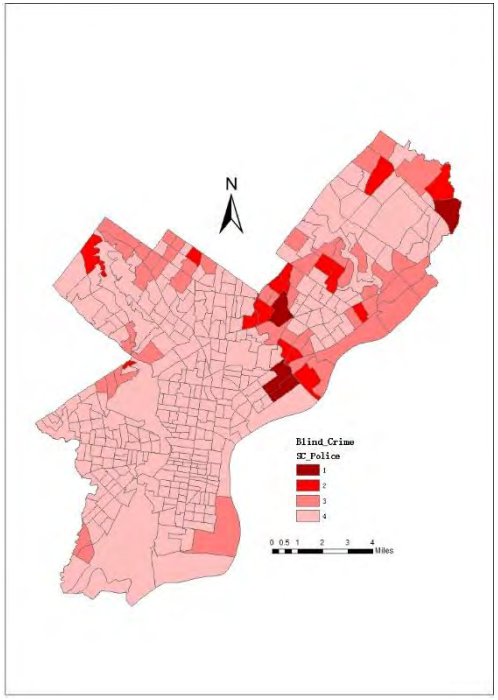
2.3 POLICE STATION

Step 6. Add a field in the attribute table for scoring. Then, using select by attributes to distinguish the 4 classes of scoring. For the selecting features, use field calculator to give them scores (More incidents, lower score!)

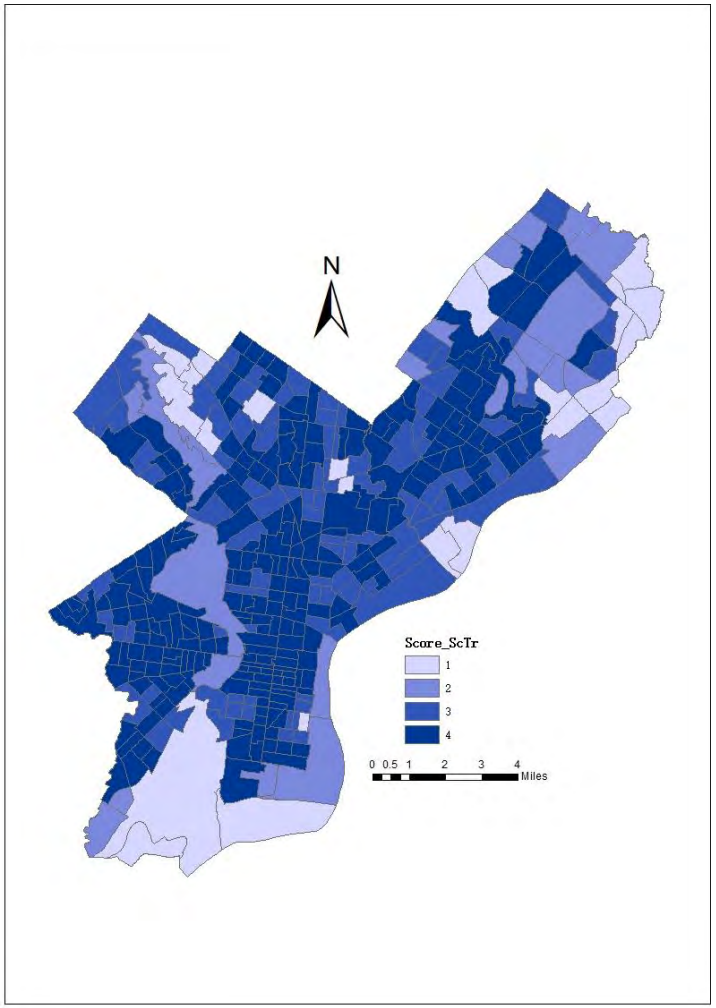


Step 7. Change Symbology: Right click on the layer → properties → Categories (Unique Values), select SC_Police in Value Field. Finally, choose an appropriate Color Ramp and add basic map elements.





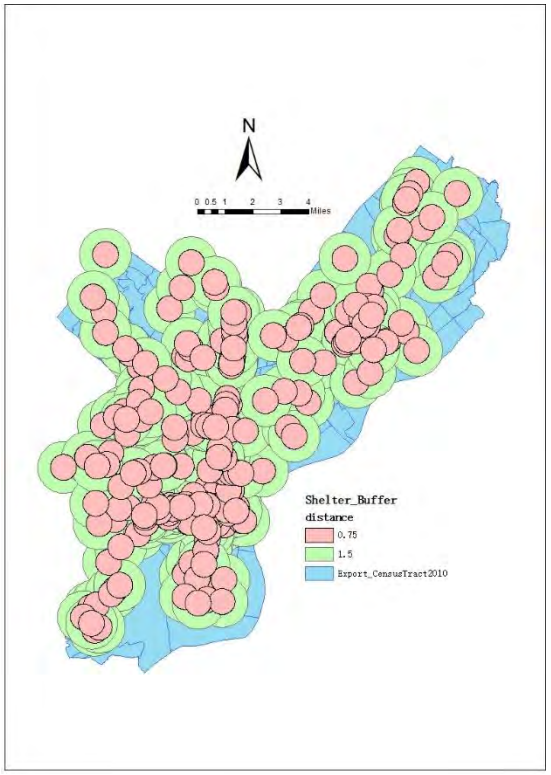
LAYER 3: ACCESSIBILITY TO BUS SHELTER AND SCHOOL

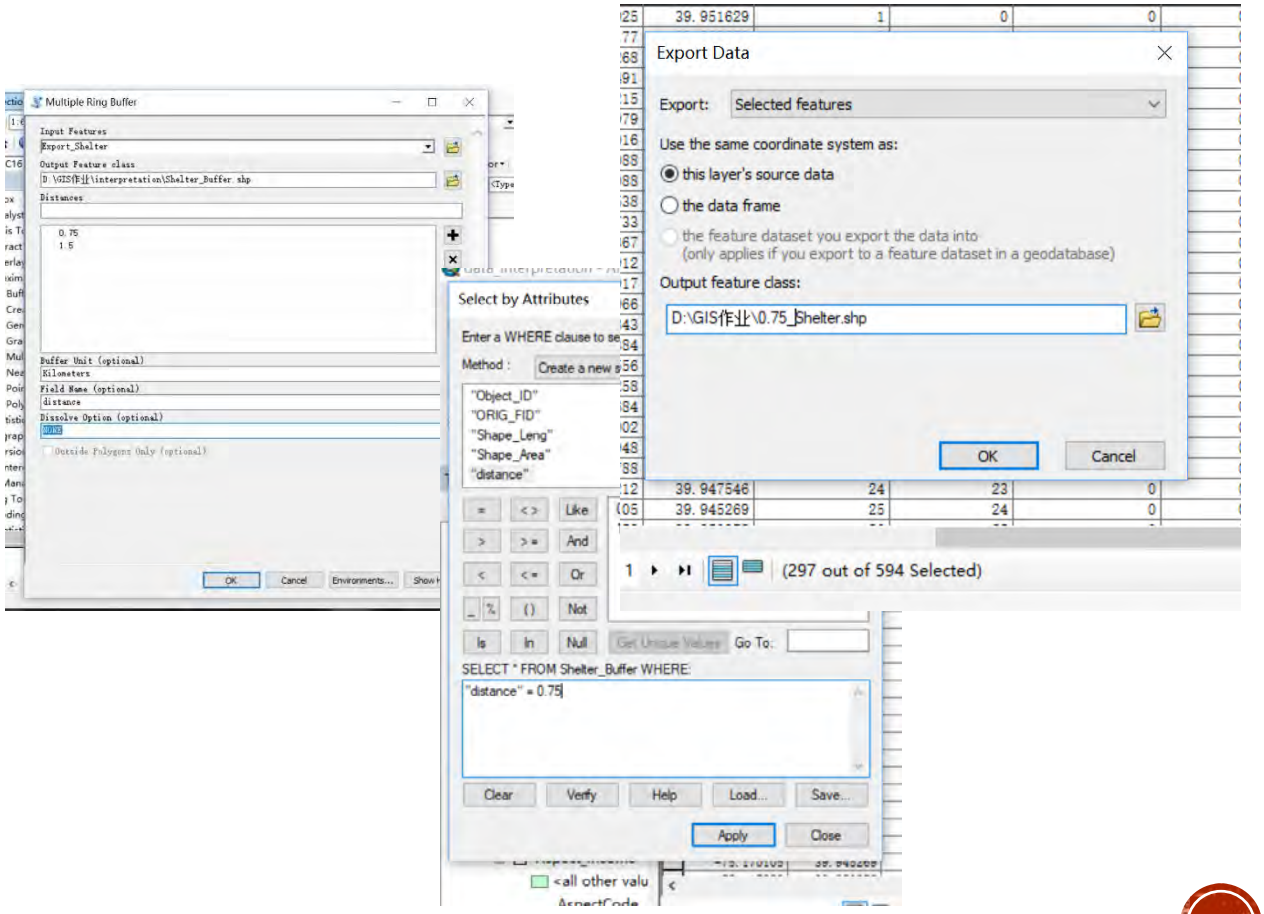


The higher accessibility to bus shelters could encourage people to get to the shop due to the convenience of using public transit. What's more, the higher accessibility to schools means a large pool of potential consumers. So, in this section, I tried to evaluate the accessibility to both bus shelters and schools comprehensively.

3.1 ACCESSIBILITY TO BUS SHELTER AND SCHOOL

Step 1. Buffer (Multiple Ring)
ArcToolbox → Analysis Tools → Proximity → Multiple Buffer Ring. Make 0.75 and 1.5 kilometers buffers of the SEPTA bus shelters of the whole Philadelphia. Then, Table Option → Select by Attributes → Data → Export Data to export data seperately.





