Analysis author: ***Kayla Aburida.***

Title: Lab 1: Average ***Nearest Neighbor Analysis: Seattle crime***

Dataset source: ***Seattle Metropolitan Police Department***

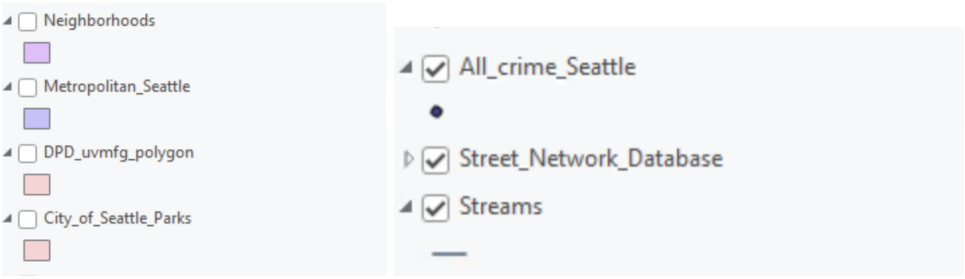
Dataset location: ***City of Seattle***

Date(s) of analysis: ***September 17th - ,2023***

Dataset time span: ***8/11/2017 - 9/11/2017***

Dataset Scale: ***City of Seattle***

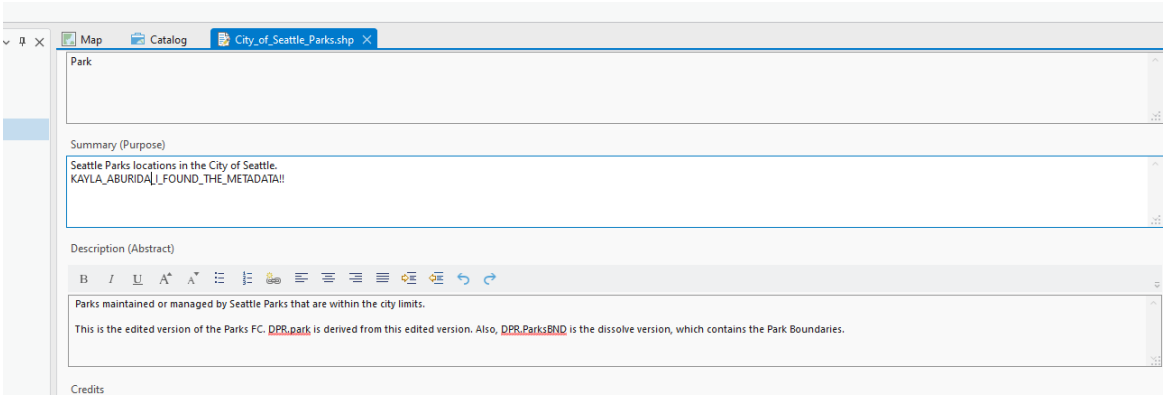
1. The layers in the map include, All_Crime_Seatlle, which includes all crime that took place 8/11/2017- 9/11/2017 in the city of Seattle, Street_Network_Database, Streams, City_of_Seattle_Parks, Metropolitan_Seattle, DPD_uvmfg_Polygon, and Neighborhoods. Once the lab was complete there was also All_crime_Seattle_Intersect, ANIMAL COMPLAINTS, ASSAULTS, ManufacturingAreas, and ManufacturingAreas_Buffer.



2. Point behavior changes as you are looking at the data at different scales. For example if you are zoomed into street level, you won't likely see clustering in this lab's data set. Although if you are looking at a couple neighborhoods at once, you will likely see clustering in this data set.

3. The Street Network Database (SND) is a representation of the physical, land based, improved, travel pathway infrastructure and associated theoretical address ranges. The pathway representation consists of Segments and Intersection elements. A Segment is a linear graphic element that represents a continuous physical travel path terminated by path end (dead end) or physical intersection with other travel paths. Segments have one street name, one address range and one set of segment characteristics. A Segment may have none or multiple alias street names. Segment types included are Freeways, Highways, Streets, Alleys (named only), Railroads, Walkways, and Bikelanes. SNDSEG_PV is a linear feature class representing the SND Segment Feature, with attributes for Street Name, Address Range, Alias Street Name and segment Characteristics objects. Part of the Address Range and all of Street Name objects are

