Utilizing 311 Service Requests as a Signature of
Urban Location in the City of Los Angeles

by

Richard Windisch

A Thesis Presented to the
Faculty of the USC Graduate School
University of Southern California
In Partial Fulfillment of the
Requirements for the Degree
Master of Science
(Geographic Information Science and Technology)

August 2019

I dedicate this paper to my parents, my sister, and to all of my family and friends who have been with me throughout this entire process. Without your guidance and support, none of this would have been possible.

Table of Contents

List of Figures

List of Tables

Acknowledgements

List of Abbreviations

ACS	American Community Survey
BSS	Building of Street Services
FOSS	Free and Open Source Software
GIS	Geographic information system
LAHSA	Los Angeles Homeless Services Authority
LAPD	Los Angeles Police Department
LASAN	Los Angeles Sanitation
MyLA311	My Los Angeles 311
PPGIS	Public Participatory GIS
SR	Service Requests
TIGER	Topologically Integrated Geographic Encoding and Referencing
VGI	Volunteered Geographic Information

Abstract

In order to increase citizen engagement, in 2013, the City of Los Angeles introduced the MyLA311 application, a smartphone app that allows residents to easily request city services. Previous service requests were funneled through four separate data service management systems and lacked transparency; the improved centralized system increases public data access and efficiency, all while ensuring a uniform tracking methodology across all departments. Citizens act as agents creating data each time a service request is made. Ease of reporting and increased use of mobile applications or digital platforms to track and monitor service requests creates huge volumes of volunteered geographic information (VGI) data. Los Angeles's shift towards open data supports data-driven decision-making regarding city services and mitigation of problems. While the original purpose behind the push towards publicly accessible information was accountability, a new purpose for the data was found in the possibility of constituents creating additional insights. The attributes of VGI provided through the MyLA311 service requests were analyzed to determine fitness for use in spatial analysis. Los Angeles experiences a great deal of spatial heterogeneity given the differences in socioeconomic attributes and local neighborhood contexts. Distinct signatures of the local urban context, similar to a neighborhood, are determined through a multivariate cluster analysis of MyLA311 Service Requests and sociodemographic data at the census tract level. This spatial analysis provides stakeholders and civic leaders with insights into which physical problems need focus in certain geographically defined communities detailed in the results and conclusion.

Chapter 1 Introduction

This thesis spatially analyzes 311 requests for service in the City of Los Angeles and uses them to identify patterns of reporting across the city. The resulting data consists of daily requests for service and are a form of volunteered geographic information (VGI), as it is spatial data directly sourced from citizens. The service request data is publicly available to promote accountability, data-driven governance and civic improvement. With freely accessible data, the opportunities for analysis are plentiful. At the time of writing, a search in Google Scholar for "MyLA311" returns not even two full pages of results. Of all of these results, only one regards a spatial analysis of the data; all remaining results mainly speak of MyLA311 as a step in the right direction for cities and open data. The analysis of 311 data herein is unique to Los Angeles and informed by previous 311 data analyses from other cities that detail the limitations and demonstrate different use cases. The purpose of this thesis is to identify spatial signatures, or a unique set of characteristics about a place, across Los Angeles through multivariate cluster analyses of 311 service requests and local sociodemographic features from census data.

With the City of Los Angeles now sharing 311 data of service requests via its online open data platform, an opportunity has arisen to perform data analysis on requests for service submitted by citizens. While many different types of service requests are available for analysis, this thesis focuses on service requests for illegal dumping, graffiti removal, homeless encampments, dead animal removal, and broken streetlights. This decision stems from the frequency of selected service request types and through the requests' shared public/visual characteristics of visibility in an urban environment. The analysis utilizes census tracts in order to provide a stable set of geographic units for the presentation of statistical data. While Los Angeles is divided into Council Districts and neighborhoods within local contexts, a multivariate

cluster analysis of service requests on the scale of the census tract provides an informative basis for determining clusters of similarity. As Waldo Tobler's first law of geography states, "everything is related to everything else, but near things are more related than distant things" (Tobler 1970). Analysis of publicly available MyLA311 data provides a use case for the introduction of viable tools for informed understanding of urban spaces and city management practices.

1.1. Motivation

The collection and management of massive data quantities about communities is no longer viewed as an impediment to the implementation of GIS services and analyses, but an accessible process through technological advances and mobile mapping technology (Novak 1995). VGI offers previously unimaginable sources and quantities of data. By reporting their own issues, citizens have the chance to play a part in managing their built environment while enacting the ideas of community betterment. With citizens as sensors, cities can explore a powerful and comprehensive approach reaching beyond the digital divide of citizens and government (Goodchild 2011). Through engaging citizens in the process of acquiring and utilizing geographic information, VGI has the potential to alter this landscape significantly and soften criticisms of citizen engagement in local government as inaccessible or difficult to accomplish (Toregas 2001).

This project builds on the recent shift in municipal government to be more data-driven and to prioritize publicly available data. The current Mayor of Los Angeles, Eric Garcetti, and the City Controller, Ron Galperin, emphasize the importance of setting benchmarks for accountability of city services, as well as constant improvement (Los Angeles City Controller 2019). The data produced by the city is then built into open access online dashboards to monitor

the progress, providing a snapshot into how the city is functioning. Both constituents and city officials can track the progress of goals online through the Mayor's Dashboard. The Los Angeles GeoHub and Open Data Portal, two publicly available online civic infrastructures, cater to city employees and residents in offering data and resources. The GeoHub is the city's online platform for exploring, visualizing, and downloading geographic information system (GIS) based data. City leaders promote "civic hacking," a term in which "hacking" refers to perverting government data's original purpose to solve a different civic problem (Tauberer 2014). The idea driving civic hacking is public engagement with the data to solve problems that matter to them. Analysis of civic data can result in direct, fact-based influences in policy with more economic, sustainable and community-based outcomes.

The MyLA311 website provides spatiotemporal data that can be used for a multitude of analyses using a GIS. Through an internal data management system, MyLA311 cost-effectively manages and collects data from citizens self-reporting through online forms, phone calls, or requests within an application on mobile devices. In addition to providing useful metrics for civic accountability, analysis of VGI from service requests provides information not previously known about an area when taken into account with other data sets. The location component of a service request is the most integral aspect of the data because it ties the spatiotemporal aspects of a service request to a unique location, qualifying it for use in a GIS analysis. Within this thesis, the temporal and spatial components of the MyLA311 data support spatial analysis while also providing a descriptive comparison of the geospatial information of the 311 data to the sociocultural context of Los Angeles neighborhoods.

1.2. Research Question

A spatial signature is a unique set of characteristics about a place, as defined in Wang et al. (2017). The 311 data analyzed is comprised of service requests for illegal dumping, graffiti removal, homeless encampments, dead animal removal and broken streetlights. The sociodemographic features from the census data focus on population diversity, education and income and employment. Spatially distributed patterns emerge through multivariate clustering of data across a defined area. An understanding of the nuances of working with VGI and MyLA311 data informs the thesis to determine differences in signatures, which may reflect the geographic constraints of similar characteristics that make up neighborhoods in Los Angeles.

After completion of this analysis, comparisons can be drawn between the distribution of cluster locations and the arbitrarily defined political boundaries existing in Los Angeles. For example, stakeholders can incorporate their own location with the resulting cluster outcomes to see how politically defined geographic political regions compare with the actual urban makeup of an area. Additionally, civic entities can observe characteristics about the people of a place and the services requested for problems in an area. While the research does not answer why different signatures exist, it supports policy and decision makers in how they can better understand the context of an area, allowing them to better determine an approach for service.

