An Investigation of Spatial Distribution of Chicago Basketball Courts

Nicholas Liagridonis

Research Question: I am interested in gaining insight on the spatial distribution of basketball courts in Chicago, and investigating the reasons for the patterns present. My central question is, given the spatial distribution of basketball courts in Chicago, which areas exhibit notable patterns of court density? What factors may contribute to the spatial distribution of basketball courts across different community areas in Chicago, and are they primarily socioeconomic or demographic?

Data: By analyzing the density and clustering patterns of basketball courts, my study seeks to identify areas that demonstrate significant concentrations or variations in court availability, thus highlighting regions of potential interest for further examination and intervention. The factors influencing the distribution of basketball courts, such as demographic characteristics and socioeconomic factors, will be explored to understand their role in shaping access to basketball courts. The data for this study consists of the locations of basketball courts in Chicago, sourced from the Chicago Data Portal. Additionally, demographic, and socioeconomic data for Chicago Community Areas in 2020 were obtained from the American Community Survey via the CMAP data portal. The spatial scale of analysis will be at the level of Chicago Community Areas, and period of analysis will be 2020, as that is when the CMAP data was collected. The basketball court data will cover all outdoor courts within the Chicago area.

Hypotheses:

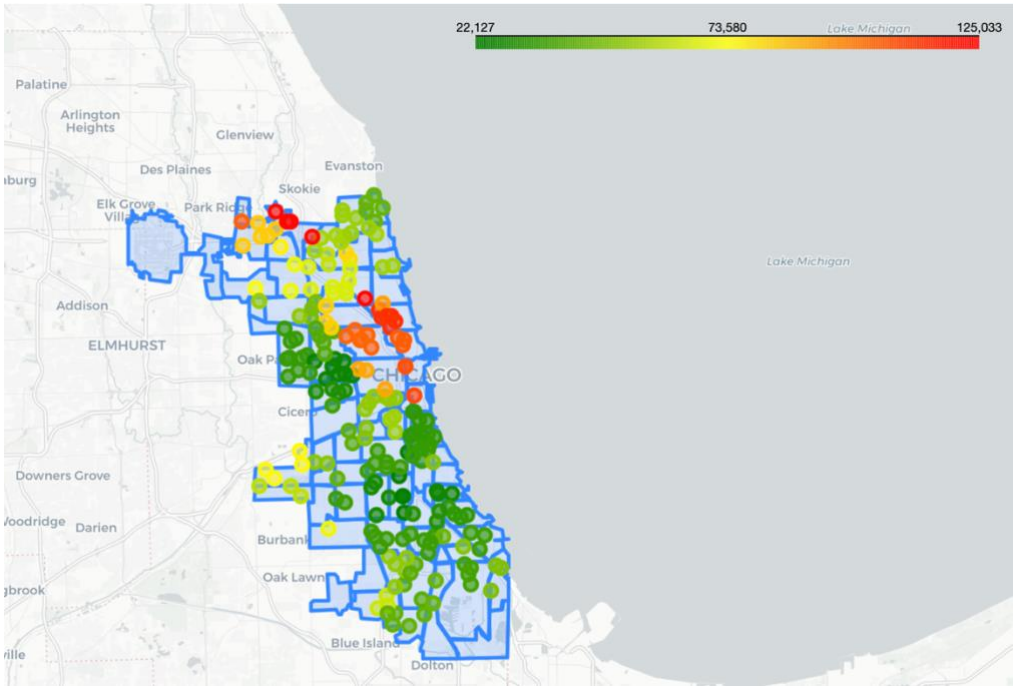
1. Community areas with higher median household incomes will have a greater density of basketball courts.