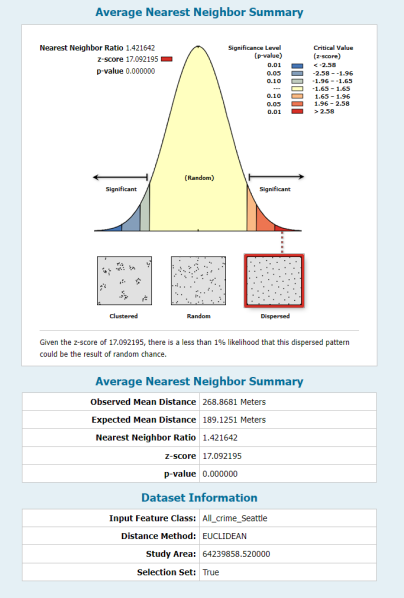
logically shared with the Discrete Address Point-Master Address File layer

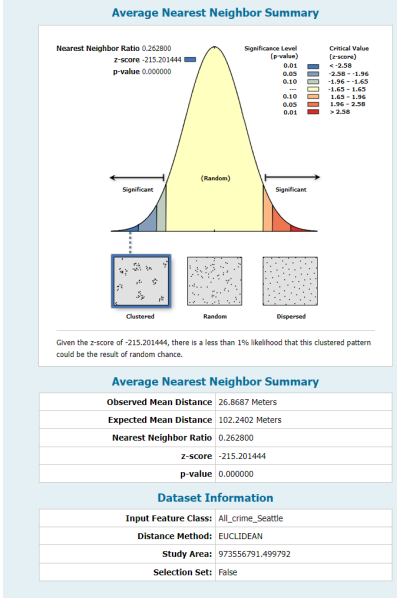


4. The Average Nearest Neighbor tool is a spatial statistics tool that analyzes patterns. Specifically, this tool calculates a nearest neighbor index based on the average distance from each feature to its nearest neighboring feature. The ANN tool computes 5 different outputs, the observed median distance, the expected distance, the nearest neighbor ratio, the z-score, and the p-value. Observed median distance (OMD) is if you were to compute the Mean Center for a compact cluster of points, the result would be a location at the center of the cluster. The expected median distance (EMD) is the average distance between neighbors in a hypothetical random distribution. The nearest neighbor ratio is the observed average distance divided by the expected average distance (with expected average distance being based on a hypothetical random distribution with the same number of features covering the same total area) ANN=Do/DE. Z-scores (aka, a standard score) indicates how many standard deviations an element is from the mean. If, for example, a tool returns a z-score of +2.5, you would say that the result is 2.5 standard deviations. The p-value is a probability. For the pattern analysis tools, it is the probability that the observed spatial pattern was created by some random process. When the p-value is very small, it means it is very unlikely (small probability) that the observed spatial pattern is the result of random processes, so you can reject the null hypothesis.

5. ANN reports

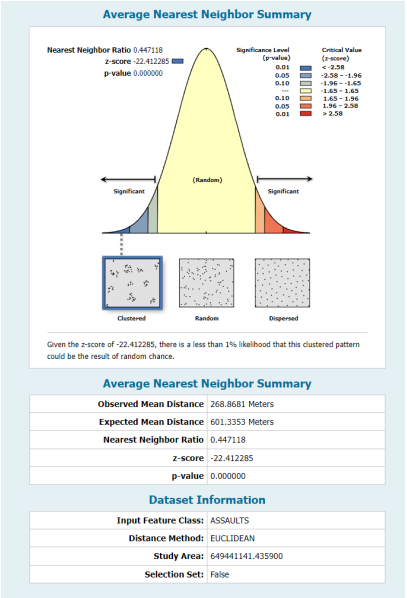
All_crime_Seattle

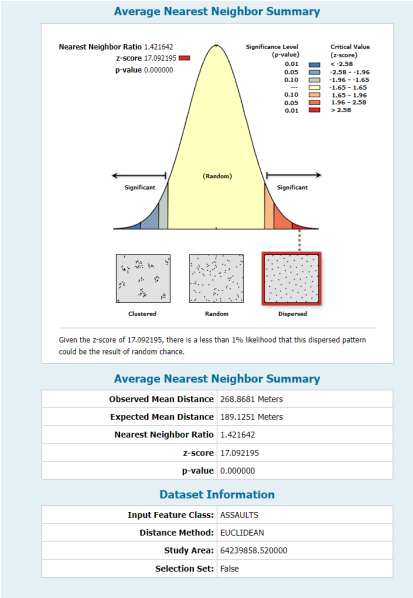




I thought that All_crime_Seattle looked pretty clustered over certain scales. For example near Elliot Bay I thought there was significant clustering. I assume the areas that I see clustering are the downtown parts of seattle. In terms of the All_crime_Seattle data with the new area number used, I am curious about the ellipsoid that this area created, because the fact that ANN is registering clustered with such a low p-value is extremely out of the realm of what I think is reality. This study is only one decimal place smaller. Again I think there are pitfalls here.

