

Commercial Segregation in Chicago

Introduction

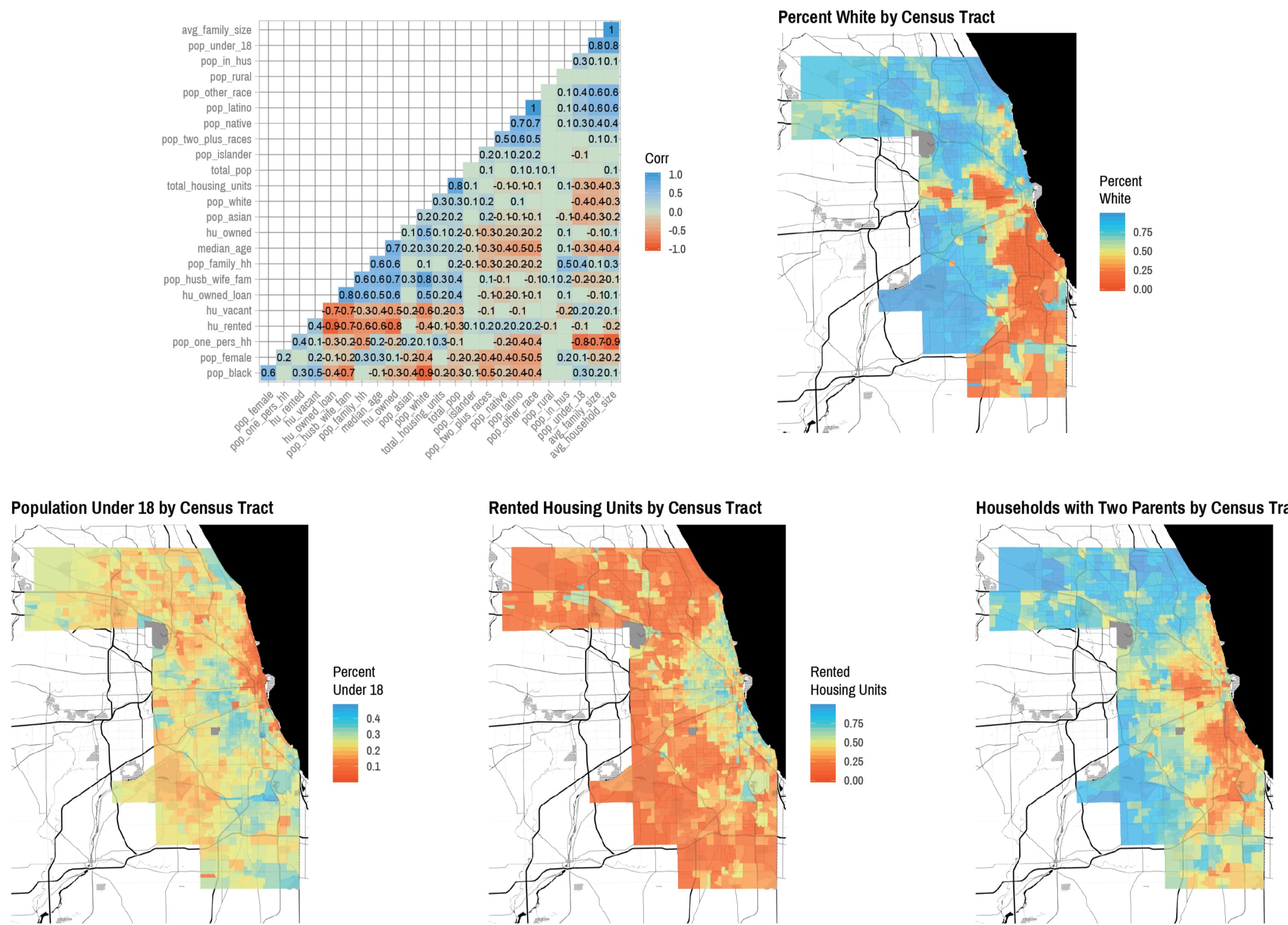
Businesses are an integral part of any community, providing goods and services that reflect the cultural needs of that area. Where businesses choose to invest can affect which resources are available or absent. This project explores how businesses vary among different Chicago communities, and the implications this presents for battling inequality in the city.