2. Community areas with a younger population demographic have a higher density of basketball courts
3. Areas with less internet access have higher densities of basketball courts

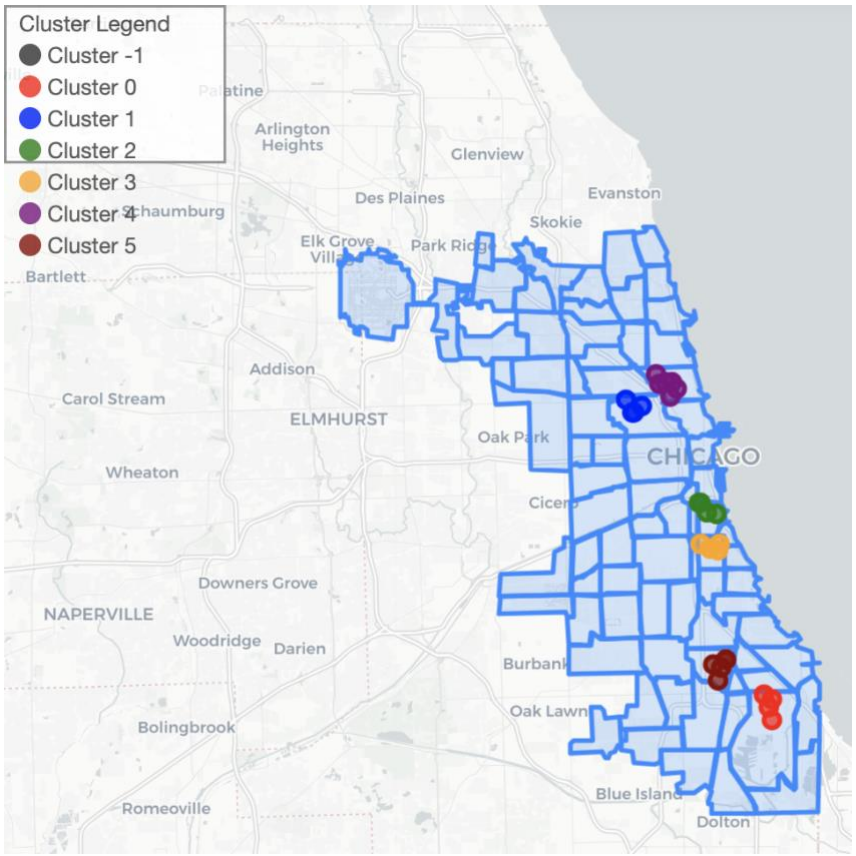
Analysis & Results:

I first completed some exploratory analysis. There are a lot of variables in the CMAP dataset, and after going through the long process of joining it with the basketball court data, I had a lot of ways I could start. I began by taking a look at the distribution of courts in Chicago, which yielded some interesting and also predictable results, the first being that the most courts were present downtown, but additionally, there was a fairly even distribution of courts across the city. Both these facts were helpful in informing the following studies. The first hypothesis I wanted to test was this: Community areas with higher median household incomes will have a greater density of basketball courts. I first investigated median household income in relation to the spatial distribution of basketball courts in the city. I first set 5 clusters, using K-means clustering initially to get a general overview of how income related to basketball court density.

Cluster	count	mean	std
0	24.0	112142.000000	7530.063277
3	15.0	84919.666667	4550.458920
4	24.0	67204.250000	4766.097221
2	61.0	48969.918033	5199.464352
1	89.0	31530.269663	4582.175738



It is clear based off the summary statistics for the clusters that as median household income increases, there is a general decrease in the size of the clusters. There are far more basketball courts in areas of lower median household income than the contrary, over 70% of the basketball courts in Chicago are in neighborhoods with a median household income below $50,000. There is an element of this that absolutely is affected by the socioeconomic state of the city, there are simply far more areas with median household incomes in the green territory than the red, and it is clear that the clusters are separated by this factor; you do not see much geographical mixing of the clusters. Standard deviation amongst these clusters did not vary considerably, even considering the fact that the mean household incomes were vastly different in each cluster. This is doubly apparent when looking at Figure 1. It is also apparent that as Chicago ventures North, the pockets of higher income pop up, with many of the parks in the South belonging to neighborhoods with lower median household incomes. The densest portions appear in the downtown areas, decreasing in density as you venture North and South. After seeing these results, I decided to try to employ DBSCAN to see which clusters may appear. I had to do quite a bit of feature engineering, starting off with far too many clusters, and then no