3.2 ACCESSIBILITY TO BUS SHELTER AND SCHOOL

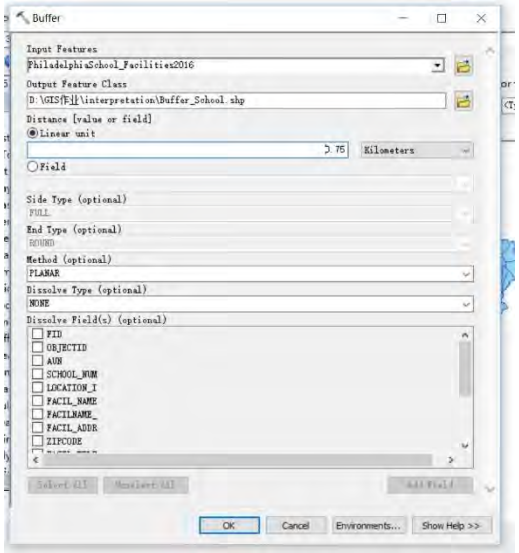
Step 2. Erase the buffers by ArcToolbox → Analysis Tools → Overlay → Erase
Then, the outcome is the buffer between 0.75km and 1.5km.

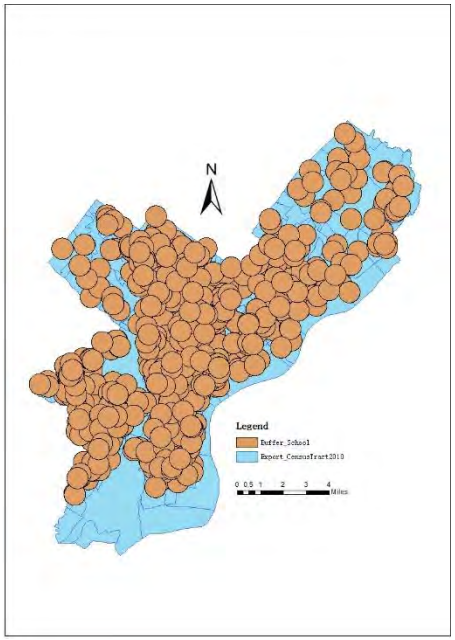
- 1.5km Buffer

- 0.75km Buffer

- 0.75-1.5km Buffer

Step 3. ArcToolbox → Analysis Tools → Proximity → Buffer. Buffer the school facilities for 0.75km.

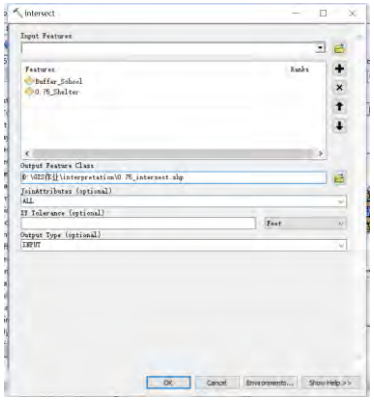


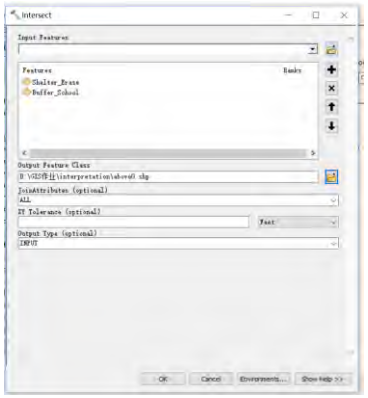


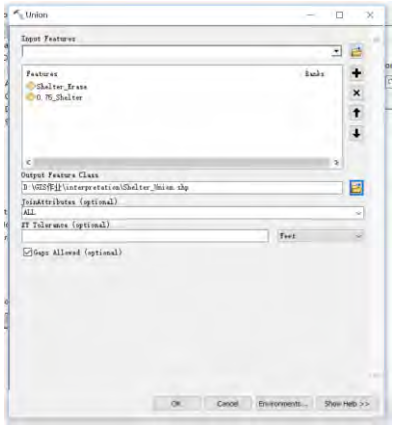
3.3 ACCESSIBILITY TO BUS SHELTER AND SCHOOL

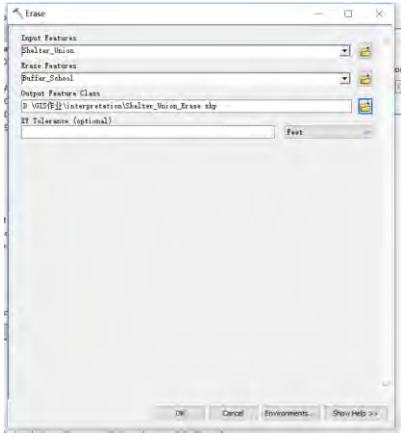
Step 4. Intersect buffers of bus shelter and school facilities: ArcToolbox → Analysis Tools → Overlay → Intersect. Then, export the intersected parts seperately.

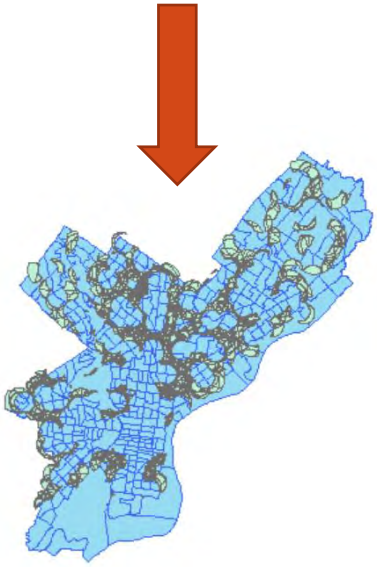
Step 5. Union 0.75 and 0.75-1.5km buffers: ArcToolbox → Analysis Tools → Overlay → Union. Then, erase the school_buffer from the unioned feature.

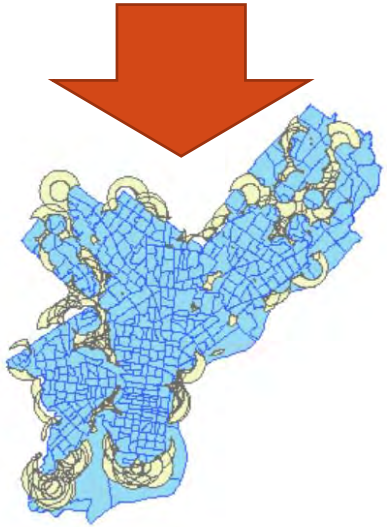












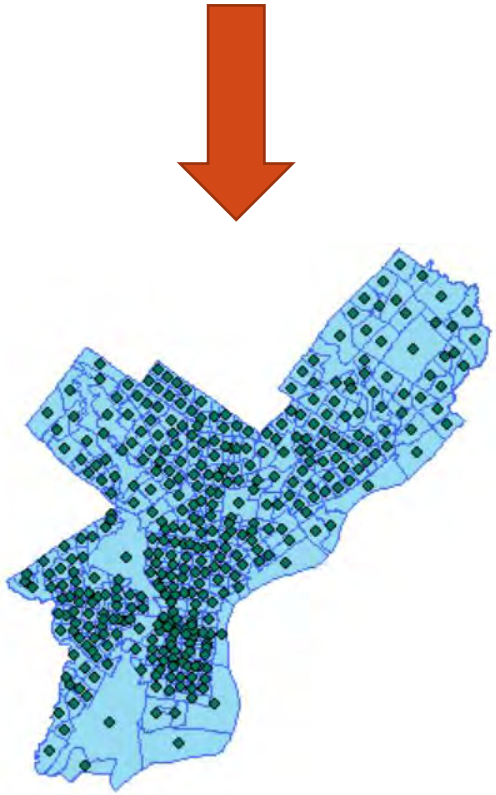
Score4: 0.75km intersect

Score 3: 0.75-1.5km intersect

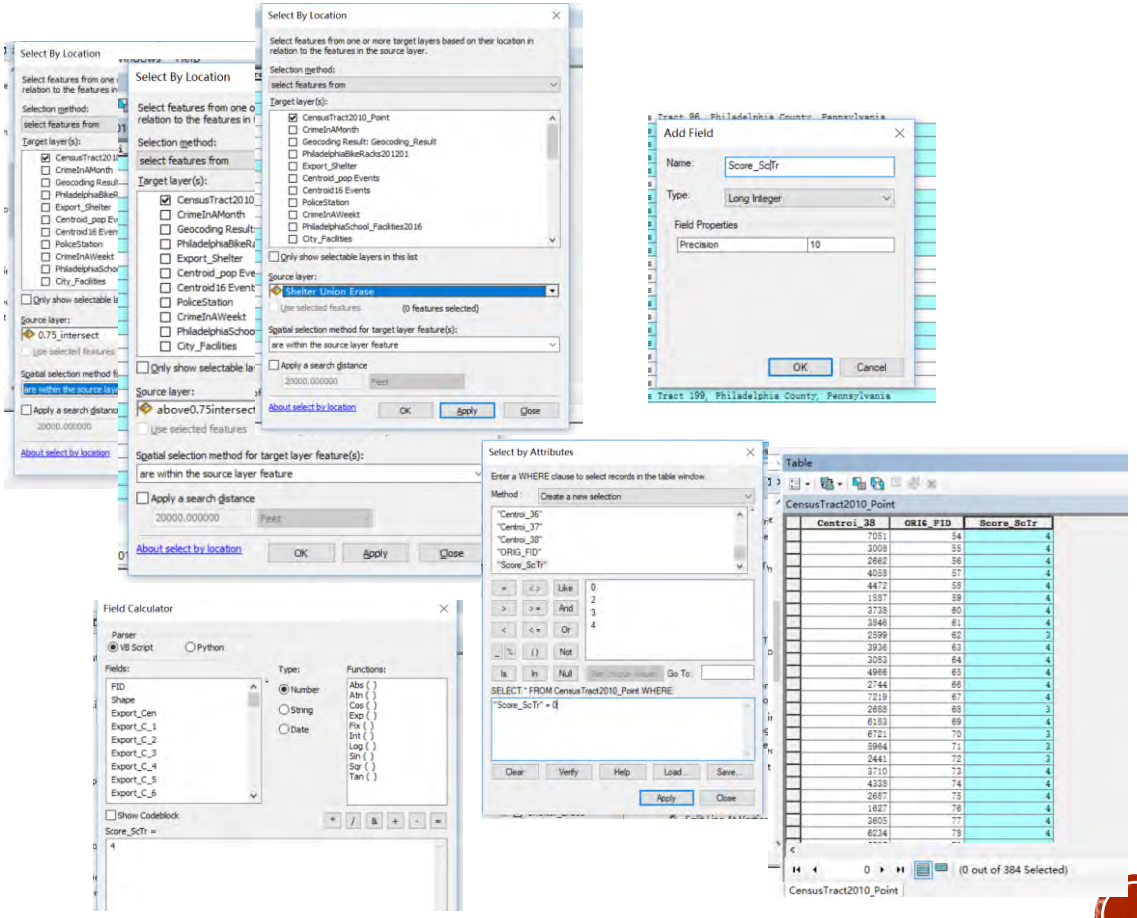
Score 2: 0.75-1.5km (out of school buffer)

