Six unique load shapes: A segmentation analysis of Illinois residential electricity consumers

Jeff Zethmayr[a], Ramandeep Singh Makhija[a]

[a]Citizens Utility Board
309 W. Washington St. Suite 800
Chicago, IL 60606
United States

Corresponding author:
Jeff Zethmayr
Citizens Utility Board
309 W. Washington St. Suite 800
Chicago, IL 60606
United States
jzethmayer@citizensutilityboard.org

Abstract
Many electricity policies have disparate impacts on consumers with different load shapes. This study applies k-means clustering to 2.5 million Illinois customers, and matches six resulting clusters with demographics. Flatter load shapes were more likely in urban and low-income areas, with high-volume, peak usage more likely in high-income/suburban areas. This highlights potential for grid cost reduction through DR programs targeting suburban areas, and illustrates potential cross-subsidization intrinsic to common electric rate designs.

Keywords
Energy efficiency; State policy; Electric utility; Rate design; Customer Segmentation; Advanced Metering Infrastructure

Vitae
Jeff Zethmayr conducts energy policy and economic research, with a focus on rate design for regulated utilities. Mr. Zethmayr's previous work includes expert testimony and analysis for utility rate cases, and research on the bill impacts of demand-based delivery rates. He holds a Masters of Public Administration from Columbia University's School of International and Public Affairs and a Bachelor of Arts from Macalester College.

Ramandeep Singh Makhija is a Data Scientist at the Citizens Utility Board, where he investigates utility customer behavior through advanced statistical analysis. He holds a Master's degree in Industrial Engineering from the University at Buffalo with specialization in Operations Research, and a Bachelor's degree in Mechanical Engineering from the University of Pune.

Declarations of interest: None

Abbreviations
ACS: American Community Survey
AIC: Ameren Illinois
AMI: Advanced Metering Infrastructure
CUB: Citizens Utility Board
ICC: Illinois Commerce Commission
LIHEAP: Low Income Home Energy Assistance Program

PLC: peak load contributions

Funding
A grant from Energy Innovation helped fund the research.

1.0 Introduction

Discussions of electric utility regulation have long recognized the impact of peak load on total system costs. According to Faruqui et al., "even a 5% reduction in peak demand in the United States could lower consumer energy costs by at least $3 billion a year."[1]

For this reason, economists have argued for the creation of time-variant rates for the various components of electric service, as an encouragement for consumers to manage their usage so as to lower their contribution to peak system load and improve system load shape. In a similar vein, demand response programs are used to induce lower usage of electricity at times of peak energy demand, and many utilities have also initiated energy efficiency programs to provide incentives to consumers to install energy saving devices and other home upgrades to reduce overall energy consumption.

To support these initiatives and enhance service reliability, many utilities have upgraded their metering systems with the implementation of Advanced Metering Infrastructure (AMI), including smart meters.[2] AMI allows the utility to collect customer usage volumes in 15 minute (or less) increments throughout the day rather than on a monthly basis.

This provides a wealth of information into the way an individual customer uses electricity and that customer's actual contribution to the system's peak load. It also allows for customer segmentation techniques that were previously unavailable to researchers. Attaining a better understanding of the differences in how consumers actually use electricity is important for several reasons.[3]

First, and most obviously, it can help improve energy efficiency, demand response, and time-variant pricing program design. [4] By targeting a particular program at those most likely to benefit from it, utilities can maximize cost-effectiveness and achieve more "bang for the buck."

This dynamic can also help with the overall political economy of distributed energy resource program advocacy. Time-variant pricing program implementation, for example, is often stalled by questions around the likely impacts of such programs on low-income customers, a question that to date has been difficult to answer conclusively given the paucity of actual load shape data at the local level.

Second–while it raises complicated policy questions around the appropriateness of system-cost socialization–big data analytics based on individual usage patterns allow for a more accurate assessment of consumers' marginal system costs impact, a key input in economically efficient tariff design. [5,6,7]

---

[1] See Faruqui et al, 2007
[2] It should be noted that more needs to be done to maximize the consumer benefits of AMI, particularly when it comes to advancing distributed energy resources.
[3] See, e.g., McLoughlin et al. 2015: "A wealth of new data exists for utilities, giving detailed electricity consumption at increased granularity for a large number of customers within the residential sector. The availability of this source of data can potentially be used by utilities to create customized electricity load Profile Classes (PC) and can assist in areas such as: improved load planning and forecasting; Time of Use (ToU) tariff design; electricity settlement; and Demand Side Management (DSM) strategies"
[4] Yu et al. 2018
[5] Chen et al 1997
[6] Burger et al 2019, "Fair, Equitable, and Efficient Tariffs in the Presence of Distributed Energy Resources"

In this paper, we apply k-means clustering, a machine-learning algorithm, to AMI data of more than 2.5 million customers of Commonwealth Edison (ComEd) and Ameren Illinois (AIC), the two largest electric utilities in Illinois. From our review of the literature, it appears that this is the largest dataset yet used in an electric consumer segmentation analysis. The k-means clustering process generates subsets of similar customers based on their average summer load shape, expressed in percentage terms as their average usage in each hour compared to their average peak usage.

We then use a logistic regression model to determine the likelihood of a customer in each of the resulting six clusters to reside in locations associated with various demographic indicators, as defined in the most recent American Community Survey (ACS) put out by the U.S. Census Bureau. In this way, we were able to construct demographic profiles for customers who consume electricity according to the six distinct cluster load shapes. We then use the same logistic regression method to isolate the load shapes most prevalent in low-income communities.

This paper describes the data we used to perform the analysis, the methodology we followed, and the cluster profiles that this process generated. Through the analysis, we identified six distinct load profiles, ranging from nearly flat to high-volume, peaky usage, and found that customers with flatter load shapes were substantially more likely to live in urban areas and low-income communities.

2.0 Theory

This paper relies primarily on anonymous electric consumption data for residential customers of ComEd and Ameren Illinois, captured through smart meters. In 2017, the Illinois Commerce Commission (ICC) approved a plan for Illinois utilities to make customers' smart meter data available to researchers and other third parties, provided the data was kept anonymous.[8] The datasets provided through this plan include daily observations consisting of hourly (in the case of AIC) or half-hourly (in the case of ComEd) interval volumes read from individual smart meters, identified only by the customer class, a random ID number, and a geographic location.

According to the rules approved by the ICC, an individual customer's data can only be included in the data release if it passes an anonymity screen, to ensure the identity of a customer cannot be determined or reverse engineered. In practice, the screen adopted in the rule requires that a customer's location cannot be provided if there are 15 or fewer customers in the given geographic area, or if they represent 15% or more of that area's load.[9]

Due to the large differences in the relative population density in the service territories of ComEd, which serves northern Illinois, and AIC, which serves central and southern Illinois, this anonymity screen results in very different levels of geographic specificity in the two datasets. For ComEd customers, we were able to use a dataset consisting of 1.5 million customers, all geo-located at the 9-digit ZIP code level (postal code). This allowed for a very precise matching of customers to demographic attributes, at the Census Block Group level.

However, so few AIC customers passed the initial anonymity screen that the vast majority of observations in that dataset are only identified at the municipal level. This required aggregating demographic

---

[7] Burger et al 2019, "The Efficiency of Distributional Effects of Alternative Residential Electricity Rate Designs

[8] See The Big Energy Data Center for information on Illinois' big data story.

[9] ComEd and AIC used different methods of implementing the screen: ComEd provided two different datasets, one containing customers who met the screen at the postal code level, and one containing all customers who met it at the ZIP code level; while AIC provided one dataset containing all customers, with location data at the smallest level where that customer met the screen.

information, which greatly reduced the significance level of socio-demographic characteristic results for the AIC territory. The AIC dataset contains just over 1 million customers.
To construct our demographic attributes dataset, we downloaded Census Block Group level data from the 2017 ACS.[10] This dataset includes local information on a variety of categories; the general categories used in our regression model are as follows:

• Age of head of household

• Construction year of residence

• Educational attainment

• Heating fuel

• Household makeup (number of residents, family vs. non-family)

• Home values

• Density (number of households in block group)

• Number of rooms in residence

These categories were selected based on an exploratory analysis of the downloaded data set from ACS. Visualization techniques were used to select the significant factors to link these factors with our clusters. In order to match postal codes to block groups, this study also relied on a commercially available dataset acquired from Melissa, a geographic data services company.[11] This dataset provided a direct concordance between postal codes and census block groups.

2.1 Data cleaning

The first step in preparing the data for analysis was downloading the AMI data from ComEd and AIC, then putting that data into a standard format and cleaning it for the cluster analysis. Because the AIC data was provided in an hourly format, it was first transformed into half-hourly intervals by dividing the hourly volumes in half, to maintain the similarity of features between the datasets. For every customer, an average was taken of their half-hourly volumes for each weekday for the months of June through September of 2018. Customers with missing data (less than 1%) were removed from the dataset.

To prevent extreme observations from skewing results, dataset customers with average daily usage higher than the 99th percentile or less than 1st percentile of customers were eliminated from the dataset. The final datasets used for clustering contained 1,540,693 ComEd customers and 1,027,321 AIC customers with 48 half-hourly usage values respectively.

To generate a dataset for demographic analysis, block group level data from the 2017 ACS was downloaded for 43 different variables, pertaining to the categories listed above.[12]

With the exception of median income, these data consisted of counts of residences within a block group who fell within defined ranges of a particular value (for example, *N* of residences with highest educational attainment of high school degree, bachelor's degree, post-graduate degree, etc.). Median income values consisted of the median annual income for residences within the block group.

---

[10] See American Fact Finder, U.S. Census Bureau
[11] See Melissa.com for the geographic data set.
[12] See Appendix 2 for complete list of variables.

These features were then converted into percentages of total households within the block group, and independent binary variables were generated by assigning a value of 1 to observations with a percentage greater than the mean of all block groups in the respective dataset. For the ComEd data, these variables were generated at the block group level; however, lack of postal code designations in the AIC data necessitated aggregating demographic counts at the municipality level. The final datasets used for demographic analysis contained 33,576 distinct block group-cluster combinations and 948 distinct city-cluster combinations with 43 features (socio-demographic characteristics) for ComEd and AIC customers respectively.

2.2 Cluster analysis

Cluster analysis is an unsupervised learning algorithm that can be used for consumer characterization based on their consumption pattern.[13] In recent years, researchers have applied new data mining and statistical techniques to characterize consumer profiles according to their usage patterns.[14] From our review of the literature on customer segmentation, we selected the k-means clustering method for this analysis.[15] Half-hourly smart meter datasets from ComEd and AIC were used to cluster residential customers in different groups based on their usage patterns.

K-means is a type of unsupervised learning algorithm, which assigns individual observations into similar subsets by minimizing the variance between those observations. The algorithm continuously generates random cluster assignments for all observations, to produce an optimal set of assignments. Rather than defining groups beforehand, clustering allows us to identify the organically formed groups in our dataset.

The inputs to the k-means algorithm are the number of clusters *k* and the dataset that is to be clustered. Iterative refinement is used by the algorithm to produce the result. At the start, the algorithm randomly selects or generates *k* centroids from the dataset. Then the algorithm iterates between these two steps to generate the final result:

• **Assignment**: Each observation of the dataset is associated with the nearest centroid and assigned a cluster. The association is done based on the squared Euclidean distance. Each data point *y* is assigned a cluster based on the following equation:

$$\underset{c_i \in C}{\arg min}\, ED(c_i, y)^2$$

where *ED( )* is the function to calculate Euclidean distance, and c_i is the collection of centroids in *C*.

• **Centroid update**: In this step the centroid of the cluster is updated by computing the mean of the data points assigned to that centroids cluster in Step a. The equation used is:

$$c_i = \frac{1}{|A_i|} \sum_{y_i \in A_i} y_i$$

where A_i is the ith cluster centroid assigned in Step a.

The algorithm iterates between the above mentioned steps until the sum of the Euclidean distances is minimized or the maximum number of iterations is reached.

---

[13] McLoughlin et al 2015
[14] Figueriedo et al 2005
[15] Al-Wakeel and Wu 2016

To determine the number of clusters to study, we used the elbow test. In this method, we ran an initial set of k-means tests with a range of values of *k*, and plotted the sum of squared errors with increasing *k*. With a higher *k* value, the sum of squared errors within each cluster diminishes, indicating a higher degree of similarity between the observations. However, higher *k* values will also lead to less differentiation between the clusters.

Therefore, the inflection point, or “elbow”, of this curve represents the *k* value where the marginal similarity of cluster observations diminishes with additional clusters, indicating the ideal *k* value that maximizes similarity within each cluster and differentiation between them, thus identifying the *k* value that produces the most useful information.

The plots for our dataset are as follows:

Figure 1: ComEd dataset

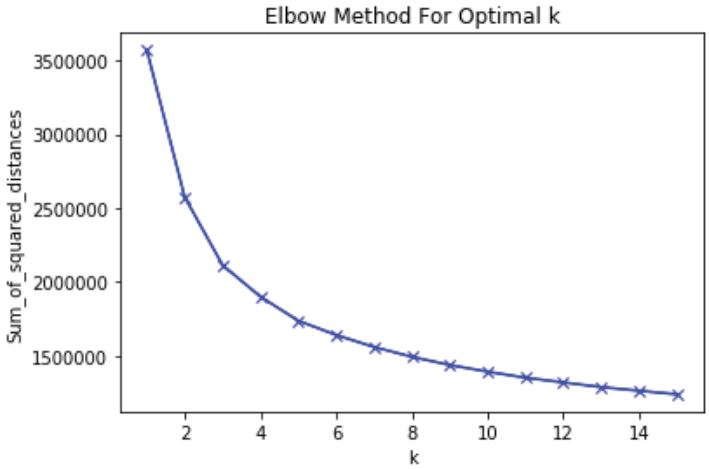


Figure 2: AIC dataset

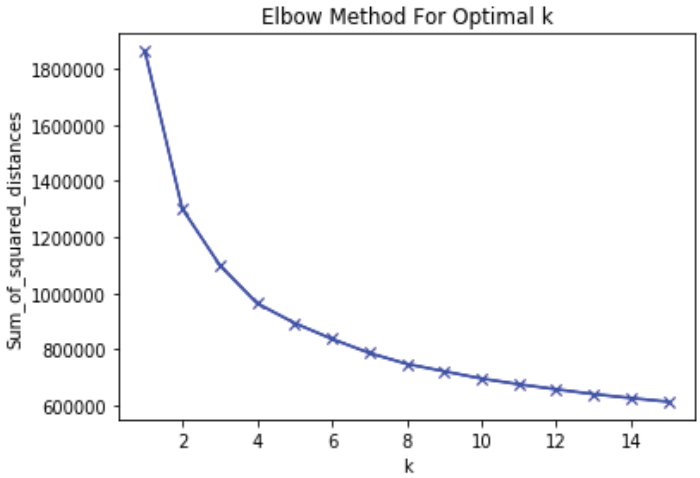
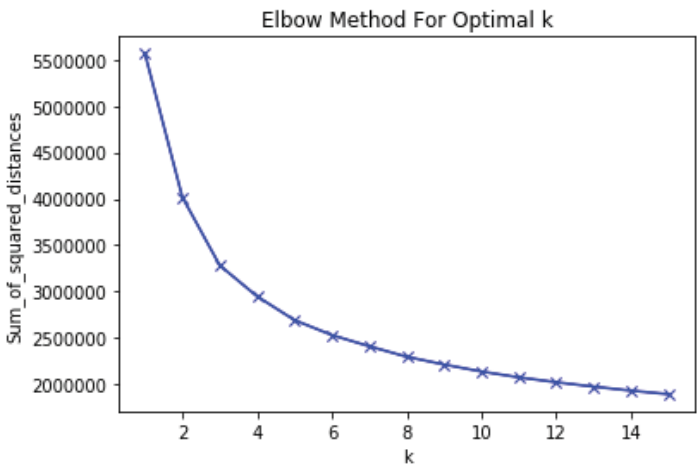


Figure 3: Combined dataset



From the above plots *k* = 6 was chosen as the best value for both the ComEd and AIC datasets.

2.3 Demographic analysis

To compare the demographic characteristics of customer clusters, we used customers' geographic identifiers (postal codes for ComEd customers and municipalities for AIC customers) to pair them with block group census data from the annual American Consumer Survey. Block groups are the smallest geographic division for which this census data is publicly reported. In our review of the literature, we evaluated several methods of associating customer clusters with demographic characteristics.[16]

We then selected multinomial logistic regression to link the clusters formed with customer demographics and household characteristics such as age, household built year, education level attainments, and income. This method uses one cluster as a basis of comparison for the other clusters, and determines the likelihood of a customer with specific demographics and household characteristics belonging to a particular cluster, as compared to the baseline. The equation used to calculate the likelihood or odds ratio $\exp(\beta)$ of belonging to a particular cluster is:

$$\exp(\beta) = \ln\left[\frac{p(y)}{p'y}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

Where, β_0 is a constant, β_1, β_2…. β_n are the regression coefficients of the independent explanatory variables X_1, X_2 …. X_n. Each explanatory variable is a binary indicator of whether or not a plurality of residents within a particular block group falls into that demographic category.[17] Our model uses 43 independent explanatory variables. The p(y) (Clusters 1-6 in our research) is the probability of a customer belonging to a cluster compared against a reference cluster $p'y$ (Cluster 1 in this case).

2.4 Low-income analysis

To further isolate the effect of lower incomes on electricity consumption, the same logistic model was used on the ComEd dataset using only a median income variable. This model used a binary variable to designate block groups with median annual incomes low enough to qualify a family of four for the Low Income Home Energy Assistance Program (LIHEAP), or $37,650. Due to the lack of postal code locations in the AIC dataset, this analysis was not applied to those customers.

3.0 Results

[16] Rhodes et al. 2014 and Beckel et al 2014

[17] The full list of variables in the model is available in Appendix 2.

