George Chen

Homework 3: Maryland Analysis

I decided to choose Maryland to research, because I grew up and spent all my life here. After I added the MD layer to my project, my goal was to narrow down the fields to housing. The reason I chose housing expenses for households is because traditionally, the DC, Maryland, Virginia areas are very sought after locations in terms of housing and also have a very diverse range of areas such as city, urban, rural, suburban, etc. I also desire to own property in Maryland in the future, making the housing expense field even more relevant to me.

My research question is whether or not there is a random pattern of housing expenses per household in Maryland. Personally, I believe this not to be true as, based on my experiences, the urban hub and surrounding areas of DC that borders Maryland traditionally have higher costs, and more densely situated housing.

My first step was to utilize Spatial Autocorrelation (Global Moran's I) in order to find the global correlations of all of Maryland's Housing expenses. I did this by searching up spatial autocorrelation in Toolboxes. I selected the specified housing field layer as my input feature class, chose 2022 Housing as my input field, and made sure to check 'generate report' before clicking run. This gave me the Spatial autocorrelation report of housing expenses in Maryland.

Based on the report, I determined that we reject the null hypothesis, (there is a random pattern of housing expenses in Maryland) because there is a less than 1% likelihood that the clustered pattern of houses are the result of random chance. The report gives a Moran's index of 0.246, which is >0 (0 implies complete random dispersion), meaning there is indeed a clustering of similar values. The report also gives a z-score of 46.46%, which means that the values are within 3 standard deviations of the mean, meaning even less of a chance that the clustered patterns are random. Finally, the report gives a P-value of 0.0, which implies to reject the null hypothesis that there is a random pattern of housing expenses in Maryland.

My second step was to utilize Cluster and Outlier Analysis (Anselin Local Moran's I) in order to identify clusters of High and low values as well as outliers. I did this by searching up Cluster and Outlier Analysis in Toolboxes. I selected the specified housing field layer as my input feature class, chose 2022 Housing for my input field, changed output feature class to Maryland_CT_Housing_Expenses_2022_Local_MoransI, and pressed run. This gave me a new layer with the aforementioned feature class as the title. The layer contained 5 different types of clusters, in terms of High High, High Low, Low High, Low Low, and not significant areas. I chose to focus on High high (High spatial auto-correlation with high housing expenses) and low low areas(High spatial auto-correlation with low housing expenses).

The High-High areas predominantly surround the DC border, with the areas adjacent to NorthEastern DC having the largest chunk of High High area. As you move to the right and

circle around DC, the High High areas gradually thin out, with the areas adjacent to southeast and southwest DC having substantial amounts of statistically insignificant areas closer towards DC. I believe this is the case, because DC is traditionally more affluent towards the Northeast and Northwest, becoming less affluent in the south east, southwest areas, thus matching how the housing expenses are clustered, affirming that there is correlation and not randomness.

Moving towards the City of Baltimore and it's surrounding area, the High High clusters are predominantly located adjacent to Northern Baltimore, and spanning to the right, merging with some of the DC metropolitan areas. The area to the west of Baltimore seems to have far more statistically insignificant areas. I believe this is due to Baltimore's proximity to DC, as there is not a major city towards the west of Baltimore, but DC is around 40 miles away. Again, this affirms that there is a correlation in terms of housing expenses in the area that is not due to randomness.

Looking at Low Low areas, the predominant areas in which are highly correlations of low housing expenses are located on the Eastern shore as well as Western Maryland (close to Appalatia). These areas are the furthest from the urban hubs of Baltimore and DC, which I believe is a big reason why these areas have less housing expenses. Eastern Maryland as well as western Maryland have large swaths of agricultural land, indicating less economic growth compared to more urban areas of Maryland.