3.4 ACCESSIBILITY TO BUS SHELTER AND SCHOOL

Step 6. Feature to Point. By use ArcToolbox → Data Management Tools → Features → Feature to Points

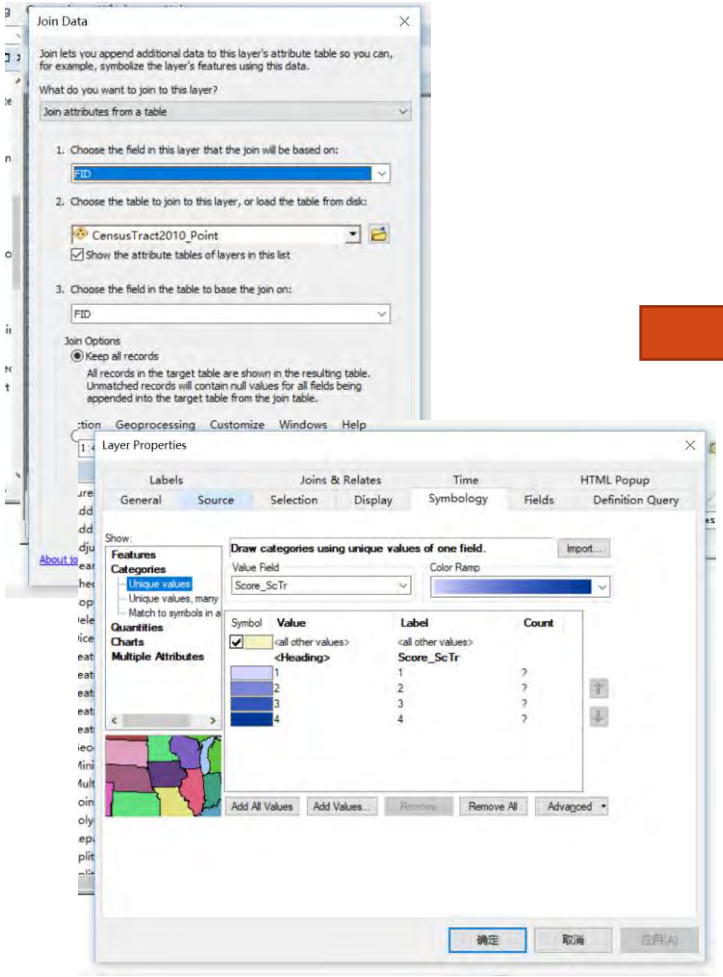


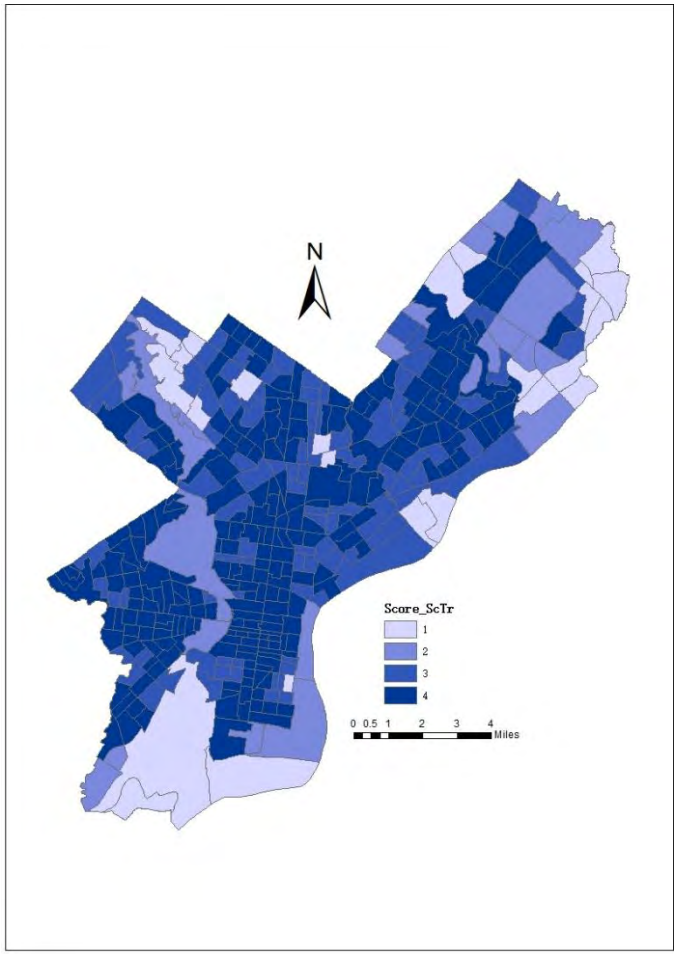
Step 7. Select by location & Field Calculation
Use selection → select by location with the target layer as census tract point shapefile, and source layer as the intersect layers did before. Then add a new field of the score of the accessibility. Give the census tract of these3 layers 4,3 and 2. The rest census tracts will get 1.



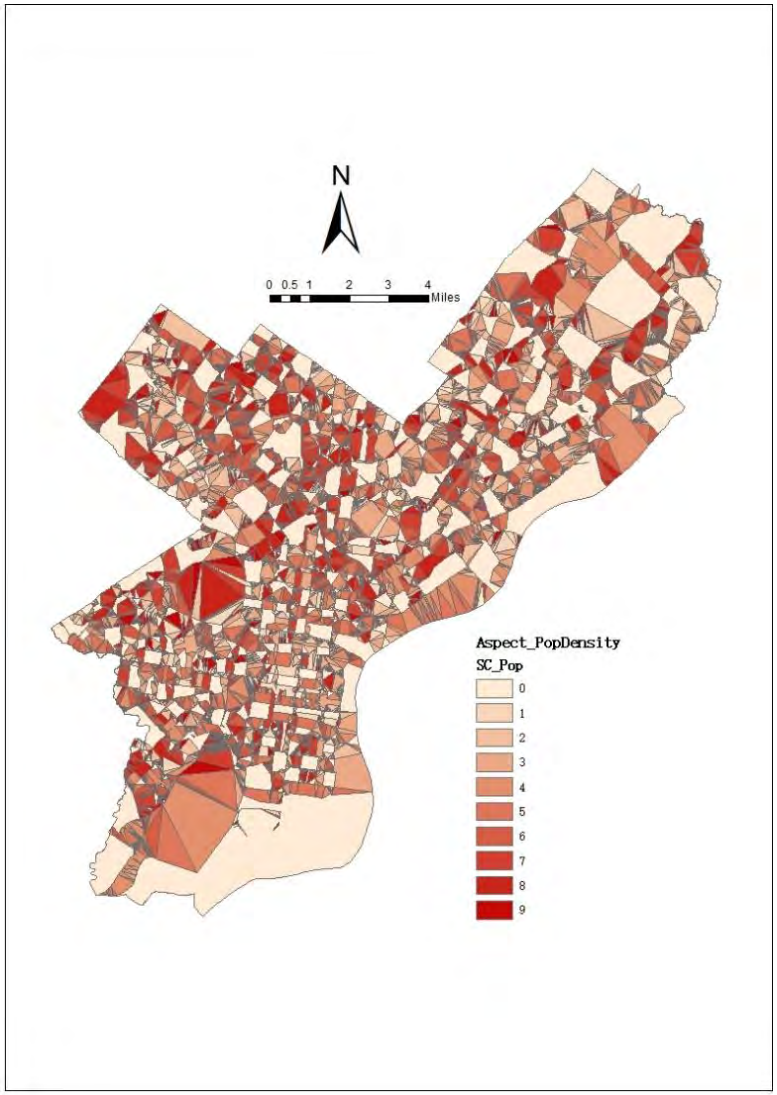
3.5 ACCESSIBILITY TO BUS SHELTER AND SCHOOL

Step 8. Use Join and relate → Join to join the census tract point table to the census tract attribute table. Then, symbologize the map based on the score of the accessibility (Score_ScTr).