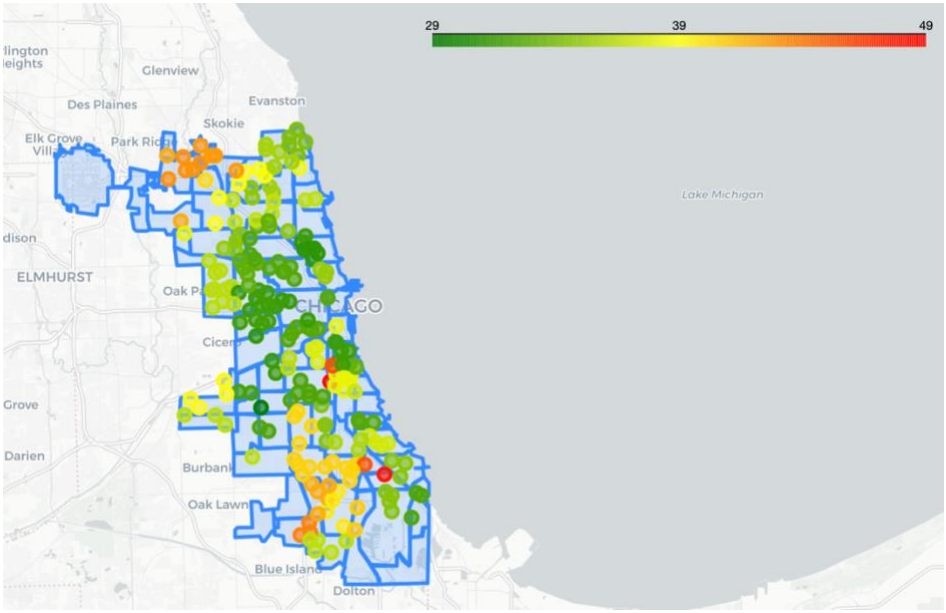
clusters, until I happened upon an epsilon and minimum samples numbers that finally created an amount of clusters that wasn't 100 or 0. I set an epsilon of 0.01, and a minimum sample of 3, and ended up with 5 clusters. These clusters ended up being far more tied to each other quite closely, which is of course due to the small epsilon value, but if the value was increased much more, the clusters would quickly increase to the hundreds. As you will see from the following output, most of the values ended up in Cluster -1, which is noise. DBSCAN was not as effective for this hypothesis; K-means are sensitive to the density of data points within clusters. It performs well when clusters have relatively uniform densities and clear separation between clusters. DBSCAN, on the other hand, is designed to identify dense regions of points separated by areas of lower density. The nature of this data seems to fit K-mean far better.

DBSCAN_Cluster_MEDINC	Cluster	Number of Points	Average MEDINC
-1	-1	186	52057.38172
0	4	6	36315.00000
1	5	5	104639.00000
2	0	4	31856.00000
3	1	4	33503.00000
4	3	4	115389.00000
5	2	4	34241.00000

Even so, although we don't get good info on court density from the DBSCAN, Northern clusters absolutely have higher average median household income. Figure 1 proves my hypothesis wrong, in fact, it appears that there are far more courts in lower income areas than the higher income areas. This is possibly due to the presence of more indoor basketball courts and gyms in higher income areas, limiting the need for outdoor courts.

