CSCI 1951A – Final Submission

bongo-bongo

May 15, 2022

William Back - wback
Dhruv Bhatia - dbhatia1
Niyoshi Parekh - nparekh1
Herbert Traub - htraubiv

1 Abstract

1.1 Hypothesis

Cities in the US have long been segregated, either through racially discriminatory landlords, mortgage and loan policies or a "long history of federal, state and local policies" [**?**]. Given this information, we were curious to see if this segregation extended beyond residential patterns, to the density of critical infrastructure – schools, libraries, hospitals, parks, supermarkets and banks. Specifically, we wanted to test the following hypothesis: majority white and higher income neighborhoods have more critical places of interest (POIs) than non-majority-white and lower income neighborhoods.

1.2 Data

Our data comes from two main sources: the Google Maps Places API and the Census Bureau.

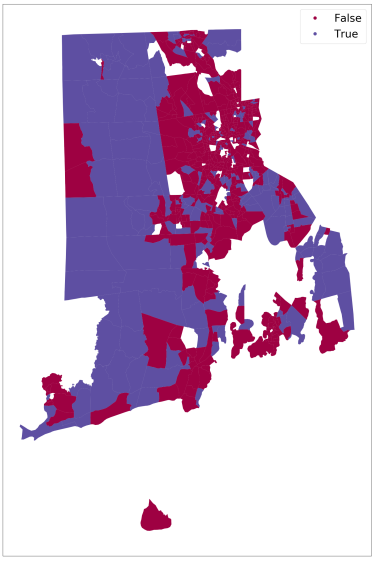
Through the Google Maps API's Nearby Search functionality, we search Google Maps by six specific keywords (hospital, school, park, library, bank, supermarket) in pre-defined circles within a city. This returns the following for each POI: place_id, name, price_level, rating, user_ratings_total, longitude, and latitude. Most places do not have a price_level. Google only guarantees that POIs have a latitude and a longitude, however most POIs have all relevant data. We then calculated connectivity scores for each type of POI based on the number and quality of them in a block group for each block group.

The Census Bureau provides demographic data for block groups (small regions) within cities. We used the following demographic data classifiers: race, ethnicity (only hispanic/non-hispanic) and median household income. The Census provides the number of people in every block group for every race and ethnicity category. We used the total population to normalize these numbers, and calculated majority_race and majority_ethnicity labels for each block group, based on the race and ethnicity categories with the highest % of people in that block group. We also labeled block groups as low income, lower middle, upper middle and high income, based on their median household income.

1.3 Findings

1. **Connectivity to social infrastructure is associated with demographic factors, especially with race, ethnicity and income.**
   We used K-Means Clustering to investigate an overview of similarities between block groups within a given city or state. We clustered twice, one clustering included demographic census data (race, ethnicity, income, commute time to work), and the other only clustered based on the connectivity scores. We then looked at the overlap between these clusterings to validate our machine learning process and do a baseline comparison of demographic and connectivity features. In the 3 cities and 1 state, we found on average a 24.75% overlap between the two clusters, which suggests an underlying similarity. To strengthen this claim and make it more specific, we set out to determine whether there was an association between a demographic categorical feature (majority race, majority hispanic/non-hispanic, and income group) and the block group's label given by the connectivity score clustering. In all cases except hispanicity in Boston, an association was found.

(a) Boston, Overlap: 28%



(b) Rhode Island, Overlap: 28%

Figure 1: Cluster Cross Validation with Overlap % between Connectivity and Demographic Clusters

Table 1: Chi-square test results, between block group demographic variables and connectivity scores

Null Hypothesis: Block group connectivity clusters and X demographic variable in Y city/state are independent.			
City/State	Majority Race	Majority Hispanic/Non Hispanic	Income Group
Boston	Null Rejected	Null Accepted	Null Rejected
Chicago	Null Rejected	Null Rejected	Null Rejected
NYC	Null Rejected	Null Rejected	Null Rejected
Rhode Island	Null Rejected	Null Rejected	Null Rejected

