Catalyzing Policing Reform with Data

Technical Appendix

Alena Stern, Ashlin Oglesby-Neal, and Kathryn L. S. Pettit

May 2020

This appendix documents the technical steps that support the report *Catalyzing Policing Reform with Data: Policing Typology for Los Angeles Neighborhoods*. In the report, we use open data on crime, arrests, stops, and calls for service to develop a typology that sheds light on the relationship between resident-initiated and police-initiated activity and how that varies across Los Angeles neighborhoods. In this document, we detail the data sources and methodology used in our analysis.

In 2018, the National Neighborhood Indicators Partnership (NNIP) and Microsoft launched a partnership to spur more data-driven and community-led criminal justice reform efforts, with the goal of building trust with law enforcement and improving public safety. As one of the activities under this partnership, we selected one city—Los Angeles, California—to explore how publicly-available police data can be analyzed to create a comprehensive measure of community-police engagement. We collaborated closely in the design, analysis and interpretation of the findings with the Microsoft Criminal Justice Reform team, the Microsoft Data Science and Analytics team, and the local NNIP partner at University of Southern California's Sol Price Center for Social Innovation. For a full discussion of the policy background, analytic results, and implications for Los Angeles and other communities, please refer to that report.[1]

Data Sources

The data used in our analysis come from three sources—policing data from the City of Los Angeles Police Department (LAPD), demographic data from the American Community Survey (ACS), and the number of addresses from the US Postal Service. We do not include data from the many other law enforcement agencies that have jurisdiction in Los Angeles.[2]

Policing Data

The data related to policing are publicly-available on the Los Angeles Open Data Portal, and include all LAPD arrests, calls for service, crimes, and stops since 2010.[3] The data are published on the Los Angeles Open Data Portal at the incident (stop, arrest, call for service, or crime) level. For each unique event, there is information on the type of incident and the demographic characteristics of the individual, if applicable and available.[4] Each observation is also coded with the police reporting district where it occurred - a geographic unit used for police operations. As of the time of the analysis, there were 1,135 LAPD reporting districts, with an average size of 0.42 square miles each. We downloaded the policing data from the Los Angeles Open Data Portal in July 2019 by exporting each data file as a CSV. While we use the data from 2010 to 2018 for use in the descriptive Power BI data visualizations that accompany our analysis, we only use the 2018 data for our analysis as discussed further in the Analytical Approach section below. We aggregate the events that occurred in 2018 to the reporting district level so that we could construct indicators for reporting districts as our unit of analysis. With the LAPD data, we create measures for analysis in three primary categories: resident-initiated activity, police-initiated activity, and reported crime.

Resident-initiated activity. One key indicator in this analysis is a proxy measure of *resident-initiated activity*, defined as calls for service made by people in Los Angeles. Calls for service are incidents in which individuals call 911 requesting police services for emergencies and non-emergencies. They can be to report a crime, traffic crash, road hazard, suspicious activity, injured person, missing person, or more. With input from LAPD, we narrowed the call for service data obtained from the Open Data Portal to include only resident-initiated calls and excluded calls that are generated by the dispatch system or police[5]. Throughout this report, we are referring to this custom definition when we mention "calls" or "calls for service." Within the resident-initiated calls for services, we distinguish calls that are for serious emergencies, defined as murder, kidnapping, robbery, battery, assault with a deadly weapon, and child abuse following the definition of a Part 1 Violent crime discussed below.

Police-initiated activity. Another contribution of this analysis is the creation of an aggregate measure of *police-initiated activity*, including stops and arrests. The police department stops data includes both pedestrian and vehicle stops, which are differentiated in the data by the *Stop Type* variable. We include total pedestrian and vehicle stops as separate variables in our analysis, but combine them for our variables of stops by race. A stop occurs when a vehicle or pedestrian is temporarily detained for investigative purposes by police. Traffic (vehicle) stops can occur for many reasons, including moving violations, equipment violations, or reasonable suspicion of criminal activity. Reasonable suspicion of criminal activity is also used to justify pedestrian stops, and can include a person matching a reliable lookout, displaying characteristics of being armed, or engaging in activity perceived as suspicious and unusual. Both pedestrian and vehicle stops can lead to protective pat downs (frisks), searches, and/or arrest.

An arrest occurs when police have probable cause to believe that a person committed a specific crime and take them into custody. The police department arrest data includes the age of the arrestee, which enables us to create separate variables for arrests of juveniles (under 18) and adults (18 or older). We combine juvenile and adult arrests for our variables of arrests by race.

Reported crime. We measure *neighborhood crime* as crime that is reported to and recorded by LAPD. Crimes may be associated with an arrest, but many crimes do not lead to an arrest. Crimes are commonly categorized based on their severity. Following the Uniform Crime Report classification system for LAPD[6], we categorize crimes into Part 1 Violent (murder, rape, robbery, aggravated assault, and simple assault), Part 1 Property (burglary, motor vehicle theft, and theft), and Part 2 (all other crime categories). One known limitation to administrative crime data is that not all crimes are reported, so this measure may be an underestimation of crime.