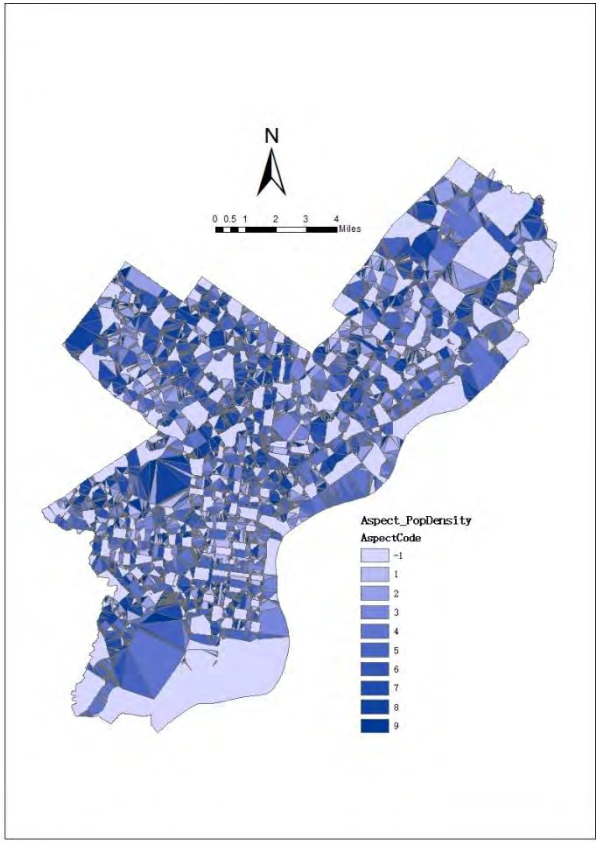
LAYER 4: POPULATION (DENSITY)



In this section, I will evaluate not only the population density itself, but also the direction of change of the population density. In fact, if a census tract has high population density and have a largely positive tendency to keep growing, the pool of the potential consumers would be very large and it's good for opening a new shop there.

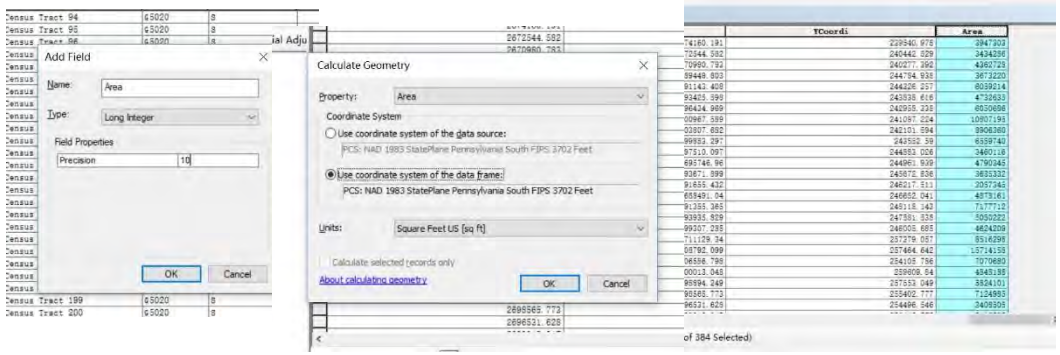
4.1 POPULATION (DENSITY)

From the surface aspect measure, I can know what direction the population density will change.

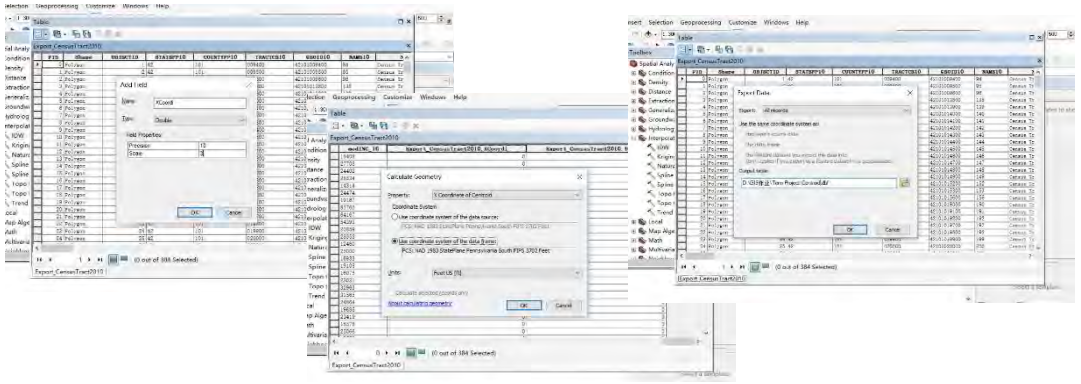


- Surface Aspect of Population Density

Step 1. **Area & Field Calculation**. Firstly, add a new field in the attribute table of census tract layer for Area of each census tract. Then, right click on the field → Calculate Geometry to get the area numbers (in square feet).

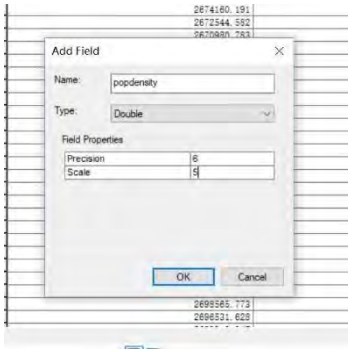


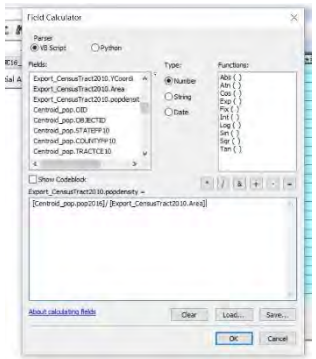
Step 2. Create **Centroid**. Firstly, add two new fields (XY coordinates) to generate the centroids. Then, right click on both fields → Calculate Geometry to get the result, and export it as a new dbf file. Finally, right click the table → Display XY data to get a map of centroids.



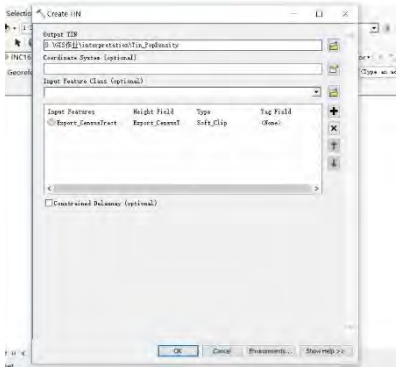
4.2 POPULATION (DENSITY)

Step 3. Add a new field in the attribute table called popdensity. By using Field Calculator, I divided population by area to get the density of population in each census tract.



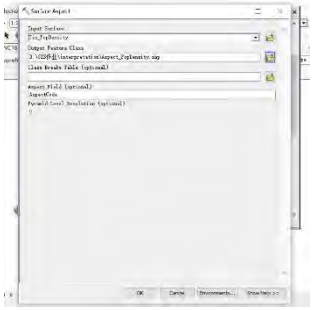


Step 4. Create TIN of pop density: ArcToolbox → 3D Analyst Tools → Data Management → TIN → Create TIN.

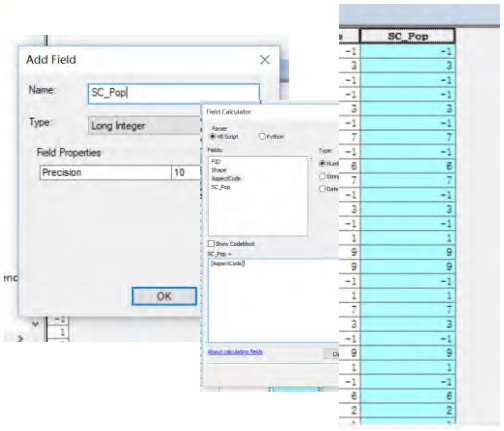


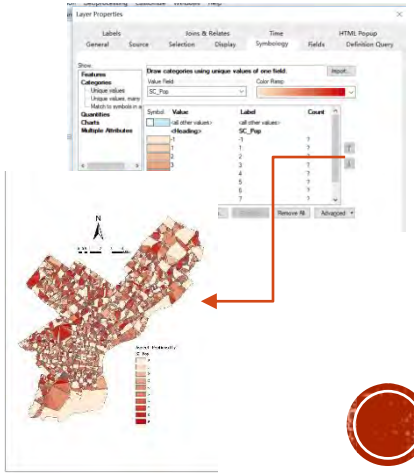
- TIN of pop density

Step 5. Based on the TIN of pop density, I used ArcToolbox → 3D Analyst Tools → Triangulated Surface → **Surface Aspect** to get the Aspect layer of pop density.