3.1 Cluster load shapes

The k-means clustering algorithm produced six average cluster load shapes, presented here in terms of percentage of maximum load and average volume.[18]

Figure 4: Average summer weekday usage, in % of max loads

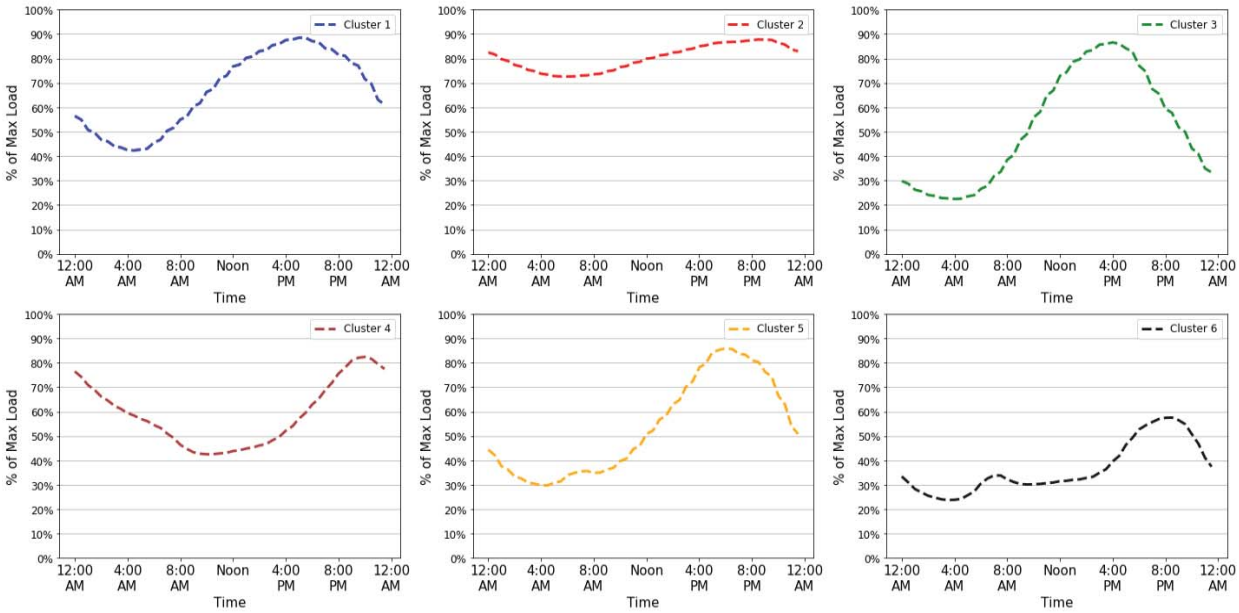


Figure 5: Average usage by customers in different clusters in KWh

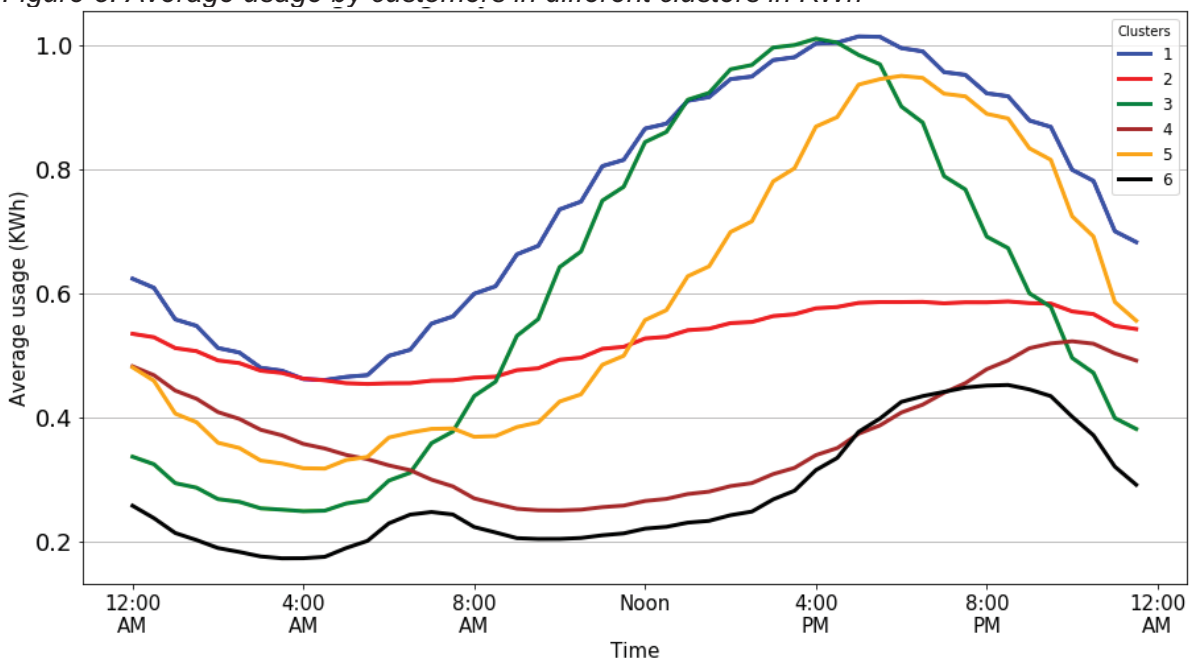


Figure 6: Cluster Composition

[18] Locations within five miles of the center of one of the 20 most populous Illinois cities outside the Chicago Metropolitan Area were considered exurban, areas greater than five miles from the same city centers were considered rural, and locations in the Chicago Metropolitan Area but outside the City of Chicago were considered suburban.

Cluster Composition			Delivery Service			Area			
Cluster	**Size**	**% of Total Data**	**Residential Single**	**Residential Multi**	**Ameren**	**Chicago**	**Chicagoland**	**Exurbs**	**Rural**
1	706,174	27.50%	238,912	140,421	326,841	195,308	76,720	95,204	338,942
2	421,782	16.42%	136,566	133,898	151,318	200,721	35,644	44,011	141,406
3	493,306	19.21%	198,446	78,375	216,485	84,867	67,904	78,485	262,050
4	282,773	11.01%	86,076	111,036	85,661	142,489	26,684	30,995	82,605
5	451,540	17.58%	169,173	88,111	194,256	116,546	54,524	67,648	212,822
6	212,439	8.27%	57,079	102,600	52,760	79,699	28,737	30,790	73,213
Total	2,568,014	100.00%	886,252	654,441	1,027,321	819,630	290,213	347,133	1,111,038

Cluster 1 contains 706,174 customers, representing 25% of ComEd customers and 32% of AIC customers. This cluster has a wide afternoon through evening peak, ranging between 40-85% of maximum load. Their lowest usage is 0.5 kWh at 4 a.m., while their highest usage reaches 1 kWh in mid-afternoon. This load shape is the most similar of all clusters to ComEd's system-wide load shape.

Cluster 2 represents 421,782 customers, accounting for 18% of ComEd customers and 15% of AIC customers. This cluster exhibits a very flat load shape, staying between 70-90% of maximum load throughout the day on average. These customers' average hourly volumes range between .45 kWh at 6 a.m. to a prolonged, flat peak near 0.6 kWh from 5-9 p.m.

Cluster 3 contains 493,306 customers, representing 18% of ComEd customers and 21% of AIC customers. These customers have the widest difference between base and peak load, ranging from 20-85% of maximum load, using 0.25 kWh at 4 a.m. and 1 kWh at their 4 p.m. peak.

Cluster 4 represents 282,773 customers, or 13% of ComEd customers and just 8% of AIC customers. This cluster exhibits a late evening peak at 10 p.m., using 85% of maximum load on average, and a late morning trough at 40% of maximum. Their average base load is 0.25 kWh, and average peak is at 0.55 kWh.

Cluster 5 contains 451,540 customers, accounting for 17% of ComEd customers and 19% of AIC customers. There is a slight mid-morning uptick in usage, followed by a steep, late-forming peak from late-afternoon through early evening. The average volume for this cluster ranges between 0.35 kWh at 4 a.m. to 0.95 kWh at 6 p.m.

Cluster 6 contains 212,439 customers, accounting for 10% of ComEd and 5% of AIC customers. This cluster has the lowest average usage of all the clusters, ranging from 0.1 kWh at 4 a.m. to 0.45 kWh at 8 PM. These customers exhibit a bimodal load shape, with a slight morning peak at 7 a.m. and an evening peak at 8 p.m. The average percentage range is between 25%-60% of maximum load.

While customers from each cluster appear throughout both the ComEd and Ameren Illinois service territories, there are strong correlations between geography and load shape. The maps below illustrate the prevalent load shapes in locations throughout the ComEd service territory. Each point in these maps represents a postal code, colored to represent the most frequently occurring cluster load shape in that location. They demonstrate the predominance of Clusters 2, 4, and 5 in Chicago and other city centers, and the frequency of Clusters 1 and 3 in suburban and rural areas.

Interestingly, the tendency of Clusters 2 and 5 to appear in central city locations extends to smaller cities, as shown in maps of Rockford and Aurora. Outside of Chicago, significantly fewer locations meet the initial anonymity screen, leading to the relative scarcity of observations in the rest of the service territory.

Figure 7: ComEd service territory

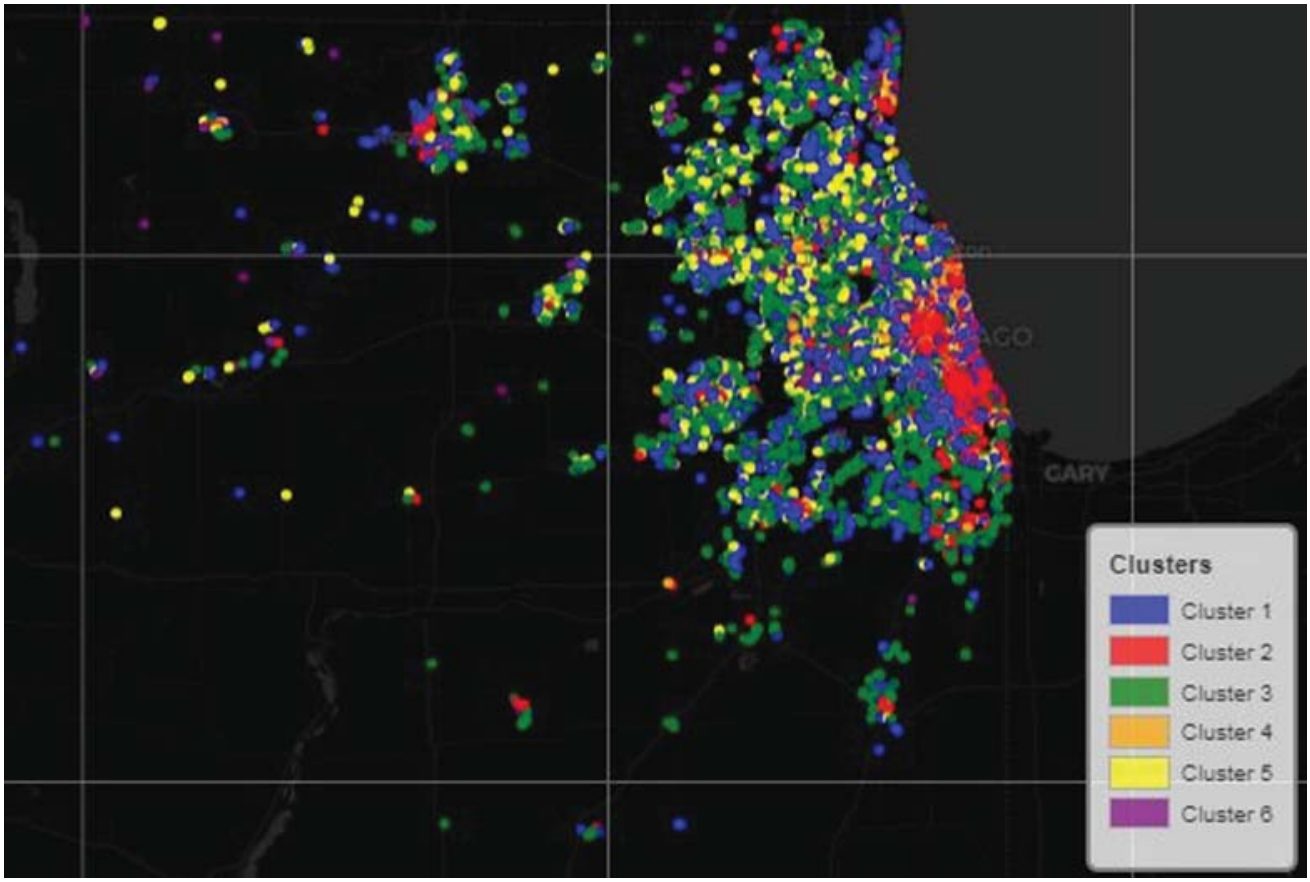


Figure 8: Chicago

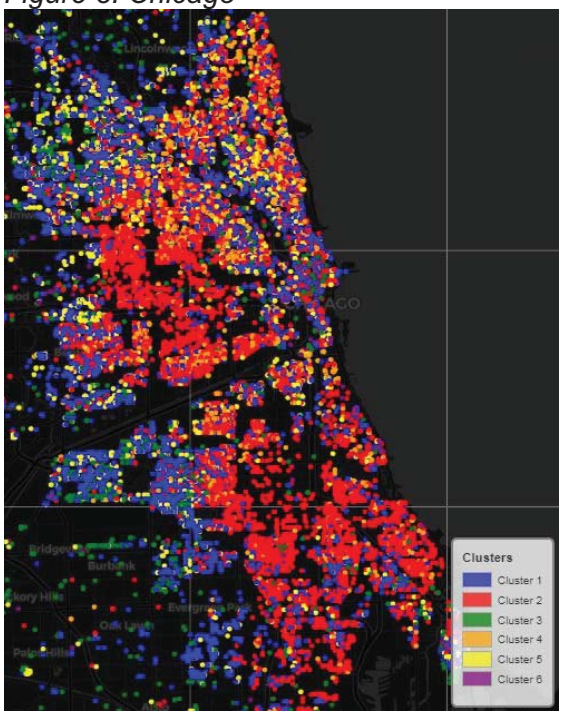


Figure 9: Rockford

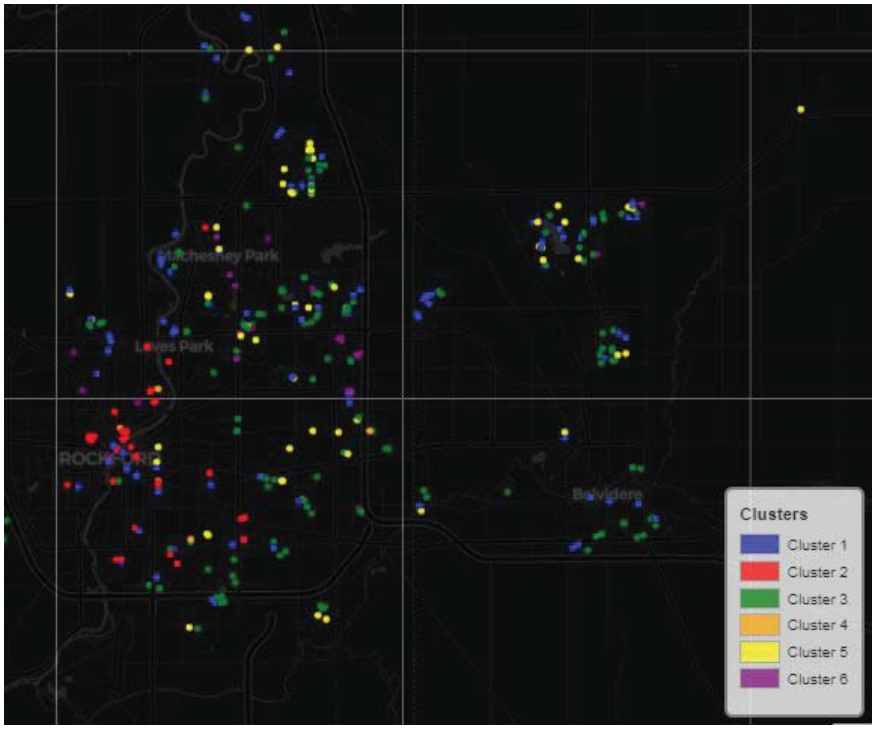
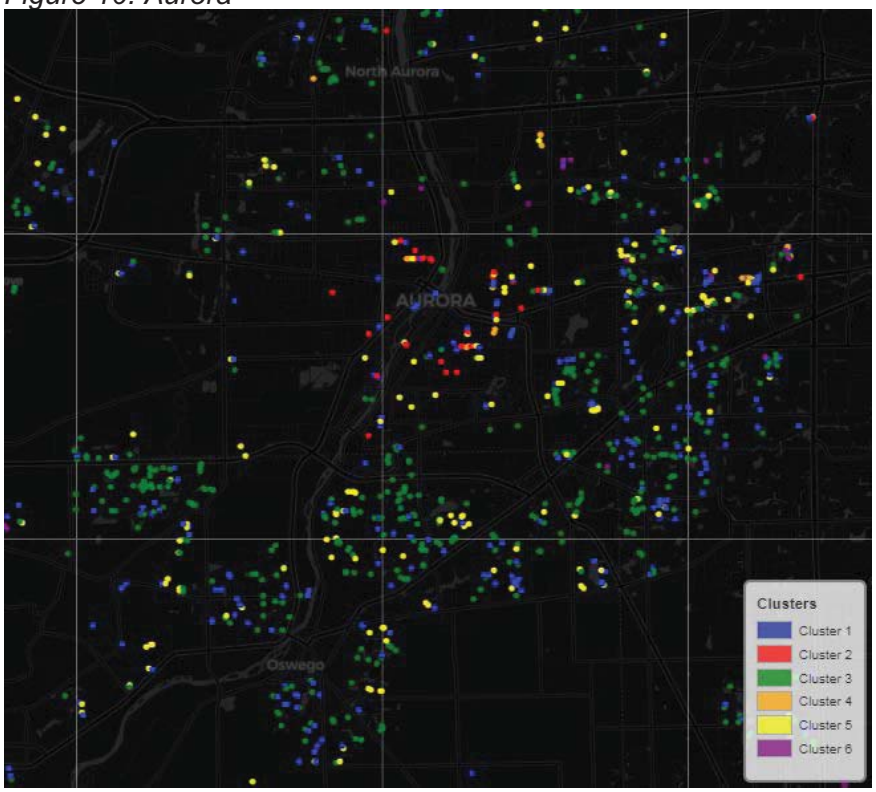


Figure 10: Aurora



3.2 Demographic analysis: ComEd

With geographic data at the postal code level, we were able to match ComEd customers to their Census block group, the smallest geographic level for which results are publicly available. This allowed for a fine-grained regression analysis that produced several statistically significant socio-demographic variables for every customer cluster, allowing for basic characterizations of these customer types.

The logistic regression model produces results in terms of relative likeliness of a cluster having a majority of customers who fit a particular demographic descriptor, as compared to Cluster 1. Cluster 1 was used as the seed for the model because its load shape is the closest to ComEd's overall load shape and it is geographically distributed throughout the footprint. Results greater than 1 indicate a cluster is more likely than Cluster 1 to be located in an area matching that variable, while results less than 1 indicate that cluster is less likely to match the variable.

Cluster 1: **ComEd baseline**
This cluster has the fewest significant demographic indicators, likely because it is the load shape closest to the overall ComEd load. For this reason, it was chosen as the seed for comparison in our logistic regression model; i.e. all of the regression coefficients for this cluster are equal to 1. However, we are able to make two inferences about these customers based on the other clusters' results.

Compared to the other clusters, these customers are relatively likely to have high income. The other clusters are all between 0.55 to 0.78 times as likely to earn $150k or more, with high levels of statistical significance. And while they are relatively well distributed throughout the footprint, they are relatively less likely to live in high density areas: Clusters 2, 3, 4, and 6 were between 1.1 and 1.5 times more likely to live in a block group with greater than 660 housing units, with Cluster 5's density result lacking significance. In other demographic categories, clusters had significant results both below and above 1, indicating Cluster 1 to be relatively "average" in those traits.

Cluster 2: **City flat** (young, urban apartment dwellers and low-income households)

Figure 11: Cluster Two

Category	Variable	Coefficient
Age	< 33	1.565***
	33 – 56	1.162*
	> 56	0.539***
House Built	Before 1951	1.767***
	1951-99	0.69***
	After 1999	0.805**
Education	No HS Degree	0.791***
	No Secondary Degree	0.904**
	AA	0.899*
	BA	0.753***
	Grad Degree	1.291***
	PhD	1.249***
Heating	Electric	0.698***
	Gas	0.74***
HH Makeup	Family - 2 person	1.097
	Family - 3-5	0.886**
	Family - Over 5	0.882***
	Non-Fam - <3	0.9
	Non-Fam - 3-5	1.024
	Non-Fam - Over 5	0.477
Home Value	< 10700	0.24
	10700 - 104999	0.929
	10500 - 15999	0.523***
	> 250000	0.343***
Housing Units	< 361	1.082
	361- 475	0.869**
	476 - 660	0.939
	> 660	1.11*
Median Income	< 50000	3.726***
	50000 - 74999	1.304***

	75000 - 99999	0.681***
	100000 - 149999	0.614***
	> 149999	0.549***
Rooms/Unit	< 4	0.903
	4 - 5	0.728***
	> 5	0.609***
Occupancy	Occupied	0.804
	Vacant	1.421
Location	Chicago	2.482***
	Chicago suburbs	0.977
	Exurb	0.67***
	Rural	0.604***

Significance Levels: ***0.01 **0.05 *0.1

Cluster 2 appears to contain two customer profiles. These customers are likely to be relatively young, educated, and live in a Chicago apartment. They are also the most likely of all clusters to earn less than $50,000 per year. These factors suggest this load shape pattern is representative of both young professionals and low-income households, which is supported by the map of Chicago; high concentrations of these customers are found in both the west and south sides of the city, areas with high levels of poverty, and the relatively affluent north side of the city, which is popular with young professionals.