Neighborhood Demographic Data

We rely on census tract-level data from the 2013/2017 American Community Survey (the most recent data available at the time of our analysis) for information about neighborhood demographics and characteristics, such as racial composition, immigrant population, poverty, educational attainment, and share of renter households.[7] A census tract has a population between 1,200 and 8,000 people, averaging at 4,000 across tracts.[8] The County of Los Angeles has 2,344 tracts[9], which we crosswalk to the LAPD reporting districts using the methodology discussed in Step 1 of the Analytical Process section below. The USC team assisted with the processing of the ACS data, which is available on the Neighborhood Data for Social Change website.[10] In this analysis, we use four primary race/ethnicity categories, non-Hispanic Asian, non-Hispanic Black, Hispanic, non-Hispanic White, and other following

the terminology used by LAPD. The other group includes American Indian/Alaskan Native, other, and unknown. LAPD uses 19 unique descent codes, which we group into these five categories.[11]

Neighborhood Address Data

We hypothesize that the number of stops in a reporting district as well as the breakdown of vehicle vs pedestrian stops is influenced by the built environment of a reporting district. We suspect that stops will be overall higher in more dense neighborhoods, and the proportion of pedestrian stops is likely to also be higher in denser, more walkable areas. We use the U.S. Department of Housing and Urban Development (HUD) aggregated census tract-level U.S. Postal Service data on address vacancy to determine neighborhood address density as a proxy for the built environment of neighborhoods.[12] We calculate the density of business, residential, and overall addresses per square mile for each reporting district.

Methodology

Analytical Approach

To create a typology of community-police activity, we conduct a cluster analysis—a type of unsupervised machine learning—of reporting districts based on their crime characteristics, resident-initiated activity, and police-initiated activity in 2018. Researchers have widely used cluster analysis to develop typologies of neighborhoods to illuminate a wide variety of patterns across these neighborhoods including commuting behavior (Manaugh et al 2009), immigrant neighborhood classes (Vicino et al 2011), social capital (Sampson and Graif 2016), social problems (Chow 1998) and socioeconomic and demographic characteristics (Li and Chuang 2009). In the criminal justice literature, researchers have used cluster analysis to develop a typology of officers by occupational attitudes and burnout types (Paoline 2004, Loo 2004). Cluster analysis is a valuable tool for generating typologies because it enables researchers to distill numerous input variables into a discrete number of types. Broadly, the goal of any cluster analysis is to create groupings of observations that maximizes both the similarity of observations in the same group and the differences between groups (Bahr et al 2011). In our case, using clustering analysis allows for reporting districts to be grouped together based on their similarities in terms of the 17 input variables on calls for service, crime, stops, and arrests provided in Table 1 below. This approach enables us to draw novel connections between similar reporting districts that would not be apparent by simply looking at geographic relationships or analyzing our input variables in isolation.

We conducted all the analysis in RStudio using R v 3.6.1 and the code to replicate our work can be found on GitHub. The general steps of the analysis are the following:

1. Merge raw data and produce analytical variables.
2. Weight and scale variables for cluster analysis.
3. Run cluster analysis with different algorithms and numbers of clusters.
4. Evaluate models and select best model.
5. Conduct stability analysis of best model.
6. Analyze cluster characteristics.

Analytical Process

STEP 1: MERGE RAW DATA AND PRODUCE ANALYTICAL VARIABLES

Prior to analysis, we merge the policing data (stops, arrests, calls for service, and crimes) published by the LAPD at the reporting district level with the demographic data originally published by the ACS at the tract level. In order to do this, we must crosswalk the 2,344 census tracts in Los Angeles County to the 1,135 LAPD reporting districts.[13] We use a population-weighted crosswalk between tracts and reporting districts. This crosswalk involves the following steps:

1. For each reporting district, we identify all of the census blocks with a geographic centroid point that falls inside the district boundaries.

2. We then calculate the proportion of each census tract's population that falls inside each reporting district. For a given tract, we identify all of the census blocks that belong to that census tract. We then use our results in step 1 to identify the proportion of each tract that falls in each reporting district by dividing the population of the blocks that fall in a given reporting district by the total tract population.

3. The proportion calculated in step 2 is then multiplied by the census tract demographic data and the resulting value is assigned to the reporting district. We apply this process across all tracts and reporting districts to calculate reporting district demographic data as a population-weighted average of the census tract data.

One result of this approach is that there are 22 reporting districts in Los Angeles for which all of the blocks with centroids inside the reporting district have zero population and therefore have no values for the ACS demographic data. This may be the case when a reporting district is a park, a commercial center, or a non-residential landmark like an airport. We decided to include these reporting districts in our cluster analysis as policing activity can occur in non-residential areas and significantly impacts the lives of LA residents that live near or use the amenities in those reporting districts. Moreover, we do not use the demographic variables to determine the cluster assignments, as we are interested in creating a typology that demonstrates how policing varies throughout the city. However, these 22 reporting

districts are dropped from our analysis of the demographic composition of each cluster as they do not have any associated demographic data.[14]

We impute all missing values for the policing data with zero as we assume that no crimes, calls for service, stops, or arrests being reported for a given reporting district is equivalent to zero instances of the relevant activity occurring in that reporting district. The variables used in the clustering analysis are all for the year 2018. While Los Angeles has published data from 2010 to 2018, we chose to only use the 2018 because we felt that the most current policing environment would drive resident and police behavior and be most salient to community conversations. Moreover, we wanted to avoid a scenario in which the same reporting district would be assigned to different clusters in different years.

We then calculate the 17 rate variables that we use for the cluster analysis given in Table 1 below (where "RD" = reporting district):

TABLE 1

Variables Used for Cluster Analysis

Variable	Numerator	Denominator
Resident-Initiated Activity		
Serious Calls for Service Rate	Serious Calls for Service in RD	Population of LA
Non-Serious Calls for Service Rate	Non-Serious Calls in RD	Population of LA
Police-Initiated Activity		
Vehicle Stop Rate	Vehicle Stops in RD	Population of LA
Pedestrian Stop Rate	Pedestrian Stops in RD	Population of LA
Asian Stop Rate	Asian Stops in RD	Asian Population of LA
Black Stop Rate	Black Stops in RD	Black Population of LA
Hispanic Stop Rate	Hispanic Stops in RD	Hispanic Population of LA
White Stop Rate	White Stops in RD	White Population of LA
Adult Arrest Rate	Adult Arrests in RD	Adult Population of LA
Juvenile Arrest Rate	Juvenile Arrests in RD	Juvenile Population of LA
Asian Arrest Rate	Asian Arrests in RD	Asian Population of LA
Black Arrest Rate	Black Arrests in RD	Black Population of LA
Hispanic Arrest Rate	Hispanic Arrests in RD	Hispanic Population of LA
White Arrest Rate	White Arrests in RD	White Population of LA
Reported Crime		
Part 1 Violent Crime Rate	Part 1 Violent Crimes in RD	Population of LA
Part 1 Property Crime Rate	Part 1 Property Crimes in RD	Population of LA
Part 2 Crime Rate	Part 2 Property Crimes in RD	Population of LA