This thesis therefore answers a two-part research question:

1. What clusters of neighborhoods can be identified by spatial clusters of common patterns of 311 service requests?

2. Do the areas identified as clusters share similar sociodemographic features?

1.3. Study Area

The study area for the analysis is the City of Los Angeles. MyLA311 only contains service requests within the Los Angeles city boundaries. This analysis would work at the county level if neighboring cities used a similar methodology for dissemination of service request data and had parallel request types. The land area of the city is 468.67 square miles and contains a population of roughly four million people (U.S. Census Bureau 2018). While Los Angeles is one large city, it is often said to be a collection of many smaller neighborhoods, each with a unique identity. These neighborhoods, as determined by crowdsourced opinions by the Los Angeles Times, are revealed in Figure 1. The names of some of these neighborhoods are used throughout this thesis to provide geographical context for the analysis.

Mapping L.A.

Los Angeles residents have traditionally formed strong neighborhood affiliations, but where any neighborhood begins or ends has long been a source of confusion and controversy, with no official boundary map in existence. In 2009, database editor Doug Smith, a lifelong Angeleno, and researcher Maloy Moore began the process of standardizing neighborhood boundaries based on historical and anecdotal definitions, civic proclamations and reader commentary. The two converted thousands of census blocks into a complete picture of Los Angeles neighborhoods, with no ambiguities, overlaps or missing pieces. Data analyst Sandra Poindexter compiled census and education data to create demographics unique to each neighborhood. Using cutting-edge open-source software, database producer Ben Welsh constructed maps and charts to display the work interactively, assisted by Stephanie Ferrell, Mark Hafer and Thomas Suh Lauder. Editors David Lauter and Megan Garvey integrated the project into The Times' overall local coverage, and copy editors Robert Browning and Kathy Gosnell meshed the language and naming conventions with Times style. Shown here are the resulting 114 neighborhoods that make up the City of Angels.

Legend:

1 Rancho Park
2 Cheviot Hills
3 Century City
4 Beverlywood
5 Pico-Robertson
6 Carthay
7 Beverly Grove
8 Hancock Park
9 Windsor Square
10 Harvard Heights
11 Arlington Heights
12 Jefferson Park
13 Adams-Normandie
14 Exposition Park
15 Leimert Park
16 Central-Alameda
17 Chesterfield Square
18 Harvard Park
19 Vermont-Slauson
20 Manchester Square
21 Gramercy Park
22 Vermont Vista
23 Broadway-Manchester
24 Green Meadows
25 Elysian Valley

Figure 1 Neighborhoods comprising Los Angeles, *source*: Los Angeles Times

1.3.1. Los Angeles Demographics

      In order to better understand the different characteristics that inform clustering in Los Angeles, it is important to have a sufficient understanding of the general makeup of the city. Common sociodemographic features representing important phenomena in population diversity, education, and income and employment within Los Angeles are examined within the analysis. Census data validates the analysis and further informs MyLA311 service request clusters to determine a proxy for the socioeconomic characteristics of neighborhoods. This analysis uses data from the U.S. Census 2017 5-year estimate of the American Community Survey (ACS) for socioeconomic and demographic information.

      The distribution of race and ethnicity within Los Angeles census tracts by concentration is demonstrated in Figure 2. The analysis uses the racial and ethnic categories of the 2017 ACS; more information discussing the reasoning is discussed in Chapter 3. The four maps in Figure 2 highlight the areas with the highest percentage of a single race and/or ethnicity. The Valley, the Westside, Malibu and Hollywood Hills have the highest percentage of Non-Hispanic White residents. African American neighborhoods are mostly located within central Los Angeles. Small pockets of Asian population exist near the Downtown and Midtown areas of Los Angeles. Hispanics or Latinos comprise the majority ethnic group in Los Angeles and have dense communities in the northern portion of the Valley, the Downtown area, East Los Angeles, and the corridor down to San Pedro Harbor.
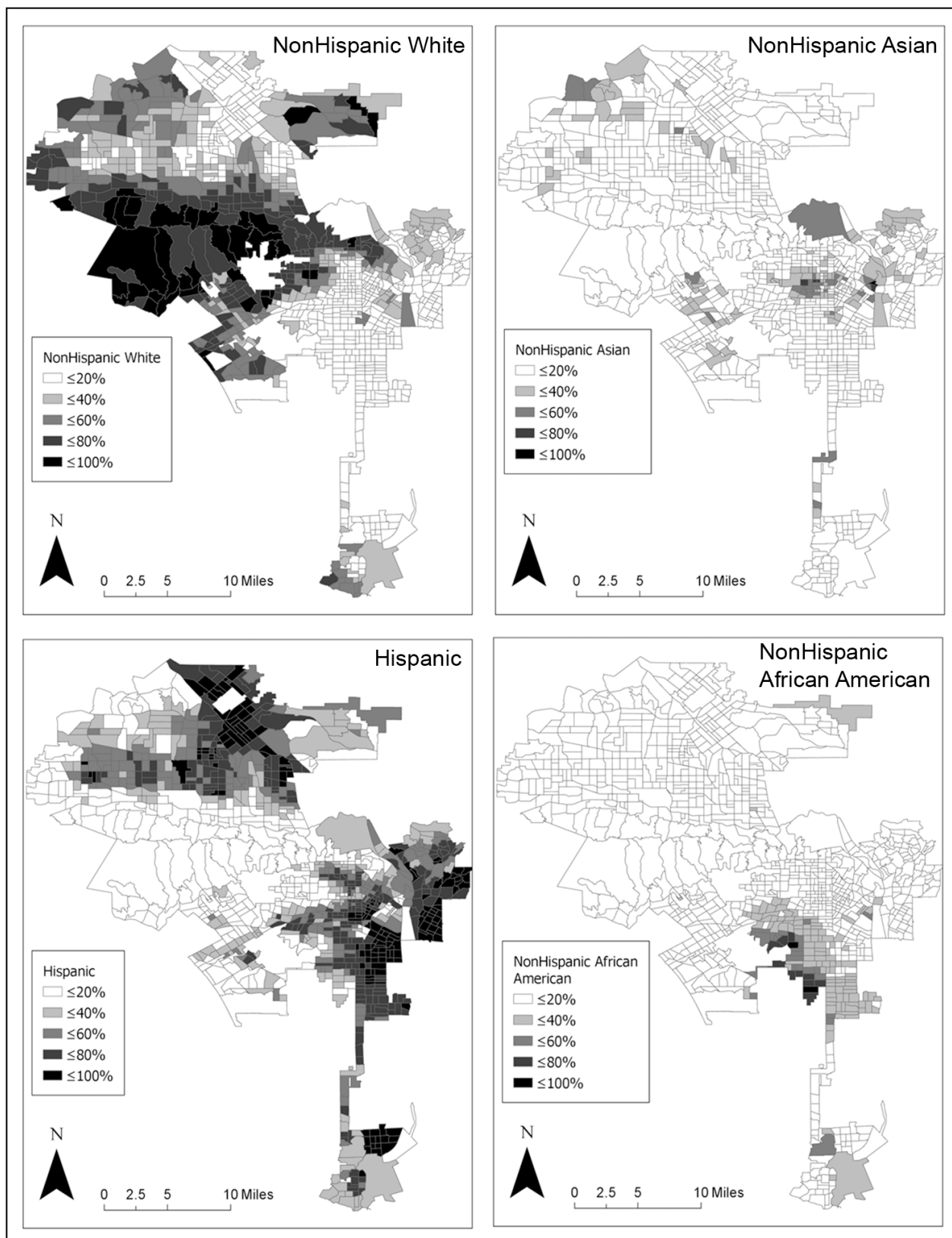
Figure 2 Distribution of race and ethnicity by census tract in Los Angeles