Cluster 2 customers are the most likely to live in the City of Chicago, nearly 2.5 times more likely than the baseline. They are 1.6 times more likely to be younger than 33, 1.2 times more likely than the baseline to be between 33 and 56, and only 0.5 times as likely to be over 56. They are relatively likely to hold a post-graduate degree (1.3 for graduate degree, 1.2 for doctorate), and the least likely to not hold a high school degree (0.8). They are likely to live in older buildings in high density block groups, and 3.7 times more likely than the baseline to earn less than $50,000. A flat overall load shape supports the likelihood of apartment dwelling without central air conditioning.

Cluster 3: Exurban retirees

Figure 12: Cluster Three

Category	Variable	Coefficient
Age	< 33	1.259
	33 - 56	1.221
	> 56	1.95**
House Built	Before 1951	0.648
	1951-99	2.062**
	After 1999	2.245**
Education	No HS Degree	1.048
	No Secondary Degree	1
	AA	0.951
	BA	0.852***
	Grad Degree	0.636***
	PhD	0.807**
Heating	Electric	1.097
	Gas	1.226**
HH Makeup	Family - 2 person	0.88**
	Family - 3-5	0.985
	Family - Over 5	0.913*
	Non-Fam - <3	0.729*
	Non-Fam - 3-5	0.639**
	Non-Fam - Over 5	1.063
Home Value	< 10700	71.483
	10700 - 104999	0.744*

	10500 - 15999	0.808
	> 250000	0.851
Housing Units	< 361	1.136
	361- 475	1.238
	476 - 660	1.385
	> 660	1.54*
Median Income	< 50000	1.528**
	50000 - 74999	1.459**
	75000 - 99999	1.503**
	100000 - 149999	1.386*
	> 149999	0.646**
Rooms/Unit	< 4	0.916
	4 - 5	1.088
	> 5	1.482***
Occupancy	Occupied	0.07
	Vacant	0.067
Location	Chicago	0.497***
	Chicago suburbs	1.35
	Exurb	2.165***
	Rural	2.064***

Significance Levels: ***0.01 **0.05 *0.1

These customers are likely to be older, live in exurban or rural areas, and dwell in buildings with five or more rooms built after 1951. This cluster is the most likely of all clusters to be over 56 years of age and the most likely to be rurally located. Those in Cluster 3 are 2.2 times more likely to live in an exurb, 2 times more likely to live in a rural area, and half as likely to live in Chicago. While they appear to be relatively distributed throughout the income range, they are the least educated cluster, as they are the least likely to hold an advanced degree as well as the most likely to hold less than a college degree. Their load shape with a high, early peak supports the profile of a larger residence that is occupied in the afternoon.

Cluster 4: City duck curve (younger urban apartment dwellers)

Figure 13: Cluster Four

Category	Variable	Coefficient
Age	< 33	1.946 ***
	33 – 56	1.202
	> 56	0.656 **
House Built	Before 1951	1.946 ***
	1951-99	0.927
	After 1999	0.85
Education	No HS Degree	1.256 ***
	No Secondary Degree	0.596 ***
	AA	0.633 ***
	BA	1.002
	Grad Degree	1.997 ***
	PhD	1.259 ***
Heating	Electric	0.721 ***
	Gas	0.731 **
HH Makeup	Family - 2 person	1.146
	Family - 3-5	0.959
	Family - Over 5	0.978
	Non-Fam - <3	0.664
	Non-Fam - 3-5	0.805
	Non-Fam - Over 5	0.009
Home Value	< 10700	0.73
	10700 – 104999	0.532 ***
	10500 – 15999	0.428 ***

	> 250000	0.327 ***
Housing Units	< 361	1.024
	361- 475	0.942
	476 – 660	1.118
	> 660	1.423 ***
Median Income	< 50000	2.553 ***
	50000 – 74999	1.605 ***
	75000 – 99999	0.837 *
	100000 – 149999	0.806 *
	> 149999	0.555 ***
Rooms/Unit	< 4	0.957
	4 – 5	0.521 ***
	> 5	0.161 ***
Occupancy	Occupied	2.83
	Vacant	0.326
Location	Chicago	1.777 ***
	Chicago suburbs	0.857
	Exurb	1.016
	Rural	0.991

Significance Levels: ***0.01 **0.05 *0.1

These customers exhibit very similar demographics to Cluster 2; they are likely to be relatively young (1.9 times likely to be under 33, 0.7 times likely to be over 56), and live in older, smaller residences in Chicago. Interestingly, these customers are both relatively likely to have no high school while also the most likely to hold a post-graduate degree. They most likely earn less than $75,000. With low average volume and a high, late peak, it is likely these customers are more likely than Cluster 2 customers to have all household residents employed.

Cluster 5: Exurban, middle class

Figure 14: Cluster Five

Category	Variable	Coefficient
Age	< 33	0.822 **
	33 - 56	0.897
	> 56	0.806 *
House Built	Before 1951	0.803 **
	1951-99	0.87
	After 1999	0.851 *
Education	No HS Degree	1.073
	No Secondary Degree	0.838 ***
	AA	1.158 **
	BA	1.26 ***
	Grad Degree	0.984
	PhD	0.929
Heating	Electric	0.792 **
	Gas	1.046
HH Makeup	Family - 2 person	0.851 **
	Family - 3-5	1.015
	Family - Over 5	0.89 **
	Non-Fam - <3	0.743
	Non-Fam - 3-5	0.741

	Non-Fam - Over 5	0.809
Home Value	< 10700	0.569
	10700 - 104999	0.794
	10500 - 15999	0.744 *
	> 250000	0.852
Housing Units	< 361	0.832 **
	361- 475	0.906
	476 - 660	0.847 **
	> 660	0.931
Median Income	< 50000	0.941
	50000 - 74999	1.045
	75000 - 99999	0.932
	100000 - 149999	0.983
	> 149999	0.66 ***
Rooms/Unit	< 4	1.019
	4 - 5	1.237 **
	> 5	1.166
Occupancy	Occupied	2.244
	Vacant	1.758
Location	Chicago	0.555 ***
	Chicago suburbs	0.784 ***
	Exurb	1.234 ***
	Rural	1.108

Significance Levels: ***0.01 **0.05 *0.1

Cluster 5 has relatively few significant demographic indicators. These customers are the most likely cluster to hold a bachelor's or associates degree, at 1.2 for an AA and 1.3 for a BA. They are 0.8 times as likely to be under 33, with a statistically weak 0.8 likelihood of being under 56.They are unlikely to earn more than $150,000, 0.7 times as likely as the baseline, and more likely to be over 56 years of age.

These customers are 0.6 times as likely to live in Chicago, and 1.2 times as likely to live in an exurb. With a load shape that exhibits a sharper, later peak and small morning peak, it is likely there is no one home during the day, suggesting families with working parents.

Cluster 6: Low-volume space heaters

Figure 15: Cluster Six

Category	Variable	Coefficient
Age	< 33	0.952
	33 - 56	0.654 ***
	> 56	0.546 ***
House Built	Before 1951	0.884
	1951-99	0.768 **
	After 1999	0.501 ***
Education	No HS Degree	0.866
	No Secondary Degree	0.561 ***

	AA	0.688 ***
	BA	0.991
	Grad Degree	1.439 ***
	PhD	1.157 *
Heating	Electric	2.515 ***
	Gas	0.756 **
HH Makeup	Family - 2 person	0.931
	Family - 3-5	0.562 ***
	Family - Over 5	0.453 ***
	Non-Fam - <3	1.709
	Non-Fam - 3-5	1.561
	Non-Fam - Over 5	0.001
Home Value	< 10700	0.847
	10700 - 104999	0.947
	10500 - 15999	0.518 ***
	> 250000	0.662 **
Housing Units	< 361	0.561 ***
	361- 475	0.572 ***
	476 - 660	0.8 **
	> 660	1.323 ***
Median Income	< 50000	0.854
	50000 - 74999	0.825 **
	75000 - 99999	0.751 ***
	100000 - 149999	0.823 **
	> 149999	0.781 *
Rooms/Unit	< 4	1.349 ***
	4 - 5	0.941
	> 5	0.516 ***
Occupancy	Occupied	2.815 *
	Vacant	2.328
Location	Chicago	0.406 ***
	Chicago suburbs	0.749 ***
	Exurb	1.091
	Rural	1.026

Significance Levels: ***0.01 **0.05 *0.1

Cluster six customers are likely to reside in smaller residences (less than four rooms) in high-density block groups that are outside of Chicago; of all the clusters, in fact, they are the least likely to live in Chicago. They are also the least likely to live with family members, and the most likely to hold an advanced degree. They are also less likely than average to earn less than $150,000. Finally, these customers are 2.5 times more likely than the baseline to have electric space heating.[19]

3.3 Demographic analysis: Ameren Illinois

[19] This is an interesting result and one we will examine further in a follow-up analysis focused on winter load shapes.

As described above, our AIC usage data had a much lower level of geographic specificity, resulting in few significant results. The first issue is low confidence in the local demographic indicators. Since the smallest area we are able to isolate in this dataset is individual cities, the initial census data must be aggregated over a much larger area. This significantly reduces the accuracy of demographic estimates.

The second obstacle this raises is a low diversity of cluster assignments from city to city. Because Cluster 4 contains a plurality of all consumers, and these customers are distributed throughout the service territory, 800 of the 860 cities included in the dataset would be considered Cluster 4 cities according to the original methodology. Cities were then reassigned cluster designations according to their second most prevalent cluster type, and restricted to cluster types that accounted for more than ten cities. The final dataset for regression included Clusters 1, 2, and 4.

Figure 16: Cluster Two

Category	Variable	Coefficient
Age	< 33	1.011
	33 - 56	0.941
	> 56	1.466
House Built	Before 1951	0.699
	1951-99	0.415 **
	After 1999	1.505
Education	No HS Degree	1.3
	No Secondary Degree	0.789
	AA	0.575 *
	BA	0.182 ***
	Grad Degree	1.136
	PhD	0.357
Heating	Electric	0.415 ***
	Gas	0.921
HH Makeup	Family - 2 person	1.032
	Family - 3-5	0.845
	Family - Over 5	1.285
	Non-Fam - <3	0.678
	Non-Fam - 3-5	1.128
	Non-Fam - Over 5	0.0002
Home Value	< 10700	1.039
	10700 - 104999	1.654
	10500 - 15999	0.623
	> 250000	0.752
Housing Units	< 361	1.771
	361- 475	1.366
	476 - 660	1.52
	> 660	0.685
Median Income	< 50000	1.845 *
	50000 - 74999	1.616 *
	75000 - 99999	0.851
	100000 - 149999	1.844
	> 149999	0.338

Rooms/Unit	< 4	0.645
	4 - 5	1.026
	> 5	1.166
Occupancy	Occupied	0.24 ***
Location	Exurb	1.86
	Rural	1.354

Significance Levels: ***0.01 **0.05 *0.1

Figure 17: Cluster Three

Category	Variable	Coefficient
Age	< 33	1.156
	33 - 56	1.095
	> 56	2.275 ***
House Built	Before 1951	0.671
	1951-99	0.974
	After 1999	2.367 ***
Education	No HS Degree	1.166
	No Secondary Degree	0.74
	AA	0.752
	BA	0.62
	Grad Degree	0.405 **
	PhD	0.638
Heating	Electric	0.733
	Gas	0.788
HH Makeup	Family - 2 person	1.341
	Family - 3-5	1.339
	Family - Over 5	1.9 *
	Non-Fam - <3	1.529
	Non-Fam - 3-5	2.064
	Non-Fam - Over 5	4.796
Home Value	< 10700	1.95 **
	10700 - 104999	1.918 **
	10500 - 15999	0.602
	> 250000	0.717
Housing Units	< 361	0.685
	361- 475	0.986
	476 - 660	1.408
	> 660	1.166
Median Income	< 50000	1.217
	50000 - 74999	1.066
	75000 - 99999	0.606 **
	100000 - 149999	0.813
	> 149999	0.737
Rooms/Unit	< 4	0.753
	4 - 5	1.454

	> 5	0.969
Occupancy	Occupied	0.56 *
Location	Exurb	0.648
	Rural	1.713

Significance Levels: ***0.01 **0.05 *0.1

As the table shows, the model produced few significant results for AIC customers. However, there are two results that both have statistical significance and are consistent with results from the ComEd analysis:

• Cluster 4 customers are the most likely to earn less than $50,000 and to earn less than $50,000-$75,000. This agrees with the ComEd result for Cluster 1, which found them significantly more likely to earn less than $50,000, and less likely to earn more than $75,000. This supports the conclusion that a low volume, flat load shape is likely associated with lower income customers.

• Cluster 3 customers are the most likely to be over 56. This agrees with the ComEd results, where Cluster 3 customers are also the most likely cluster to be over 56. However, while ComEd Cluster 3 customers had no significant correlation with household makeup, AIC customers in this cluster are most likely to live with 5+ family members.

3.4 Low-income analysis

Our analysis of low-income cluster assignments in ComEd showed that low-income households were significantly more likely to exhibit lower overall volumes and flatter load-shapes. In fact, the single most likely cluster for low-income households is Cluster 2, with a percentage of maximum load difference of only 20%, average half-hourly volume between only 0.45 – 0.6 kWh, and a late peak at 9 PM. Low-income customers were 3.5 times more likely than average to exhibit this usage profile.

The next most likely usage profile for these areas is Cluster 4, with a similar late peak but much lower proportional usage during the day, with average usage ranging between 0.25 and 0.5 kWh per half-hour. Low-income customers were 2.5 times more likely than average to exhibit this load shape. Customers in both of these clusters are highly likely to live in relatively dense areas of Chicago. All told, more than 50% of ComEd low-income customers have a flat, low-volume usage profile.

Figure 18: Low-income regression results

Demographics and household characteristics	Clusters									
	Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
	Exp(β)	Std. Error	Exp(β)	Std. Error	Exp(β)	Std. Error	Exp(β)	Std. Error	Exp(β)	Std. Error
LIHEAP	2.472***	0.0077	0.666***	0.0102	1.774***	0.0088	0.702***	0.0102	0.973***	0.0111
Constant	0.637***	0.0027	0.749***	0.0026	0.489***	0.0029	0.699***	0.0026	0.415***	0.0031

Figure 19: Low-Income Cluster Summary

Cluster Composition			Delivery Service		Area			
Cluster	**Size**	**% of Total Data**	**Residential Single**	**Residential Multi**	**Chicago**	**Chicagoland**	**Exurbs**	**Rural**
1	30884	20.68%	11341	19543	24037	3574	2440	833
2	48599	32.55%	17185	31414	42367	3141	2362	729
3	15405	10.32%	5477	9928	10020	2838	1910	637
4	26806	17.95%	9095	17711	23880	1652	920	354
5	15152	10.15%	5852	9300	11929	1652	1096	475
6	12469	8.35%	2634	9835	9658	1632	947	232
Total	**149315**	**100.00%**	**51584**	**97731**	**121891**	**14489**	**9675**	**3260**

To further illustrate the correlation between load shape and income, these maps show the location and local median income for members of each cluster, throughout the service territory as well as within Chicago. These maps indicate postal codes where each cluster is the most prevalent load shape, with each location colored according to the median income level.