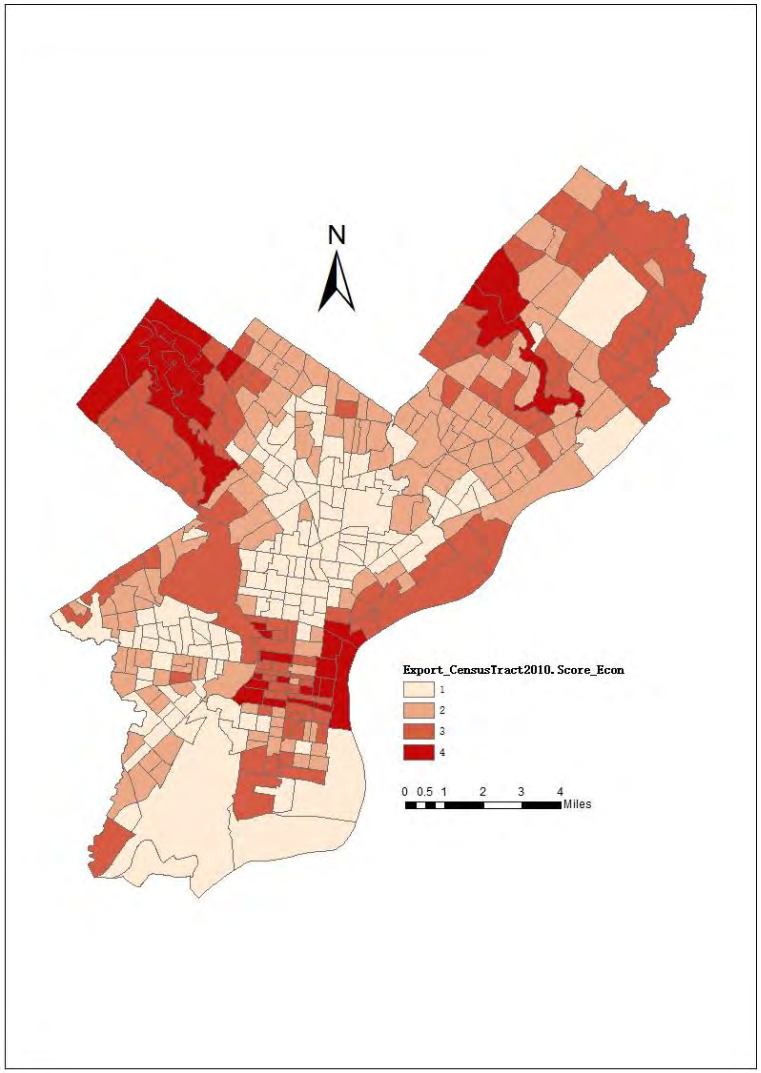


Step 6. Through the aspect, I know what direction the population density will change. So, I score the population density based on the legend of the aspect layer except set -1 to 0. More dense the area is, larger market and more opportunities there is. Add a new field of the score and set the score equal to the aspect number. Change Symbology based on the score.





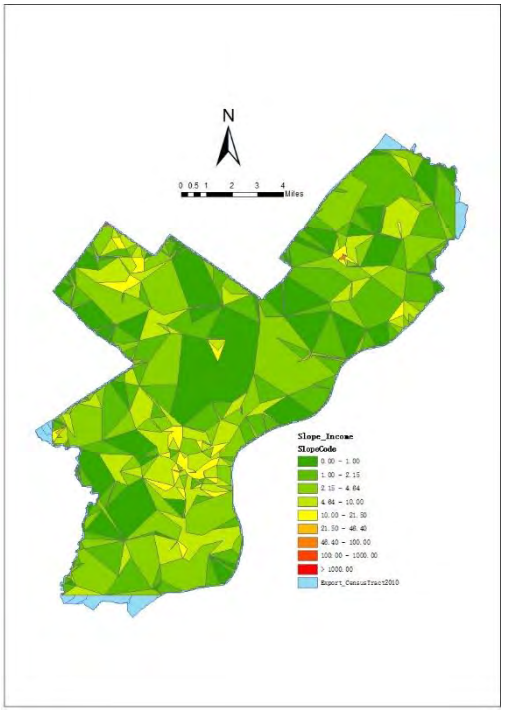
LAYER 5: ECONOMY (MEDIAN HH INCOME)



In this section, I will evaluate the economic level of each single census tract in Philadelphia. Although many factors could be used, I would only use median household income to evaluate the specific number for each census tract, and the extent and direction it would change. In fact, if a census tract has high median household income and have a largely positive tendency to keep growing, the purchase power would be very high among the local residents. So, it would be easy to sell the goods.

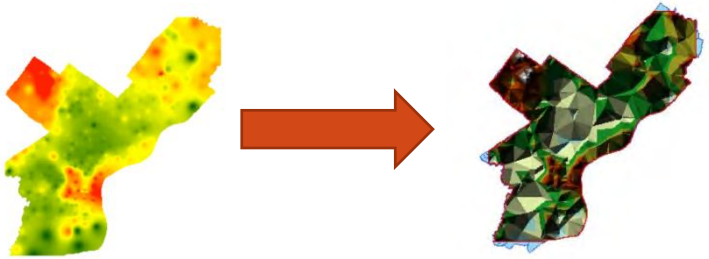
5.1 ECONOMY (INCOME): SLOPE AND ASPECT

From the surface slope measure, I can know what extent the income varies and how the extent changes.

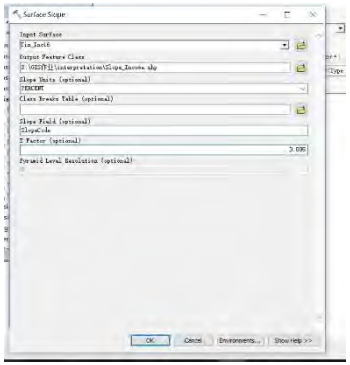


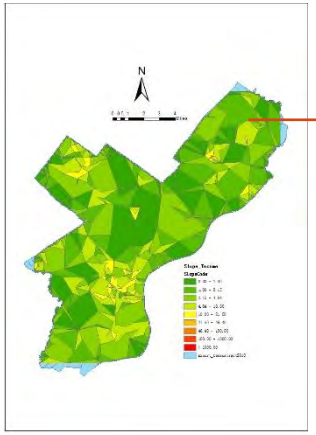
- Surface slope of Income

Step 1. **Create TIN from IDW of median household income layer**. After getting the clipped IDW map of Income from data preparation, use ArcToolbox → 3D Analyst Tools → Conversion → From Raster → Raster to TIN to create the TIN graph for median household income.



Step 2. Based on the TIN layer, I used ArcToolbox → 3D Analyst Tools → Triangulated Surface → **Surface Slope** to get the Slope layer of median household income of all census tracts of Philadelphia.

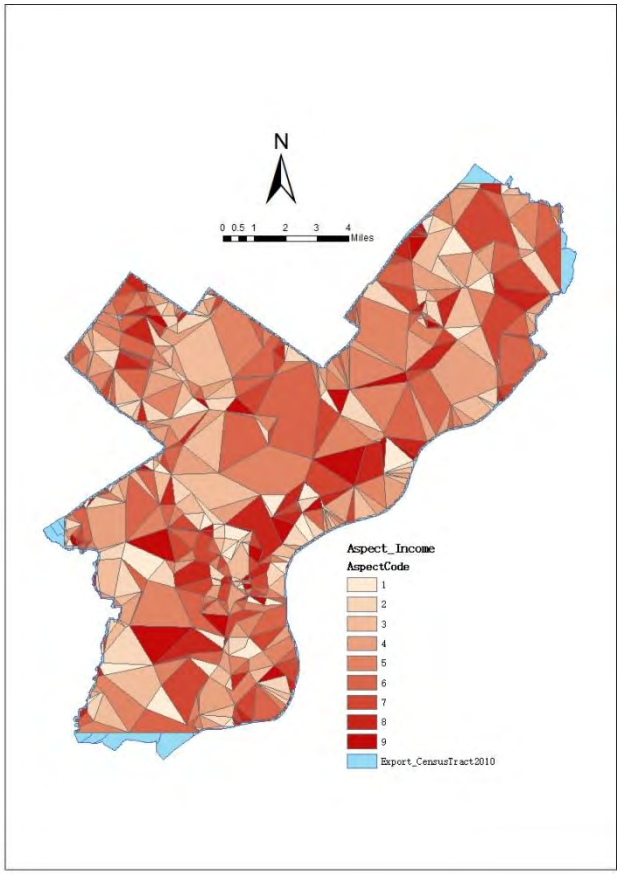




The larger the slope is, the larger the extent would change!!!

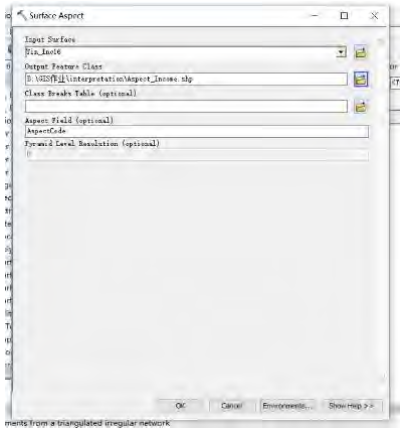
5.2 ECONOMY (INCOME): SLOPE AND ASPECT

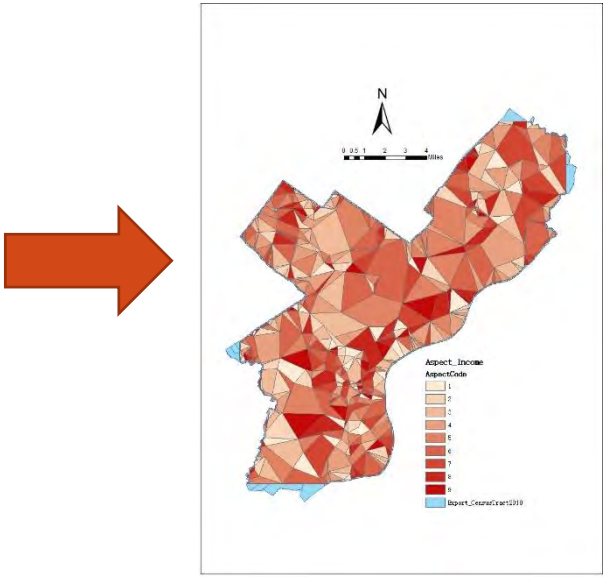
From the surface aspect measure, I can know what direction the Income will change.



- Surface Aspect of Income

Step 3. Based on the TIN layer, I used ArcToolbox → 3D Analyst Tools → Triangulated Surface → **Surface Aspect** to get the Aspect layer of median household income of all census tracts of Philadelphia.