For all rates used in the analysis, we use the relevant population in all 1,135 Los Angeles reporting districts as a whole as the denominator (e.g. total population, juvenile population, Asian population). We do not use the population of the reporting district because the calls for service, arrests, and stops may not involve the people who live in that specific area. Moreover, using the reporting district population as the denominator yielded extreme outlier rate values in cases where number of stops or arrests vastly

exceeded the residents living in a given reporting district, whether for our total rates or those by race or age. Accordingly, we felt that this approach to rates would yield the best results for both the clustering and our analysis of the resulting clusters.[15]

STEP 2: WEIGHT AND SCALE VARIABLES FOR CLUSTER ANALYSIS

Once the analytical data frame is prepared in Step 1, we apply a weighting procedure to the data prior to running the cluster analysis. To account for the fact that there are more police-initiated contact variables than resident-initiated contact ones, this weighting procedure makes certain variables more important (up-weight) or less important (down-weight) in determining cluster assignment. A common form of this pre-analysis weighting is to reduce the dimensionality of data through Principal-Component Analysis (PCA) (e.g. Sampson and Graif 2009, Vicino et al 2011, Owens 2012). This approach essentially reduces the variables into the fewest number of components that explain the maximum variation in the data. In practice, this means that if multiple variables are highly correlated and explain the same variation in the data, they would likely be reduced into a single factor. The downside of the PCA approach is that interpretation is difficult as the resulting components do not map cleanly to the original variables. Moreover, PCA does not allow for the flexibility to up-weight variables based on their substantive importance. To address both of these shortcomings, we use a different weighting approach that applies weights by making different numbers of copies of each variable. The weighting approach operates as follows:

1. Prior to weighting, we identify variables that we would like to more heavily weight for substantive reasons. In our case, we choose to up-weight the two calls for service variables. We do this because there are fewer resident-initiated contact variables than police-initiated contact variables in our analytical dataset (2 vs 15). By up-weighting the calls for service variables, we ensure that resident-initiated contact drives cluster assignment on par with police-initiated contact. We choose to give serious calls for service more weight than the non-serious ones as these calls are likely more salient and concerning for residents.[16]

2. At the outset, we make four copies of each variable in the analytical dataset. We then calculate the pairwise correlations between each variable in the analytical dataset and every other variable in the analytical dataset. Based on these correlations, we iteratively remove copies of variables wherein highly correlated variables have more copies removed than less correlated variables. We ensure that all variables have at least one copy remaining at the end of this process. This achieves a similar result as PCA in which highly correlated variables that explain similar variation in the underlying data are individually given less importance so that, when these variables are taken together, they are given similar weight in the clustering as other variables that explain different components of the variation in the data.

3. We then apply the defined weights for the variables identified in the first step of the weighting process above. The result is a dataset in which all of the original variables have at least one copy, but some variables have multiple copies, which gives those variables more influence over the final cluster assignment.[17]

4. Finally, we scale all the variables in the dataset by subtracting the variable mean from each value and dividing by the variable standard deviation. This ensures that all variables used for clustering are on similar scales, with a mean of 0 and standard deviation of 1. This scaling is important because some of the clustering algorithms we use in our analysis use measures of distance between points to form clusters. Therefore, variables with larger ranges will have more influence on the clustering. For example, if we had one variable that is a fraction and one that is a count, the "distances" between observations in the count variable will generally be larger than the fraction variable, giving the count variable more influence over cluster assignment. Scaling the variables avoids this.

STEP 3: RUN CLUSTER ANALYSIS WITH DIFFERENT ALGORITHMS AND NUMBERS OF CLUSTERS

We then run the cluster analysis with three different clustering algorithms—k-means clustering, agglomerative hierarchical clustering, and gaussian mixture models (GMM) clustering—and different numbers of clusters between 2 and 20. We try different clustering algorithms because they each have different strengths and weaknesses based on the clustering behavior present in the underlying data, which is unknown prior to running the cluster analysis. Therefore, an iterative approach of trying different combinations of algorithms and numbers of clusters enables us to find the model that best fits our data. We present the clustering algorithms we used along with a discussion of the strengths and weaknesses below (our discussion draws from Seif 2018):[18]

- ***K-means:*** First, the k-means clustering algorithm initializes a defined number of centroids; second, assigns each point to the cluster of the closest centroid; third, re-calculates the centroids of each cluster based on this assignment; and finally, repeats steps two and three until the cluster assignments are stable and the algorithm converges. This widely used clustering algorithm is both very fast and simple, though it is very sensitive to outliers in the data and the randomly initialized centroid points.[19] It also assumes that clusters are circular in shape.
- ***Agglomerative hierarchical clustering:*** This algorithm first starts with each point as a separate cluster; second, combines two clusters into one based on a defined linkage[20] and distance metric; and third, repeats step two until the set number of clusters is remaining. This algorithm gives the unique benefit of being able to see the hierarchical relationship between sub-groups but is slower than k-means and sensitive to the choice of the linkage method between clusters.