2. **Race and ethnicity influence connectivity to social infrastructure more than median household income does, with some inter-city variations.**
We ran two-tailed independent t-tests to determine whether there was a difference in the connectivity scores for each type of POI between 3 different demographic groups – majority Black/African American and majority White block groups, majority Hispanic and majority Non-Hispanic block groups and high income and low income block groups. The group with lower connectivity was derived from the sign on the t-statistic. Figure 2 shows the results of these t-tests. Overall, the income group comparison showed the least significant differences between the groups. Boston is especially interesting, with significant racial and ethnic differences in connectivity, but none based on income. The colors indicate whether the results match our original claim (green) or not (red), that majority White, Hispanic and higher income neighborhoods have greater connectivity.

Recording which has lower connectivity between Black and White neighborhoods, and the t-value						
City	**Supermarket**	**Bank**	**Library**	**Park**	**Hospital**	**School**
NYC	No difference	Black	Black	Black	Black	Black
Boston	Black	Black	Black	Black	Black	Black
RI	No difference	No difference	White	No difference	No difference	No difference
Chicago	Black	Black	Black	Black	No difference	Black
Recording which has lower connectivity between Non Hispanic and Hispanic neighborhoods, and the t-value						
City	**Supermarket**	**Bank**	**Library**	**Park**	**Hospital**	**School**
NYC	Hispanic	Hispanic	Hispanic	Hispanic	No difference	Hispanic
Boston	No difference	No difference	Hispanic	Hispanic	Hispanic	Hispanic
RI	No difference	No difference	Non Hispanic	No difference	Non Hispanic	Non Hispanic
Chicago	Non Hispanic	Non Hispanic	Hispanic	Hispanic	Hispanic	Non Hispanic
Recording which has lower connectivity between Low Income and High Income neighborhoods, and the t-value						
City	**Supermarket**	**Bank**	**Library**	**Park**	**Hospital**	**School**
NYC	No difference	Low Income	Low Income	Low Income	Low Income	High Income
Boston	No difference	No difference	No difference	No difference	No difference	No difference
RI	High Income	No difference	High Income	No difference	No difference	No difference
Chicago	Low Income	Low Income	Low Income	Low Income	Low Income	High Income

Figure 2: Independent T-Test Results showing groups with lower connectivity across race, ethnicity and income.

Some interesting results:

- In NYC, majority Black block groups have lower connectivity to schools than majority White block groups. However, High Income block groups also have lower connectivity to schools than Low Income block groups. Figures 3 and 4 show larger school connectivity scores concentrated in Brooklyn, Manhattan and the Bronx. Manhattan is mainly majority White and High Income, but Brooklyn is majority White and Low Income, while the Bronx is a mix of majority Black and Some Other Race.