Figure 20: Income Groups by Cluster

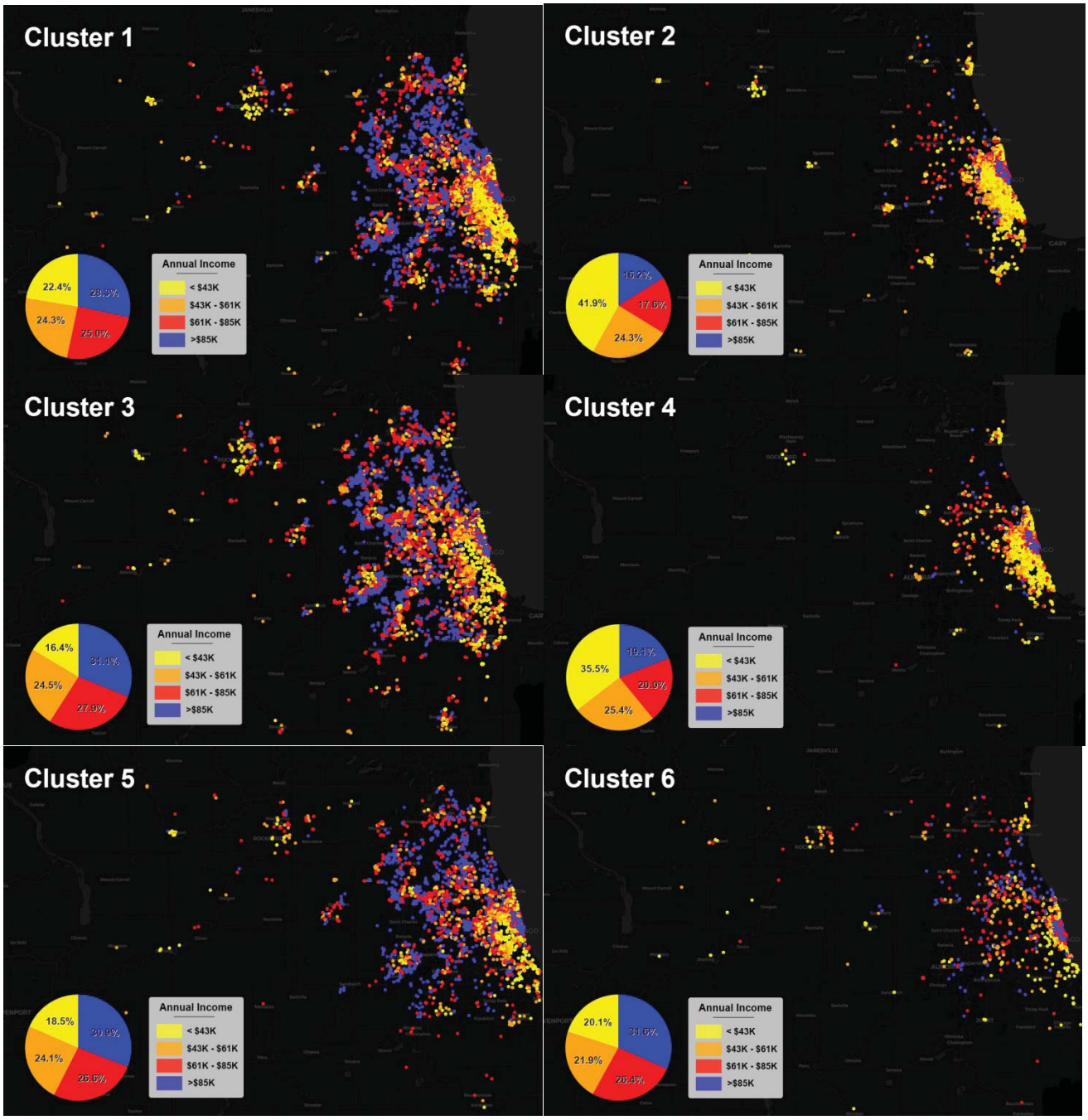
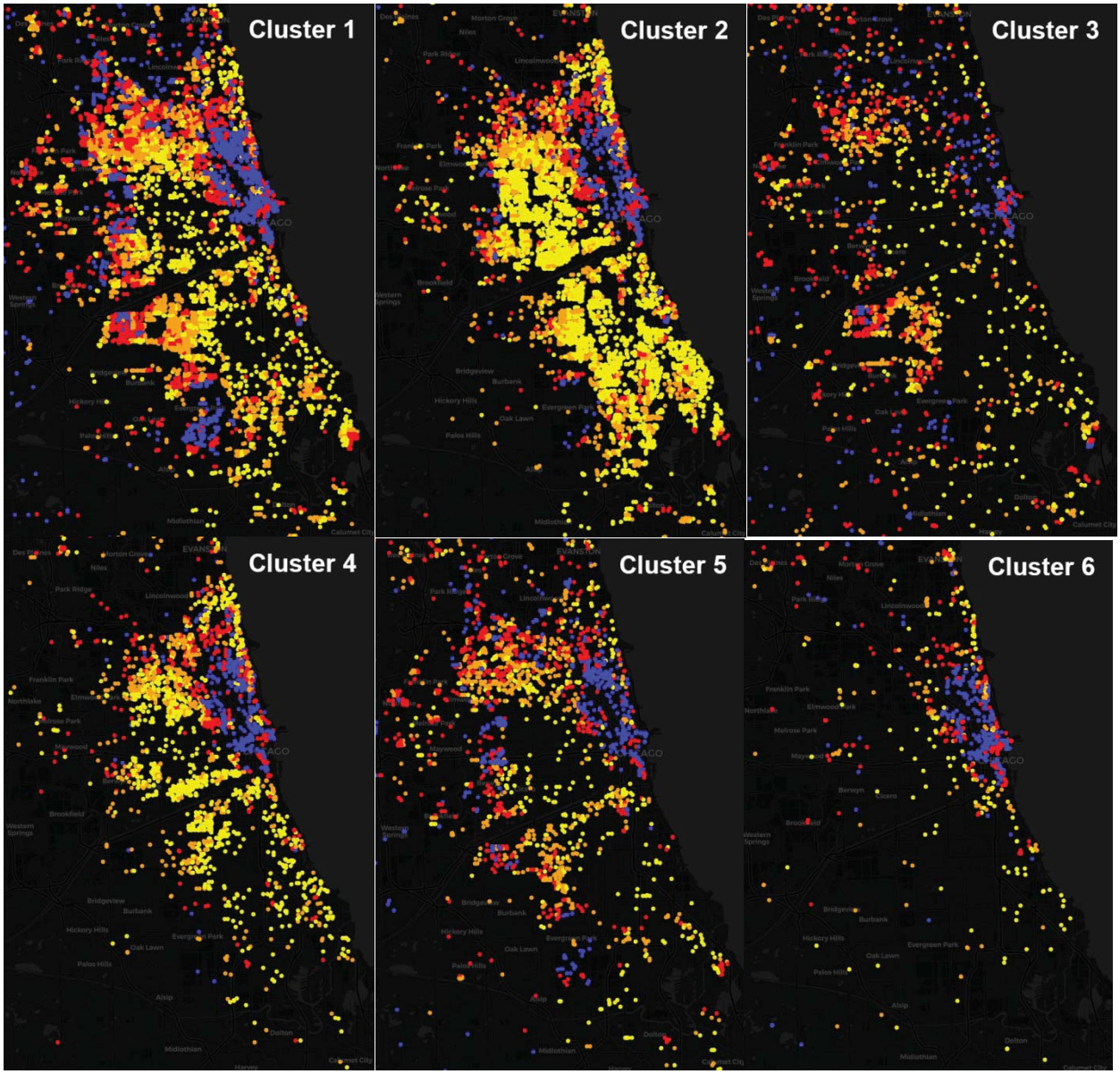


Figure 21: Income Groups by Cluster, Chicago



The service territory maps illustrate the significantly higher occurrence of clusters 1, 3, and 5 both in areas earning between $61,000 and $84,000, and $85,000 or more. By contrast, clusters 2 and 4 are highly concentrated in areas earning less than $61,000 and less than $43,000. This disparity is particularly salient in Chicago. The second set of maps show Cluster 2 customers highly concentrated in areas earning less than $43,000, on the west and south sides of the city: areas that have long suffered the effects of endemic poverty, racial discrimination, and disinvestment.

4.0 Conclusion

While previous studies seeking to segment electricity consumers had access to survey data and personal identifiers, this study sought to perform meaningful segmentation on these customers without the help of that information. By clustering anonymous data and applying a logistic regression model to a dataset using only geographically associated Census data, we were able to discover statistically significant results that marked clear demographic delineations between different types of electricity consumers.

This information can be used to improve the effectiveness of energy efficiency programs and dynamic rate designs by helping to target those initiatives at those customers whose participation would have the biggest impact on the system, as well as those customers who would benefit from them the most. It also

can facilitate more informed policy analysis, by helping to determine the potential impacts of rate design decisions on vulnerable communities, and to evaluate the equity of cost allocations.

High peak usage is a large driver of system costs, requiring more robust distribution and transmission grids, and higher capacity requirements in areas with capacity markets.[20] From this perspective, customers with high peaks relative to their overall volume likely pay less in their electricity bills than the system costs that they actually cause, while customers with flatter load shapes have higher volume relative to their peak usage, leading to them paying more than their fair share.

That we find such a high correlation between flat load usage and lower incomes suggests that this cross-subsidization has particularly harmful consequences, considering low-income households already pay a higher proportion of their income on utility bills. This finding should encourage utilities and utility commissions to adopt a wider offering of dynamic rate designs that may more accurately reflect customers' cost of service, reducing this potential cross-subsidization.[21]

Another conclusion we can draw from these results is the high value proposition of energy efficiency and distributed energy resources in reducing system costs. Programs encouraging energy efficiency adoption and distributed energy resources investment in urban areas are important and beneficial for low-income communities. However, expansion of demand response and price responsive demand programs in suburban and ex-urban areas may have a greater overall impact on system costs, reducing bills for all customers.

As a result of lower-density areas being excluded due to the anonymity screen, our ComEd dataset has a higher proportion of urban customers than the actual population of ComEd customers. One consequence of this, and of our inability to do demographic analysis on Ameren customers, may be a low sampling of rural low-income customers. One avenue for further research would be to perform this analysis using ComEd's ZIP code dataset, which contains significantly more customers at a lower level of geographic granularity. This would likely increase our subset of rural low-income communities, at the cost of lower confidence in our demographic estimates.

Finally, a crucial piece of usage data that is missing from our dataset is customers' peak load contributions (PLCs). With customer PLCs, we could perform more detailed cost of service estimates and rate design evaluations, with a higher degree of confidence. Theoretically, it would be possible to determine these values through calculation. However, in the case of ComEd, customers are assigned new anonymous identifiers every two months, preventing the time-series analysis that would be required. Utility commissions with data access policies such as Illinois' should consider requiring PLCs in the usage data, which can be done while still preserving individual customer anonymity.

References

Al-Wakeel, A., Wu, J., 2016. K-means Based Cluster Analysis of Residential Smart Meter Measurements. Energy Procedia Vol. 88, 754-760. https://doi.org/10.1016/j.egypro.2016.06.066

Beckel, C., Sadamori, L., Staake, T., and Santini, S., 2014. Revealing household characteristics from smart meter data. Energy Vol. 78, 397-410. DOI: 10.1016/j.energy.2014.10.025

Big Energy Data Center, Citizens Utility Board. https://www.citizensutilityboard.org/welcome-big-energy-data-center/ **(accessed 22 May 2019)**

[20] This is a particular problem in the ComEd service territory, where PJM capacity market charges now constitute roughly 21%of the average residential customers' bill. See Monitoring Analytics, LLC. Q1 State of the Market Report for PJM.

[21] The implications for rate design and cost of service regulation is a thorny area that warrants further investigation. In future studies, we intend to examine more of the details and implications.

Burger, S., Knittel, C., Perez-Arriaga, I., Schneider, I., vom Scheidt, F., 2019. The Efficiency of Distributional Effects of Alternative Residential Electricity Rate Designs. NBER Working Paper No. 25570. http://www.nber.org/papers/w25570

Burger, S., Schneider, I., Botterud, A., Perez-Arriaga, I., 2019, Fair, Equitable, and Efficient Tariffs in the Presence of Distributed Energy Resources. Consumers, Prosumers, Prosumagers: How service innovations will disrupt the utility business model https://doi.org/10.1016/B978-0-12-816835-6.00008-5

Chen, C., Hwang, J., Huang, C., 1997, Application of Load Survey Systems to Proper Tariff Design, IEEE Transactions on Power Systems Vol. 12 No. 4. DOI: 10.1109/59.627886

Faruqui, A., Hledik, R., Newell, S., Pfeifenberger, J., 2007. The Power of Five Percent, The Electricity Journal, Volume 20, Issue 8. https://www.smartgrid.gov/files/The_Power_Five_Percent_How_Dynamic_Pricing_Can_Save_35_Billi_200709.pdf **(accessed 22 May 2019)**

Figueiredo, V., Rodrigues, F., Vale, Z., Gouveia, J., 2005. An Electric Energy Consumer Characterization Framework Based om Data Mining Techniques. IEEE Transactions on Power Systems Vol. 20 No. 2. DOI: 10.1109/TPWRS.2005.846234

McLoughlin, F., Duffy, A., Conlon, M., 2015. A clustering approach to domestic electricity load profile characterization using smart meter data. Applied Energy. 141, 190-199. https://arrow.dit.ie/cgi/viewcontent.cgi?article=1063&context=dubenart **(accessed 22 May 2019)**

Melissa Inc., Geo*Code, Melissa Direct, https://www.melissa.com/direct/reference-data/#3

Monitoring Analytics, LLC. Q1 State of the Market Report for PJM, 2018 http://www.monitoringanalytics.com/reports/PJM_State_of_the_Market/2018/2018q1-som-pjm.pdf

Rhodes, J., Cole, W., Upshaw, C., Edgar, T., Webber, M., 2014. Clustering analysis of residential electricity demand profiles. Applied Energy 135, 461-471. DOI: 10.1016/j.apenergy.2014.08.111

US Census Bureau, 2017 ACS Estimates, American FactFinder, https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml

Yu, Y., Liu, G., Zhu, W., Wang, F., Shu, B., Zhang, K., Astier, N., Rajagopal, R., 2018. Good Consumer or Bad Consumer: Economic Information Revealed From Demand Profiles, IEEE Transactions on Smart Grid Vol. 9 No. 3, 2347-2358. DOI: 10.1109/TSG.2017.2662684

Appendix A: Full Regression Results

ComEd

Demographics and Household characteristics	Clusters									
	Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
	Exp(β)	Std. Error	Exp(β)	Std. Error	Exp(β)	Std. Error	Exp(β)	Std. Error	Exp(β)	Std. Error
age_less_than_33	1.565***	0.083	1.259	0.315	1.946***	0.134	0.822**	0.095	0.952	0.118
age_33_56	1.162*	0.082	1.221	0.314	1.202	0.133	0.897	0.093	0.654***	0.117
age_more_than_56	0.539***	0.113	1.950**	0.319	0.656**	0.17	0.806*	0.121	0.546***	0.145
built_before_1951	1.767***	0.081	0.648	0.316	1.946***	0.132	0.803**	0.092	0.884	0.124
built_1951_1999	0.690***	0.078	2.062**	0.314	0.927	0.128	0.87	0.088	0.768**	0.118
built_after_1999	0.805*	0.094	2.245**	0.316	0.85	0.14	0.851*	0.096	0.501***	0.121
less_than_high_school	0.791***	0.048	1.048	0.053	1.256***	0.07	1.073	0.057	0.866	0.105
some_college_no_degree	0.904**	0.046	1	0.045	0.596***	0.066	0.838***	0.051	0.561***	0.096

associate_degree	0.899*	0.057	0.951	0.056	0.633***	0.094	1.158**	0.06	0.688***	0.118
bachelor_degree	0.753***	0.059	0.852***	0.05	1.002	0.079	1.260***	0.055	0.991	0.089
graduate_or_prof_school_degree	1.291***	0.065	0.636***	0.055	1.997***	0.082	0.984	0.057	1.439***	0.086
PhD	1.249***	0.071	0.807**	0.092	1.259***	0.078	0.929	0.079	1.157*	0.08
electric_heating	0.698***	0.096	1.097	0.083	0.721***	0.122	0.792**	0.101	2.515***	0.116
gas_heating	0.74***	0.099	1.226**	0.09	0.731**	0.124	1.046	0.108	0.756**	0.116
family_2_person	1.097	0.06	0.880**	0.057	1.146	0.085	0.851**	0.063	0.931	0.112
family_3_5	0.886**	0.053	0.985	0.052	0.959	0.076	1.015	0.058	0.562***	0.107
family_more_than_5	0.882***	0.047	0.913*	0.053	0.978	0.066	0.890**	0.056	0.453***	0.125
nonfamily_less_than_3	0.9	0.189	0.729*	0.19	0.664	0.25	0.743	0.205	1.709	0.408
nonfamily_3_5	1.024	0.191	0.639**	0.195	0.805	0.252	0.741	0.208	1.561	0.408
nonfamily_more_than_5	0.477	0.651	1.063	0.23	0.009	5.762	0.809	0.312	0.001	11.457
value_less_than_10700	0.24	10.099	71.483	5.228	0.73	14.595	0.569	12.395	0.847	19.288
value_10700_104999	0.929	0.125	0.744*	0.162	0.532***	0.168	0.794	0.187	0.947	0.197
value_105000_159999	0.523***	0.106	0.808	0.15	0.428***	0.124	0.744*	0.167	0.518***	0.166
value_more_than_250000	0.343***	0.112	0.851	0.154	0.327***	0.13	0.852	0.171	0.662**	0.168
h_units_less_than_361	1.082	0.066	1.136	0.294	1.024	0.109	0.832**	0.078	0.561***	0.139
h_units_361_475	0.869**	0.0631	1.238	0.237	0.942	0.104	0.906	0.073	0.572***	0.115
h_units_476_660	0.939	0.061	1.385	0.236	1.118	0.101	0.847**	0.07	0.800**	0.102
h_units_more_than_660	1.110*	0.06	1.540*	0.236	1.423***	0.099	0.931	0.069	1.323***	0.094
income_less_than_50000	3.726***	0.066	1.528**	0.194	2.553***	0.096	0.941	0.073	0.854	0.097
income_50000_74999	1.304***	0.062	1.459**	0.191	1.605***	0.092	1.045	0.063	0.825**	0.085
income_75000_99999	0.681***	0.068	1.503**	0.191	0.837*	0.098	0.932	0.062	0.751***	0.085
income_100000_149999	0.614***	0.077	1.386*	0.193	0.806**	0.103	0.983	0.07	0.823**	0.091
income_more_than_149999	0.549***	0.144	0.646**	0.212	0.555***	0.18	0.660***	0.104	0.781*	0.139
rooms_less_than_4	0.903	0.104	0.916	0.157	0.957	0.106	1.019	0.145	1.349***	0.103
rooms_4_5	0.728***	0.081	1.088	0.087	0.521***	0.094	1.237**	0.097	0.941	0.094
rooms_more_than_5	0.609***	0.096	1.482***	0.094	0.161***	1.141	1.166	0.105	0.516***	0.129
Occupied	0.804	0.434	0.07	2206	0.283	0.799	2.244	0.515	2.815*	0.576
Vacant	1.421	0.433	0.067	2.206	0.326	0.798	1.758	0.513	2.328	0.573
Chicago	2.482***	0.067	0.497***	0.237	1.777***	0.105	0.555***	0.072	0.406***	0.099
Chicagoland	0.977	0.069	1.35	0.236	0.857	0.111	0.784***	0.071	0.749***	0.096
Exurb	0.670***	92	2.165***	0.237	1.016	0.131	1.234***	0.076	1.091	0.103
Rural	0.604***	0.087	2.064***	0.236	0.991	0.127	1.108	0.073	1.026	0.099
Constant	0.981	0.212	2.999	0.939	1.535	0.365	0.595**	0.251	0.340***	0.319

Ameren

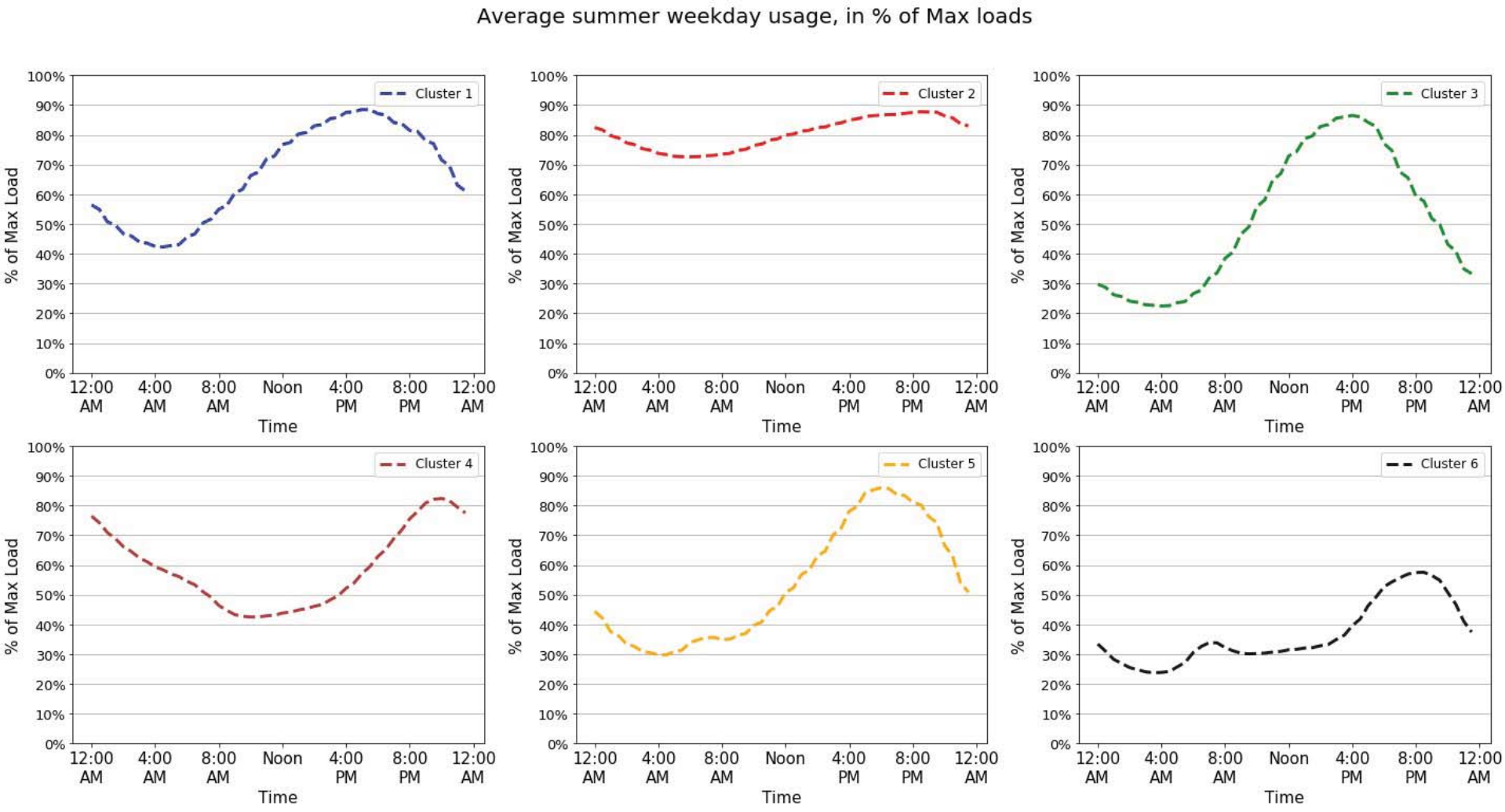
Demographics and Household characteristics	**Clusters**			
	Cluster 2		**Cluster 3**	
	Exp(β)	**Std. Error**	**Exp(β)**	**Std. Error**
age_less_than_33	1.011	0.312	1.156	0.232
age_33_56	0.941	0.293	1.095	0.221
age_more_than_56	1.466	0.32	20275***	0.246
built_before_1951	0.699	0.379	0.671	0.286
built_1951_1999	0.415***	0.352	0.974	0.253
built_after_1999	1.505	0.45	2.367***	0.316
less_than_high_school	1.3	0.403	1.166	0.329
some_college_no_degree	0.789	0.342	0.74	0.25
associate_degree	0.575*	0.326	0.752	0.247
bachelor_degree	0.182***	0.543	0.62	0.317
graduate_or_prof_school_degree	1.136	0.724	0.405**	0.448
PhD	0.357	0.96	0.638	0.54
electric_heating	0.415***	0.318	0.733	0.229
gas_heating	0.921	0.267	0.788	0.205
income_less_than_50000	1.845*	0.326	1.217	0.257
income_50000_74999	1.616*	0.278	1.066	0.222
income_75000_99999	0.851	0.3	0.606**	0.227
income_100000_149999	1.844	0.409	0.813	0.283
income_more_than_149999	0.338	0.756	0.723	0.402
family_2_person	1.032	0.387	1.341	0.288
family_3_5	0.845	0.385	1.339	0.282
family_more_than_5	1.285	0.461	1.900*	0.38
nonfamily_less_than_3	0.678	1.587	1.529	1.257
nonfamily_3_5	1.128	1.531	2.064	1.217