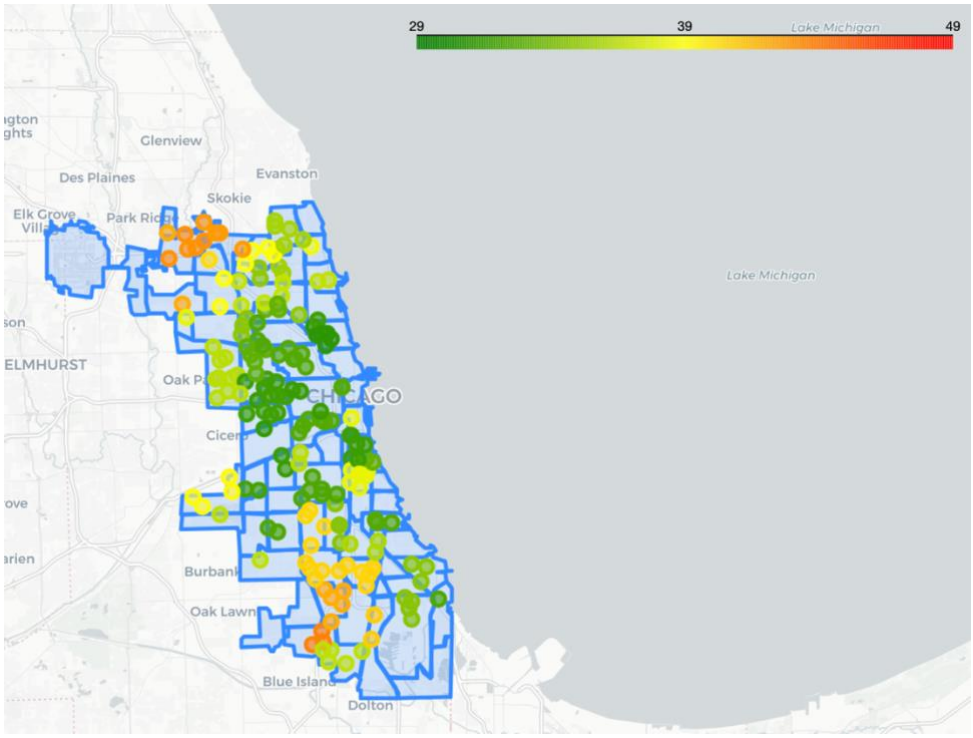
I then decided to test my second hypothesis, which was that community areas with a younger population demographic would have a higher density of basketball courts. I decided to use K-means again.

Clustering by age shows a very clear pattern with respect to density, the younger areas seem to exhibit a far denser collection of basketball courts, the older areas are considerably more spread out. The denser areas, downtown, are also younger than the areas farther on the outskirts of Chicago.



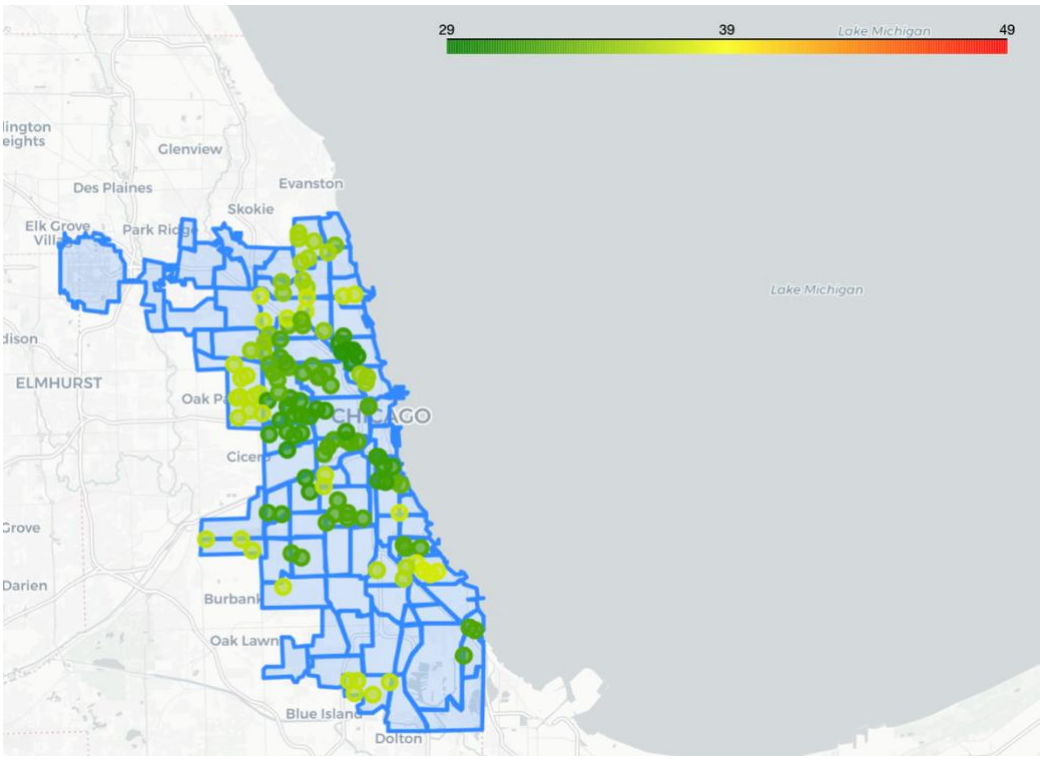
	Cluster	Number of Points	Average MED_AGE
3	3	23	43.291304
0	0	33	39.787879
2	2	47	36.740426
4	4	35	34.640000
1	1	75	31.866667