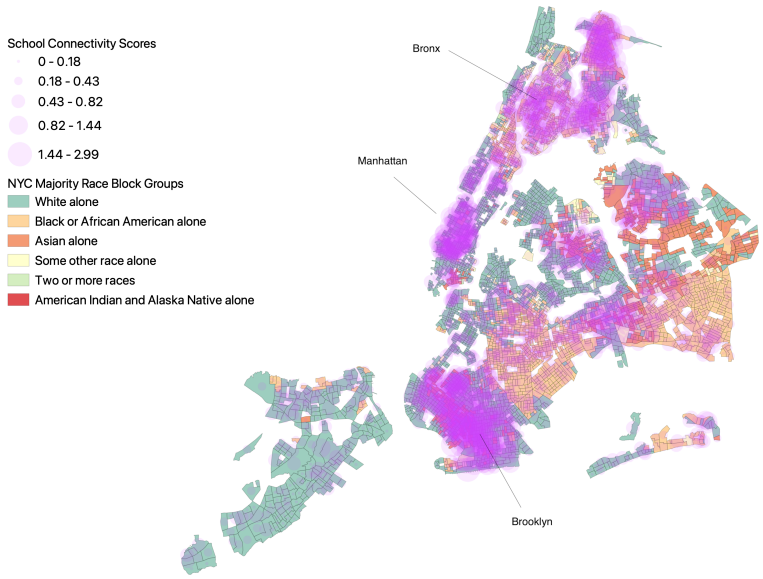


Figure 3: NYC School Connectivity Scores over Majority Race Groups Map

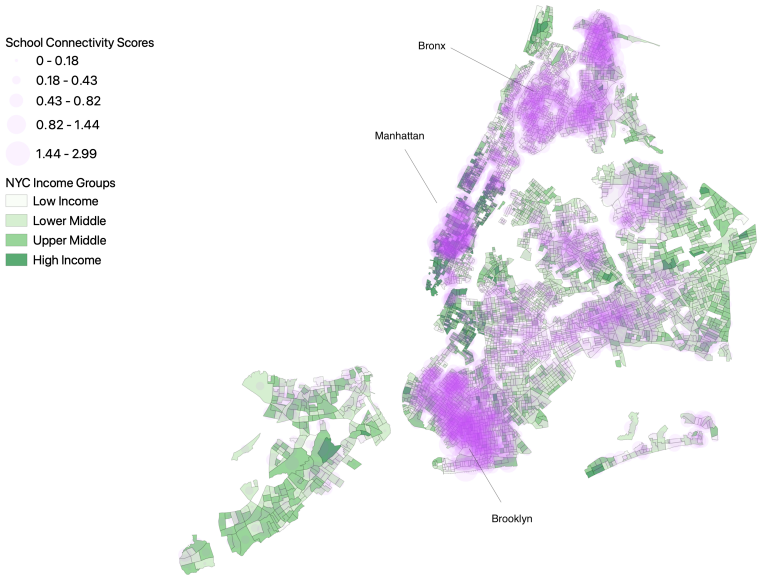


Figure 4: NYC School Connectivity Scores over Income Groups Map

- Boston has no difference in connectivity to any of the POIs between low and high income block groups. However, majority Black block groups have a lower connectivity to all POIs than majority White block groups. In Figures 5 and 6, we can see that Boston is largely segregated by race, and within the racial groups there is income variation, which explains our results.