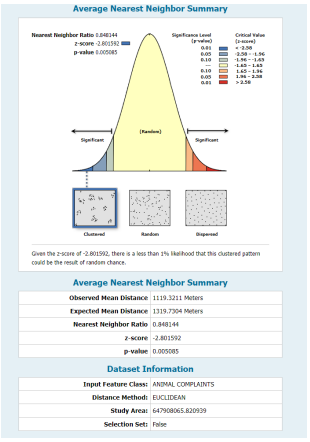
ASSAULT

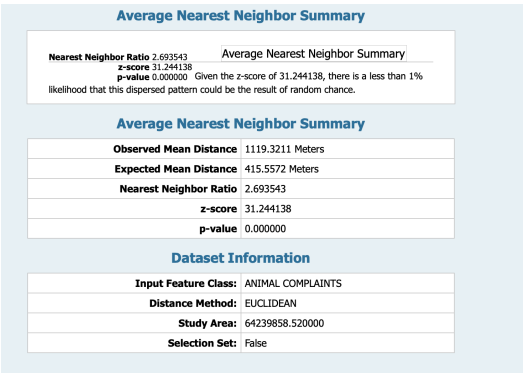




Just visually, assault definitely looks clustered around Elliot bay. Therefore it makes complete sense to me that the ANN function also registered the data as clustering as well. Alternatively, ASSAULT with the new area number used seems with this new area the ANN function is not *functioning*. Assault was definitely centered around Elliot bay. Again, for the p-score to be at zero, the new area must be messing up this function.

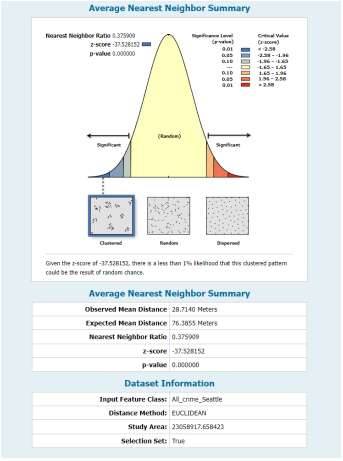
ANIMAL COMPLAINTS

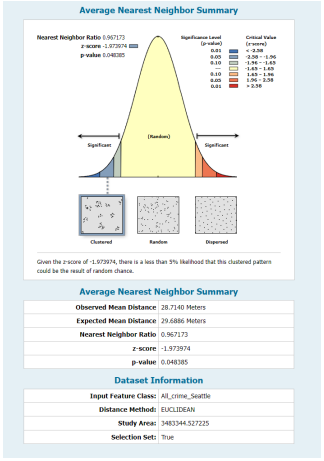




In ANIMAL COMPLAINTS, I can see clustering around elliot bay and Brighton neighborhood. Therefore, I agree with the ANN process to describe Animal complaints in Seattle as clustered. The ANIMAL COMPLAINTS with the new area number used ANN report shows dispersed, but again I think that this is part of the caveat of adding an area into the ANN function, I'm not entirely sure why this is, but it is incorrect.

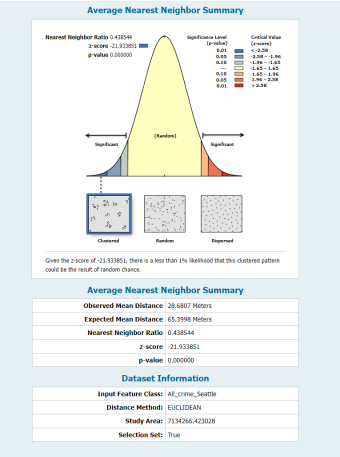
QUEEN ANNE

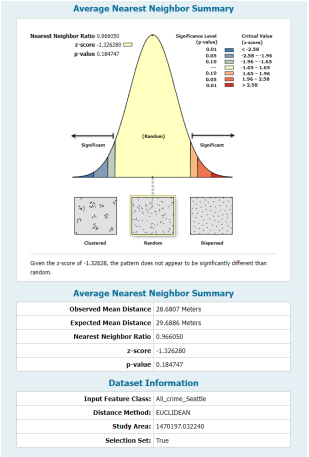




The ANN report of the Queen Anne details that Crime in the Queen Anne neighborhood is clustered, which I agree with partially. For the most part the data points are clustered to the south of the Queen Anne neighborhood. Although the degree of clustering is not high since there are also a significant amount of crime data points outside of just the southern portion of Queen Anne. The QUEEN ANNE ANN report with the new area number used output clustered.This is the first time when using the ANN function with a new prescribed area the function gave a similar clustering pattern. I wonder why that is? I agree with this ANN report.