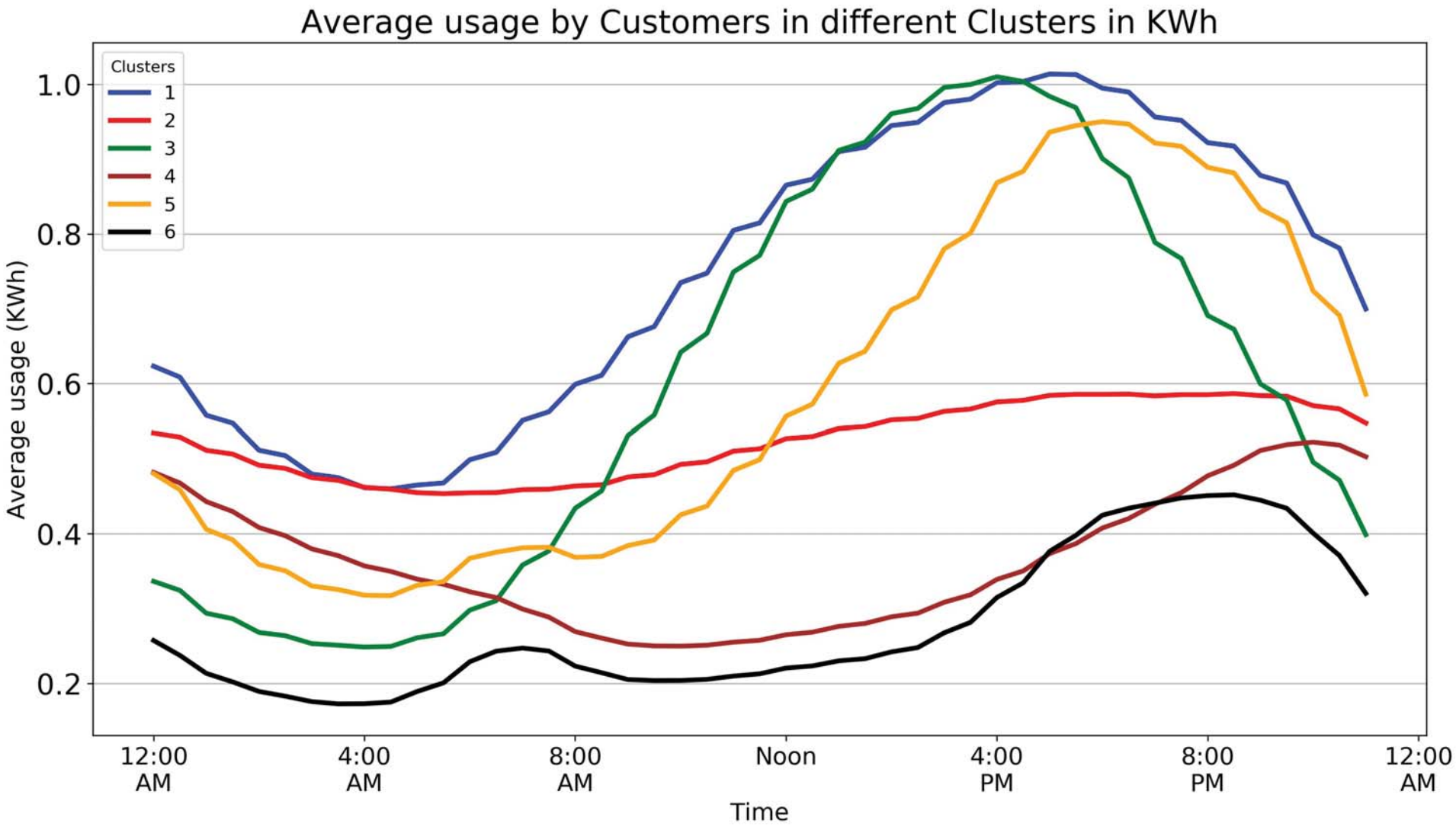
nonfamily_more_than_5	0.0002	124.084	4.796	1.22
value_less_than_10700	1.039	0.514	1.95	0.433
value_10700_104999	1.654	0.371	1.918**	0.279
value_105000_159999	0.623	0.303	0.602**	0.216
value_150000_250000	0.715	0.43	0.789	0.279
value_more_than_250000	0.752	0.581	0.717	0.39
h_units_less_than_361	1.771	0.378	0.685	0.289
h_units_361_475	1.366	0.344	0.986	0.253
h_units_476_660	1.52	0.341	1.408	0.248
h_units_more_than_660	0.685	0.587	1.166	0.313
rooms_less_than_4	0.645	0.483	0.753	0.365
rooms_4_5	1.026	0.356	1.454	0.269
rooms_more_than_5	1.166	0.382	0.969	0.29
exurb	1.86	0.601	0.648	0.464
rural	1.354	0.545	1.713	0.425
occupied	0.240***	0.345	0.560*	0.306
constant	2.519	1.036	1.11	0.812

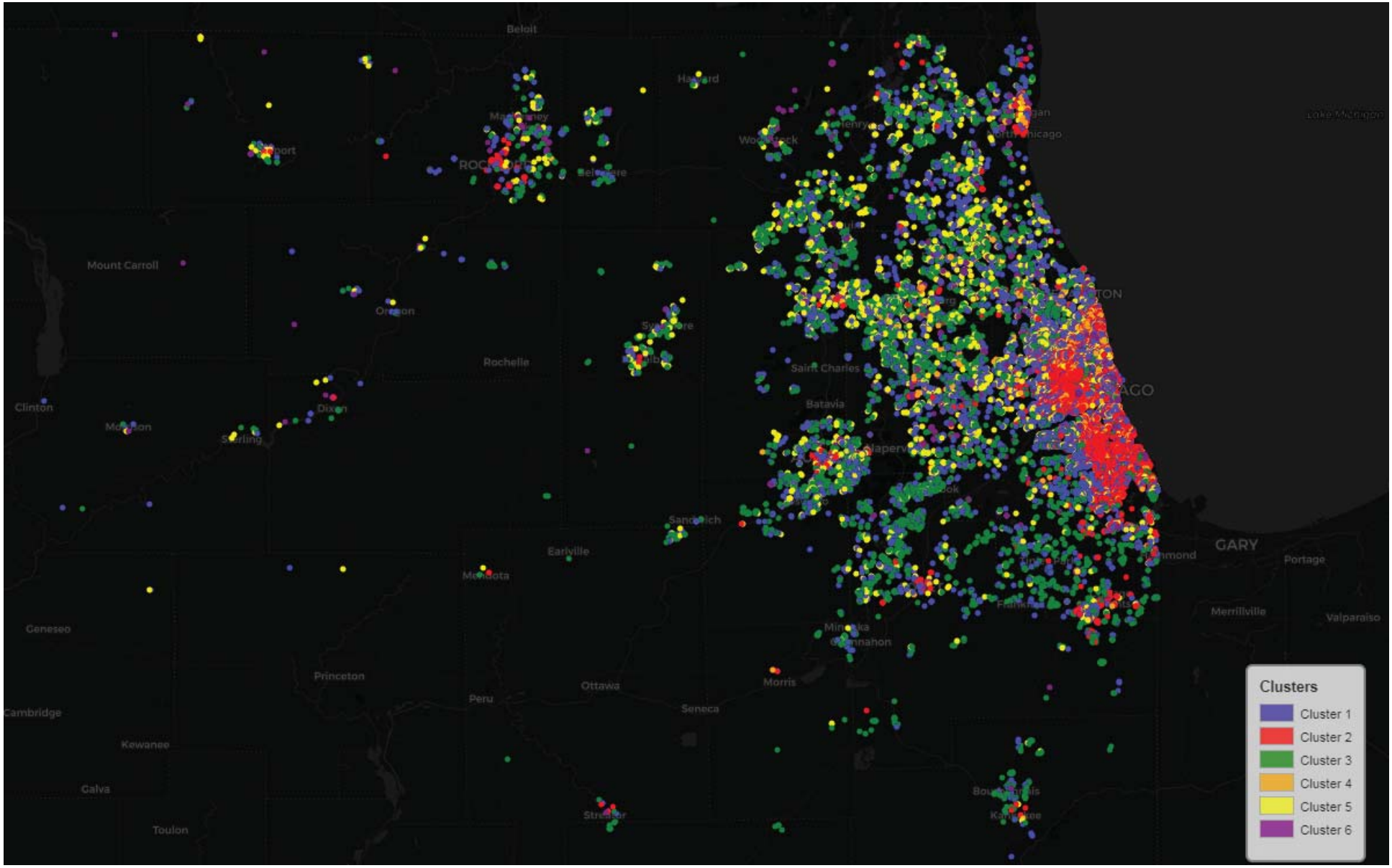
Appendix B: Model Variable Definitions

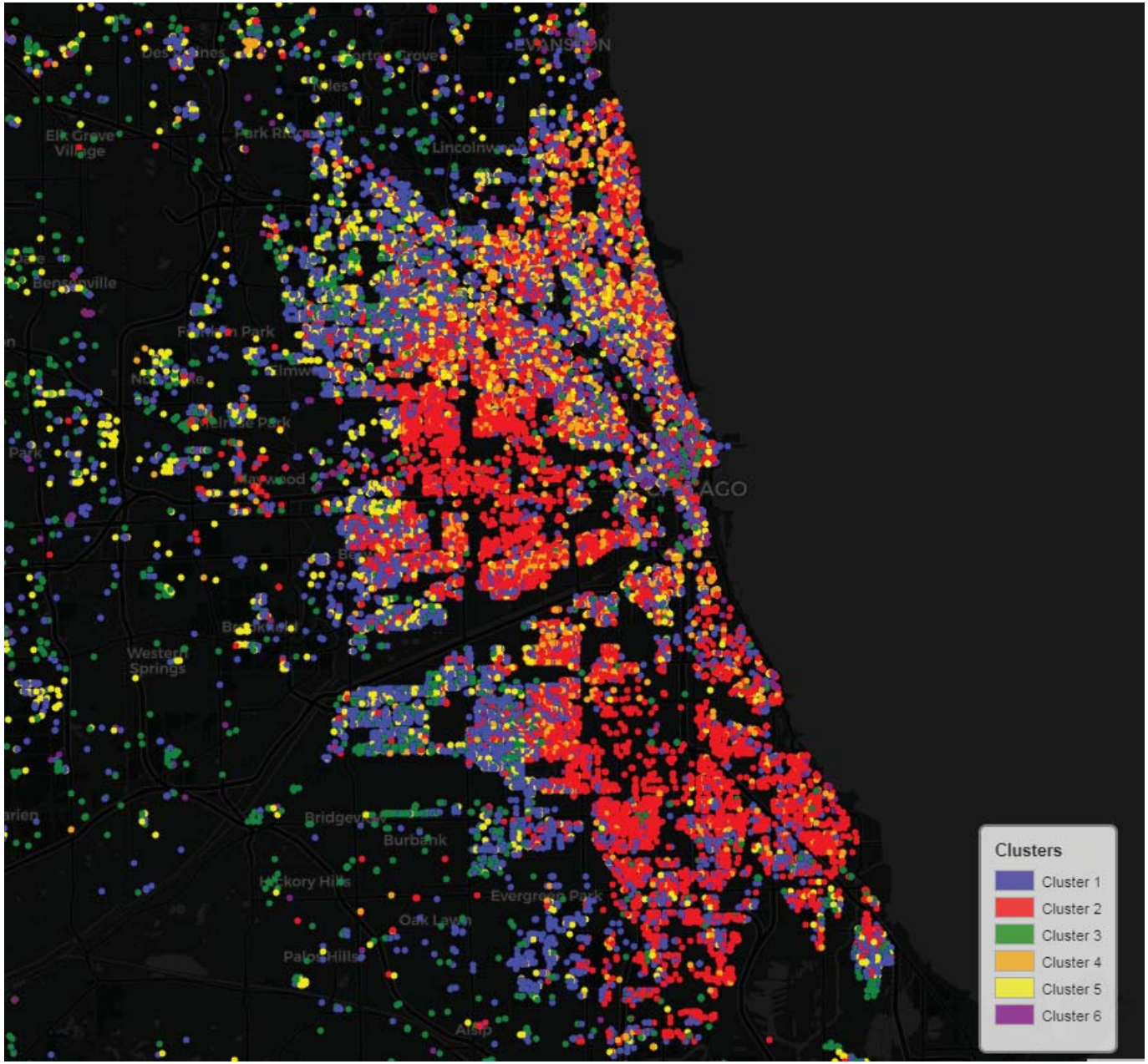
age_less_than_33	Average head of household less than 33 years of age
age_33_56	Average head of household between 33 and 56
age_more_than_56	Average head of household greater than 56
built_before_1951	Most residences built prior to 1951
built_1951_1999	Most residences built between 1951 and 1999
built_after_1999	Most residences built after 1999
less_than_high_school	Average household has no residents hold high school degrees
some_college_no_degree	Average household has resident attended college, but did not graduate
associate_degree	Highest level of educational attainment within average household is associates degree
bachelor_degree	Highest level of educational attainment within average household is bachelor's degree
graduate_or_prof_school_degree	Highest level of educational attainment within average household is masters or other professional degree
PhD	Highest average level of educational attainment is doctorate
electric_heating	Average household uses electric heating
gas_heating	Average household uses gas heating
income_less_than_50000	Average annual household income is less than $50k
income_50000_74999	Average annual household income is between $50k and $75k
income_75000_99999	Average annual household income is between $75k and $100k
income_100000_149999	Average annual household income is between $100k and $150k
income_more_than_149999	Average annual household income is greater than $150k
family_2_person	Average household consists of one or two related or married individuals
family_3_5	Average household consists of three to five related or married individuals
family_more_than_5	Average household consists of more than 5 married or related or married individuals

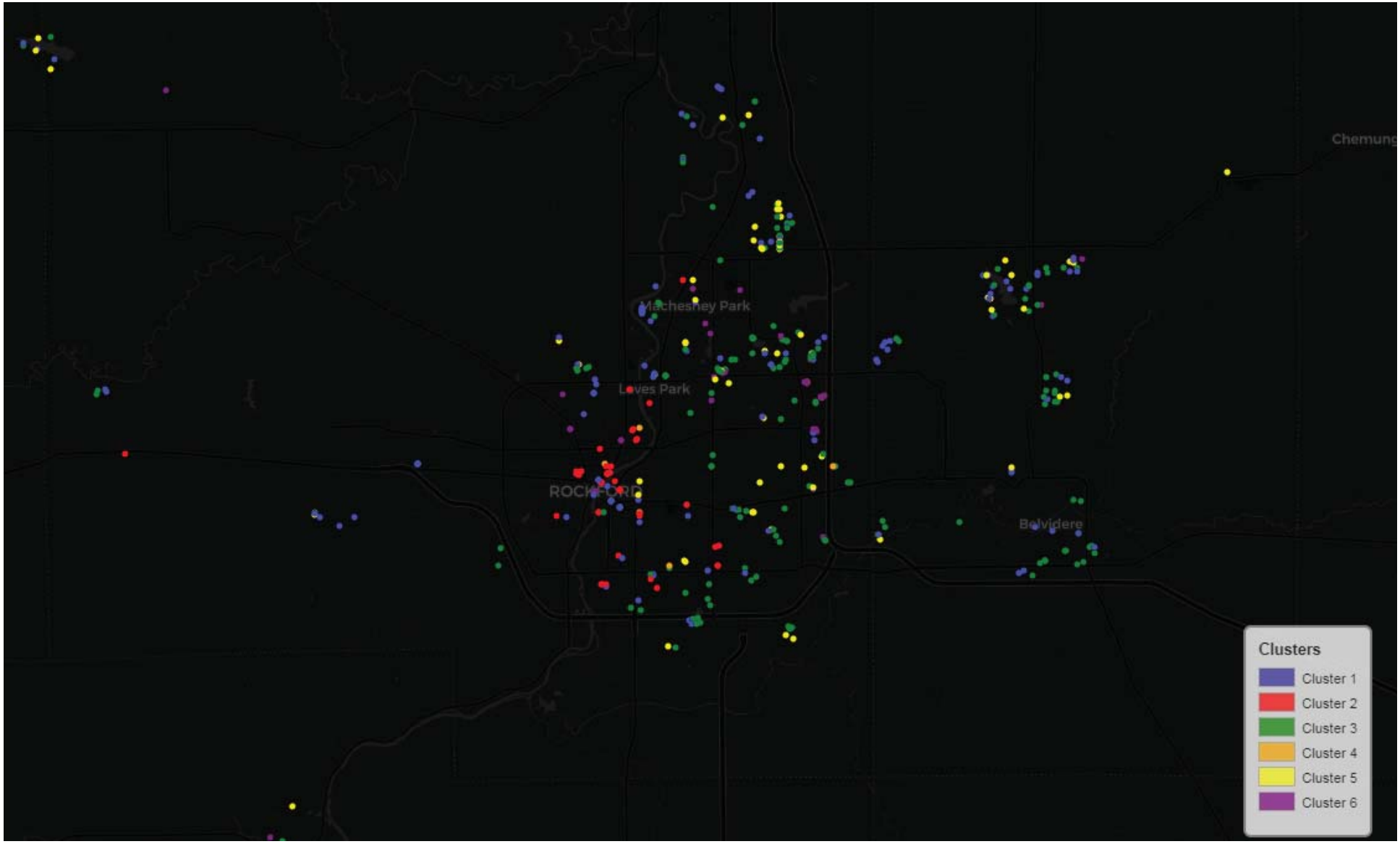
nonfamily_less_than_3	Average household consists of one or two unrelated individuals
nonfamily_3_5	Average household consists of three to five unrelated individuals
nonfamily_more_than_5	Average household consists of greater than five unrelated individuals
value_less_than_10700	Average home value is less than $10,700
value_10700_104999	Average home value is between $10,700 and $105k
value_105000_159999	Average home value is between $105k and $160k
value_150000_250000	Average home value is between $150k and $250k
value_more_than_250000	Average home value is greater than $250k
h_units_less_than_361	Block group contains fewer than 361 housing units
h_units_361_475	Block group contains between 361 and 475 housing units
h_units_476_660	Block group contains between 476 and 660 housing units
h_units_more_than_660	Block group contains greater than 660 housing units
rooms_less_than_4	Average home has less than four rooms
rooms_4_5	Average home has four or five rooms
rooms_more_than_5	Average home has more than five rooms
chicago	Block group is within Chicago city limits
chicagoland	Block group is outside of the City of Chicago but within the Chicago Metropolitan Area
exurb	Block groups is located within five miles of the center of one of the 20 most populous cities in Illinois, besides Chicago
rural	Block groups is outside of five miles from any of the 20 most populous cities in Illinois
occupied	Majority of housing units in block group are occupied