Percentages of the highest educational attainment of each census tract are displayed through the maps in Figure 3. Graduated symbology on a scale from the minimum percentage of education attainment to the maximum percentage is used to account for the disparities between each level. Attainment of only a high school level of education is most prevalent in the northern portion of the Valley, and then from Downtown to the San Pedro Harbor area. College education at the levels of associates and bachelor's degrees, are more equally distributed throughout the city, while the Westside of Los Angeles and Malibu have a higher proportion of residents who have completed some type of graduate school.
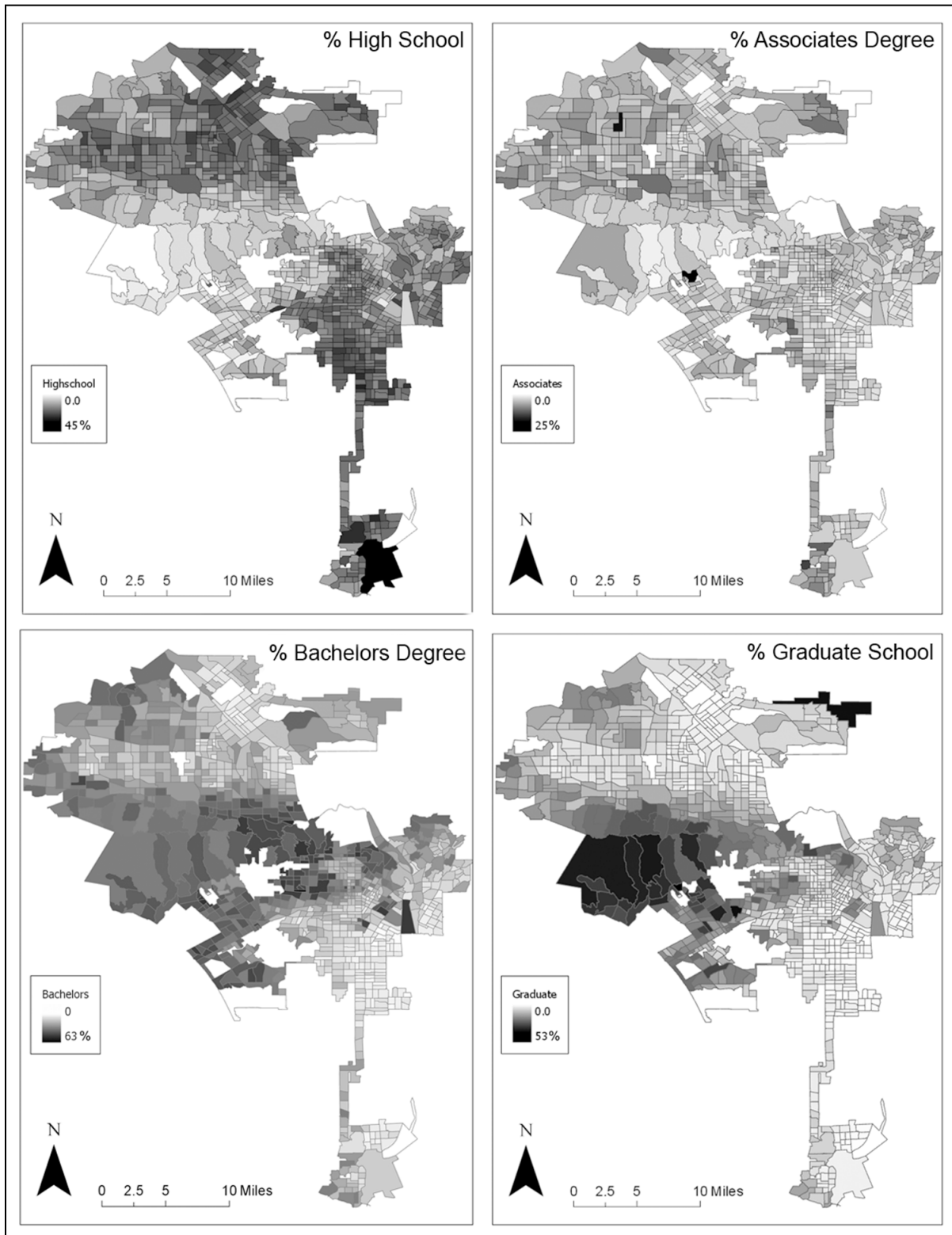
Figure 3 Distribution of highest educational attainment in Los Angeles census tracts

Two different symbolizations were used to display the Mean Income (Figure 4). The same data is represented in two different ways due to the disparity in income levels present with unclassed data symbolization detailing percentages from zero up to $430,000. Through unclassed symbolization, the wealth appears to be only concentrated in the Hollywood Hills. However, once the symbolization is by standard deviation, characteristics of overall income in the city become more apparent. While the Westside and the Hollywood Hills remain with the highest average incomes, the perimeters of the Valley, East Los Angeles, and pockets in the Harbor area and Downtown are now visible. The Downtown pocket of higher mean income parallels the recent shift towards higher incomes in that area as high-end residential development gentrifies the area (Collins 2016). Otherwise, the general Downtown and Harbor Corridor are in the lower end of the average income bracket.
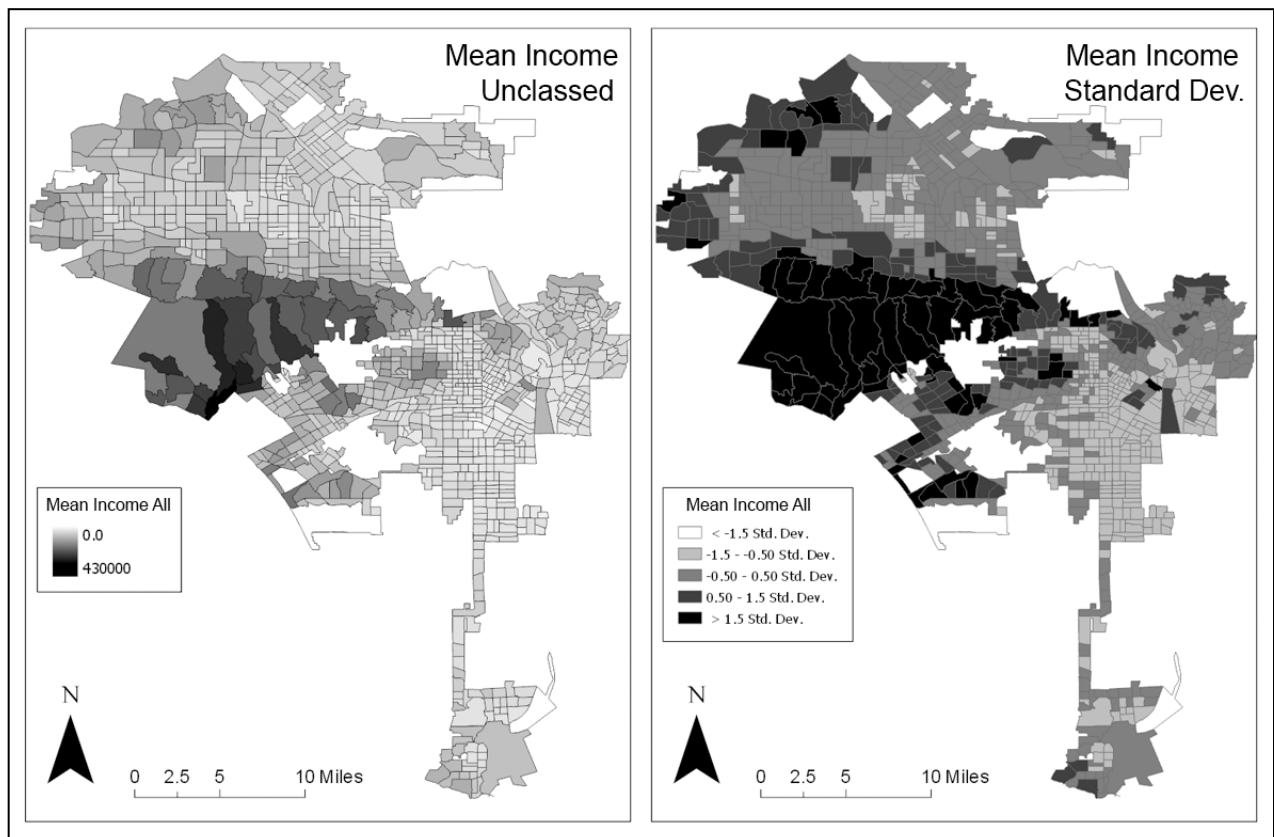
Figure 4 Maps of both the unclassed (left) and standard deviation (right) of mean income in Los Angeles census tracts

The percentages of those who are uninsured and those living below the poverty line are determined through different metrics, yet there is an overlap present between the two groups. The percentages of residents who are uninsured and who are below the poverty line mirror each other in some regards (Figure 5). This may be attributed to insurance not being seen as a high priority when money must go towards other immediate necessities when one is living below the poverty line (Bundorf 2006). The Census Bureau establishes placement below the poverty line by a set of income thresholds varying by family size and composition. If a respondent's stated total income is less than the family's determined threshold, then that family unit and each individual in it is considered living below the poverty line (U.S. Census Bureau, 2018).