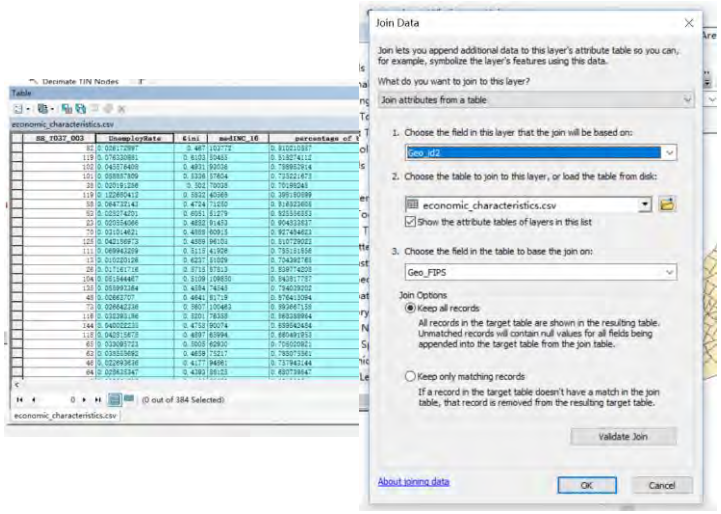


Step 4. **Change to appropriate symbologies** by right click the layers → properties → symbology to both the slope and aspect layers. According to the AspectCode, income will change to the direction that the color becomes darker.

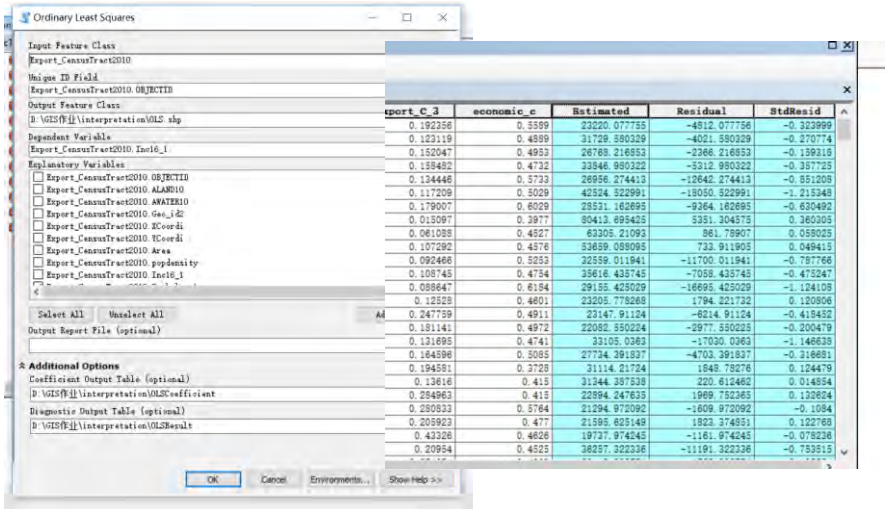
5.3 ECONOMY (INCOME): REGRESSION

In order to assess the economy of the census tracts of Philadelphia, I choose Median Household Income as the standard. However, there are many other characteristics also may affect the economic level of the county. So, I ran an OLS regression model to check if the median household income is highly correlated to them, which are unemployment rate, Gini Index (economic inequality) and percentage of people who attained bachelor or high for education.

Step 1. Getting the csv files of selected economic characteristics from Social Explorer, and combining them together. Then, add the csv file to the attribute table.

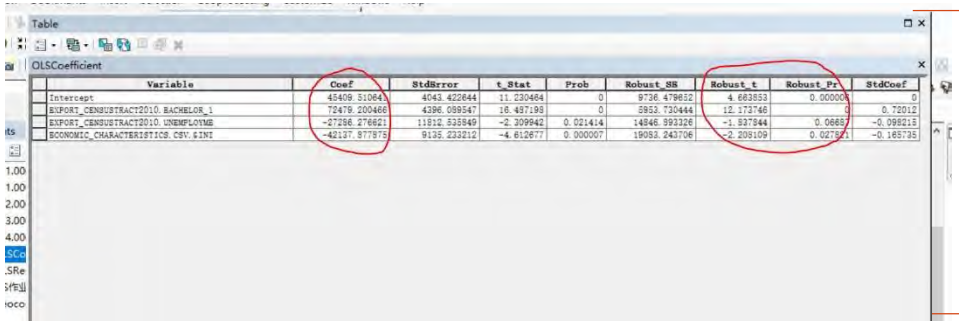


Step 2. Run the OLS Regression by ArcToolbox → Spatial Statistics Tools → Modeling Spatial Relationship → Ordinary Least Square. Set median household income as dependent variable and unemployment rate, Gini Index, Percentage of people who get bachelor degree or higher as explanatory variables. Then, I got the estimated value, residuals and standard deviations.



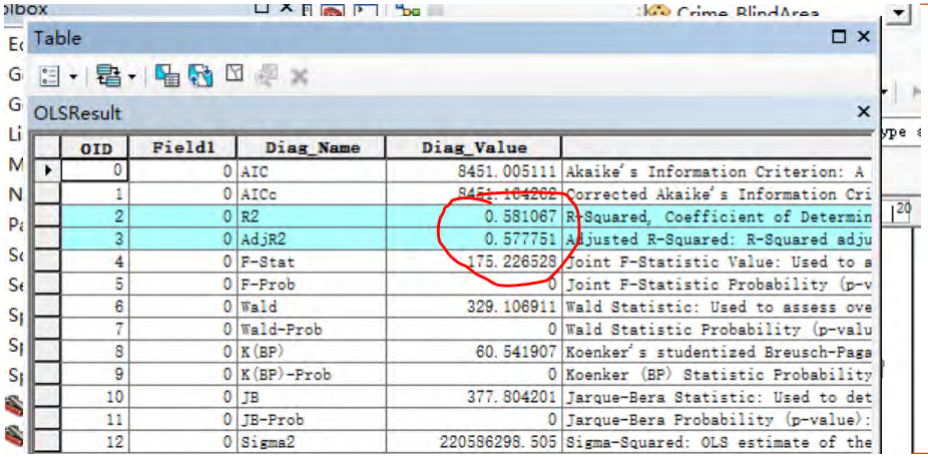
5.4 ECONOMY (INCOME): REGRESSION

- OLS Coefficient Output Table



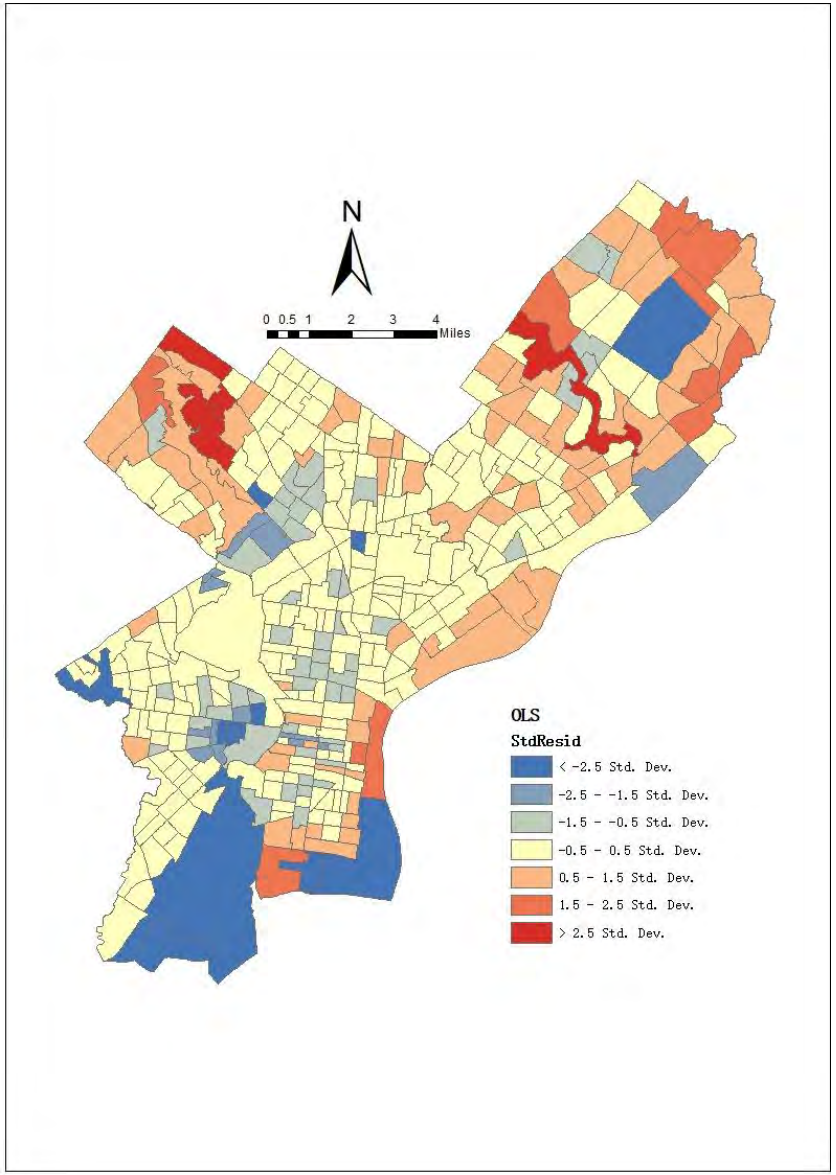
From the OLS Coefficient Output Table, I got the coefficient of the explanatory variables to the dependent variable, which shows that unemployment rate and Gini Coefficient have largely negative relations to median household income, but the education attainment has a largely positive relation to income. What's more, I also got the values of robust standard error, robust T-test and robust probability in the table.