My final step was to use Optimized Hot Spot Analysis in order to get an unweighted layer of correlated housing expense areas in Maryland. For the input feature I chose the original aforementioned housing field specified layer, for the output feature, I typed: Maryland_CT_Housing_Expenses_2022_OptimizedHotSpotAnalysis1, and for my analysis field I again chose 2022 Housing. Pressing run, a new layer was created with Cold spots that determine areas of low housing expense at 99, 95, and 90 percent confidence, as well as Hot spots that determine areas of high housing expense at 99, 95, and 90 percent confidence. The results gave me a similar map to the Cluster and Outlier Analysis layer, except more refined/detailed. The majority of the DC metro area in Maryland seems to be hot spots with 99% confidence while all Western Maryland seems to have cold spots with majority 95-99% confidence. Again, this affirms the rejection of my null hypothesis that the clusters of housing expenses are random, as there is very high confidence that urban areas are hot spots and rural areas are cold spots.

Finally, looking at the Moran's scatterplot, the scatterplot supports meaningful clustering, because most of the values are on the positive side of the z-score field. If most z-values are

positive, it implies less of a chance that these clusters were random, thus further affirming my null hypothesis.

I conclude that, based on my research using ArcGis Pro and the Maryland database, the pattern of housing expense clusters in Maryland is not based on random chance, but rather a reflection of the economic standings of each area. Based on the local Moran's data, the cluster areas closer to DC and Baltimore (somewhat) tended to have higher housing expenses while the clustered areas further from DC and Baltimore, such as Eastern and Western Maryland tended to be more rural and have lower housing expenses. This is also accurately proven by the Hotspot layer in which we see the same areas of significance from local Moran’s I have a 95-99% confidence in terms of the likelihood the cluster data is accurate.

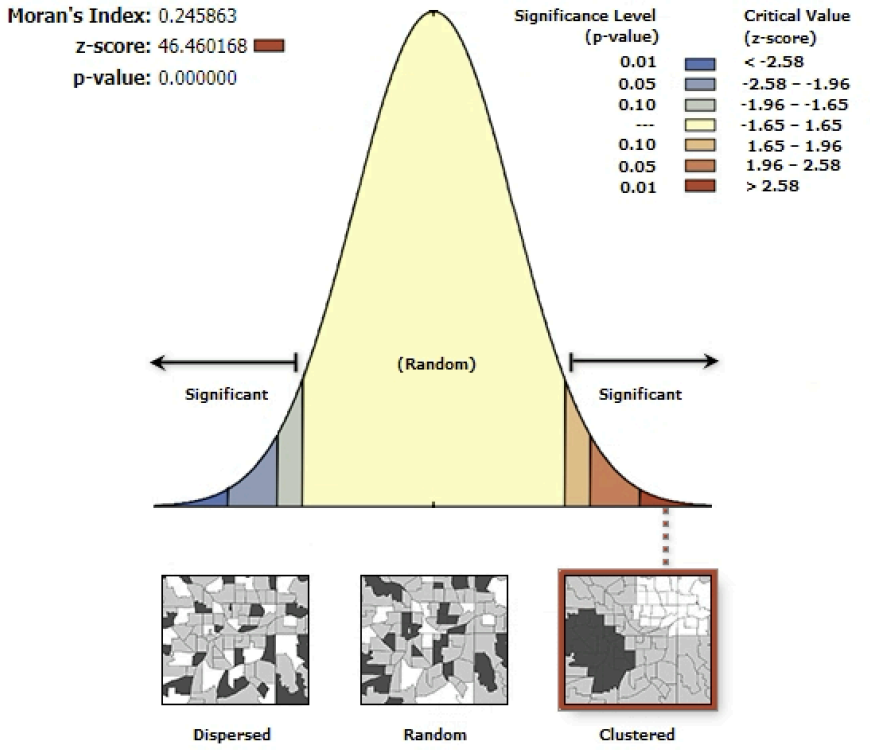
George Chen

Homework 3: Maryland Analysis

Maps:

Global Moran's I Spatial Autocorrelation Report:

Spatial Autocorrelation Report



Given the z-score of 46.460168, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