Hypothesis: Businesses in Chicago follow similar clustering patterns as residents.

Data

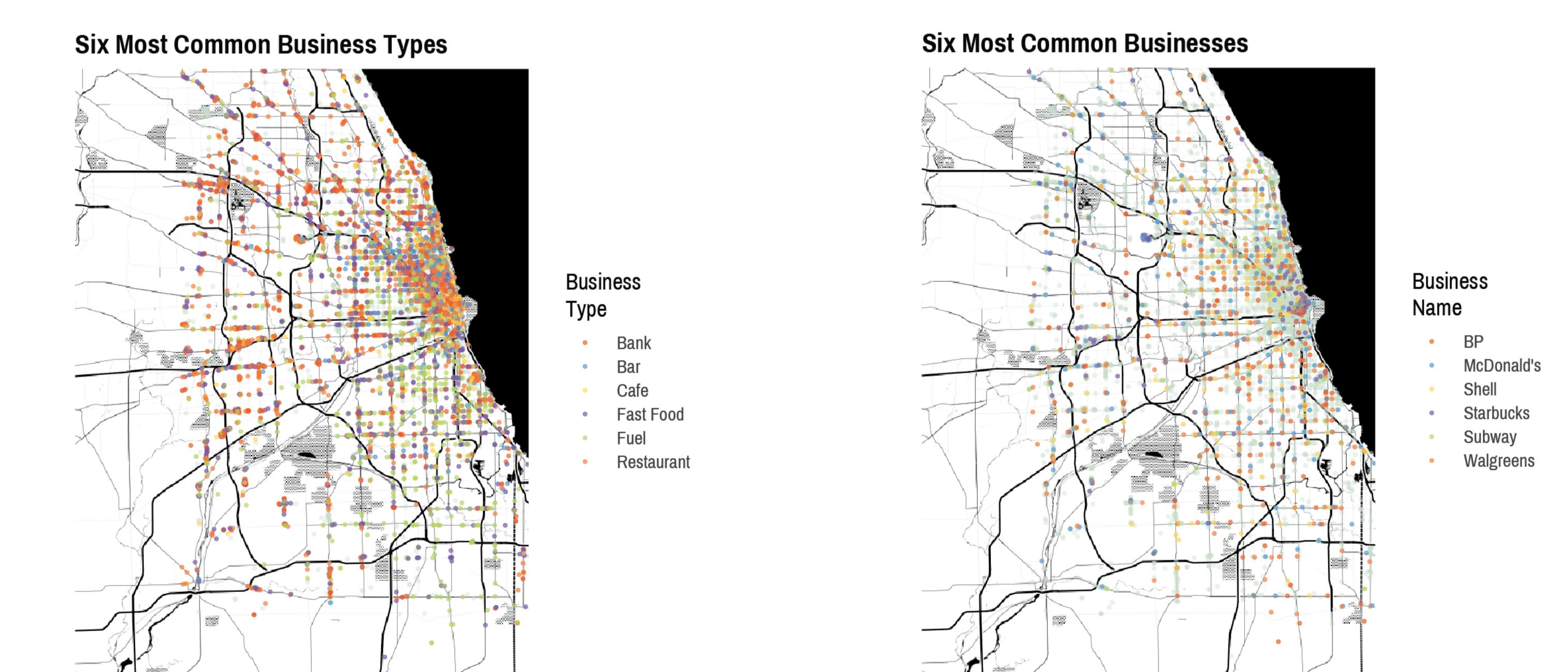
2010 Decennial Census (1,319 census tracts)

Contains 24 demographic variables such as race/ethnicity, age, gender, family status, or housing status



Open Street Maps (10,359 businesses)

