Jianjin Chen
Professor Burnicki
GEOG3500Q
27 May 2021

Geographic data analysis report

Introduction:

This Geographic data analysis report is the successive part of the exploratory spatial analysis report published in Mar.14. The previous exploratory report focuses on the State of Connecticut's electric vehicle registration situation. The previous report comes up with three major founding. Firstly, the distribution of electric vehicle registration in Connecticut displays an unimodal skewed trend. In the other words, electric vehicles are fair behind a popular choice among Connecticut citizens. Secondly, the percent of electric vehicle owner are related to median household income. Thirdly, the relationship between electric vehicle registration and charging stations count is uneven. Specifically, only 50% of Connecticut Zip Codes have charging stations and 89% Connecticut Zip Codes have electric vehicle registration record.

Research Questions:

The previous report left three questions and this report focuses on answering those questions.
Question1: How does intensity of charging stations vary across the state? To answer this question , this report should focus intensity variation within the Connecticut territory.
Question2: Trying to further analysis localized spatial pattern of median household income and electric vehicle registration. The previous exploratory report indicated the percent of electric vehicle owner are related to income. This report should proving the previous hypothesis through geographic data analysis.
Question3: Trying to find the relationship between count of charing stations and number of electric vehicle registration. The previous exploratory report pointed out relationship between electric vehicle and charging stations is uneven. This report will analysis the potential relationship between them through geographic data analysis.

Methods:

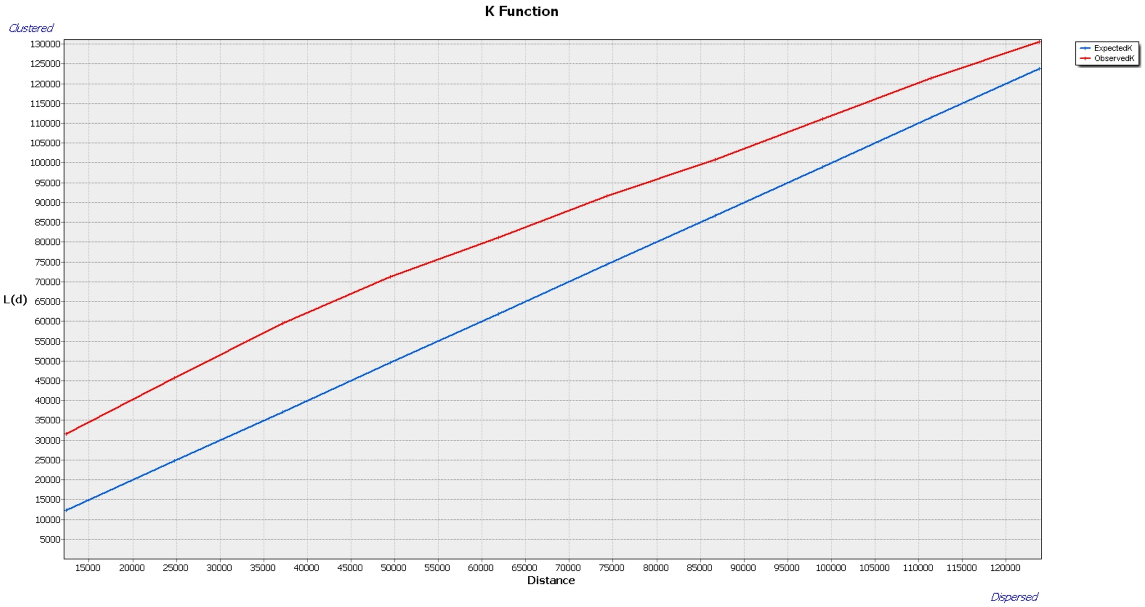
For question 1, a point pattern analysis is necessary. The report utilizes ArcMap to execute the point pattern analysis. The first step is examining the first order properties. Kernel density tool in ArcMap is an ideal choice to examine it. Specifically, it is an intensity estimation of electric vehicle charging stations within Connecticut. Based on the dataset's area(The State of Connecticut), ArcMap recommends default cell size at 1563.42. The report modified the output cell size to 1500 in order to fitting flexible research bandwidth. The report tried three bandwidths: 15000, 30000, and 35000. After that, the report clipped these kernel density layers to fit the Connecticut area. The report consequently chooses 30000 bandwidth for best intensity estimate performance. The kernel density map is presented as Map 1 in page 3.