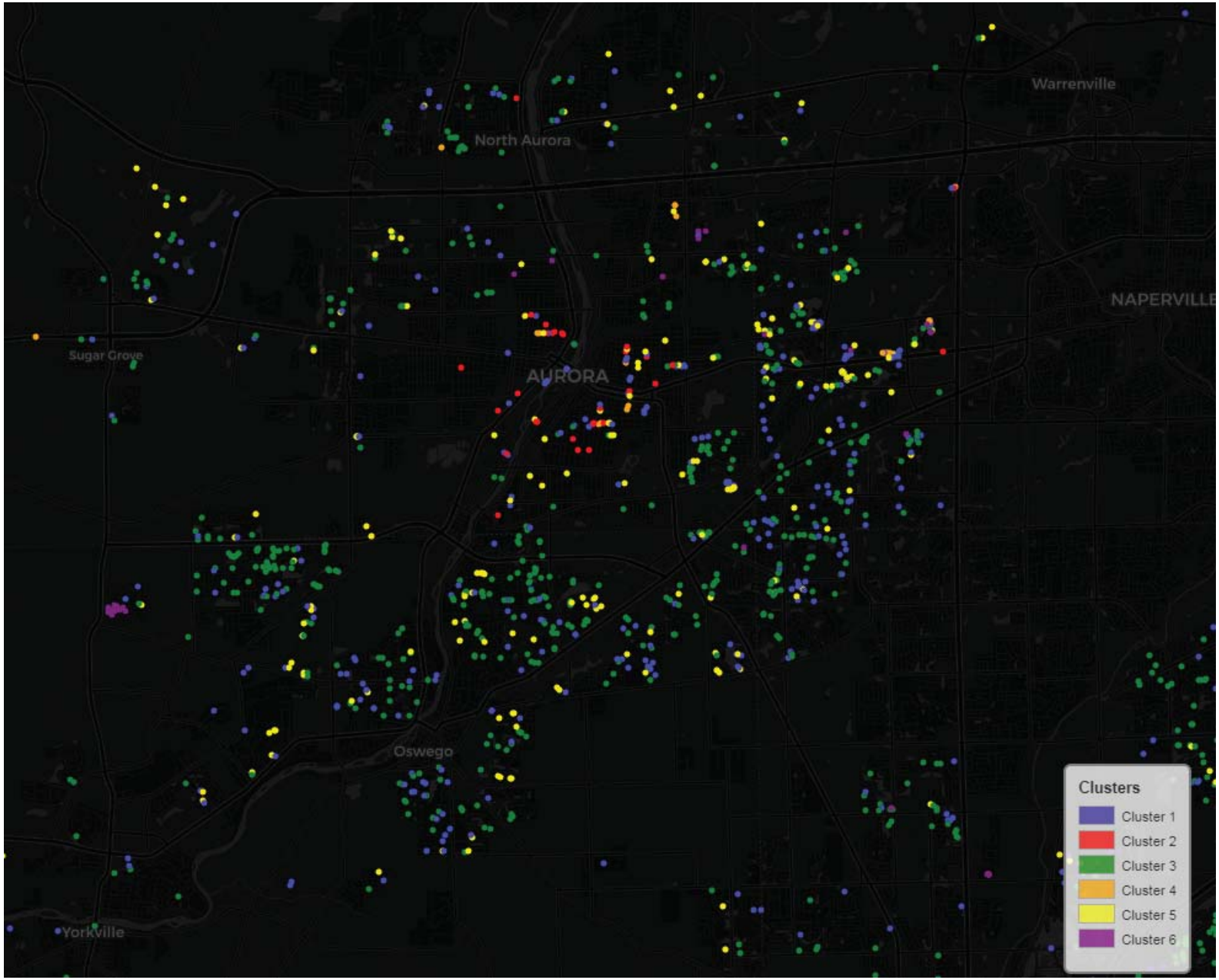

Average summer weekday usage, in % of Max loads
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
% of Max Load
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
12:00 AM
4:00 AM
8:00 AM
Noon
4:00 PM
8:00 PM
12:00 AM
Time

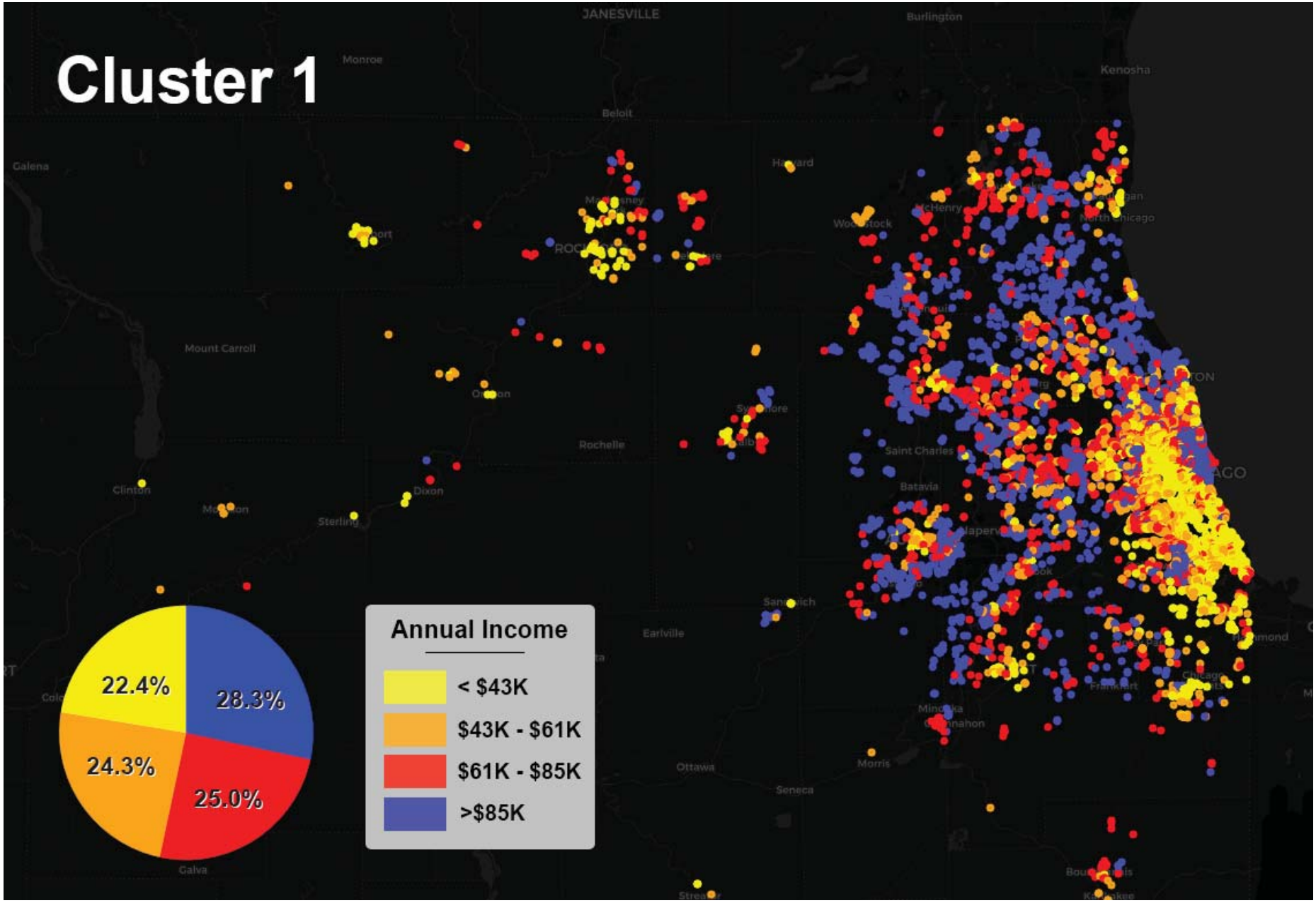

Average usage by Customers in different Clusters in KWh
Clusters
1
2
3
4
5
6
1.0
0.8
0.6
0.4
0.2
Average usage (KWh)
12:00 AM
4:00 AM
8:00 AM
Noon
4:00 PM
8:00 PM
12:00 AM
Time

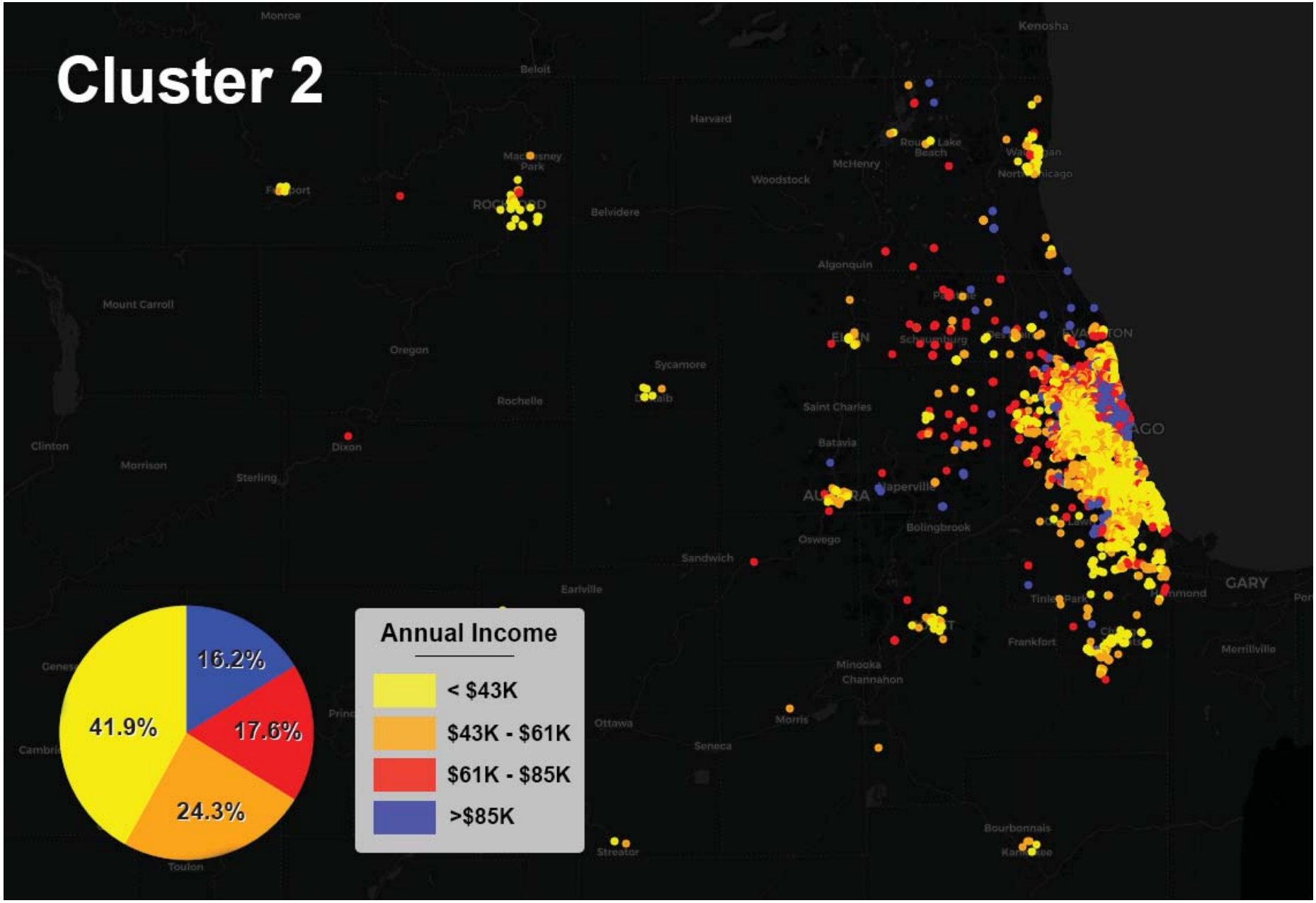

Beloit
Mount Carroll
Oregon
Rochelle
Clinton
Dixon
Sterling
Saint Charles
Batavia
Sandwich
Earlville
Mendota
Geneseo
Princeton
Peru
Ottawa
Seneca
Morris
Cambridge
Kewanee
Galva
Toulon
GARY
Portage
Merrillville
Valparaiso
Clusters
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6

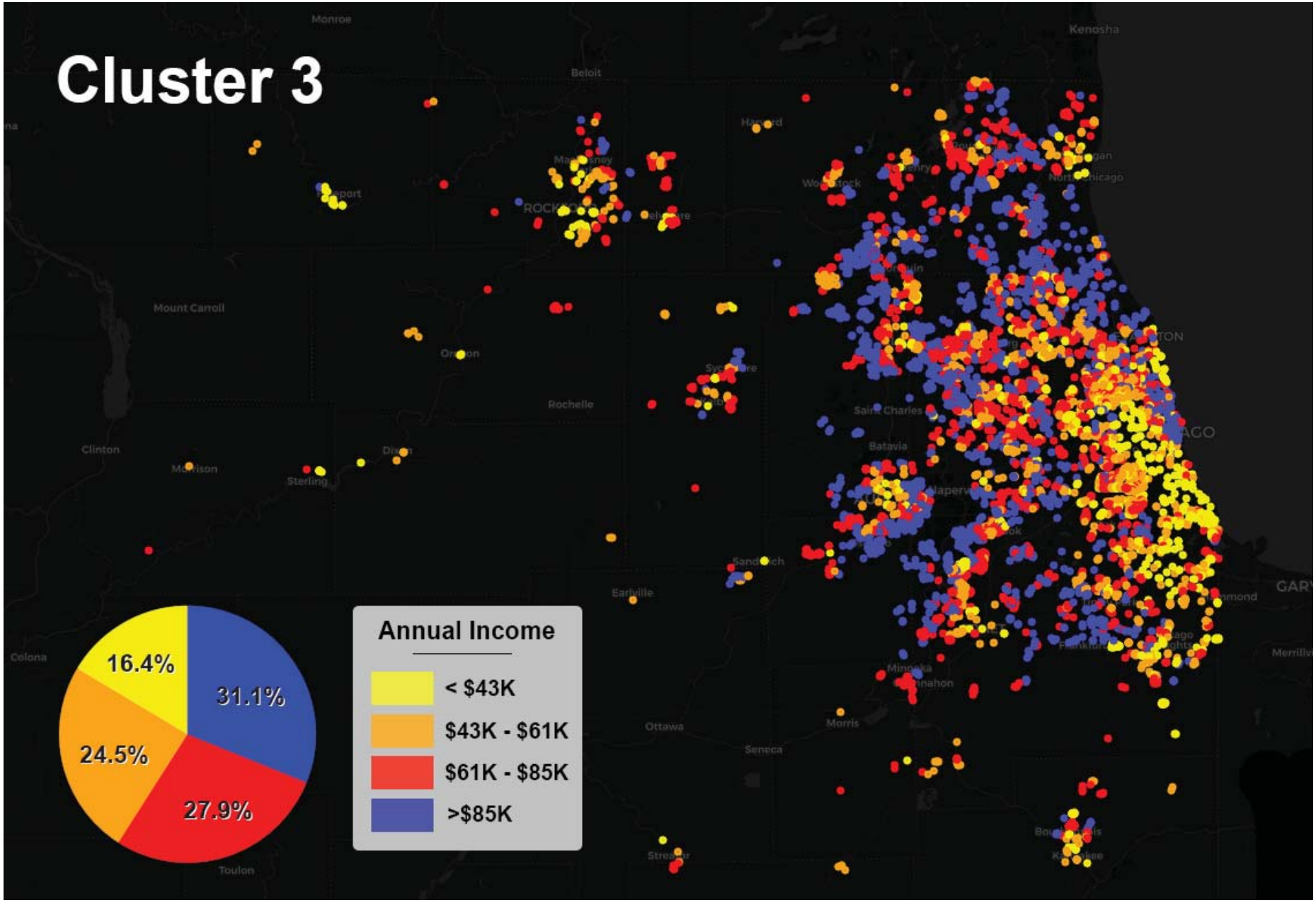

Clusters
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Elk Grove Village
Lincolnwood
Bensenville
Franklin Park
Melrose Park
Maywood
Western Springs
Bridgeview
Burbank
Hickory Hills
Evergreen Park
Oak Lawn
Palos Hills
Alsip


Machesney Park
Loves Park
ROCKFORD
Belvidere
Clusters
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6