Contains 4 variables on business location, name, and type of business

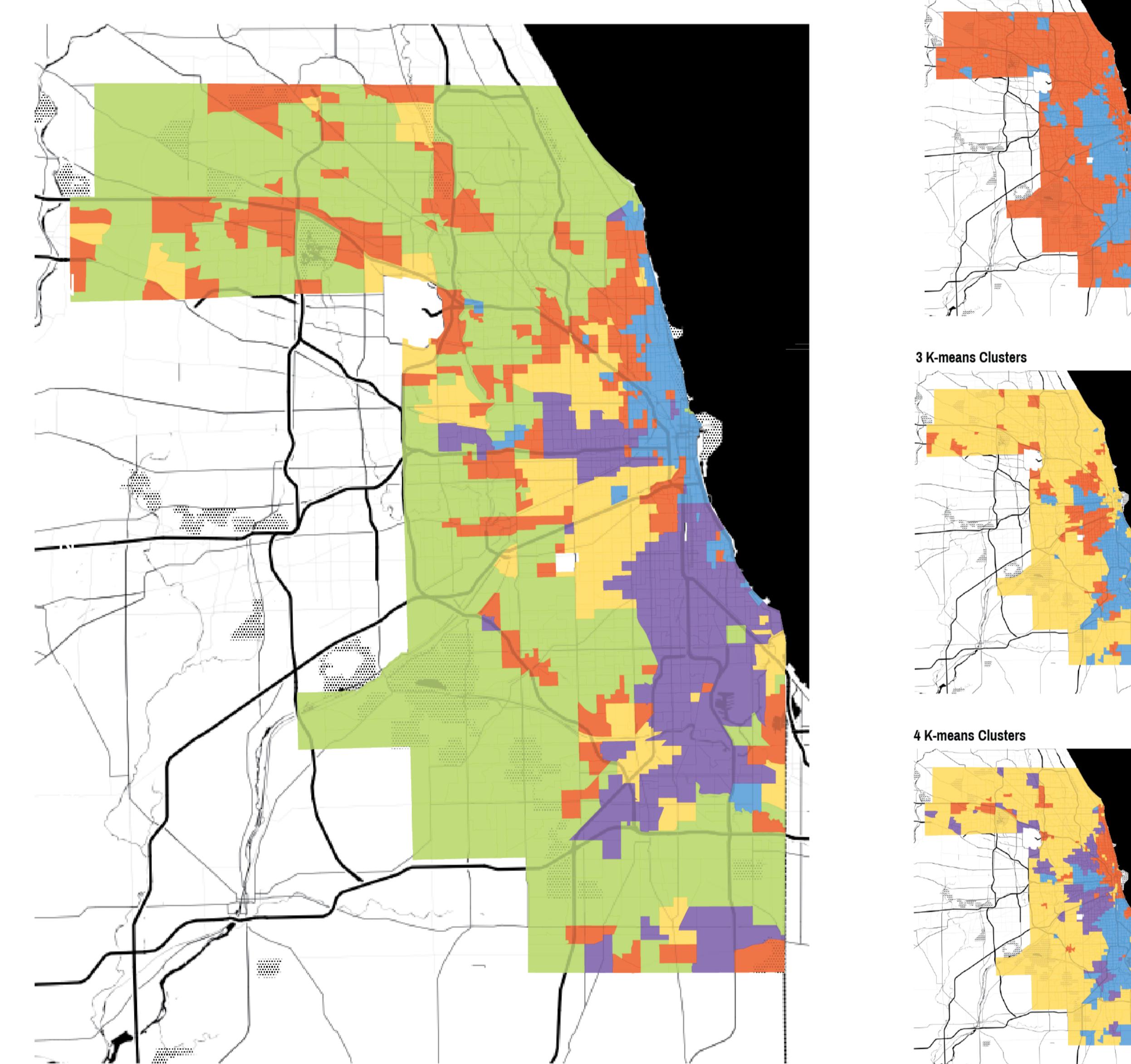


Methods

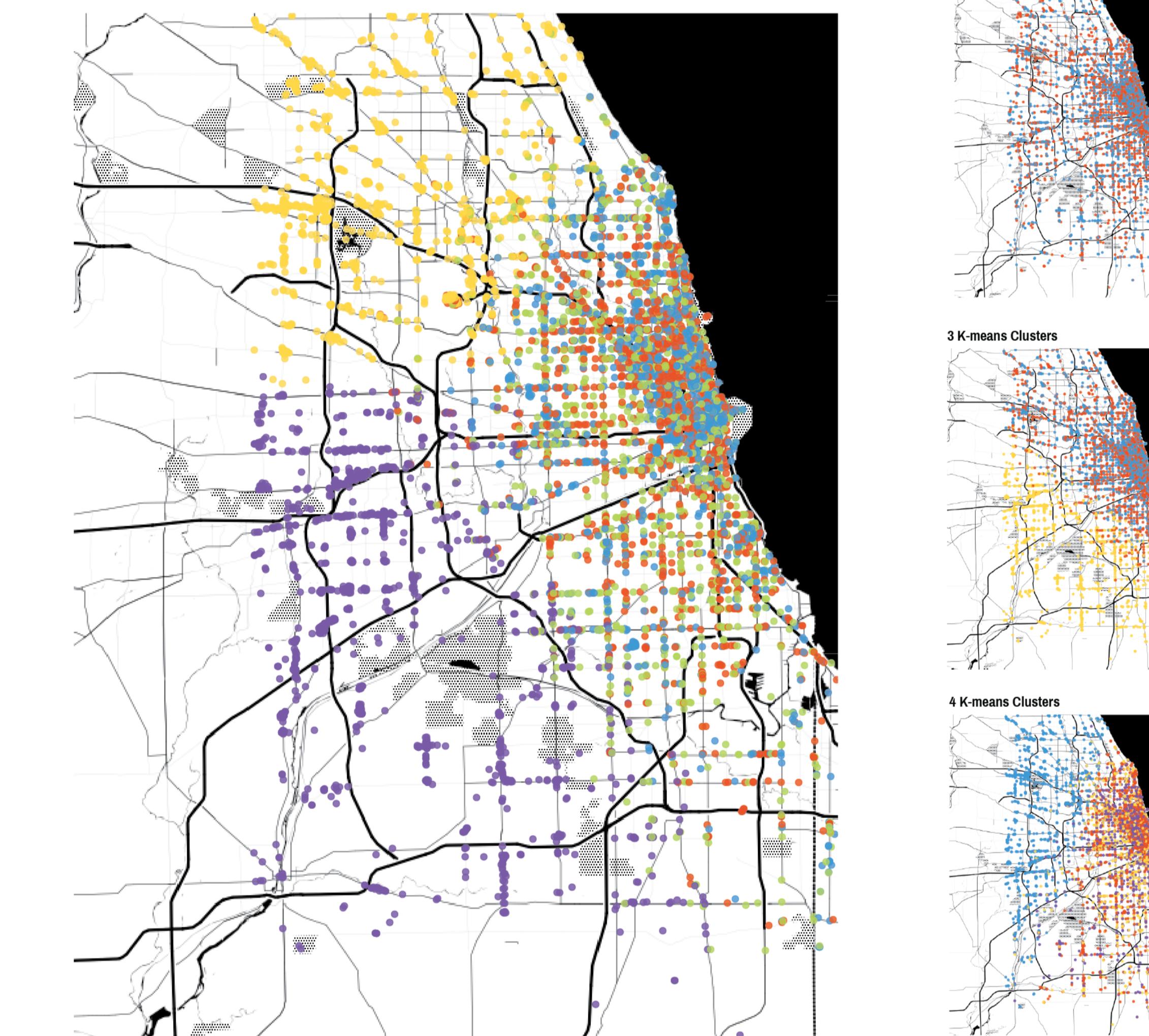
This project takes advantage of unsupervised clustering methods to create groupings of data purely based off that data. For both datasets K-means and Hierarchical Clustering models were built for 2 to 6 clusters, with total within sum of squares, average silhouette width, and gap statistics calculated to assist in selecting the optimal clustering model. Decision tree surrogate models with repeated 10-fold cross validation were used to evaluate the clustering models further.