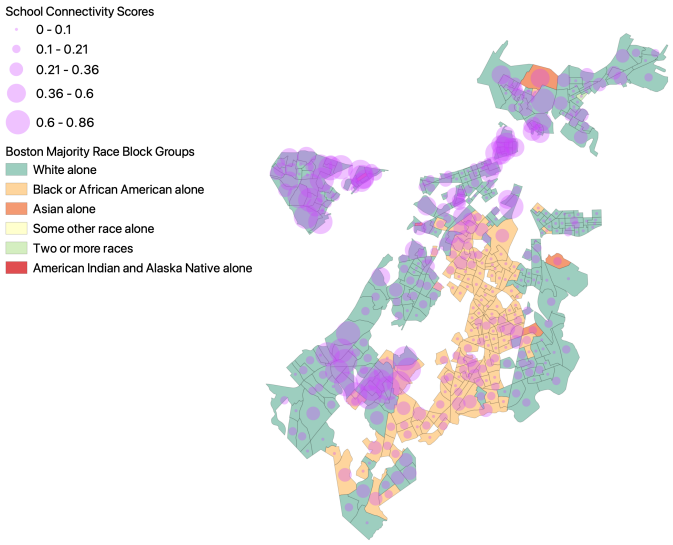


Figure 5: Boston School Connectivity Scores over Majority Race Groups Map

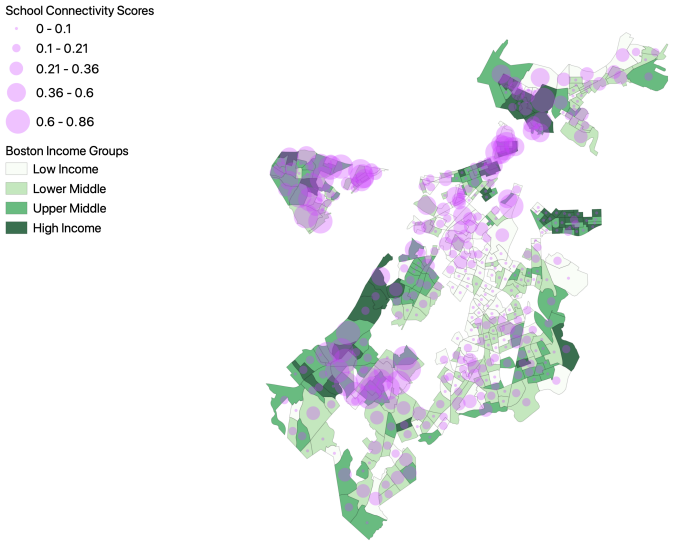


Figure 6: Boston School Connectivity Scores over Income Groups Map

RI's results are different from the rest, often contradicting our original claims, potentially due to the effects of suburbanization, with majority White, Non-Hispanic and High Income populations living in the suburbs, and thus having lower connectivity scores than Black, Hispanic and Low Income populations living closer to the city centers, where more amenities are located. This provides a point for further investigation into differences in connectivity between suburbs and cities.

3. **Connectivity to supermarkets is least influenced and library and school connectivity is most influenced by demographic factors of race, ethnicity and income.**
In addition to the groups compared in above t-tests, we conducted t-tests on 3 more pairs of groups: Low and Lower Middle income, Lower Middle and Upper Middle income, and Upper Middle and High income, to test whether there was more significant influences of income on connectivity between adjacent income groups, than there was between the two extremes of low and high income groups. We continued to find low numbers of significant differences between income groups, however this allowed us to more comprehensively look at which connectivity variables were most influenced by demographic factors, across geographies. From all the pairs of groups compared across all 4 geographies, the least number of significant differences were found between groups when comparing their supermarket connectivity scores, and the most were found when comparing their library and school connectivity scores (reading Figure 2 vertically, instead of horizontally). Figures 7 and 8 show the supermarket connectivity and school connectivity scores on income maps of Chicago.

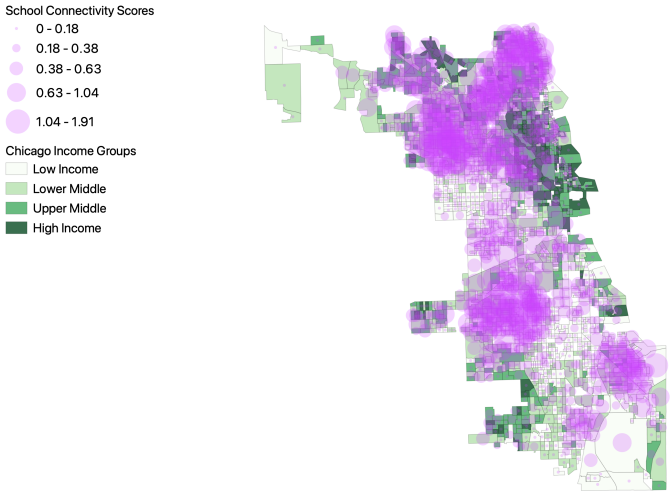


Figure 7: Chicago School Connectivity Scores over Majority Race Groups Map

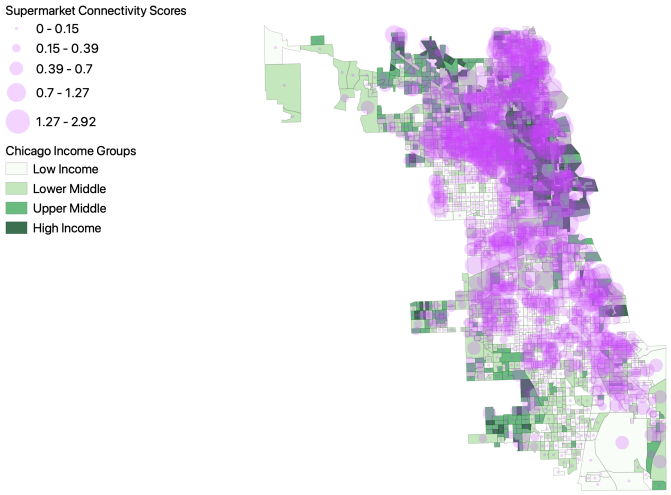


Figure 8: Chicago Supermarket Connectivity Scores over Income Groups Map