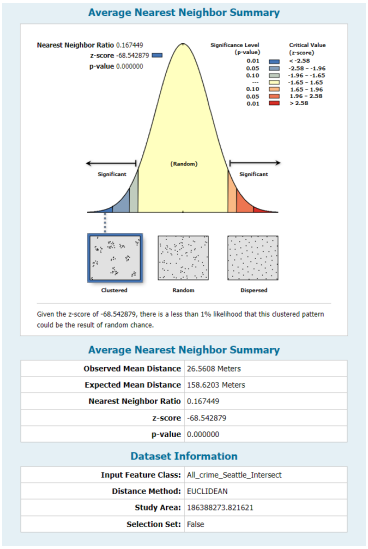
FREMONT





I thought that the data points in the Fremont neighborhood were somewhat clustered. What I found interesting about this pattern is that there was a “whole” of data points meaning one small area of this neighborhood is apparently devoid of crime data. I don’t know if I agree with the ANN function. The main clustering is near the manufacturing zones and the NE corner of the neighborhood. The FREMONT with the new area number used ANN report with this area size shows random dispersal. I also don’t agree with this unless I was zoomed into a scale in portions of Fremont, but because there is a “hole” of data in the Fremont neighborhood, I wouldn’t agree that the dispersal is random.

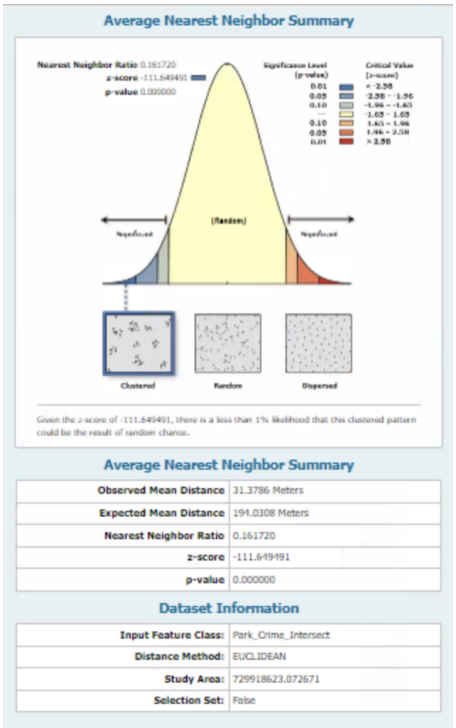
Buffered Manufacturing



I agree that the data from crime data intersected with the buffered manufacturing areas. The crime data is occurring on the outlying part of the manufacturing zones. I think this is because, no one is “reportedly” committing crime in the manufacturing zones, maybe because there are machines, or it's not a place where people can be. This is a huge drawback of the ANN function because in the manufacturing zones there's clustering likely because of land use. Though, the ANN function does not see this issue therefore it gives an output without the nuance of reality.

Buffered Parks

Because buffered parks are essentially like little islands, I'm not how the ANN function works here. I see high concentrations of crime in certain parks in central Seattle mainly near the western bay.



6. Conclusion

I saw various patterns in the crime data that would indicate clustering, random, or dispersed data without using the ANN functions. In instances like ANIMAL COMPLAINTS with the new area number used, the points looked clustered, but when I ran ANN resulted in a non clustered statistic. I believe this is because of the Area that Arc used while running this function. The topic of scale while running ANN is something we've talked about in class. Based on how zoomed into a data set you are you can be more or less clustered. Therefore I think in the instance I talk about above Arc used a scale that didn't represent the scale I was looking at the data with. There are many instances where this same sort of thing occurred, all where the new area gave seemingly the wrong report. ANN is a great function, for the most part. I do think though that because scope is a big part of it and there is not a normalized area function of the ANN function, the results the equation gets can be misleading. The equations used to calculate the average nearest neighbor distance index and z-score are based on the assumption that the points being measured are free to locate anywhere within the study area. ANN negates the areas within the study area that shouldn't have data at all that the equation is using. A common example is when looking at points on land in a coastal region, we typically wouldn't consider the

water to be part of the area we are studying but the ANN function doesn't have a way to negate that space. Therefore these ANN reports might indicate less or more clustering when in reality that's not the case. I observed that when you give the ANN function an area that it doesn't produce naturally, it can give a result that I don’t necessarily think is true. I don’t know the reason for this, but I assume it has something to do with the scope that the function is reporting on. I think the ANN function can be a very powerful tool in some cases. Although, there needs to be a decisive say when it’s used or not. For example, I wouldn’t use ANN to see if there were clustering of data points showing car accidents in a marshland with one road going through. ANN is better used when seeing if data’s clustering is significant or if the points are dispersed in a continuous area.