- OLS Diagnostic Output Table



From the OLS Diagnostic Output Table, I got the values of R-square and Adjust R-square which are both around 0.58, which means that about 58% of dependent variables data could be explained by the explanatory variable data. Thus, to some extent, the independent variables fit the model well, and it's a good model.

5.5 ECONOMY (INCOME): REGRESSION

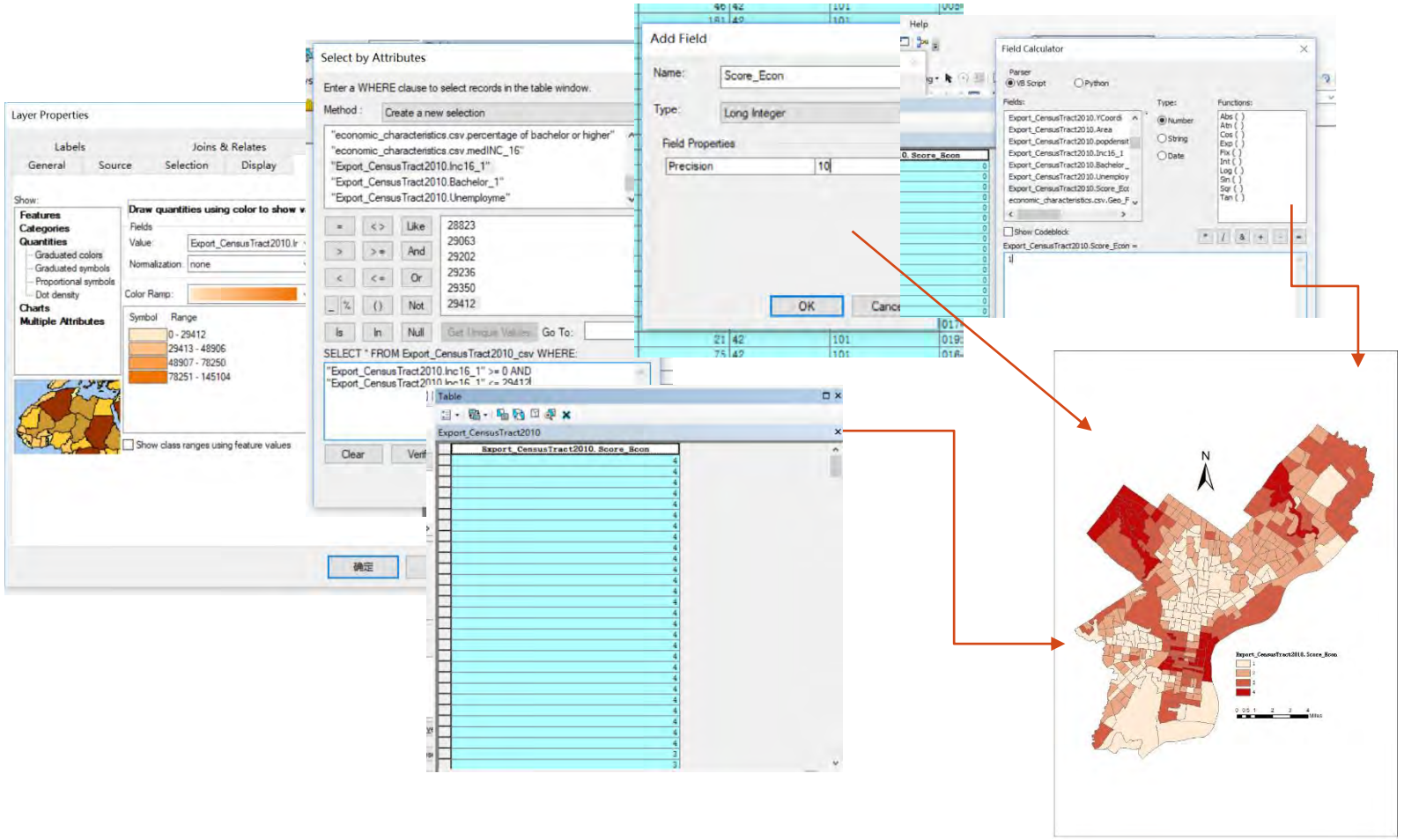


Summary

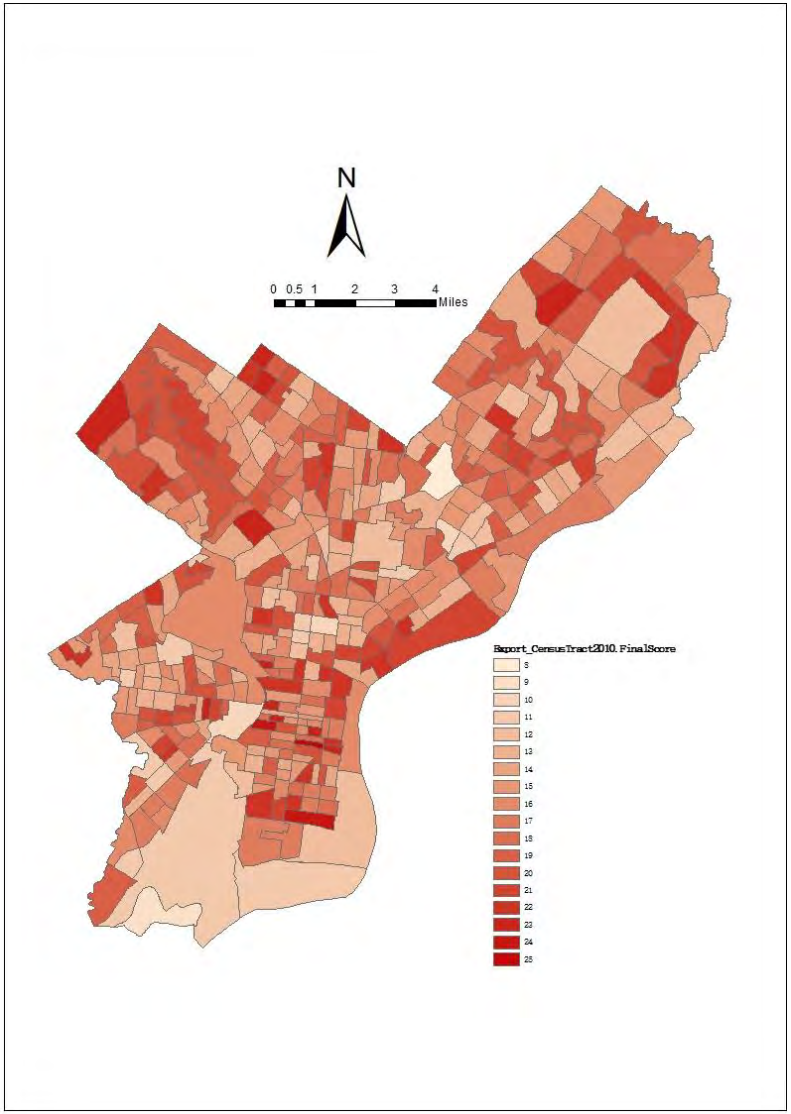
It's the outcome map for OLS regression! The symbology is based on the distance (how many standard deviations) between the data to the mean value.

5.6 ECONOMY (INCOME): SCORING

As did before in the section of crime and accessibility to bus shelters and schools, I firstly use layer properties to determine the standard of scores. Then, I add a new field for scoring in the attribute table and then use select by attribute and field calculator to give scores to the field. Finally, based on the scores, I change the symbology for the map.



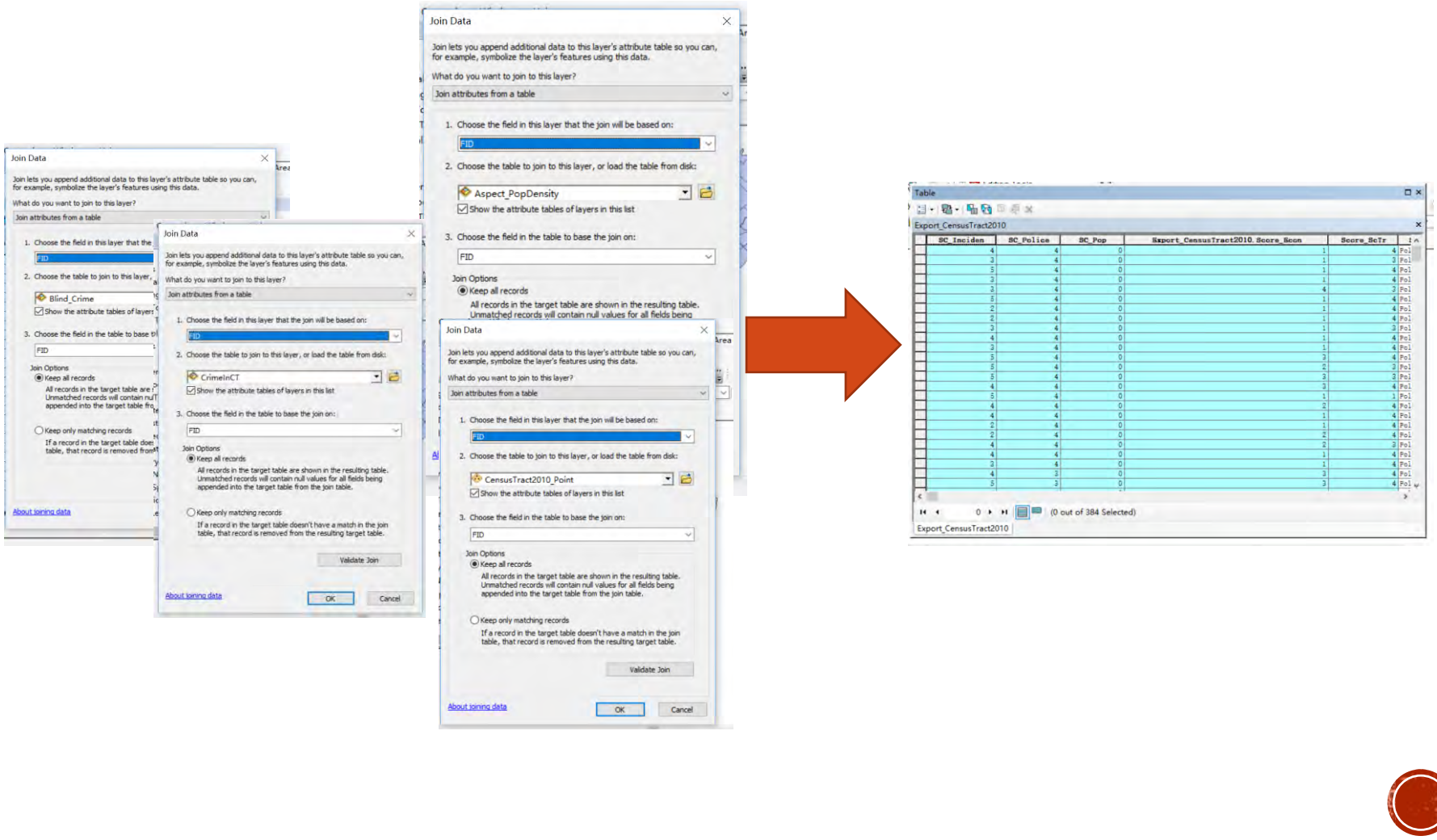
LAYER 6. RESULT (FINAL SCORE)