Results

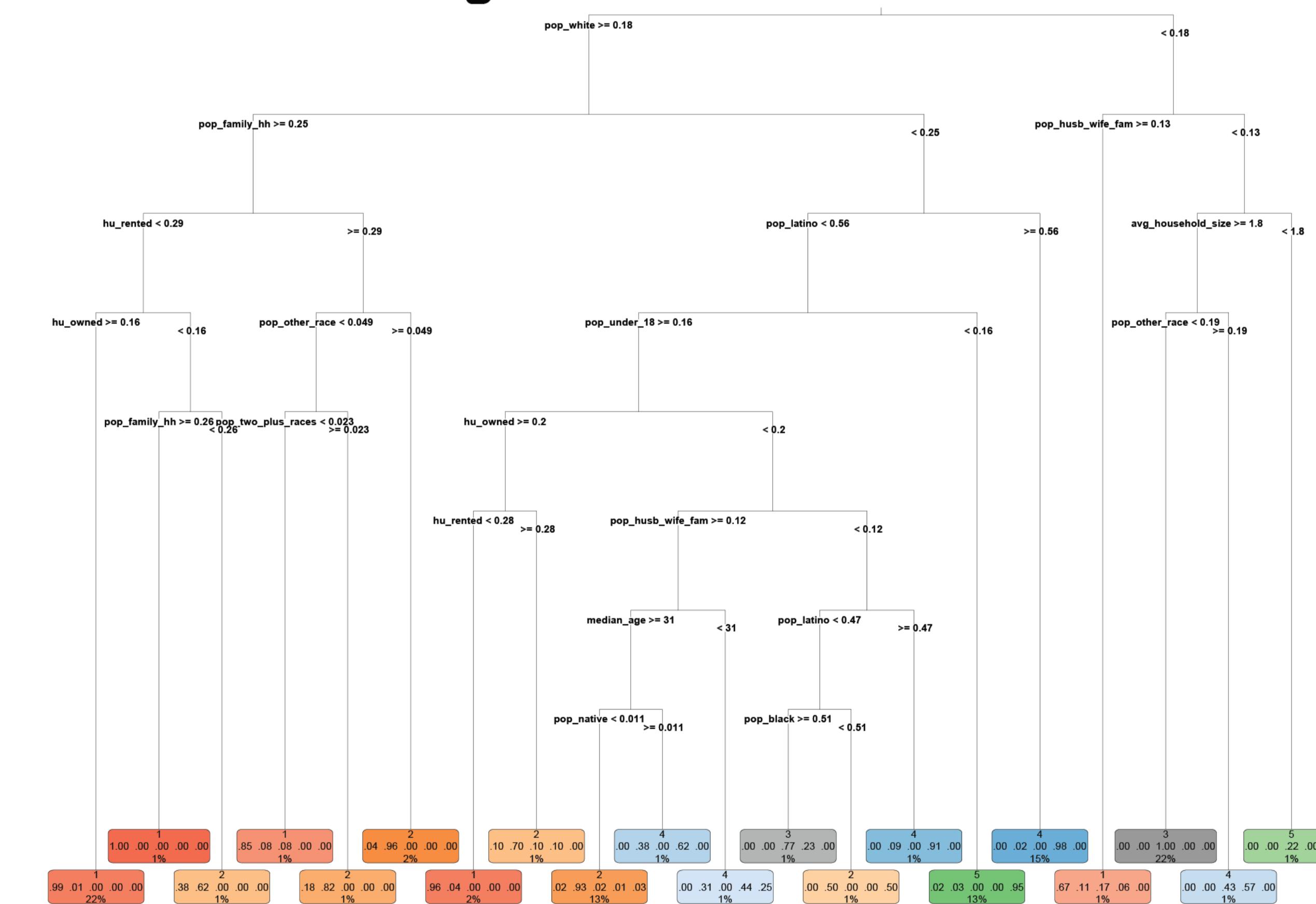
K-means Clusters of Census Tracts



K-means Clusters of Businesses

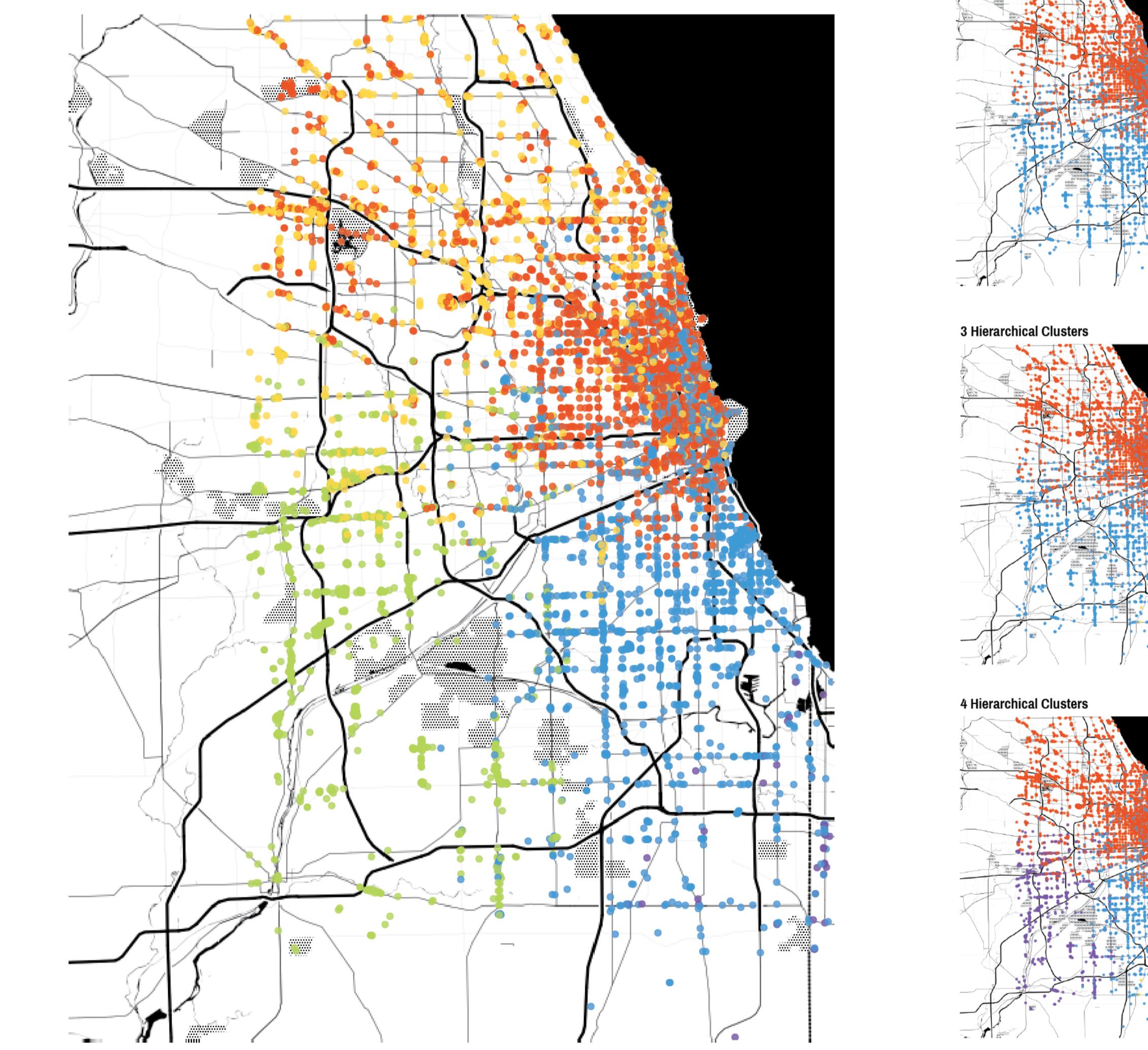


Decision Tree Surrogate Model of Census Tracts



The Census data appears to cluster well using K-means with 5 clusters. Race is a significant component of residential segregation in Chicago, but there are other factors as well, such as age and family or household characteristics. The clustering of businesses appears similar at 5 clusters, with clear differentiation of the southwest side from the rest of Chicago.

Hierarchical Clusters of Businesses



The hierarchical model more closely matches the clustering of Census data, clearly separating South Side businesses from the Loop and North Side, which clustering algorithms have a harder time separating. This suggests that businesses are segregated in measurable and meaningful ways.