The model output exhibits a pattern not dissimilar from median household income, the size of the cluster correlates with age as well, as age decreases, there appears to be a general increase in the size of the cluster, which backs up the visual observations we can make from the plot. I decided to try HDBSCAN with this plot, as there are varying density structures in place, those of which were not well accounted for with DBSCAN.



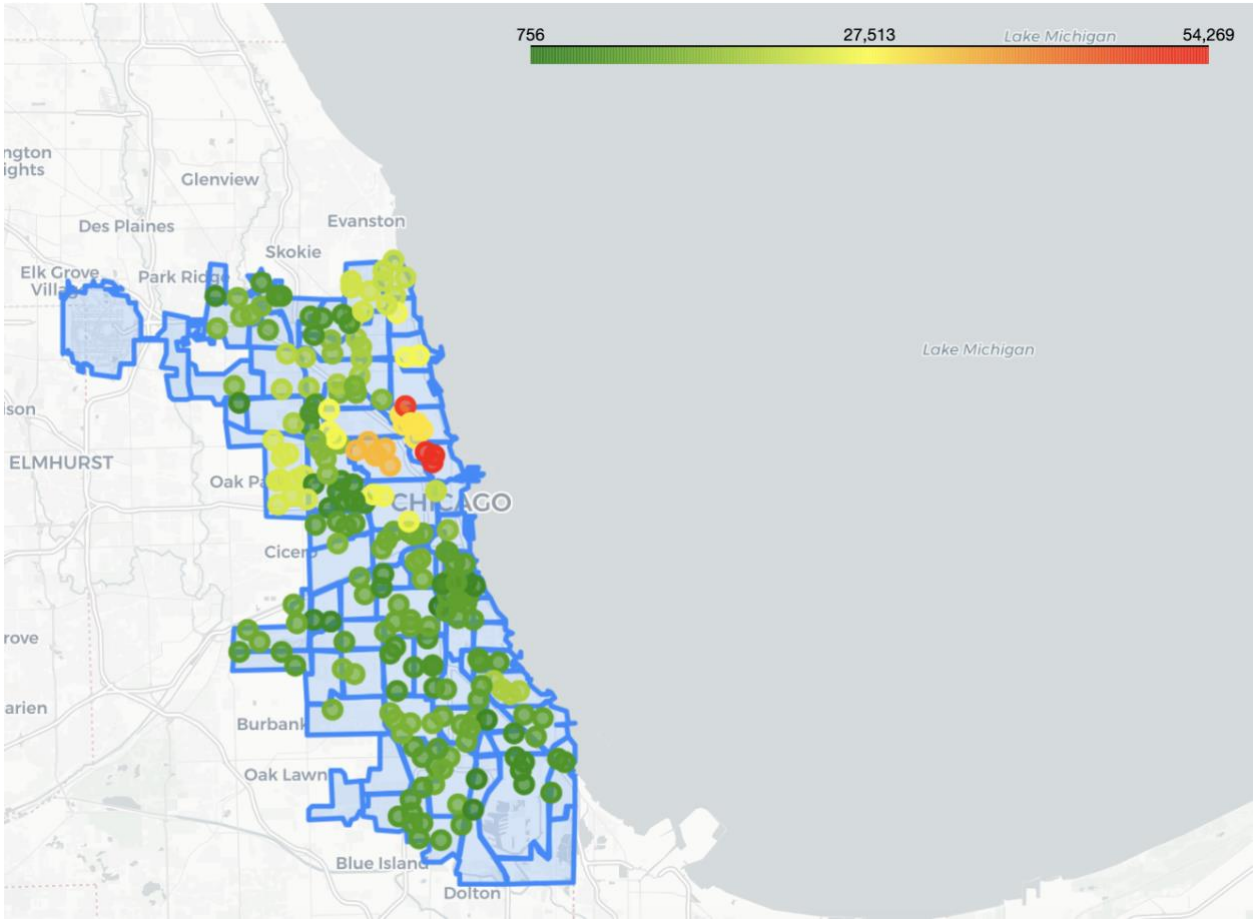
HDBSCAN creates a near identical plot, but leaves out some of the outliers, which you can see from the highest and lowest age ranges decreasing in density. My hypothesis that community areas with younger population demographics will have a higher density of basketball courts appears to be the case, with both clustering techniques showing the older regions to have more spread basketball courts, and the younger areas having more dense collections of basketball courts. This is even more apparent in HDBSCAN if I continue to improve the minimum cluster