In this section, I would combine the scores of the 5 topics I discussed before to generate the final score for opening a new shop within each census tract. The higher the score, the better to open the shop.

6.1 RESULT (FINAL SCORE)

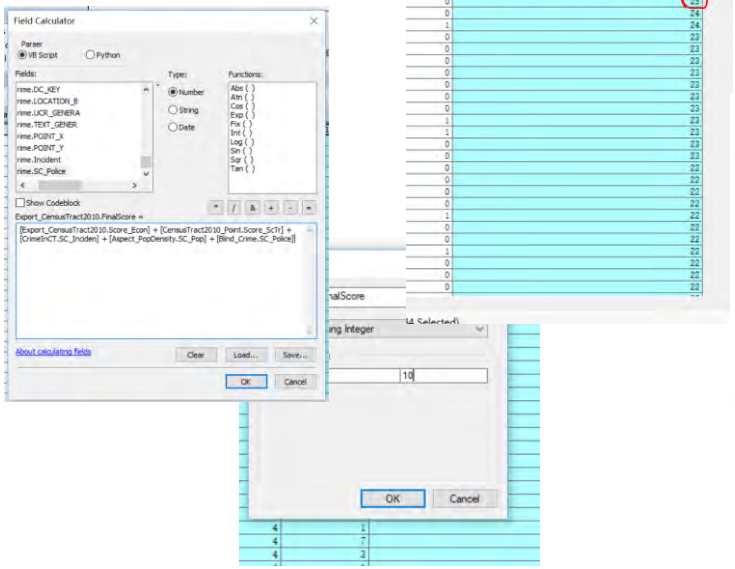
Step 1. Join all tables with related scores to the attribute table census tract shapefile to get scores of Crime Incidents, Population Density, Accessibility to Bus Shelters and Schools, Police Power (Police Stations) and Economy (Median Household Income).

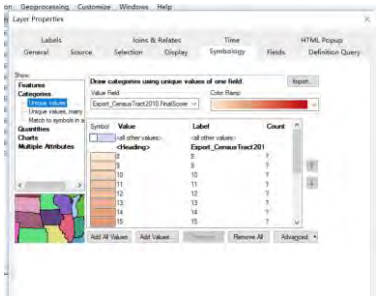


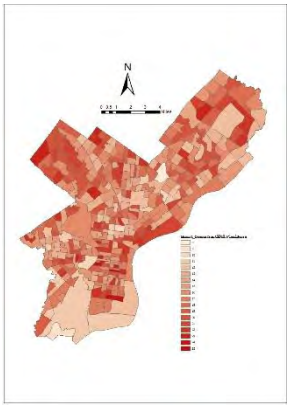
6.2 RESULT (FINAL SCORE)

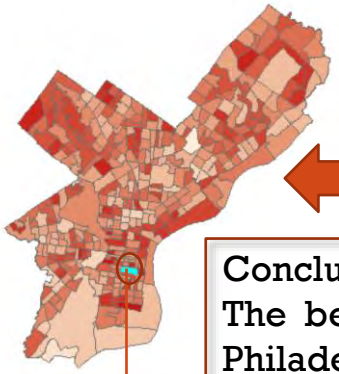
Step 2. Add a new field for Final Score. Then, use Field Calculator to get the sum of the scores of these 5 classes, which is the final score of each census tract. According to the range of each score criteria, the range of the final score would be 4-26. The highest score among all census tract is 25 and only one census tract get this score!

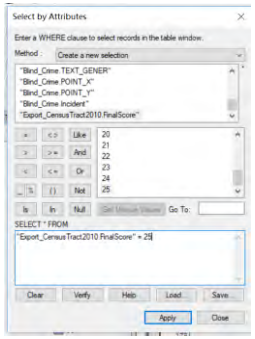
Step 3. Based on the final score, properties → symbology to make a good map for the Final Score. After that, use Select by Attribute and set the final score = 25 to find the location of that "best" census tract for opening a new shop.



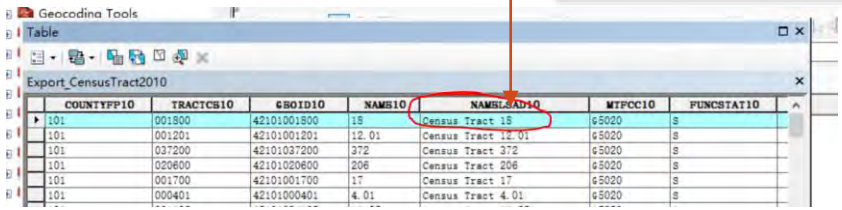






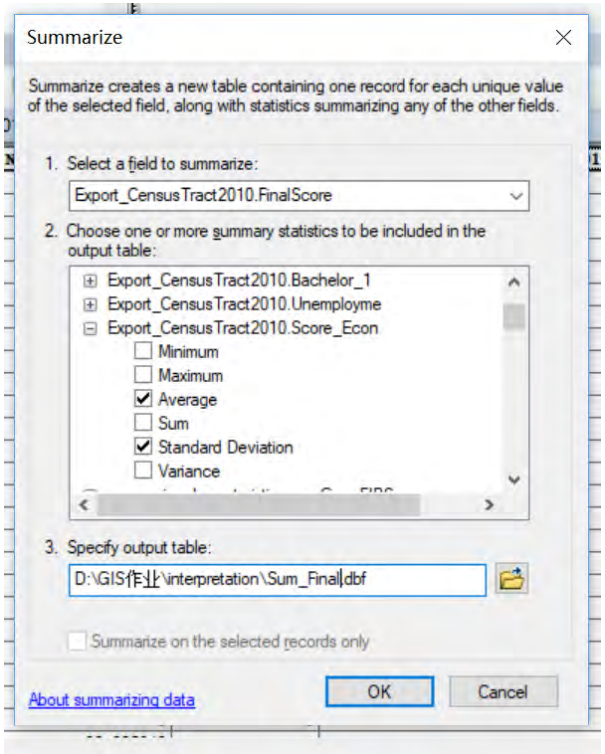


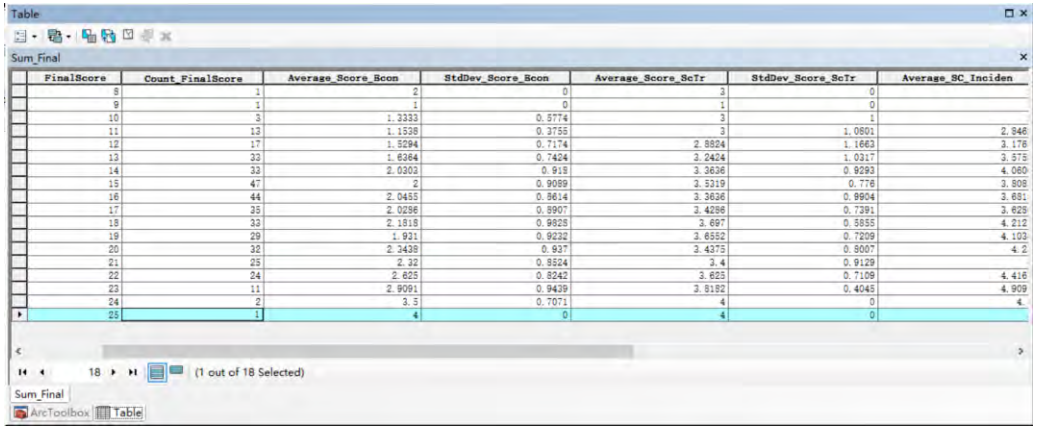
Conclusion:
The best census tract in Philadelphia for opening a new shop is **Census Tract 18**!!!



6.3 RESULT (FINAL SCORE): SUMMARIZE

Finally, I want to statistically **summarize** the output of the final score. Right click on the head of the field of final score → summarize to get the mean values and standard deviations of the 5 scores, and also the situation of the final score. What's more, from the field Count_FinalScore, I get the information as before that there's only one census tract which get the highest score: 25.





Thanks for Reviewing!

Zizhao He

Source: Wikipedia