Economists established trends demonstrating that health insurance is generally affordable for between one quarter and three quarters of adults who are not insured (Los Angeles County Department of Public Health 2017). More recently and specific to Los Angeles, LA County Public Health (2017) determined a significant reduction in the number of uninsured individuals from 2011 to 2015. The general geographic locations for higher populations both percentage below the poverty line and percentage uninsured are in the northern portion of the Valley and the corridor from Downtown Los Angeles to the San Pedro Harbor.
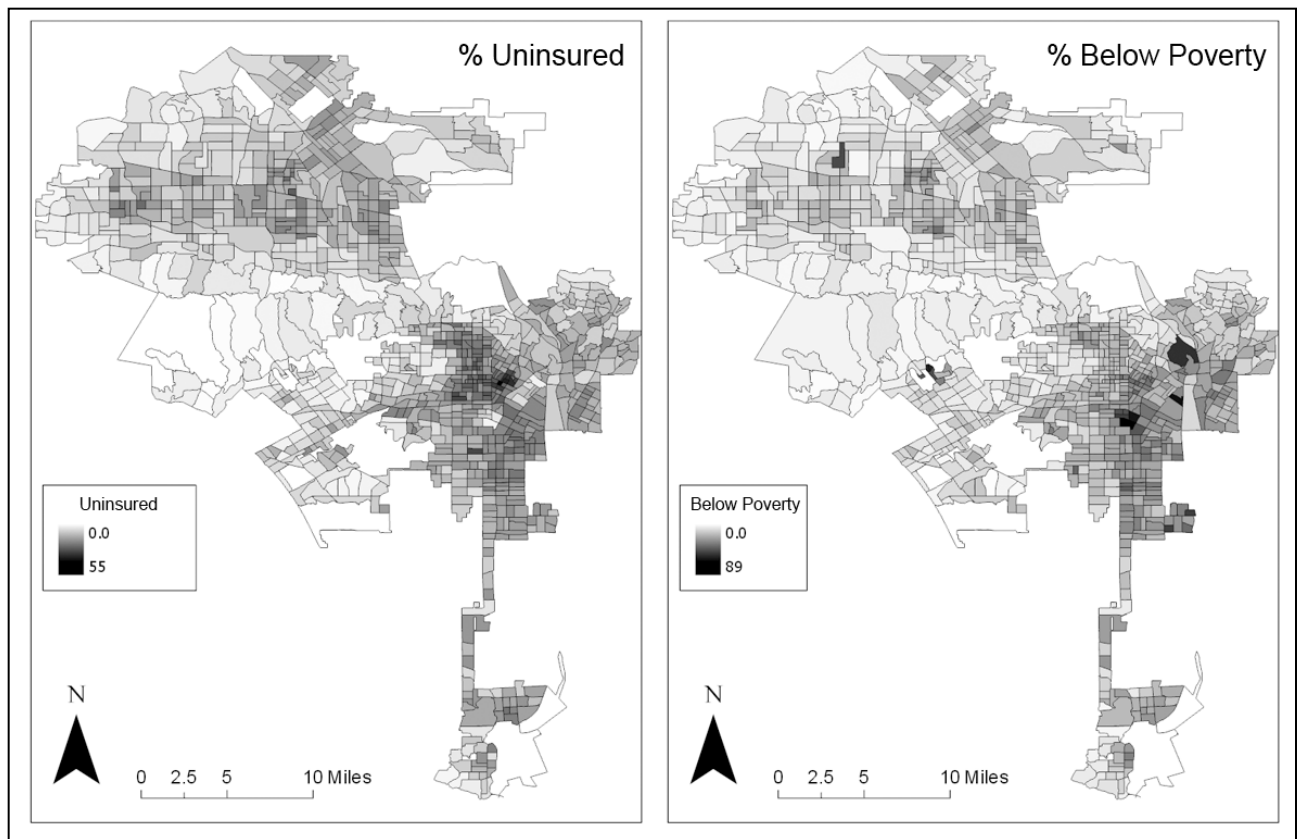
Figure 5 Distribution of uninsured and people living below poverty in Los Angeles census tracts

The final census variable observed is the unemployment rate, displayed in Figure 6. The unemployment rate is the most widely known labor market indicator and reflects the number of unemployed people as a percentage of the labor force. The official definition of unemployment encapsulates people who are jobless, actively seeking work and available to take a job. This statistic is formulated by number of unemployed as a percentage of the labor force (the sum of the employed and unemployed). Within the map, the darkest spots appear to be outliers: Cal State Northridge in the Valley, UCLA on the Westside and USC in South Central all have high rates of unemployment based upon the high residency of students in the area. The darkest, large census tract north of Downtown is Griffith Park, and the darkest census tract east of Downtown is Skid Row, known for a high population of homeless individuals.

Figure 6 Unemployment rate in Los Angeles by census tract

1.4. Origins of 311 Systems

Prior to the existence of 311 services, citizens only had 911 as an easy to remember option to call if they needed help, or they could call city departments directly with service requests if they had the correct phone numbers available. As a result, 911 received an overload of non-emergency calls, which led to delays in the response time for emergency services attending to actual emergencies (Schwester 2009). An alternative number for non-emergencies was created to alleviate the congestion of calls to 911, and 311 was chosen as the number. A successful trial run of a 311 system in Baltimore resulted in the Federal Communications Commission reserving 311 as a national help number for non-emergency situations, as deemed in Federal Communications Commission Report 97-51 (1997). The introduction of 311 systems in cities

14

provided an easily accessible, single point of entry to local government information and services and a "real" person to speak to. Customer service-oriented call centers and the creation of service data gained momentum to become a catalyst for modernizing city operations. This section describes the initial implementations of 311 systems in New York and Philadelphia, both cities Los Angeles sought to emulate in its own implementation.

1.4.1. New York

In 2002, New York City's Mayor Michael Bloomberg sought out the creation of a customer service focused initiative to provide access to GIS integrated information about city complaints from 15 major categories, including air and water quality, construction, noise, animals, snow, streets and sidewalks, and transit and parking (Nam 2012; Nadeu 2011). New York, with a population double that of Los Angeles, needed to implement a robust 311 system with the goals of allowing government departments to focus on respective core missions and efficiently manage workloads.

Now a modernized and transparent system, NYC 311 provides accurate and consistent data tracking and analysis of all service requests, with links to other civic resources for residential engagement. New York's 311 system is representative of one of the most significant correlations of citizens engaging with local government, as more than 8 million service requests are created annually (Kontokosta 2017). Data-driven mindsets changed internal views of the 311 system as a method for complaints into an opportunity for larger-scale solutions; it could now become an integration of all civic culture, influencing organizational structure, technological components and workers (Nam 2012). In 2010, Wired Magazine included a feature analyzing New York's 311 data and detailing the resulting policy changes. For example, once geographically identified patterns of service requests for handling excessive noise became the

top service request, the administration instituted noise-abatement programs and quelled the problem (Johnson 2010). As Los Angeles readies itself for scaling up its 311 system and digital technologies, as noted in Mayor Garcetti's Directive No. 3 (2013), and it can look to New York as an example for city service request management and resulting policies.