However, first order properties is not enough to accurately analysis charging stations' spatial pattern in Connecticut. So, the second step of point pattern analysis is second order properties.The report utilizes Average Nearest Neighbor tool in ArcMap as part of second order properties. The nearest neighbor analysis examines distances between events to assess pattern. In this report, the average nearest neighbor tool specifically compares observed average distance among charging station to expected average distance for random point pattern. The ArcMap tool generates a report based on default distance method (euclidean distance). Referring to ArcMap report, the null hypothesis is CSR, the observed average neighbor distance is 5006.89 feet and expected average neighbor distance is 11031.97 feet. The nearest neighbor index is 0.4539. So it suggests a clustered pattern. The Z-Score is -20.29, P-Score is 0.00. This outcome rejects the null hypothesis(random pattern). To conclude, the ArcMap report indicates charging stations are distributed in clustered pattern across Connecticut. Form 1 in page 2 represents the ArcMap report's nearest neighbor summary section. The third step is analysis the charging stations spatial pattern at a range of scales based on a Ripley's K function. The report chooses Multi - Distance Spatial Cluster Analysis tool in ArcMap in order to calculate the Ripley's K function. In the data preparation section, the report set number of distance bands to 10, compute confidence envelope to 0 Permutations. Form 2 in page 3 represents the Ripley's K function. The function indicates that charging stations are distributed in cluster pattern regardless the scale although it displays a trend from cluster to random.

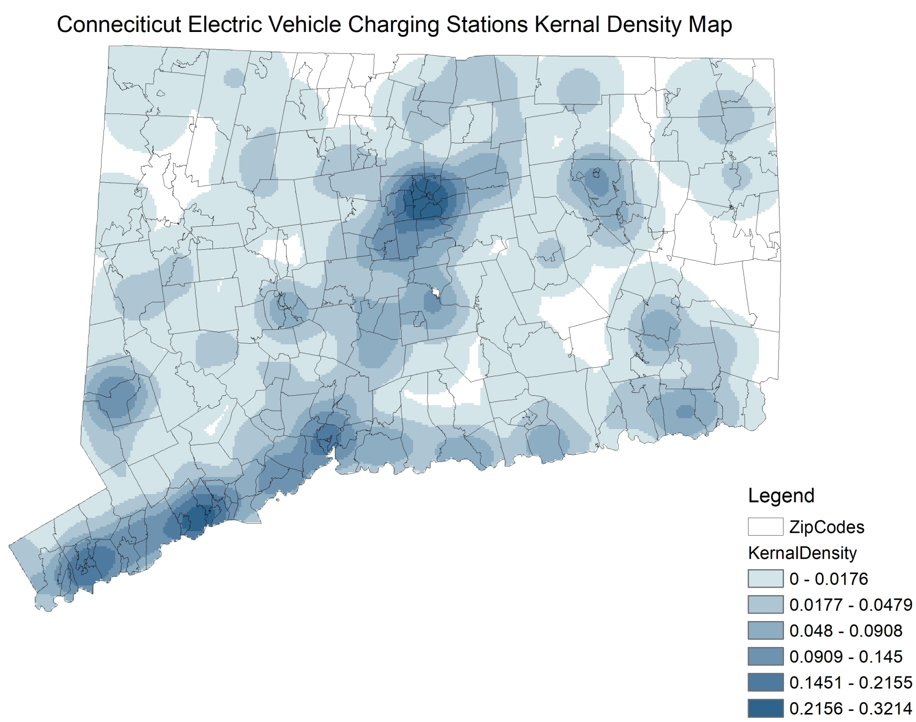
Average Nearest Neighbor Summary

Observed Mean Distance:	5006.8895 US_Feet
Expected Mean Distance:	11031.9651 US_Feet
Nearest Neighbor Ratio:	0.453853
z-score:	-20.286689
p-value:	0.000000