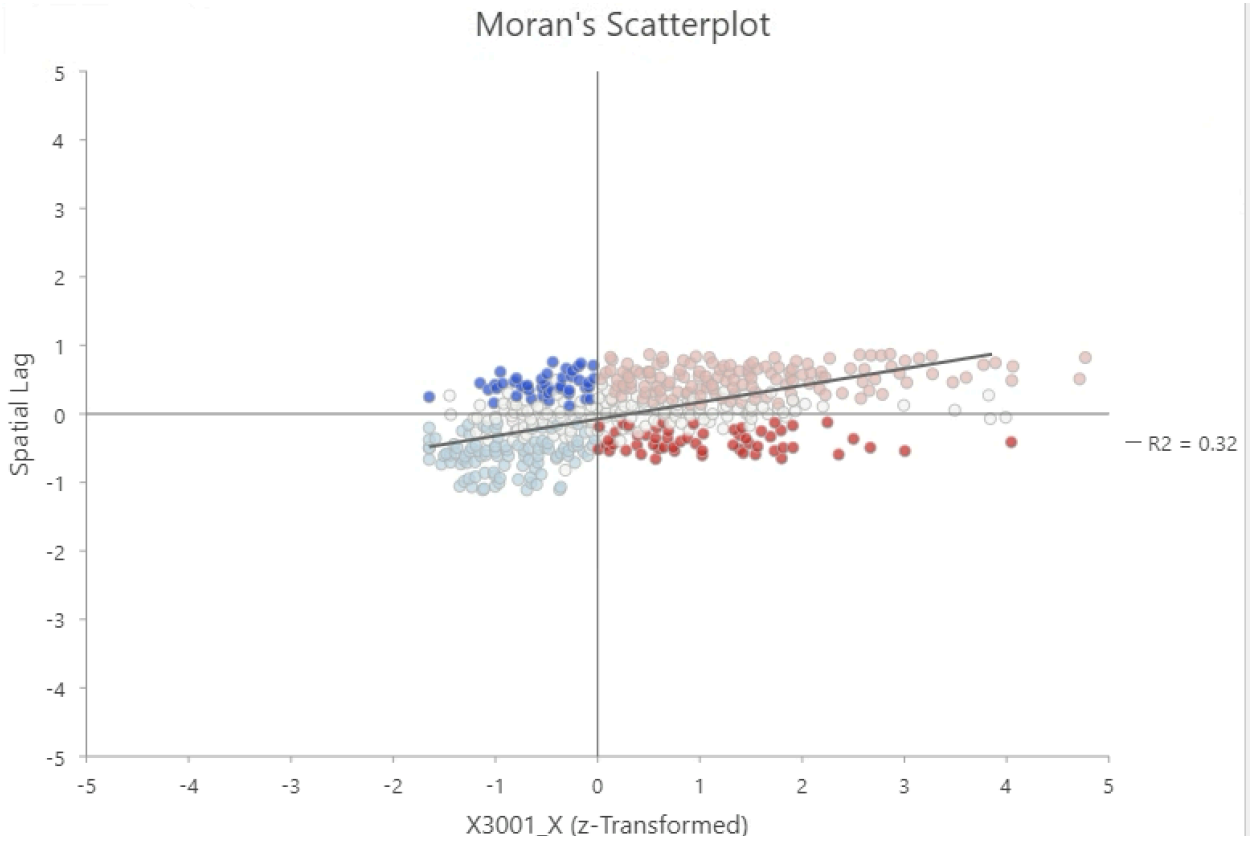
Global Moran's I Summary

Moran's Index:	0.245863
Expected Index:	-0.000684
Variance:	0.000028
z-score:	46.460168

George Chen

Homework 3: Maryland Analysis

Scatterplot of Local Morans I:



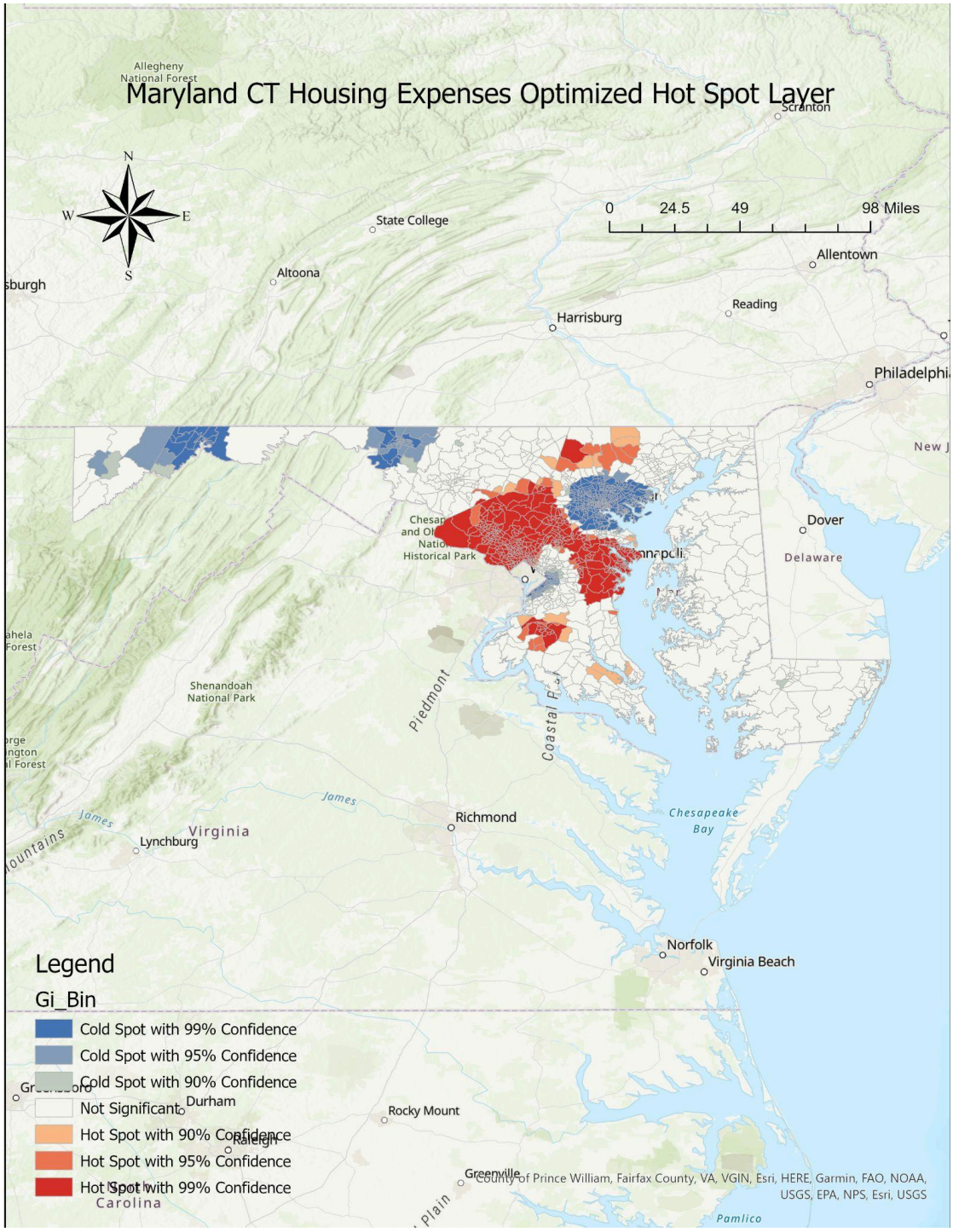
George Chen

Homework 3: Maryland Analysis

Hot Spot Layer:

George Chen

Homework 3: Maryland Analysis



Cluster/Outlier (Local Moran's) Layer:

George Chen

Homework 3: Maryland Analysis