1.4.2. Philadelphia

According to an analysis by the Pew Charitable Trust using U.S. Census American Community Survey data, Philadelphia was determined to be the poorest large city in the United States (Trinacria 2018). In 2008, Philadelphia Mayor Michael Nutter unveiled Philly 311 as a modernized government-to-constituent customer service model for accountability and to improve life for the city's residents. Creating a database of feedback under a single entity was the goal; previously, the city had multiple customer service hotlines, however each of them was located in singular compartmental silos. Rather than forcing city departments to merge with the system, the city made it optional in order to ease the adoption process. Philly 311 had a clear process for maintenance and corrections in order to maintain its knowledgebase, further bolstering the need for interdepartmental collaboration, transparency, and integration (Nam 2012). Philly 311 (Figure 7) differentiates itself from the MyLA311 system dashboards by providing a live map feed of the reported problems.
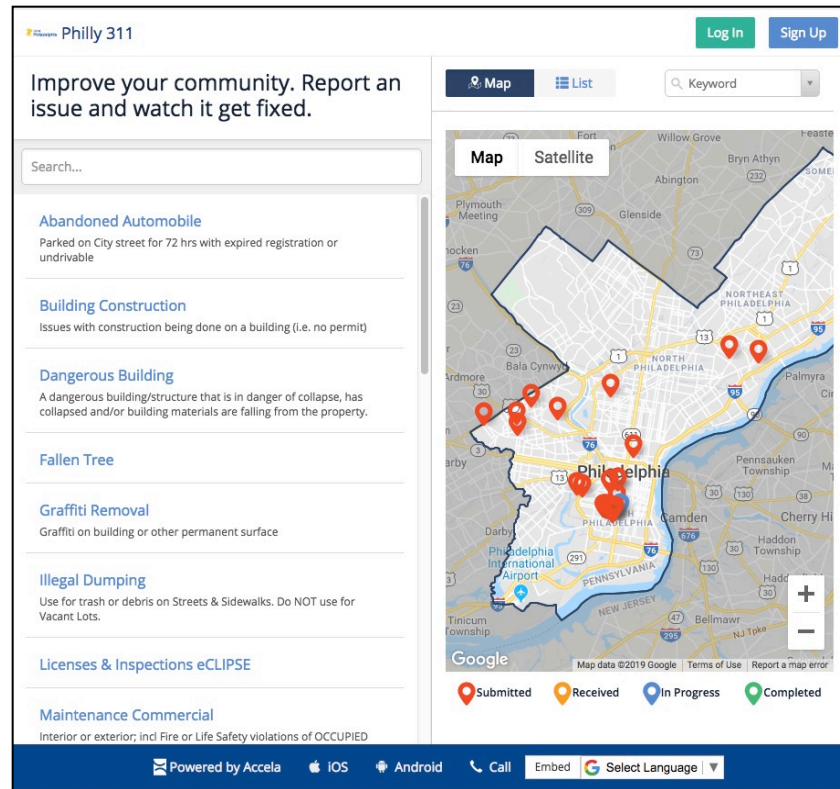
Figure 7 Philly 311 website interface for reporting 311 service requests

Philly 311 enabled data-driven management in the city and its data became essential in managing, tracking, and monitoring organizational performance. The Philadelphia City council reported resources became used more effectively, saving money and time of legislative members to direct towards other important needs of constituents (Nam 2014). While the system was an improvement, city personnel still struggled to modify the system and reported difficulty in data management practices. Multiple internal studies of Philly 311 lead to changes and updates in how the service would be run and continue to adapt to the city's changing needs. In 2015, Philly 311 modernized to a cloud-based system, effectively eliminating antiquated technology, upgrading communications, and becoming more "scalable, resilient, transparent and responsive to the needs of constituents" (Bengfort 2019). Philadelphia's 311 system provides a model for Los Angeles to modernize the MyLA311 system.

1.5. MyLA311 and Selected Service Requests

Los Angeles' 311 Call Center provides general information relating to city run programs and services. Citizens have the option to submit city service requests as needed and are connected directly with specific city departments, bridging communication between constituents and the city and simplifying operations with the City of Los Angeles. The system provides an extensive list of options available for submission as service requests through MyLA311. This section provides context of case studies on use of service requests within GIS analyses and the reasoning behind reporting relating to illegal dumping, graffiti removal, homeless encampments, dead animal removal, and broken streetlights. The reasoning behind reporting specific problems lends insight to the urban signature, characteristics and values of a location.

1.5.1. MyLA311

Individual Los Angeles city departments had veteran service request systems, yet these compartmentalized, department-specific silos of information lacked transparency and hindered achievement of data-driven government. The four custom built existing systems were designed prior to the development of APIs and system interoperability as required features of online services. The different systems were unable to exchange information and resulted in duplicates of service request IDs and miscommunication. Introduced in 2013, the newly comprehensive MyLA311 system was produced by 3Di Systems. Adaptations to new technological trends, namely call centers, mobile applications and open data, necessitated updated technological infrastructure and removal of older system architecture, which obstructed innovation. The new system interface (Figure 8) utilizes the Oracle WebCenter Portal for enterprise portal, Oracle WebCenter Content for Customer Management System, and Oracle Siebel for Customer

Relationship Management to provide an enterprise portal and user interface capable of serving

constituents and their customer service requests through an integrated system.

Figure 8 City of Los Angeles MyLA311 Service website

In order to meet the needs of individuals from a variety of backgrounds, many methods of

submitting a MyLA311 service request exist. The current options include through a phone call,

city attorney, city council office, self-report by driver, e-mail, fax, letter, request through the

Mayor's Office, queue initiated customer call, radio request, TDD/Nex Talk (service for deaf and

blind) request, Twitter complaint, voicemail, walk-in to city department, online web form, the

MyLA311 online portal and MyLA311 Mobile Application submission (Figure 9). The

significance of the source of service requests can be identified through analysis of the MyLA311

data to determine which sources are used most frequently for submission of specific service

request types.
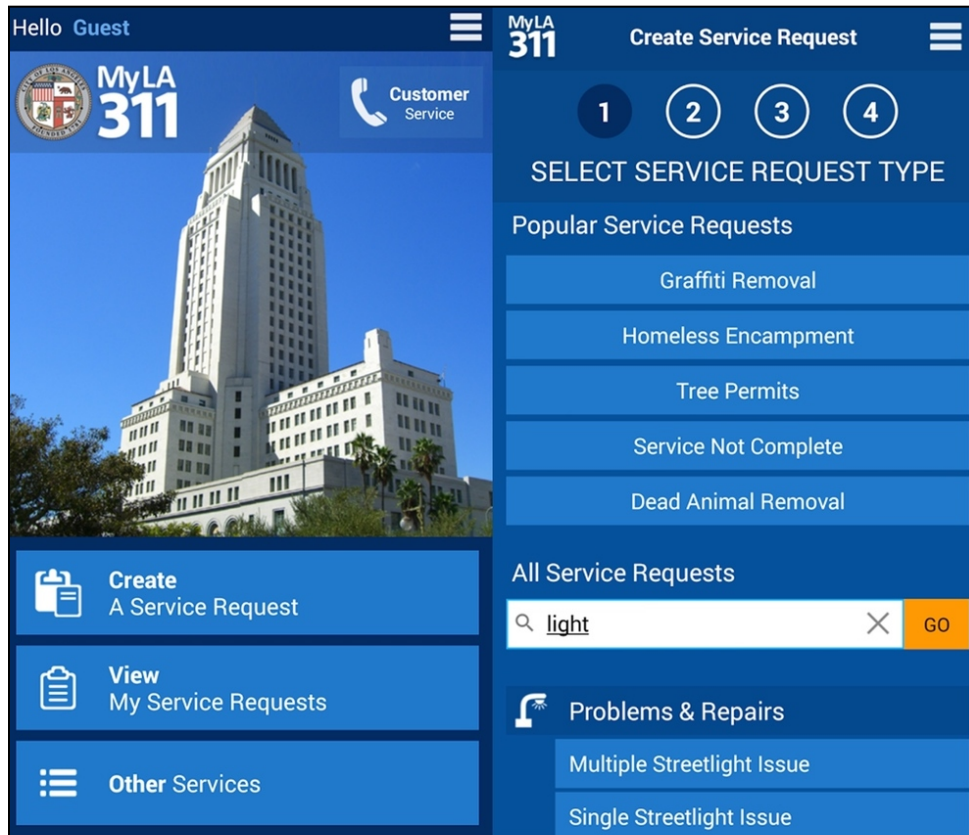
Figure 9 MyLA311 mobile phone application interface