- ***Gaussian mixture models (GMM):*** The GMM algorithm first randomly initializes different means and standard deviations for the defined number of clusters, second, computes the probability of each point in the dataset belonging to each cluster based on a Gaussian distribution, third, recalculates the cluster means and standard deviations to maximize the probabilities that the data points assigned to each cluster would fall within that cluster, and finally repeats steps two and three until the clusters are stable and the model converges. This model has the advantage of relaxing the assumption that clusters are circular in shape which makes it less sensitive to outliers, though it does assume an elliptical shape and that the data is normally (Gaussian) distributed.

All three of the algorithms we tested require the researcher to define the number of clusters up-front. For each of the clustering algorithms used, we ran the algorithm with different numbers of clusters ranging from 2 to 20. In the report, we refer to the resulting clusters as "groups" to be more understandable to non-technical readers.

STEP 4: EVALUATE MODELS AND SELECT BEST MODEL

We evaluate each algorithm-number of clusters combination to identify the best overall model for our data. First, we identify the best number of clusters for each individual algorithm using evaluation metrics that assess the 'quality' of the clusters. For hierarchical and k-means clustering, we use the same evaluation metrics:

- **Within Sum of Squares (WSS):** This metric evaluates the quality of clusters by taking the total sum of the squared distance between each data point and its assigned cluster centroid. Of course, this value is minimized in the case where every point is its own cluster and, by definition, decreases with each additional cluster added to the total number of clusters. Therefore, the optimal number of clusters by this metric is the "elbow" point in which the marginal decrease in WSS with each additional cluster reduces significantly.
- **Silhouette Analysis**: This metric evaluates the quality of clusters by assessing the degree of separation between clusters. For each observation, we calculate the average distance from all data points in the same cluster as well as the average distance from all data points in the closest neighboring cluster. We use these distances to calculate a 'silhouette width' coefficient that ranges from 1 (if the sample is far away from the neighboring cluster) to -1 (if the sample is closer to the neighboring cluster than its assigned cluster, i.e. assigned to the wrong cluster. Coefficients close to 1 suggest high-quality clustering. We take the average of the coefficient

across all points in the dataset. The optimal number of clusters by this metric has the highest average silhouette value.[21]

Taken together, these two metrics look for cluster assignments in which the points in a given cluster are most similar to each other and most different from other clusters.

For GMM, we also use Silhouette Analysis but replace WSS with the **Bayesian Information Criterion (BIC)** metric. BIC measures how effectively a given model predicts the observed data while penalizing more complex models to prevent overfitting. In our case, this means that models with a greater number of clusters are penalized. The model with the lowest BIC value is preferred. We then plot the gradient of the BIC scores curve, which essentially shows the marginal decrease in BIC as each additional cluster is added. Because BIC decreases with each additional cluster, we expect the values of the gradients to always be negative, and the optimal number of clusters is the "elbow" point in the curve at which the marginal decrease in BIC markedly declines. Because GMM allows for various elliptical cluster shapes, cluster compactness measured by WSS is not necessarily the appropriate criterion for measuring cluster quality, making BIC is a more appropriate metric.[22] The results for each clustering algorithm are shown below:

FIGURE 1

K-Means Clustering Evaluation

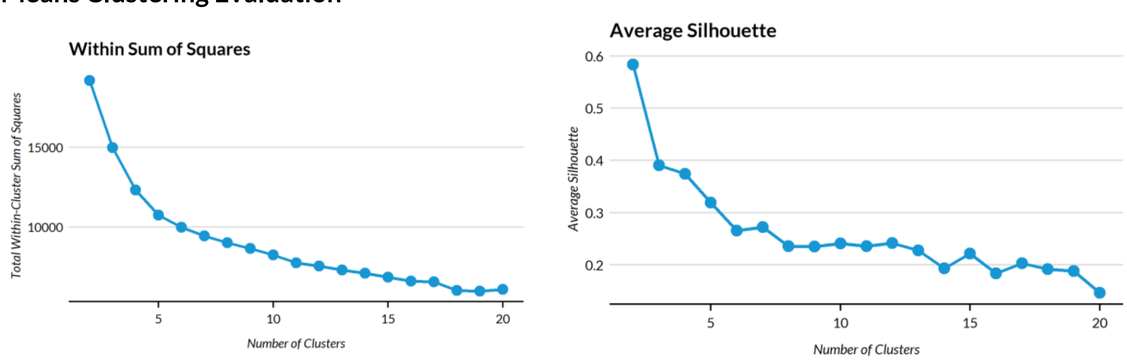


FIGURE 2

Hierarchical Clustering Evaluation

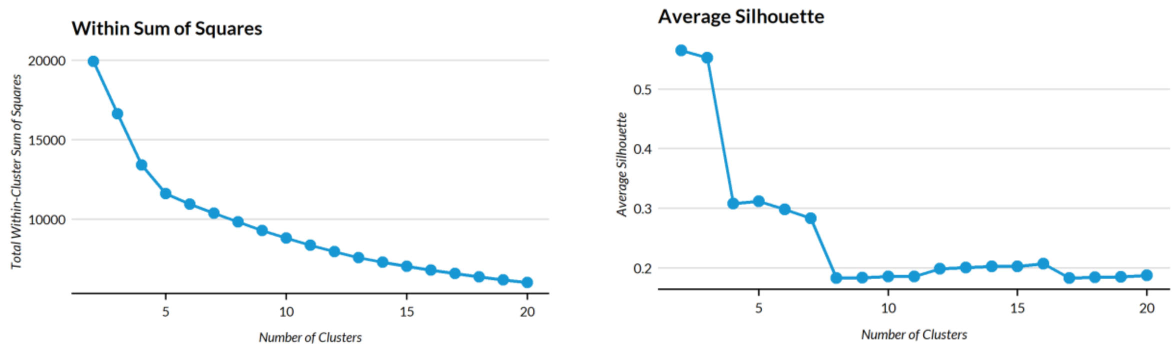
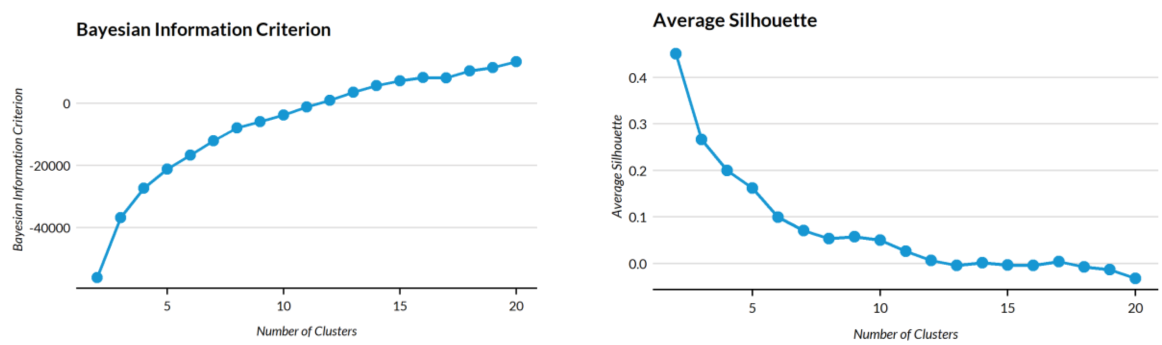