Form 1


K Function
Clustered
Dispersed
L(d)
Distance
ExpectedK
ObservedK

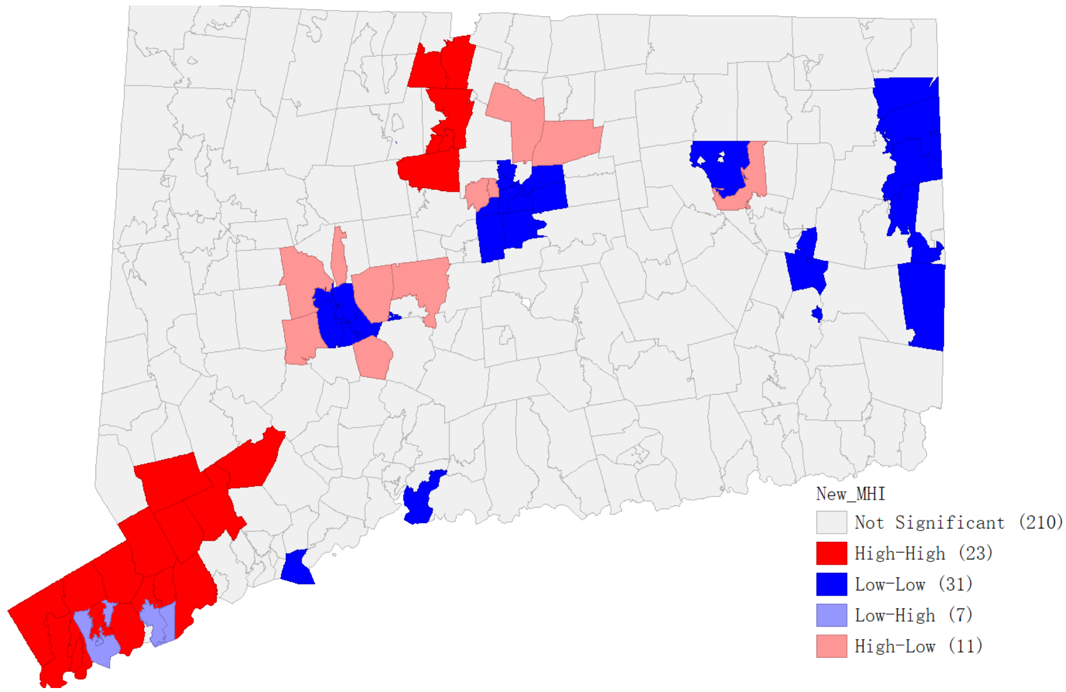
Form 2


Conneciticut Electric Vehicle Charging Stations Kernal Density Map
Legend
ZipCodes
KernalDensity
0 - 0.0176
0.0177 - 0.0479
0.048 - 0.0908
0.0909 - 0.145
0.1451 - 0.2155
0.2156 - 0.3214