In order to access the data produced from MyLA311, it is necessary to go through a separate website hosted by the city called Los Angeles Open Data. This Socrata powered website hosts a variety of data, with much of it produced by standard city processes and services. MyLA311 data is available going back to 2013. The MyLA311 publicly available data is not the same as internal city MyLA311 data. Publicly available datasets from MyLA311 only include service requests of bulky items, dead animal removal, electronic waste, feedback, graffiti removal, homeless encampments, illegal dumping pickup, metal/household appliances, multiple streetlight issue, other, report water waste, and single streetlight issue. The full list of MyLA311 data available through access to the city's backend is in Table 1; within the table, the five service request types analyzed in this thesis are underlined. These service request types were selected

because the reports are expected to come from people who did not make or cause the problem being reported.

Table 1 All Possible Service Request Types from MyLA311

311 Theme	Service Request Type
Animal Related Services	Request animal related services including <u>dead animal removal</u>, loose or confined animals.
Investigations	<u>Homeless encampment</u>, Illegal auto repair, Illegal construction, Illegal construction fence, Illegal discharge of water, Illegal dumping in progress, Illegal excavation, Illegal sign removal, Leaf blower violation, News rack violation, Non-Compliant Vending, Obstructions and Tables , chairs obstructing and Report Water Waste
Parks	Park facility and field maintenance, trash and cleanliness issues, security and park tree/animal/bug issues
Problems & Repairs	<u>Report graffiti or issues with streetlights</u>
Refuse & Pickups	Bulky item, Containers, Electronic waste, Metal / household appliances, <u>Illegal dumping</u>, Service not complete and etc
Sanitation Billing	Bulky Item fee, Extra capacity charge, Solid resource fee and Sewer Service Charge Adjustments. For recycLA billing issues, go to recycLA.com or call the Customer Care Center at 1-800-773-2489.
Street Problem/Repair	Barricade removal, Bus pad/landing, Curb repair, Flooding, General street inspection, Guard/warning rail maintenance, Gutter repair, Land/mudslide, Pothole, Sidewalk repair and Street sweeping
Transportation	Dockless Mobility Enforcement
Trees/Vegetation	Bees or beehive, Median island maintenance, Overgrown vegetation/plants, Palm fronds down, Street tree inspection, Street tree violations, Tree emergency, Tree obstruction, Tree permits and Weed abatement for pvt parcels
Other	To be used ONLY if the issue being reported does not fit into any of the SR Types available on this list. Select Radio buttons above to see more SR Types.
Feedback	Used for commentary

Another difference in the publicly available versus internal MyLA311 data is the level of detail and commentary. The nature of MyLA311's standardized open data does not include the unique comments or photos relating to each service request; as a result, additional circumstances of each request cannot be assumed, and the service request must be accepted as a baseline service request called in by a citizen. This internal cleaning of the data before publishing it serves to standardize the information and remove personally identifiable information.

1.5.2. Illegal Dumping

Service requests for illegal dumping seek to rectify the illegal disposing of appliances, barrels, construction debris, electronics, furniture, household trash, leaking liquids, tires, or yard debris. Within Los Angeles city boundaries, citizens have the option to call Los Angeles

Sanitation and coordinate the pickup of the above items through a free service for either electronic or bulky item pickup. Service request types for bulky item or electronic pick-ups were created as a service to individuals removing their own belongings, whereas the illegal dumping service request type is used for disposal of items in the public realm without clear or legal ownership. The city and council offices often implore residents to assist in cleaning through reporting illegal dumping through outreach, such as Figure 10, over social media or pamphlets.

Figure 10 LA Sanitation image on Facebook promoting reporting illegal dumping

If citizens are unable to dispose of their waste in a manner considered easy and low-cost, illegal dumping is often the outcome, as "dumping is sensitive to the cost of legal waste management and the threat of enforcement" (Matsumoto 2011). A contributing factor to illegal dumping may be the shortage of proper, or "accessible," waste treatment facilities. People are driven to dispose of their waste through illegal means to avoid the cost of paying for these services (Ichinose 2011). Consistent monitoring for illegal dumping would be inefficient and costly in terms of resources, lending weight to the inclusion of illegal dumping as something citizens can report through community participation. A cleaner environment leads to an increased quality of life for inhabitants.

1.5.3. Graffiti Removal Requests

Service requests for graffiti removal seek to rectify the illegal writing or drawings made on a wall or other surface, often without permission and within public view. Graffiti occurs on a variety of surfaces and at a variety of scales, as exemplified by Figure 11. The standardization of public MyLA311 service requests removes any detail or characteristics of the graffiti. The MyLA311 graffiti removal request data only entails the graffiti that people care about enough to report. The visual characteristics of graffiti impact how individuals feel about public safety, economic development and businesses, the quality of life for individuals in an area and feelings about neighborhoods.

Figure 11 Graffiti on a building in Los Angeles

Some cultural components of the context surrounding why graffiti occurs extend beyond borders. Megler (2014) comprehensively analyzes the spatial relationship of graffiti in San Francisco using a combination of census data and 311 data. She identified geographic and sociodemographic factors with significant correlation to graffiti reports. Most graffiti occurs along streets that are arteries because there is the most visual impact possible of graffiti being

seen. In regard to a trend with sociodemographic characteristics, more reports will be generated from a higher income community even if there is the same amount of graffiti in a lower income community, all other variables equal. This relates to the tolerance for graffiti in an area as part of an urban signature, yet tolerance cannot be specifically tested for through the graffiti removal service request data.

1.5.4. Homeless Encampments

A homeless encampment entails one or more shelters consisting from lean-tos made of cardboard, to tents, to more elaborate structures that are used by homeless individuals. An encampment has no safe way to store and clean food, attracts disease-carrying vermin through accumulating trash and often contains biohazards resulting from human waste or makeshift fuel sources. These poor hygiene conditions are ripe for spreading a multitude of diseases, further endangering the already marginalized population. Encampments bring about problems in the homeless population, the environment and the larger community through unhealthy and unsafe the conditions.

Encampments do not have a set temporal definition or location, despite environmental factors and the built environment potentially influencing prevalence and longevity of an encampment. Citizens have the option to report a homeless encampment represented by an individual repeatedly occupying the same location or establishing a shelter, potentially creating a public hazard or nuisance such as in Figure 12. After a homeless encampment cleanup service request is filed, a report is sent to the Bureau of Street Services Investigation and Enforcement Division, LASAN, Los Angeles Housing Services Authority, and Los Angeles Police Department. Initial contact must be made by LAHSA offering services to the individuals in an encampment. Notices of the dates of encampment clean-ups are posted in order to provide a

24

warning of the occurrence and give individuals a chance to relocate. Due to strict adherence to policy, procedure and scheduling, the process between reporting an encampment and the clean-up itself can take upwards of 75 days. The city seeks to enforce measures intended to protect a citizen's right to safe, healthy and accessible public spaces in an equitable and responsible manner.