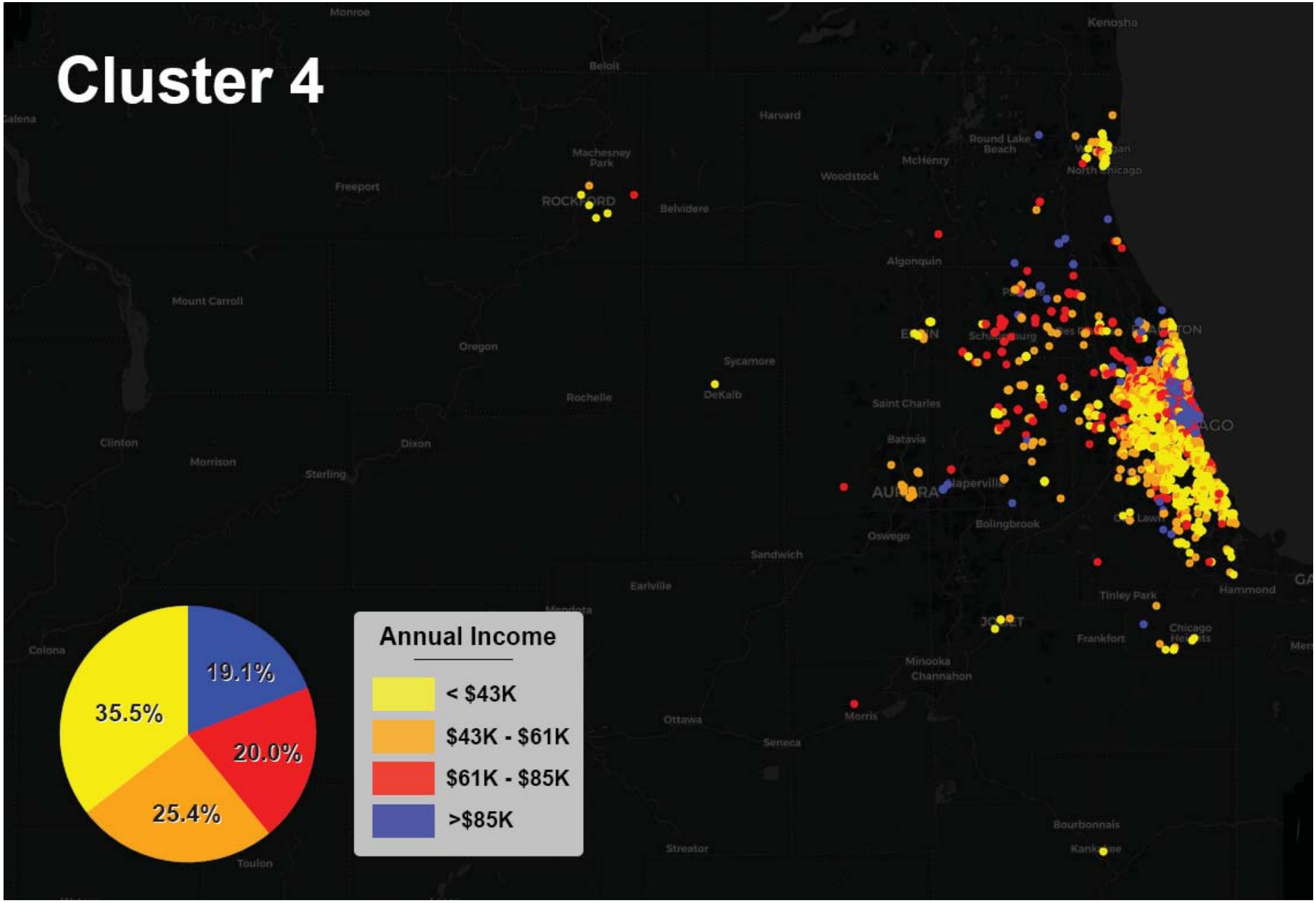

Warrenville
North Aurora
NAPERVILLE
Sugar Grove
AURORA
Oswego
Yorkville
Clusters
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6

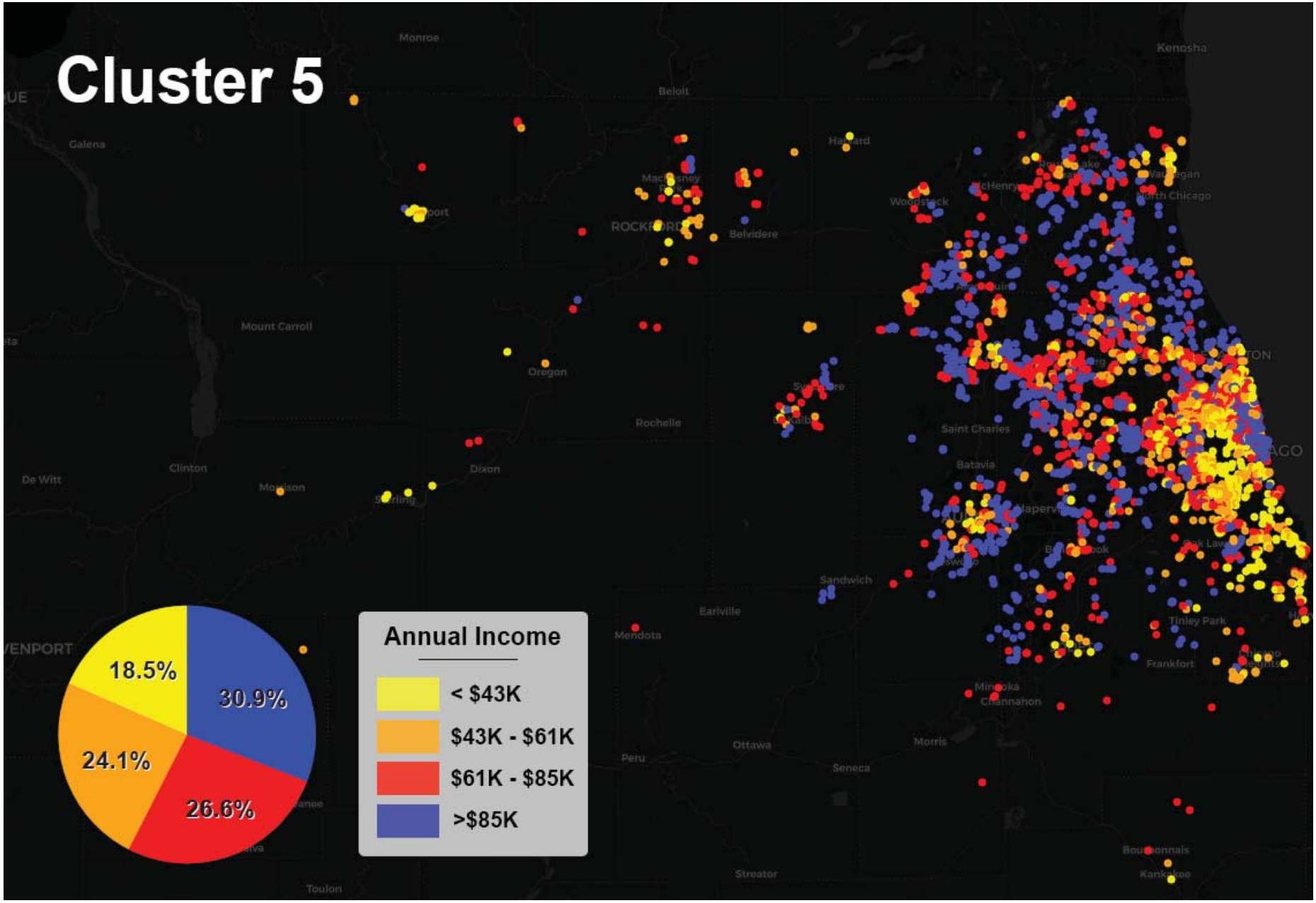

Cluster 1
JANESVILLE
Burlington
Monroe
Kenosha
Beloit
Galena
Mount Carroll
Rochelle
Saint Charles
Batavia
Clinton
Dixon
Sterling
Earlville
Ottawa
Morris
Seneca
Galva
Annual Income
< $43K
$43K - $61K
$61K - $85K
>$85K
22.4%
28.3%
24.3%
25.0%

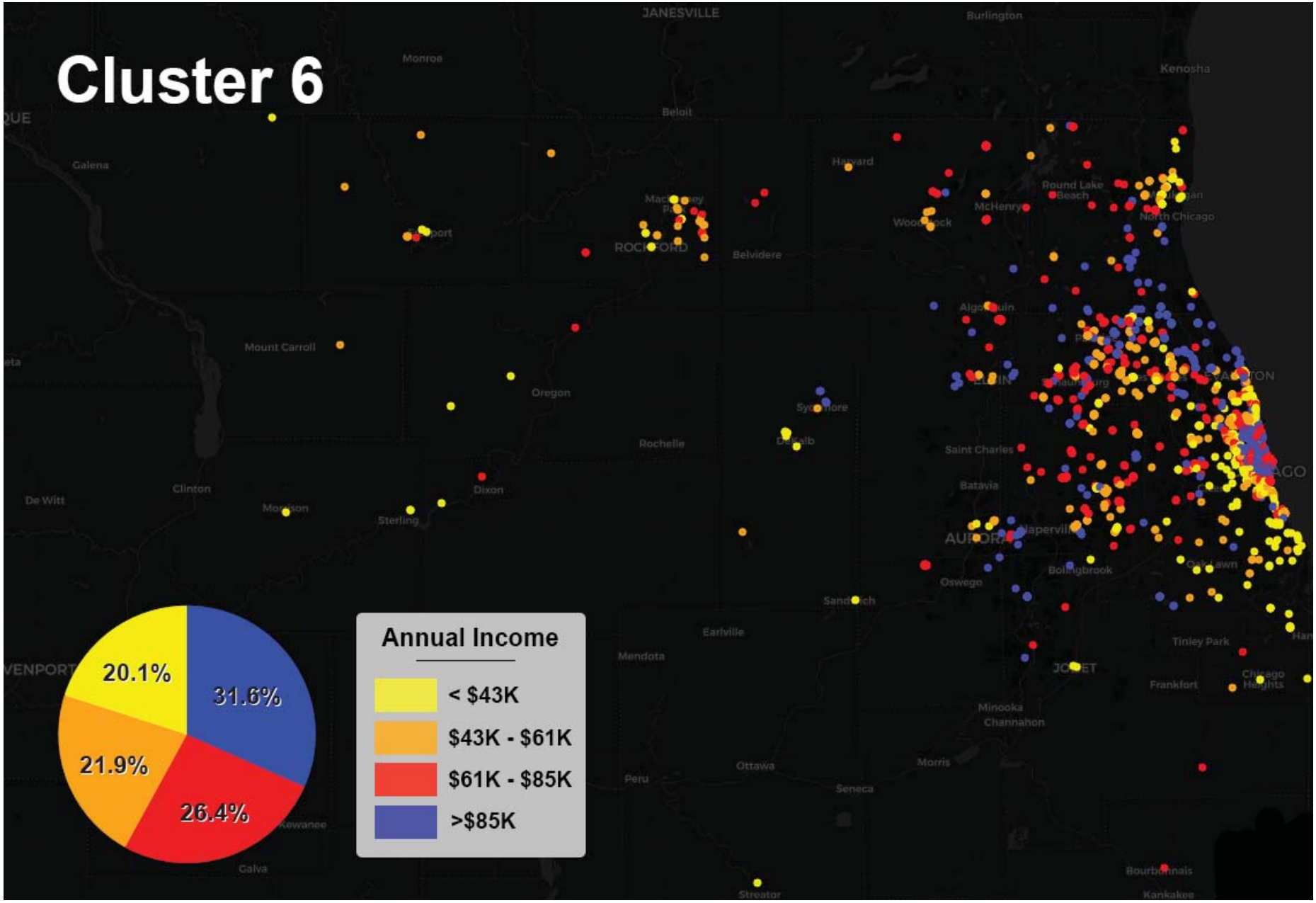

Cluster 2
Monroe
Kenosha
Beloit
Harvard
Woodstock
McHenry
Belvidere
Algonquin
Mount Carroll
Oregon
Sycamore
Rochelle
Saint Charles
Batavia
Clinton
Morrison
Sterling
Dixon
Naperville
Bolingbrook
Oswego
Sandwich
Earlville
GARY
Frankfort
Minooka
Channahon
Morris
Ottawa
Seneca
Merrillville
Bourbonnais
Toulon
Annual Income
< $43K
$43K - $61K
$61K - $85K
>$85K
16.2%
17.6%
24.3%
41.9%

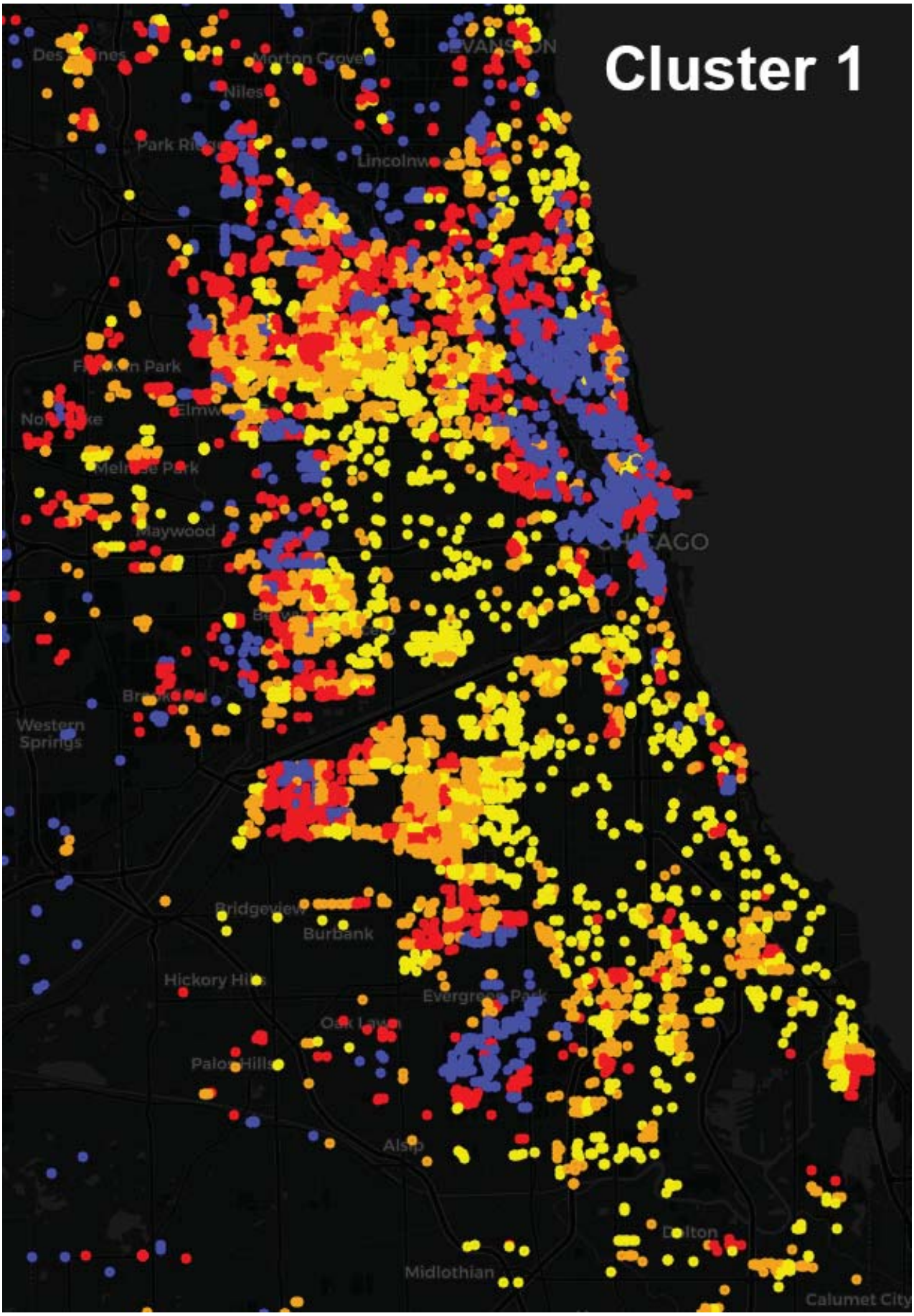

Cluster 3
Annual Income
< $43K
$43K - $61K
$61K - $85K
>$85K
16.4%
31.1%
24.5%
27.9%
Monroe
Beloit
Kenosha
Harvard
Mount Carroll
Oregon
Rochelle
Clinton
Morrison
Sterling
Dixon
Earlville
Sandwich
Batavia
Saint Charles
Naperville
Minooka
Morris
Ottawa
Seneca
Colona
Toulon
Streator
Woodstock
Belvidere
Marengo
Rockford
Freeport
Merrillville