FIGURE 3

GMM Evaluation



Looking at the evaluation plots for each model, one overarching finding is that the two metrics do not always point to the same number or clusters for each algorithm. While WSS and BIC improve as the number of clusters increases, the average silhouette score generally gets worse. Therefore, in evaluating the “optimal” number of clusters for each algorithm, we looked to identify the number that maximized quality across the two metrics while maintaining few enough clusters for interpretability and relevance to community-police conversations. In all cases, we identified 5 as the optimal number of clusters based on these criteria. We saw the “elbow” in the WSS and BIC charts at 5 in all cases, as well as seeing a 5 as a value that achieved a reasonable silhouette value before a significant decline.

Once we identified 5 as the best number of clusters for each of our algorithms, we then looked at the utility of the groupings produced by each algorithm for 5 clusters. When we looked at the distribution of reporting districts across the clusters identified by each algorithm (Table 2), we found

that the reporting districts were more evenly distributed across the five clusters in GMM than in k-means and hierarchical.

TABLE 2

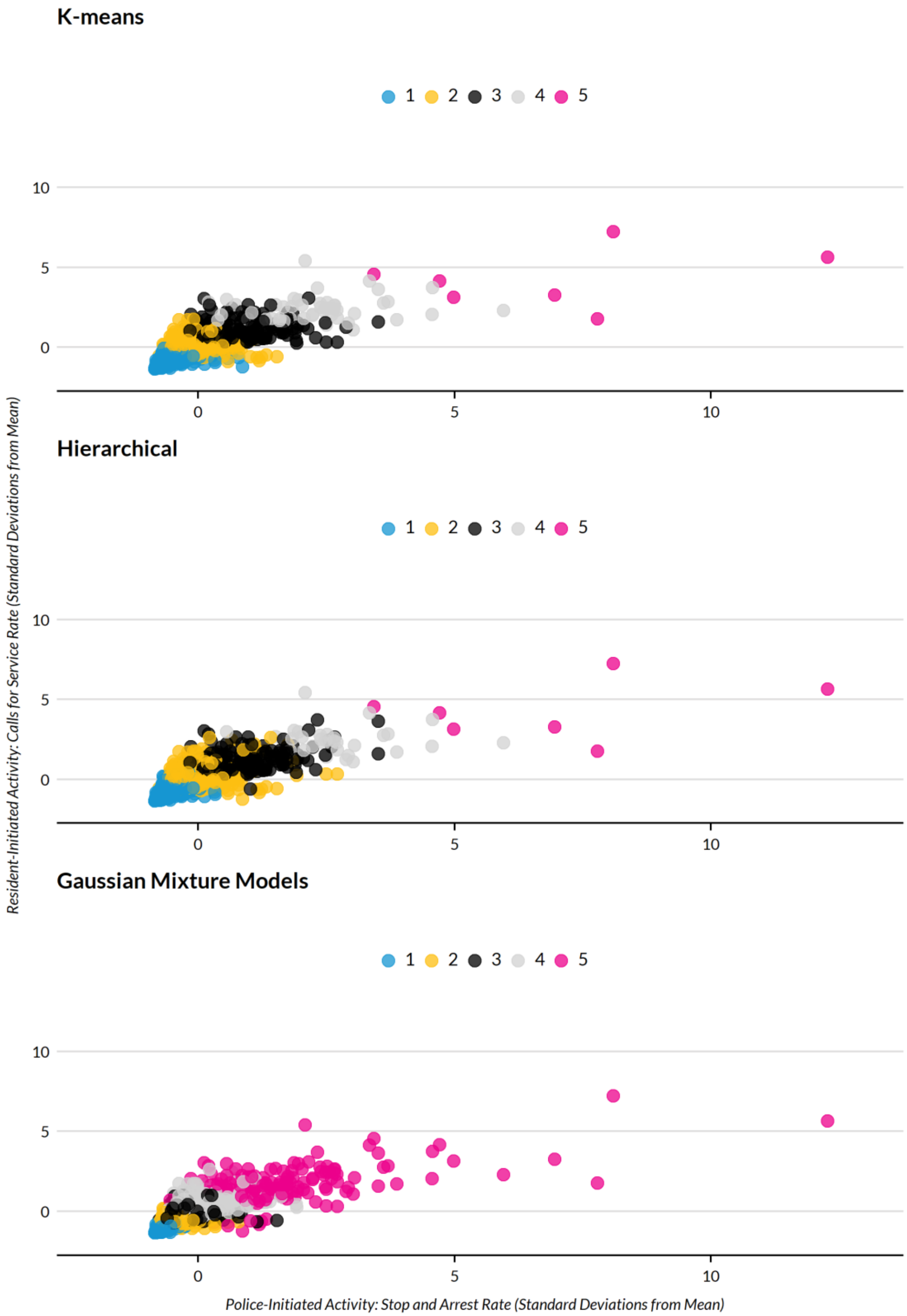
Number of Reporting Districts Assigned to Each Cluster by Algorithm