Figure 12 Homeless encampment on public right of way in Los Angeles

Homelessness is a publicly visible effect of poverty and vicious inequality. As more individuals experience homelessness, their presence becomes more integrated into the physical structure of a city. Homelessness is associated with a variety of sociodemographic and spatial factors. When mapped over a geographic location, certain characteristics of homeless individuals may tend to stand out. For example, in terms of household characteristics, the "presence of children, age of the head and the head's drug and alcohol problems" were found to be significantly associated with the probability of being homeless (Early 2005). Once out on the street, the location that a homeless individual will frequent may shift towards a location where their needs can be met, either through outreach, shelters, or policy more favorable to their

presence. This leads to the presence of encampments near food, alcohol, employment (or crime) opportunities, and shelter from the elements (Early 2005).

Generally, reports of encampments stem from opportunity to increase the sense of personal safety in an area. The community feels the impact of increased transient presence as impeding on daily life. Locations adjacent to transient encampments experience "higher levels of petty and serious crime unrelated to [self-described] 'routine behaviors,' such as drug dealing and usage, disturbance, theft, prowling, burglary, panhandling, fighting, vandalism, armed robbery, rape, and aggravated assault" (Chamard 2010). Public perception of transient activity may incite feelings of lawlessness in an area, dissuading people from spending time or money in an area with heavy homelessness as an illegitimate use of public space. Chamard (2010) finds residents living in close quarters to homeless encampments to "suffer disproportionately from crime committed by transients." McCormack (2010) understands that within the characteristics of urban spaces, "most personal safety concerns mentioned in studies were associated with the presence of undesirable users," effectively labeling the homeless as "others." In this regard, people may be more likely to report homeless encampments in areas they frequent outside the realm of their residence or job location in order to push them out of their immediate location due to concerns about increasing safety.

While an individual's reasoning behind reporting a homeless encampment is not detailed in the MyLA311 data, individualist perspective models may offer some insight as to why:

> These pathology-based models explain homelessness as a consequence upon deviant behavior and values arising from an individual's mental illness, chronic substance abuse, family disorders, disaffiliation from community, or willful participation in a "culture of poverty" (Koenig 2007).

Within this perspective, individuals reporting homeless encampments are attempting to prevent the values they associate with homelessness as values characteristic of the location they reside in.

26

While extreme variations of perspectives on homelessness exist, a general air of NIMBY-ism, where people would like the problem of homelessness to be solved but at a distance from where they reside, is attributed to their requests for removal of the encampments (Dear 1992).

1.5.5. Dead Animal Removal

The inclusion of dead animal removal as service requests relates to a desire for a clean city and lessening potential health risks associated with the carcasses of dead animals. Urban spaces are constructed in a way to prioritize human mobility, not mobility of animals; as a result, spatial and constitutive characteristics of a place are aligned in a practical manner for human purposes through infrastructure (Lulka 2013). Physical contact with a dead animal carcass poses the chance of transferring diseases detrimental or deadly for humans and domesticated pets, making timely removal of the dead animal through submission of service requests extremely important to both the environment and inhabitants.

Data on dead animal removal is "highly indicative of a general pattern of urban relation between humans and nonhumans" (Lulka 2013) in an urban physical environment. Automobiles are the medium for the most often occurring interaction of human-caused death of animals. Mobilization in a car without the direct physical, human bodily movement results in disconnect between humans and the outside occurrences of a vehicle; it is in this space where road kill occurs (Lulka 2013). Should a civilian hit an animal, it is unlikely that a driver will take the time to submit a service request for removal of the dead animal, as the main priority in the moment is traveling to the destination.

Lulka (2013) finds size and type of the animal can influence whether a civilian will submit a service request for its removal. Outside the realm of road kill, civilians also may submit a service request for removal of a dead pet in order to avoid paying the fees associated with

proper cremation or burial (Wessel 2016). While the removal of a domestic animal may differ

from the removal of road kill in terms of emotional attachment, the public mechanism for

removal disposes of it in "the same bureaucratic dehumanized fashion" (Lulka 2013).

1.5.6. Broken Streetlights

Street lighting contributes to the built environment and context of a location. Depending

on the nature of implementation, lights impact and determine the "aesthetics, cleanliness, traffic

and crime safety and community support or cohesion" of an area (Williams 2007). Residents or

visitors of an area will spend more time outside if they feel the area is conducive to activity and

has higher walkability characteristics, characterized by lighting. In this regard, street lights act as

social support features that promote or deter activity in the built environment.

Lighting is rapidly being considered in a shift towards situational crime prevention where

environmental design improvements restrict opportunities for crime and suppress fear-invoking

environments (Pain 2006). Lighting and visibility are subconsciously associated with an

increased sense of safety. Separate from feelings relating to personal safety, an increase in

"community pride and sense of ownership of the local area" may also drive people to report

broken streetlights (Pain 2006). An online platform for fixing broken street lights in France

called Signalez-nous saw continual use and 311 reporting still occurring after five years of

introducing the application; potentially providing evidence for the belief that the impact of

environmental determinants has a direct effect on citizen's physical and psychological health,

which is why reports occur (Composto 2016). It is worth noting that for studies focusing on

streetlights and crime that the presence of working street lights may not have actually reduced

crime, but only reduced the fear of crime occurring (Pain 2006).

1.6. MyLA311 as Starting Point for City's Open Data

Los Angeles Mayor Eric Garcetti signed a 2013 Executive Directive designed to promote transparency and accountability by providing raw data in easy-to-find and accessible formats. In addition to a transparency effort, the directive aimed to modernize city government and operations management as a policy goal. The mayor emphasized how a significant goal of the directive was to "foster creative new thinking about solving our most intractable challenges through public-private partnerships and promoting a culture of data sharing between our own City departments and other civic resources" (Office of Los Angeles Mayor Eric Garcetti 2013).

The city has been slowly building up its data and technology resources in an effort to set a new standard for cities intersecting data and policy. The shift towards open data sought to go beyond simply sharing information but providing opportunities for the data to transform into tools with tangible uses for constituents. An Open Data Guide for the city specified that only data which "increases public knowledge about department operations, furthers the mission of the department, creates economic opportunity, or responds to a need for public information" should be placed in the open data repository (Currie 2016). While the city does make some data public, much of the city's data is only accessible internally to city employees; for example, the majority of the service requests displayed previously in Table 1 are not available through the Open Data Portal. While including MyLA311 data is a start to truly achieving transparency and accessibility of city data, Los Angeles still must still make strides towards providing more city produced data to the public.

In 2016 the City of Los Angeles entered another contract to develop an open repository for the city's geospatial data with Esri, a GIS company known for creating powerful mapping & spatial analytics software, for an open data website called the GeoHub. The GeoHub offers the

capacity to provide file types conducive specifically to GIS work and was the source of the city

boundary data used in this thesis. According to Esri, the partnership had the goal of making each

department's data "available online in real time (or near real time) to boost efficiency and

eliminate the information bottleneck" (Esri 2016). Much of the initial public spatial datasets

shared came from existing city data repositories, such as MyLA311, with the purpose of

increasing citizen engagement and analysis of city data. Sharing data can lead to democratic

action within a community and overcomes bureaucratic obstacles while providing increased

levels of transparency into city services and operations.

1.7. Thesis Structure

This thesis is comprised of five chapters including this introductory chapter. Chapter Two

provides a comprehensive literature review examining research related to volunteered geographic

data, 311 systems, and reasons for reporting service requests. Chapter Three details the data and

the methodology used analysis following a case study detailing the relationship between service

requests and demographics as they relate to multiple cluster analysis methodologies. Chapter

Four details the results of the analysis with maps, charts, and graphs of the results of multivariate

clustering. Chapter Five interprets the multivariate clustering results of the 311 data and

discusses the benefits from an analysis using MyLA311 data. Chapter Five also includes a

broader interpretation of the overall results of this thesis project.

Chapter 2 Literature Review

To better understand the relevance and significance of the spatial analysis of 311 data, it is imperative to provide background information and discuss previous literature on the topic. The chapter begins with a discussion of VGI and how MyLA311 data is related, following the shift towards modernizing civic operations and embracing open data. Next, the 311 data and service requests are explored more regarding GIS analyses and the background behind certain service request types. The chapter ends with a brief examination of characteristics related to those who request 311.