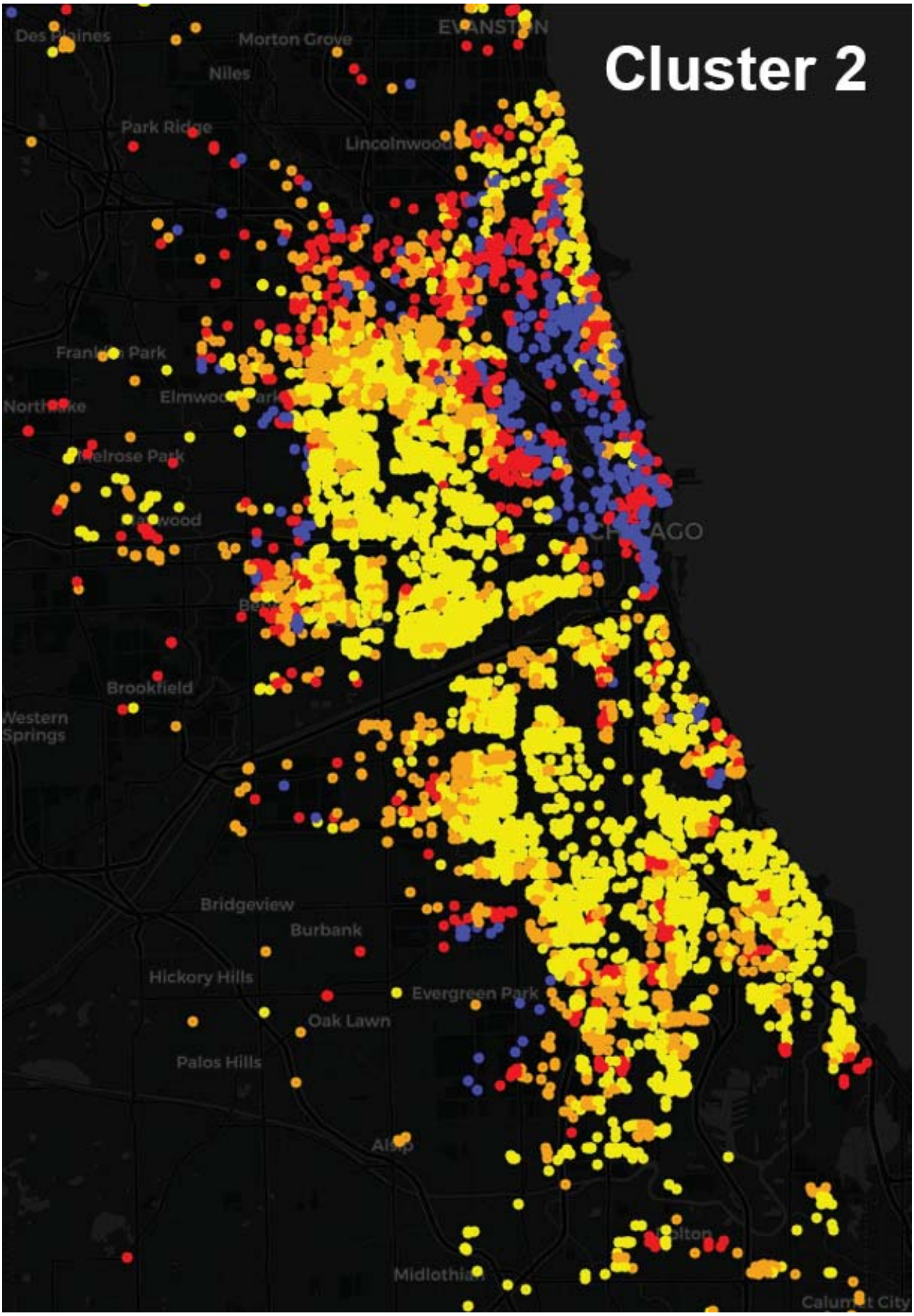
Cluster 4

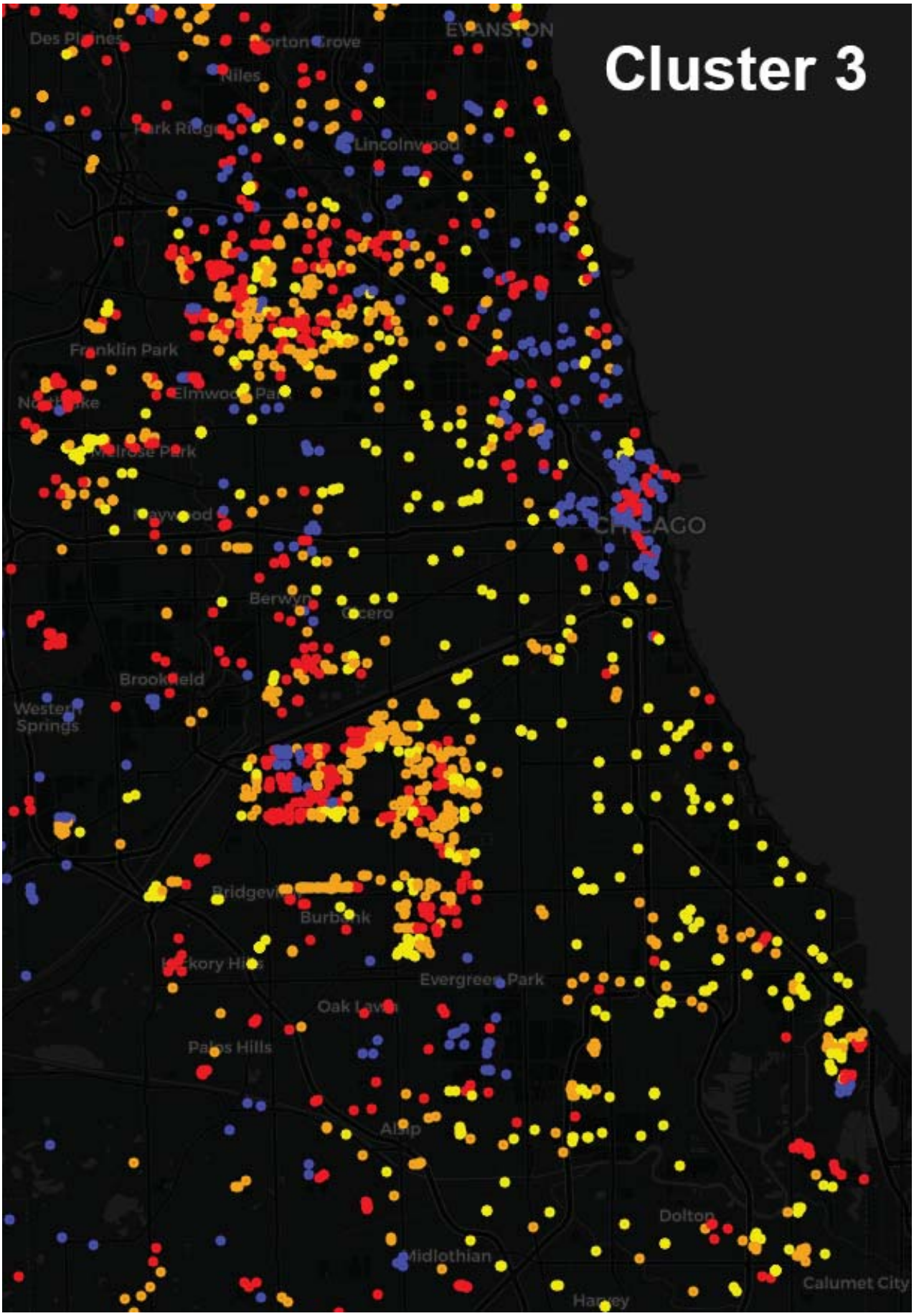

Monroe
Kenosha
Beloit
Harvard
Galena
Round Lake Beach
Machesney Park
McHenry
Woodstock
North Chicago
Freeport
ROCKFORD
Belvidere
Algonquin
Mount Carroll
Oregon
Sycamore
DeKalb
Rochelle
Saint Charles
Batavia
Clinton
Morrison
Sterling
Dixon
AURORA
Naperville
Bolingbrook
Oswego
Sandwich
Earlville
Tinley Park
Hammond
Frankfort
Chicago Heights
Colona
Minooka
Channahon
Ottawa
Morris
Seneca
Bourbonnais
Streator
Toulon
Annual Income
< $43K
$43K - $61K
$61K - $85K
>$85K
35.5%
19.1%
20.0%
25.4%

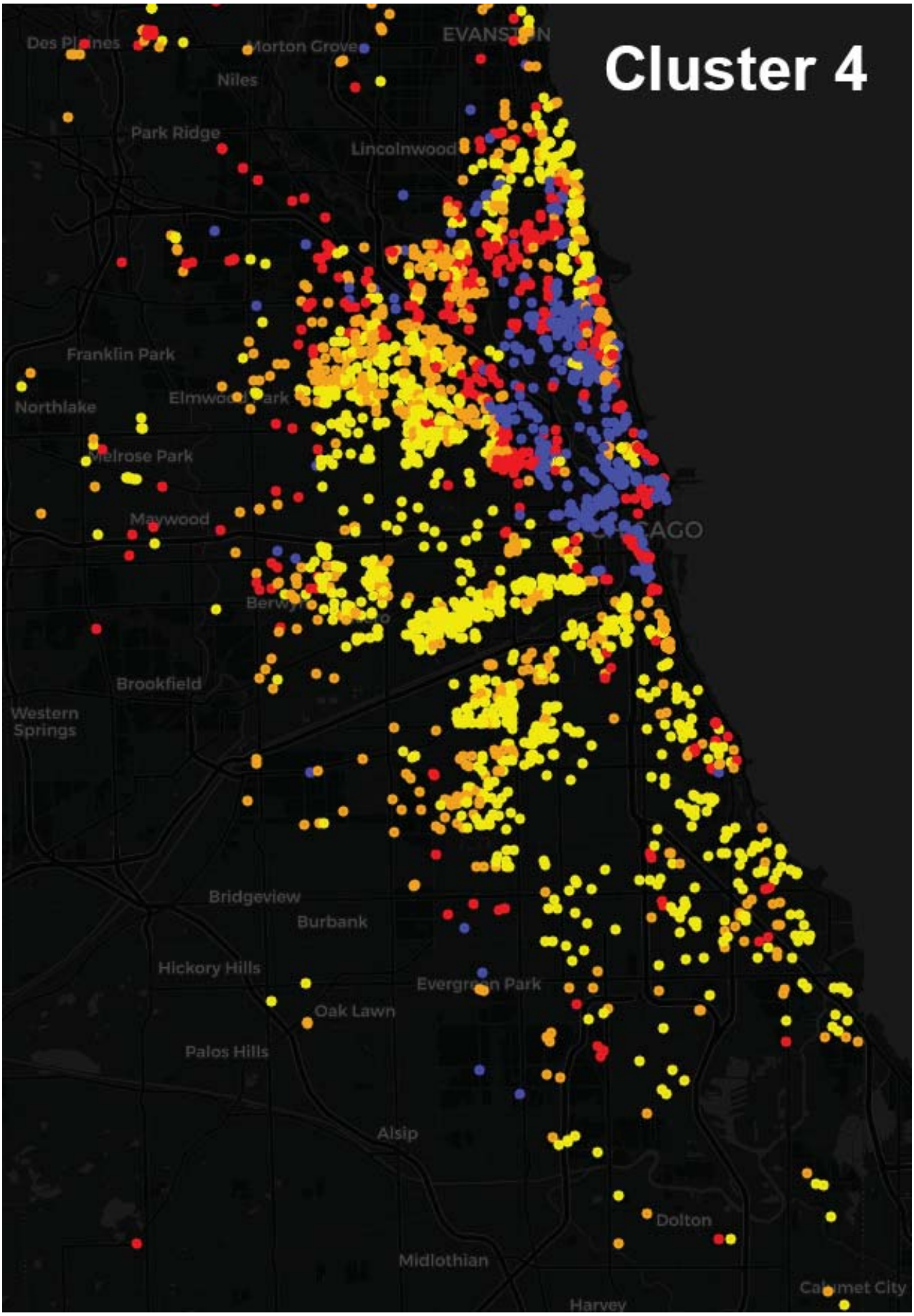

Cluster 5
Annual Income
< $43K
$43K - $61K
$61K - $85K
>$85K
18.5%
30.9%
24.1%
26.6%
Monroe
Kenosha
Beloit
Galena
Hebron
Mount Carroll
Oregon
Rochelle
Belvidere
Woodstock
Saint Charles
Batavia
Dixon
Clinton
De Witt
Morrison
Sandwich
Earlville
Mendota
Tinley Park
Frankfort
Minooka
Channahon
Ottawa
Morris
Peru
Seneca
Streator
Toulon
Kankakee

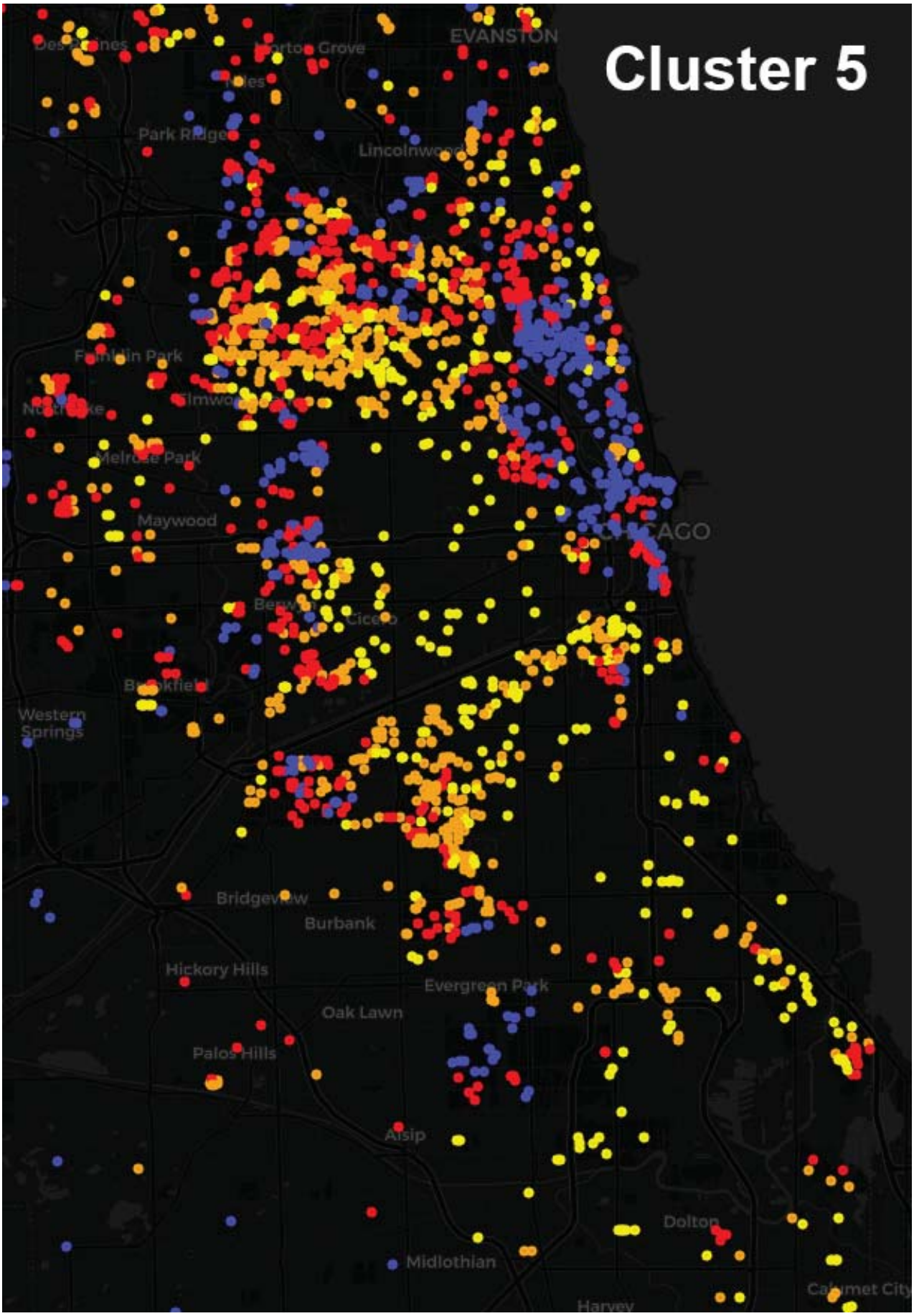

Cluster 6
JANESVILLE
Burlington
Monroe
Kenosha
Beloit
Galena
Harvard
Round Lake Beach
McHhenry
North Chicago
Belvidere
Mount Carroll
Oregon
Rochelle
Saint Charles
Batavia
Clinton
De Witt
Dixon
Sterling
Oswego
Sandwich
Earlville
Mendota
Tinley Park
Frankfort
Minooka
Channahon
Morris
Ottawa
Peru
Seneca
Galva
Bourbonnais
Kankakee
Streator
20.1%
31.6%
21.9%
26.4%
Annual Income
< $43K
$43K - $61K
$61K - $85K
>$85K

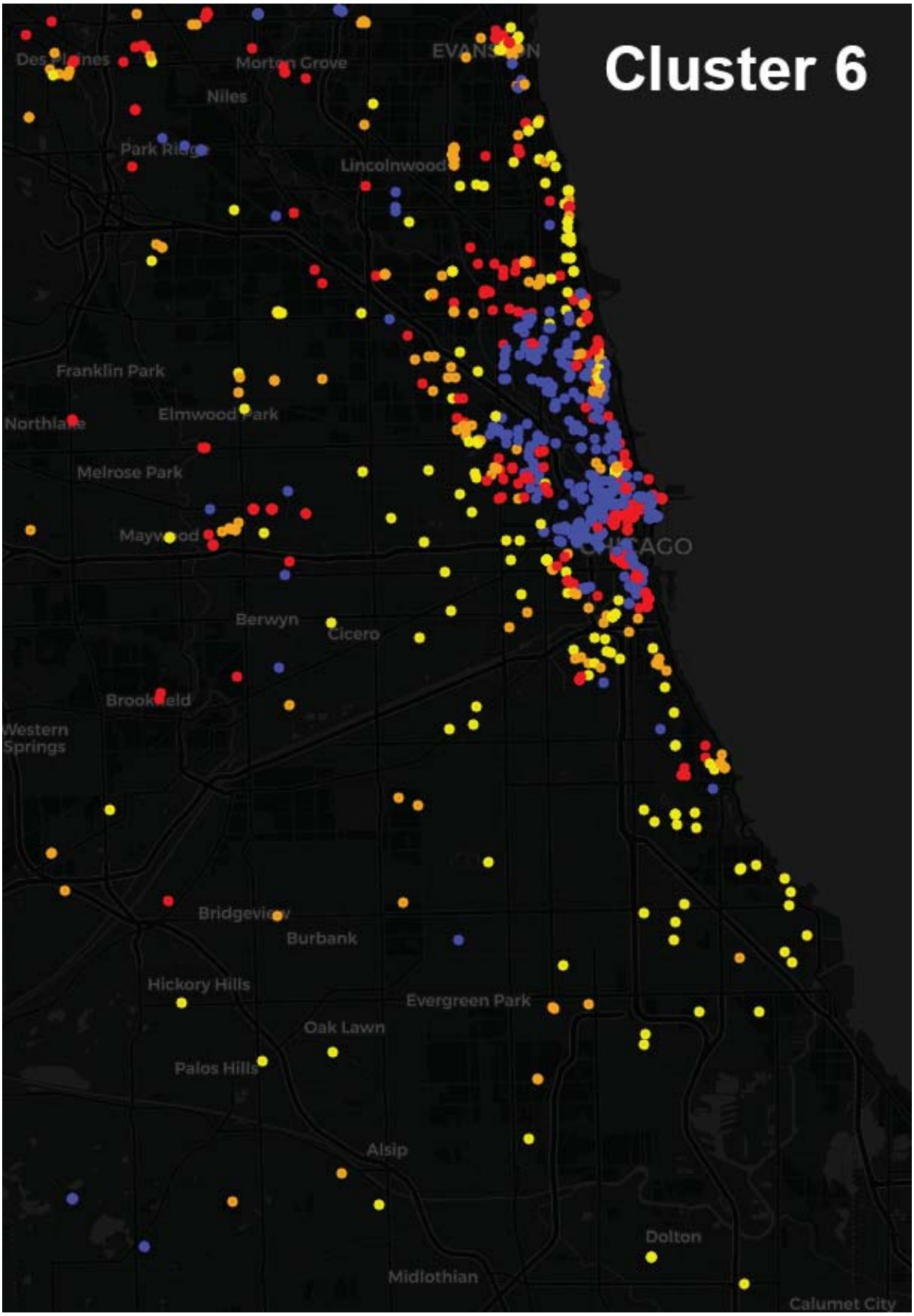

Cluster 1
Morton Grove
Niles
Lincolnwood
Maywood
Western Springs
Bridgeview
Burbank
Hickory Hills
Evergreen Park
Oak Lawn
Palos Hills
Alsip
Midlothian
Calumet City


Cluster 2
Des Plaines
Morton Grove
Niles
Park Ridge
Lincolnwood
Franklin Park
Elmwood Park
Melrose Park
Brookfield
Western Springs
Bridgeview
Burbank
Hickory Hills
Evergreen Park
Oak Lawn
Palos Hills
Alsip
Midlothian


Cluster 3
Des Plaines
Niles
Lincolnwood
Franklin Park
Melrose Park
Berwyn
Cicero
Brookfield
Western Springs
Bridgeview
Burbank
Hickory Hills
Evergreen Park
Oak Lawn
Palos Hills
Alsip
Dolton
Midlothian
Calumet City
Harvey


Cluster 4
Des Plaines
Morton Grove
EVANSTON
Niles
Park Ridge
Lincolnwood
Franklin Park
Northlake
Elmwood Park
Melrose Park
Maywood
CHICAGO
Brookfield
Western Springs
Bridgeview
Burbank
Hickory Hills
Evergreen Park
Oak Lawn
Palos Hills
Alsip
Dolton
Midlothian
Harvey


Cluster 5
EVANSTON
Park Ridge
Lincolnwood
Franklin Park
Melrose Park
Maywood
CHICAGO
Cicero
Western Springs
Bridgeview
Burbank
Hickory Hills
Evergreen Park
Oak Lawn
Palos Hills
Alsip
Dolton
Midlothian
Harvey


Cluster 6
Des Plaines
Morton Grove
EVANSTON
Niles
Park Ridge
Lincolnwood
Franklin Park
Northlake
Elmwood Park
Melrose Park
Maywood
CHICAGO
Berwyn
Cicero
Brookfield
Western Springs
Bridgeview
Burbank
Hickory Hills
Evergreen Park
Oak Lawn
Palos Hills
Alsip
Dolton
Midlothian
Calumet City

Six unique load shapes: A segmentation analysis of Illinois residential electricity consumers

Jeff Zethmayr[a], Ramandeep Singh Makhija[a]

[a]Citizens Utility Board
309 W. Washington St. Suite 800
Chicago, IL 60606
United States

Corresponding author:
Jeff Zethmayr
Citizens Utility Board
309 W. Washington St. Suite 800
Chicago, IL 60606
United States
jzethmayer@citizensutilityboard.org

Vitae

Jeff Zethmayr conducts energy policy and economic research, with a focus on rate design for regulated utilities. Mr. Zethmayr's previous work includes expert testimony and analysis for utility rate cases, and research on the bill impacts of demand-based delivery rates. He holds a Masters of Public Administration from Columbia University's School of International and Public Affairs and a Bachelor of Arts from Macalester College.

Ramandeep Singh Makhija is a Data Scientist at the Citizens Utility Board, where he investigates utility customer behavior through advanced statistical analysis. He holds a Master's degree in Industrial Engineering from the University at Buffalo with specialization in Operations Research, and a Bachelor's degree in Mechanical Engineering from the University of Pune.