Cluster	K-means	Hierarchical	GMM
1	483	535	231
2	442	415	271
3	153	146	284
4	50	32	201
5	7	7	148

Table 2 and Figure 4 below illustrate the sensitivity of both the k-means and hierarchical clustering algorithms to outlier data points. Both algorithms yield highly imbalanced cluster sizes due to the presence of reporting districts with outlier values of the input variables. Figure 4 illustrates this by plotting resident-initiated (calls for service) and police-initiated (stops + arrests) contact and showing the cluster assignments by color. The flexibility of GMM in the cluster shape much better accommodates outliers and provided a better fit to the distribution of our data. Based on this analysis, we identified GMM with 5 clusters as our final model for analysis.

FIGURE 4

Resident-Initiated by Police-Initiated Activity for Each Reporting District by Cluster for Each Algorithm



STEP 5: CONDUCT STABILITY ANALYSIS OF BEST MODEL

After identifying our final model, we conduct a stability analysis to assess which variables most drive the clustering behavior. We perform this analysis by dropping each of our 17 variables one at a time, re-running the cluster analysis for our final model without this variable, and calculating how many of the 1,135 reporting districts included in our analysis are assigned to the same cluster. Lower numbers of reporting districts assigned to the same cluster when a given variable is dropped suggest that variable was more important in driving the cluster formation in our final model. Some results are predictable, such as serious calls for service being highly important as we up-weighted that variable and violent crime being important as it drives both serious calls for service and policing activity. Other results were more surprising, such as the relative greater importance of the white and Asian stop rates relative to Hispanic and Black stop rates and all arrest rates. While some variables had more influence than others, we were comfortable that the model was not unduly driven by a single factor.

TABLE 3

Number of Reporting Assigned to Same Cluster by Left-Out Variable

Variable	Reporting Districts in Same Cluster
Serious Calls for Service Rate	700
Asian Stop Rate	715
White Stop Rate	824
Asian Arrest Rate	1073
Hispanic Arrest Rate	1088
Black Arrest Rate	1092
Non-Serious Calls for Service Rate	1096
Pedestrian Stop Rate	1105
Crime Rate – Part 1 Property	1107
Hispanic Stop Rate	1108
Vehicle Stop Rate	1110
White Arrest Rate	1112
Crime Rate – Part 2	1112
Black Stop Rate	1115
Adult Arrest Rate	1117
Crime Rate – Part 1 Violent	1117
Juvenile Arrest Rate	1118

It is important to note that a lower importance by this measure does not necessarily mean that a given variable does not drive patterns of police-resident interactions; instead it could mean that the variable is highly correlated with another variable that captured its impact on police-resident interactions even when that variable is dropped from the model.

STEP 6: ANALYZE CLUSTER CHARACTERISTICS

After selecting our final model, we then conducted several analyses to explore the characteristics of the reporting districts that comprise each of the 5 clusters. Prior to conducting this analysis, we re-labeled the clusters based on the average serious calls for service rate in each cluster, so that cluster 1 had the lowest average serious calls for service rate and cluster 5 had the highest. As shown in the scatter plot above, we found that police-initiated activity generally trends with resident-initiated activity (with some exceptions that we discuss below) so we refer to cluster 1 as the lowest-activity cluster across both police-initiated and resident-initiated activity and cluster 5 as the highest-activity cluster. The full analysis is provided in the report, but we want to discuss a few components of this analysis here:

Outliers: In our analysis, we explored outliers in groups 4 and 5 that did not follow the overall trend of resident-initiated contact increasing with police-initiated contact. We calculated the difference of the scaled total resident-initiated contact and the scaled total-police initiated contact (resident-police contact difference), where we scale each value by subtracting the variable mean and dividing by the variable standard deviation. We then take the average of the resident-police contact difference across all reporting districts and identify reporting districts that with a resident-police contact difference at least one standard deviation above the mean (more resident-initiated contact for the level of police-initiated contact) and one standard deviation below the mean (more police-initiated contact for the level of resident-initiated contact).[23] Thirty-four reporting districts (3.0 percent) have greater than expected resident-initiated contact for their level of police-initiated contact. In comparison to the rest of the reporting districts, the high resident-initiated contact outliers have a significantly higher share of White and Asian individuals, adults with college degrees, households who rent, and property crime. In contrast, thirty-two reporting districts (2.8 percent) have greater than expected police-initiated contact for their level of resident-initiated contact.[24] The outliers with high police-initiated contact have significantly more businesses and violent crime. They also have a significantly smaller immigrant population (both citizen and noncitizen) and a smaller population of people with incomes less than 200 percent of the poverty line, suggesting these are more affluent commercial areas. For example, a few of the outliers are the areas around the Los Angeles International Airport, Hollywood Walk of Fame, and Venice Beach Boardwalk. The high police-initiated contact outliers also include major university areas like the University of Southern California and the University of California, Los Angeles and Skid Row in Downtown LA, an area with a large concentration of unsheltered people experiencing homelessness.