Map 1

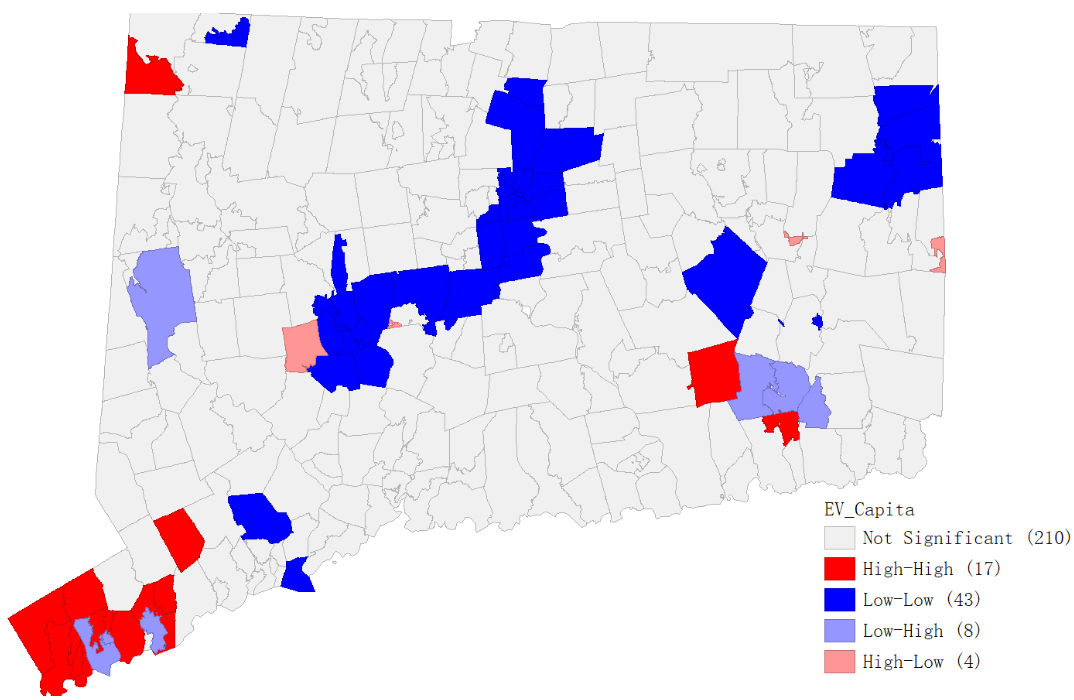
For question 2, a spatial autocorrelation and cluster analysis is necessary. The report utilizes Geoda to execute the spatial autocorrelation analysis. In order to proceed this analysis, the first step is constructing spatial weights matrices. It will determine the neighbors for each raster data cell. Neighbors the raster data cells sharing common boundary with a specific data cell, or raster data cells within certain distance with the raster specific data cell. In this report, the spatial weight matrices analysis will determine the neighbors for each Connecticut Zip code. As the result of contiguity based spatial weights matrices, such as Queen's contiguity or Rook's contiguity; a zip code with large area and its neighbors are likely to occupy much area. Meanwhile a small area zip code and its neighbors are likely to occupy little area. Because the Connecticut Zip codes are vary in area. Due to the same reason, Standardizing spatial weights matrices such as K nearest neighbor will lead to similar situation as contiguity based spatial weights matrices. The report need to standardize the total neighborhood area regardless the zip code area variation. So, a distance threshold spatial weights matrix is a better choose than other matrices mentioned above. The threshold distance is set to 36502.36. Which is the minimum distance to ensure every zip code have at least one neighbor. The second step is calculating Moran's I. The report uses a revised dataset instead of previous the dataset from previous exploratory report data sources. Because the new dataset modified an error in median household income section. Referring to the same calculation method in previous report, this report adds electric vehicle holding rate per capita section into the new dataset. The report choose Geoda Univariate Moran's I tool to execute the global Moran's I analysis. For the median house income section, the observed Moran's I value is 0.194. It is higher than expected Moran's I value(-0.0036). So the median house income has positive spatial correlation. Since Z-score is significantly positive, so values clustering spatially. For the electric vehicle holding rate per capita section, the observed Moran's I value is 0.1326. It is higher than expected Moran's I value(-0.0036). So this section has positive autocorrelation as well. The Z-score is also significantly positive, which indicates clustered spatial pattern. However, the global Moran's I analysis is not enough to uncover potential relationship between two sections. So the report runs local Moran's I for each section as the third step. It is an analysis method to detect significant clusters within study area. The report chooses Univariate Local Moran's I tool to detect clusters. Map 2 in page 5 represents clusters in the median house income section and Map 3 in page 5 represents clusters in the percent of electric vehicle owner per capita section.