size and minimum samples parameters, all that remain are the younger demographic areas, which is shown in the plot below.



Up to this point it has seemed like both socioeconomic and demographic factors affect court density. I tried to choose a third factor and hypothesis that I thought could build upon the studies I'd done so far. Considering what I've found with age, and income, my third hypothesis was that areas with higher internet access have fewer basketball courts. Older areas tend to have less basketball courts, but they may have less of a reliance on internet access. Areas with higher income tend to have less basketball courts, but it is likely that areas with higher income have better access to the internet. I wanted to see if this information could help contextualize the socioeconomic and demographic affects, I'd studied prior. Additionally, perhaps areas with lesser internet access would feel they had more of a need for a basketball court. However, as you can

see with the figure, it is clear that this data is nearly identical to the data for median household income. This one is interesting because it is measured by the total number of people that have internet access, so this may be why you see some larger values near the cities, simply due to the sheer number of people that live there. However, the basketball court pockets with the most people with internet access remain in the highest income neighborhoods in the plot, and the higher densities of basketball courts remain in the lower income, lower internet areas.



Discussion:

In this study, the spatial distribution of basketball courts in Chicago was investigated with a focus on identifying patterns and understanding the underlying factors contributing to these patterns. The analysis explored the relationship between basketball court density and various socioeconomic and demographic factors within Chicago's community areas. The initial hypothesis regarding the influence of median household income on basketball court density was refuted, revealing that areas with lower median household incomes tended to have a higher density of basketball courts. Subsequent analysis of age demographics corroborated the hypothesis that areas with a younger population tend to exhibit a higher density of basketball courts. This relationship suggests a potential correlation between the demand for recreational facilities like basketball courts and the demographic composition of the community. The investigation into the relationship between internet access and basketball court density provided further insights, revealing a notable similarity between the spatial distribution patterns of basketball courts and areas with higher median household incomes. While areas with higher internet access may be associated with higher income levels, my analysis suggests that internet access alone may not be a significant determinant of basketball court availability. Overall, the study highlights the complex interplay between socioeconomic and demographic factors in shaping the spatial distribution of basketball courts in Chicago.