Cluster Rates: For several analyses in the report, we calculate rates by cluster (see figures 5 and 6). To calculate these rates, we use the sum of incidents of the activity across reporting districts in the cluster divided by the sum of the relevant population across reporting districts in the cluster. For example, the

Hispanic stop rate in cluster 1 is the sum of Hispanic stops across all reporting districts assigned to cluster 1 divided by the sum of the Hispanic population across all reporting districts assigned to cluster 1. While it remains true that a stop that occurs within a cluster is not necessarily of an individual who is a resident of a reporting district within that cluster, we calculated the cluster rates in this manner because the clusters vary considerably in number of reporting districts, total population, average population per reporting district, and population by race as shown in Table 4 below. For example, 20.1 percent of the population in cluster 5 is Black compared to 3.4 percent of the population in cluster 1. Using each cluster's population as the denominator lets us account for these differences between cluster and see that the differences between clusters are not just a function of the differences in underlying population.

TABLE 4

Mean Count of Key Metrics by Cluster

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
n	231	271	284	201	148
Total population of all reporting districts in Cluster	372,418	826,689	1,144,751	930,867	675,305
Average population in reporting district	1,612	3,051	4,031	4,631	4,563
Asian population %	13.4	14.3	12.0	10.5	7.4
Black population %	3.4	4.2	6.9	8.3	20.1
Hispanic population %	22.9	39.8	51.5	58.0	56.2
White population %	56.5	38.3	26.8	20.9	14.1

Address Density: To calculate the address density metrics shown in Figure 7 of the report, we first crosswalk the HUD-USPS address vacancy data at the tract level to the reporting district using the same process described under Step 1 of the Analytical Process above to calculate the total and business addresses in each reporting district. We then calculate density by dividing the total and business addresses in each reporting district by the reporting district area in square miles. We calculate population density by dividing total reporting district population by the area.

Conclusion

We launched this project with the belief that neighborhood-level analyses of policing activity can help support conversation between community members and police. We hope that our analysis can be used to provide a starting point and context to further deliberations on policing reform and how to strengthen LA communities' trust in police. From NNIP's experience, having a common set of facts helps to change the nature of conversations to focus more on solutions. Our analysis confirms the importance

of digging down below citywide data to small areas to understand the range of experiences across LA's neighborhoods. For those interested, the report provides examples and resources for how stakeholders in Los Angeles might use the analysis. We also discuss how local data organizations and criminal justice advocates could implement this approach in their communities.

Notes

1 The report is available at https://www.urban.org/research/publication/catalyzing-policing-reform-data.

2 LAPD has the most sworn personnel out of the law enforcement agencies in the County of Los Angeles, and only has jurisdiction within the City of Los Angeles. However, the Los Angeles Sheriff's Department is nearly as large as LAPD and has jurisdiction throughout the rest of the County of Los Angeles. Within the city, there are many smaller agencies based out of localities (e.g. Santa Monica, Pasadena) or universities (e.g. UCLA, USC). There are also state and federal agencies that operate in LA (e.g. California Highway Patrol, FBI). Our analysis only looks at LAPD activity within LAPD reporting districts. However, other agencies may conduct activities in those reporting districts that we are unable to measure.

3 LAPD open data are available at https://data.lacity.org/browse?Data-Owner_Department=LAPD.

4 Age, sex, and race/ethnicity information is available for arrests and sex and race/ethnicity information is available for stops. Age, sex, and race/ethnicity information is also available for victims of crime, but we do not use this information in our analyses. No demographic information is available for calls for service.

5 Across 2010-2018, our definition of resident-initiated calls for service captures 80% of all calls for service. That is, 20 percent of calls for service are those generated by the dispatch system or police, not by residents.

6 Information on how LAPD classifies crimes according to the UCR categories is available at https://data.lacity.org/api/views/63jg-8b9z/files/fff2caac-94b0-4ae5-9ca5-d235b19e3c44?download=true&filename=UCR-COMPSTAT062618.pdf

7 Because our unit of analysis is a reporting district and ACS data are reported at the census tract level, we create a crosswalk that weights the ACS information according to the reporting district size.

8 For more information on census tracts, see https://www2.census.gov/geo/pdfs/education/CensusTracts.pdf

9 This number is per the LA County eGIS data accessed at https://egis3.lacounty.gov/dataportal/2011/07/19/census-tracts-2010/ on March 9th, 2020

10 To explore demographic data about the neighborhoods of LA, visit https://ladata.myneighborhooddata.org/

11 LAPD has 19 descent codes that we group into five race/ethnicity categories - Asian, Black, Hispanic, Other and White. Asian includes Chinese, Cambodian, Filipino, Guamanian, Japanese, Korean, Laotian, Pacific Islander, Samoan, Hawaiian, Vietnamese, Asian Indian, and Other Asian. Black only includes Black. Hispanic is defined as Hispanic/Latin/Mexican in LAPD's descent code. Our Other category includes American Indian/Alaskan Native, Other, and Unknown. White only includes White.

12 The U.S. Department of Housing and Urban Development aggregated USPS data on address vacancy reports quarterly data on total, vacant, and no-stat residential and business address counts at the tract level. We use the average counts of total, residential, and business addresses across the four quarters of 2018 for our analysis.

13 Some LAPD data files have observations that are associated with a reporting district that is not part of the 1,135 reporting districts under their jurisdiction. This may be a data entry error or reflect that the event actually occurred outside of their jurisdiction. We exclude these observations from analysis.

14 Of the 22 reporting districts, 18 are in group 1 and 4 are in group 2.

15 For the purposes of the clustering itself, using the same denominator across all reporting districts to calculate each rate variable is equivalent to using the raw count variables in the cluster analysis. We ran our final cluster model on the raw counts and confirmed that the results are equivalent. However, we felt that using the rates was

valuable for our post-clustering analysis to better contextualize the instances of police-initiated activity on a given sub-population relative to the size of that sub-population in the city of Los Angeles.