Median Household Income Clusters in Connecticut



Map 2

Electric Vehicle Holding Rate Per Capita in Connecticut

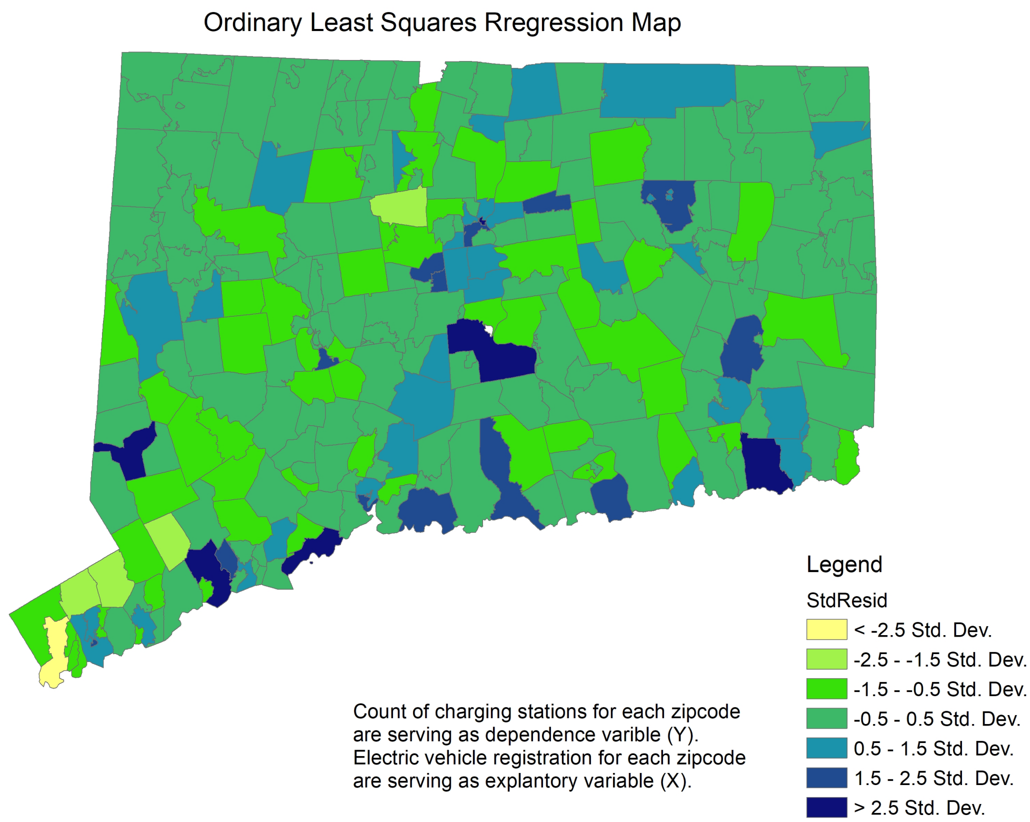


Map 3

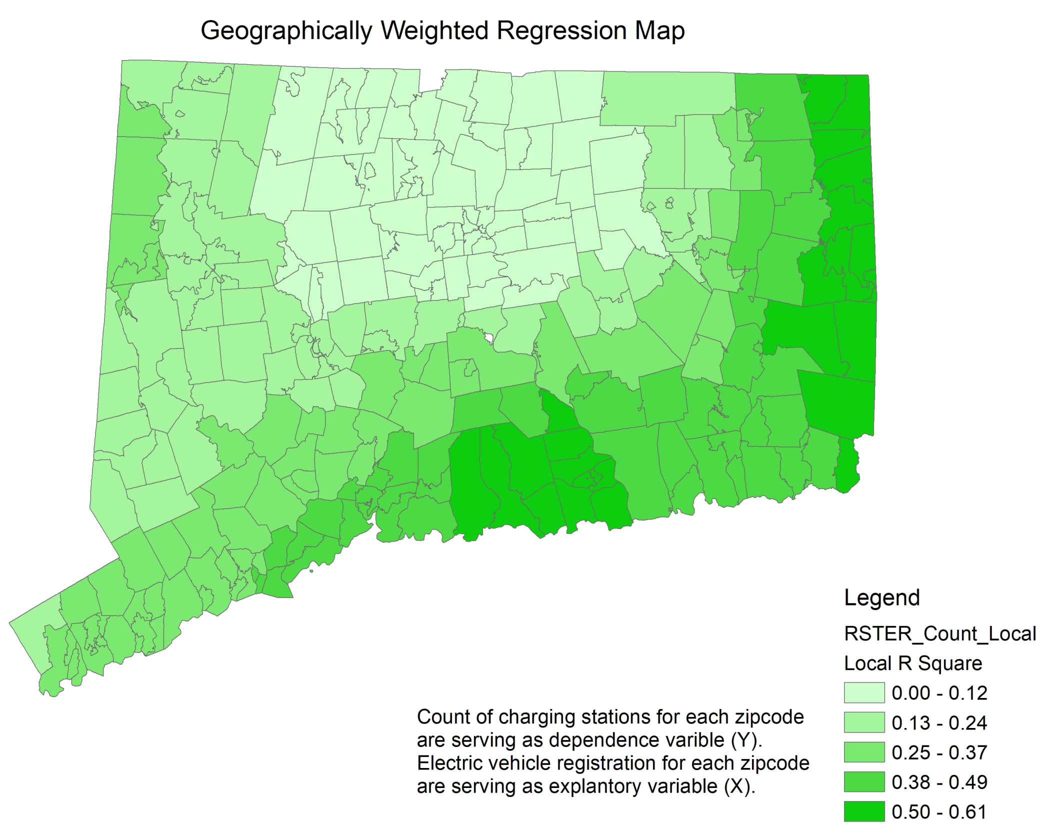
For question3, a regression analysis is necessary. Regression is a bivariate analysis examines how values of one variable are related to those of another variable[1]. One variable is response variable(Y) and the other is explanatory variable(X) The regression analysis assumes their relationship as a regression line Y = a + bX. "a" refers as intercept, which is the point on the line which intersect with the vertical axis and "b" refers as slope is the value change intensity of Y associate with one unit change of X. In this report, the two variables are electric vehicle registration value for each zip code and charging stations count value for each zip code. The report expects the electric registration is the explanatory variable and the charging stations as the dependence variable. This decision is based on a rational reasoning: high electric vehicle registration creates high charging stations demand. In the other words, the count of charging stations for each zip code is depends on the electric registration value. The first step of regression analysis is the ordinary least squares regression analysis. This analysis assumes a regression line which is expected to has a points set with smallest squared errors sum. In the other words, the line should close to observed points in order to have a small sum. The report utilizes ordinary least square tool to execute this process. Map 4 in page 8 represents the ordinary least squares regression analysis outcome. The analysis tool also generates a report regarding to these variables. Which calculates the R squared value among variables is 0.261644. It represents a poor relationship between variables. The ArcMap report indicates the bivariate relationship has little structure and seems random. However, the regression situation may vary depends across the general study area. So, the second step of regression analysis is execute a geographically weighted regression. Which a regression mode with a weights matrix for each location. The report utilizes geographically weighted regression tool in ArcMap to execute this analysis. This ArcMap tool generates a map to represent the local regression. The report modifies the default symbology in order to display the relationship more clearly. The report classifies the local R squared value into 6 classes with equal interval classify method as the map symbology. Map 5 in page 8 represents the geographically weighted regression analysis outcome.

---

[1] UConn Spring 2021 GEOG3500Q Course, Lecture Slides Session 12, part 1.