16 The manually assigned weights are 12 for serious calls for service and 3 for non-serious calls for service.

17 After this process, serious calls for service has 12 copies, non-serious calls for service has 3 copies, and all other variables have 1 copy. This indicates that the input variables for our analysis are considerably correlated.

18 See George Seif's 2018 article for Towards Data Science for an excellent discussion of different clustering algorithms https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68

19 The kmeans() function we use in R addresses the sensitivity of the k-means algorithm to the initial cluster centroid points by initializing multiple different sets of random starting cluster centroid points. The number of sets is established with the "nstart" parameter, which we set to 25 for our analysis. For more information, see the function documentation here: https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/kmeans

20 We use the ward.D2 linkage method as our initial tests of different linkage methods found the best performance with ward.D2 linkage. From the R documentation, "two different algorithms are found in the literature for Ward clustering. The one used by option "ward.D" (equivalent to the only Ward option "ward" in R versions <= 3.0.3) does not implement Ward's (1963) clustering criterion, whereas option "ward.D2" implements that criterion (Murtagh and Legendre 2014). With the latter, the dissimilarities are squared before cluster updating." For more information, see https://stat.ethz.ch/R-manual/R-devel/library/stats/html/hclust.html

21 For more detail on the Silhouette Method, see https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a

22 For an in-depth discussion on the BIC metric for GMM, see https://www.stat.washington.edu/sites/default/files/files/reports/2009/tr559.pdf

23 The average resident-police contact difference is 0.15 and the standard deviation is 1.121. This average can be interpreted as reporting districts' level of resident-initiated contact is, on average, 0.15 standard deviations greater from the mean than police-initiated contact.

24 We define the outliers as those reporting districts that are one standard deviation above the mean in the difference in scaled resident-initiated contact and scaled police-initiated contact.

References

Bahr, P., R. Bielby, and E. House. 2011. "The Use of Cluster Analysis in Typological Research on Community College Students." *New Directions for Institutional Research* (S1): 67–81. https://doi.org/10.1002/ir.417.

Li, Y. and Y. Chuang. 2009. "Neighborhood Effects on an Individual's Health Using Neighborhood Measurements Developed by Factor Analysis and Cluster Analysis." *Journal of Urban Health* 86 (5).

Loo, R. 2004. "A Typology of Burnout Types among Police Managers." *Policing: An International Journal of Police Strategies and Management* 27 (2): 156–65.

Manaugh, K., L. Miranda-Moreno, and A. El-Geneidy. 2010. "The Effect of Neighbourhood Characteristics, Accessibility, Home-Work Location, and Demographics on Commuting Distances." *Transportation* 37 (4): 627–46.

Owens, A. (2012). "Neighborhoods on the Rise: A Typology of Neighborhoods Experiencing Socioeconomic Ascent." *City & Community* 11 (4): 345–69.

Paoline, E.. 2004. "Shedding Light on Police Culture: An Examination of Officers' Occupational Attitudes." *Police Quarterly* 7 (2): 205–36.

Sampson, R., and C. Graif. 2009. "Neighborhood Social Capital as Differential Social Organization: Resident and Leadership Dimensions." *American Behavioral Scientist* 52 (11): 1,579–1,605. https://doi.org/10.1177%2F0002764209331527.

Vicino, T., B. Hanlon, and J. Short. 2011. "A Typology of Urban Immigrant Neighborhoods." *Urban Geography* 32 (3): 383–405. https://doi.org/10.2747/0272-3638.32.3.383.

About the Authors

Alena Stern is a data science fellow at the Urban Institute studying policy solutions to advance equity and inclusion in cities. Before joining Urban, she worked as a senior program manager with AidData, an Open Cities fellow at the Sunlight Foundation, and a graduate research assistant at the Center for Data Science and Public Policy, where she used machine learning, natural language processing, statistical analysis, and geospatial data to inform the design of government policies and international development programs.

Ashlin Oglesby-Neal is a research associate in the Justice Policy Center at the Urban Institute, where her research includes developing and validating risk assessment tools, and evaluating criminal justice policies and programs. Her work includes multisite data collection and management, collaboration with criminal justice practitioners, quantitative data analysis, and data visualization.

Kathryn L. S. Pettit is a principal research associate in the Metropolitan Housing and Communities Policy Center at the Urban Institute, where her research focuses on neighborhood change and how communities use data for more effective and equitable decisionmaking. Pettit is a recognized expert on several small-area local and national data sources and on the use of neighborhood data in research, policymaking, and program development. Pettit directs the National Neighborhood Indicators Partnership, a network of three dozen local organizations that collect, organize, and use neighborhood data to inform local advocacy and decisionmaking.

Acknowledgments

This report was sponsored by Microsoft. Kevin Miller and Elizabeth Grossman were invaluable thought partners throughout this project. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission. The Microsoft Data Science and Analytics team—Rahul Dodhia and Allen Kim—helped in data exploration, advice on the methodology, and in producing the Power BI data visualization that allows us to share the data in a user-friendly way.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is

available at urban.org/fundingprinciples recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

Dean Obermark, formerly of the Urban Institute, participated in the early part of the NNIP project and identifying Los Angeles as our partner site. Tracey Lloyd, also formerly of Urban, formed the research questions focused on exploring resident-initiated versus police-initiated activity. At USC's Sol Price Center for Social Innovation, Caroline Bhalla and Elly Schoen collaborated in data preparation and indicator construction, helped us understand the geography of Los Angeles, and gave us ideas on how the analysis might interest residents. We also greatly appreciate the insights of our external reviewers: Gary Painter of the USC's Sol Price Center for Social Innovation, Lieutenant Christopher Chase of the Los Angeles Police Department and Erroll Southers and Jessie Redd of the USC Safe Communities Institute. Internally, the report was improved by the review of Claire Bowen, Leah Hendey, Nancy La Vigne, and Dan Lawrence of the Urban Institute.

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.