Ordinary Least Squares Rregression Map
Legend
StdResid
< -2.5 Std. Dev.
-2.5 - -1.5 Std. Dev.
-1.5 - -0.5 Std. Dev.
-0.5 - 0.5 Std. Dev.
0.5 - 1.5 Std. Dev.
1.5 - 2.5 Std. Dev.
> 2.5 Std. Dev.
Count of charging stations for each zipcode are serving as dependence varible (Y). Electric vehicle registration for each zipcode are serving as explantory variable (X).

Map 4


Geographically Weighted Regression Map
Legend
RSTER_Count_Local
Local R Square
0.00 - 0.12
0.13 - 0.24
0.25 - 0.37
0.38 - 0.49
0.50 - 0.61
Count of charging stations for each zipcode are serving as dependence varible (Y). Electric vehicle registration for each zipcode are serving as explantory variable (X).

Map 5

Results:

As the first result of this report, as well as the answer for question 1, the electric vehicle charing nations represents a clustered spatial pattern and vary intensity across Connecticut. Based on the point pattern analysis and Map 1, the kernel density of charging stations are vary. Stations are clustered in southeastern Connecticut and the central area within the state. Meanwhile there are a number of areas do not have any charging stations at all. The following second order properties analysis also proved the observed spatial pattern above.

As the second result of this report, as well as the answer for question 2, the percent of electric vehicle holding rate per capita is related to median household income in some Connecticut zip codes. However it is not a common relationship across the whole state. Based on the spatial autocorrelation and cluster analysis; Southwestern Connecticut actually represents as the high-high cluster areas for both median house income and percent of electric vehicle owner per capita although part of this region represents as low-high clusters for both sections. Which indicates that some Southwestern Connecticut zip codes with high median household income also has high electric vehicle holding rate although. There are also some zip codes in this area have low income and low electric vehicle holding rate compare to zip codes next to them. Some zip codes in Central Connecticut and Western Connecticut are represent as low-low clusters for both sections. Which indicates that these area have low median household income and low electric vehicle holding rate. However, there are also some zip codes with low median household income have relative high electric vehicle holding rate compare to their neighbor zip codes. To conclude, the question 2 hypothesis is proved true in some Connecticut zip codes. Yet the median household income is not the sole factor which influences electric vehicle registration.

As the third result of this report and the answer for question 3, the count of charging stations have weak relationship with electric vehicle registration. Referring to the ordinary least squares regression analysis, the ArcMap indicates the relationship is weak based on relatively low R squared value. Map 4 displays a number of zip codes with larger than 2.5 residual standard deviation. Meanwhile the regression model also fits to a number of Connecticut zip codes. The geographically weighted regression analysis also indicates similar spatial pattern. Referring to Map 5, there are a number of Connecticut town has lower than 0.24 local R squared value and the highest local R sacred value is 0.61. Based on these analysis results, the charging stations are related to electric vehicle registration in part of Connecticut. However it is not a common relationship across the state and a number of zip codes have little or zero relationship between those variables.

Conclusion:

The report notices that it is unlikely to have a common relationship or spatial pattern in large scale dataset; such as the state of Connecticut. The relationships are vary depends on specific location within study area. Meanwhile, it is also unlikely to analysis a variable in bivariates scale. One variable is unlikely to act as the sole element which leads to the change of the other variable. As a reminder, the question 1 expect to receive a specific answer about charging station distribution pattern instead of a general trend. The question 2 expects this report to prove a relationship which is noticeable through exploratory observation. The question 3 is inspired by a rational reasoning: an area with many electric vehicle is likely to have many charging stations. The report answered question 1; yet fails to proves the relationship in question 2 mention and also fails to prove the reasoning in question 3. To be instead, the report proves that median household income influences electric vehicle popularity in some Connecticut zip codes. The number of electric vehicles also influences the number of charging stations in some Connecticut zip codes. As an advanced answer for question 2 and 3, the report indicates the number of electric vehicle registration for each Connecticut zip code is not solely influenced by one sector. To be instead, the value is impacted by multiple factors at various level. Answering three questions are far behind the comprehensive analysis. Metaphorically, there are lots of missing parts of this electric vehicle puzzle.