2.1. Volunteered Geographic Information

VGI is geographic data provided voluntarily by individuals or crowdsourced from user generated content (Goodchild 2007; Sangiambut 2016). The growth of VGI is discussed in this section, followed by the utility and ease of acquisition of user-generated data representative of an area. Finally, as the data is created by amateurs, the spatial data quality of VGI is discussed in relation to submitting service requests.

2.1.1. Ease of Acquisition and Collection of Data

Acquisition of data informs and drives GIS work. VGI offers an alternative source for data that is both cost-effective and allows for community-driven input in a timely manner. VGI can engage the public while acknowledging a low barrier for entry regarding capital and expertise. Advances in technology of both accessible mobile devices and web-based applications helped to popularize mechanisms for VGI data collection (Sangiambut 2016). The reduced cost and increased accuracy of GPS receivers alongside an abundance of wireless networks propels a multitude of readily obtained geographic information (Lu 2016). Previously, mass information

from individuals have been collected in more traditional and timely methods, such as through "telephone, fax, email, and direct meetings with the local administration or the mass media, which are often time-consuming and unsuccessful" (Brovelli 2015). The ease of both collecting and supporting VGI opens doors for new types of research and provision of spatial data while providing insight to scientific and policy-based research.

A VGI system has a directly participatory component to it, as users are generally aware that shared information contains a locational component. Many applications on devices take into account user information, with location as a key descriptor of a user (Whitaker 1980). Collection of this VGI can come from users putting georeferenced information or data in publicly accessible spaces, such as social media posts, or applications that take into account a location in the backend of an interface. Data made publicly available can then be freely downloaded as crowd-sourced information, even if users are not notified about their posts' information being downloaded for use by someone else. In joining a platform for social media and agreeing to a certain set of privacy settings, users legally, and sometimes indirectly, provide consent to sharing their produced VGI from the medium they shared information on. A more active, user-focused version of VGI entails a user running applications that gather multimedia data from device sensors, georeference the location and publicly disseminate it over the Internet (Brovelli 2015). Ease of sharing geographic information or data leads to more available data, therefore providing the possibility of more insight. This is not to say that all VGI is high quality, however, as it places responsibility of data quality on the users; increased quantity and assured quality of data leads to greater benefit.

2.1.2. Spatial Data Quality

The quality of VGI is dependent upon the context of users or source of the information. In 1980, the US Federal Government promulgated that a geospatial data standard must adhere to the five dimensions of positional accuracy, attribute accuracy, logical consistency, completeness and lineage (Goodchild 2012). As GIS technology advanced, a revised set of data quality standards sought to emphasize completeness, logical consistency, positional accuracy, temporal quality, thematic accuracy and usability (Antoniou 2015). These characteristics were chosen to ensure that all GIS data produced, including VGI, can measure the totality of features, provide adequate background to the context of data collection, entails a certain degree of details and has no discrepancy between the actual attributes of real-life occurrences and the respective coded attributes in the data. Metadata from VGI can be designed and combined to garner information about the users creating the data in an effort to attribute ownership and assess the data quality (Fonte 2015). The party behind the interface processing the data can design the interface in such a way to prioritize and provide assurance of higher quality data. As a result, those seeking VGI integrate design parameters adhering to the spatial data quality standards in order to produce the best product (Goodchild 2012).

Through contributions to VGI, people drive the collection of data from the masses. More participation ultimately means a larger sample of the population is able to contribute, providing more information to accurately describe a space. VGI replaces costly or time consuming methods for crowd-sourcing data through mass-produced methods at the individual level. Additionally, crowd-sourcing data entails referring the data to people without respect to their qualifications (Goodchild 2012). The data must be analyzed for participation patterns of contributors and on the content created to identify preferences and participation bias within a study (Antoniou 2015).

Knowledge of a community and the human component behind making contributions to VGI aids in understanding the level of different data quality elements (Antoniou 2015).

Quality can differ depending upon individual preferences from user input. As a collection of sources, a VGI dataset is greatly influenced by "spatial preferences, feature type preferences, and mapping behaviors" (Bégin et al. 2013). In the context of this research, an individual's spatial feature type preferences may determine an inclination to only submit service reports regarding homeless encampments or graffiti removal yet abstain from reporting illegal dumping or broken street lights. Citizens inherently, and sometimes unknowingly, have different ranking levels of prioritization in creation of VGI. These priorities may change dependent on the proximity of a problem from a place a user spends time or how strongly a user feels about solving the problem; as a result, VGI is subjective with regard to the user (Bégin et al. 2013; Lu 2016). While the 311 service requests cannot capture every occurrence of the aforementioned problems in each census tract, the quantity of data amassed over a year within a coherent, high quality data format can provide insight that could not be found otherwise through analysis.

In the case of 311 data, the service requests are a VGI byproduct collected under the assumption that people act directly to report factual problems to the city in order to alleviate them. While more active forms of VGI, such as OpenStreetMap, have a user base with the intention of accurately mapping and providing information, more passive forms resulting in VGI as byproduct of users may not have the same spatial literacy (Neis 2012). Analysis of community characteristics for reporting requires the sociocultural context for individual contributions to VGI repositories in Los Angeles to be observed. City service requests do not take an individual's requirements for credible information into account or provide commentary on the background of the reporter. Understanding the amount and type of service requests coming from an area and

incorporating sociodemographic information alongside 311 data analysis can provide clues to the

connection between service requests and the characteristics of communities reporting them.

VGI raises implications for spatial data quality, including the credibility of the source and

the accuracy of the attributes and positional information (Goodchild 2011). The motivation for

supplying VGI data can suggest a more or less potential for bias or deception, resulting in

implications regarding credibility. Flanagin (2008) seeks to determine the credibility of VGI and

how suppliers in a market should address the sensitivity of VGI users to inaccuracies and

misperceptions. Generally, the context of the source of VGI must be taken into account. As GIS

data shifts to more user-generated content, the potential for politicization or manipulation of the

data is ever more present. Research and analysis involving VGI must explicitly state the source

of VGI and address any possible bias to avoid arriving at false conclusions of VGI data.

Resulting spatial analysis combined with an understanding of the context of VGI's metadata and

sourcing is then more likely to provide credible information and results.

With the prevalence of mobile devices and increasing accessibility of cloud applications,

such as through applications like MyLA311, every citizen has the potential to be a sensor,

creating data about their environments (Goodchild 2007). Data is now accumulating faster than

ever before as a result of technological advances. It begs asking what drives people to provide

data, how accurate is the data, what the threats towards individual privacy are, or which

information providers are trustworthy (Flanagin 2008; Goodchild 2011). Cities have the

opportunity to provide mutually beneficial means of acquiring this data through 311 services.

VGI takes on the role of that of amateur geographic observation, at times even providing

a comparison for professionally attained geographic information. While VGI may not always

collect every single point of existing data, or in this case every location of a service request, it

provides a valuable opportunity on a cost-effective platform to perform geospatial analyses. Accessibility, either directly through citizen engagement or indirectly through mass data created by citizens utilizing third-party applications in the process of acquiring and using geographic information, VGI has the potential to alter GIS landscape significantly and soften criticisms regarding its use and the quality of produced data (Goodchild 2011).

2.2. Open Data

The VGI provided through the City of Los Angeles MyLA311 website falls into the category of open data. Open data is publicly available information accessible without having to make a request to the government (Scruggs 2013). Influenced by President Barack Obama's national "Open Government Directive" in 2009, local governments started to provide public information resources available in "machine-readable" formats (Carrizosa 2013). The US federal government further emphasized data as a valuable national asset through Obama's 2013 Executive Order providing a template for data quality and management. Electronic governance (e-governance) and new forms of civic engagement became popularized through the increased use of advanced technologies of the government.

2.2.1. Shifting Towards Open Data

Dissemination of government data potentially finds use as a neutral guide for solving the cities problems and allocation of resources, all while informing citizens. As local governments gradually adopted new technology, they accumulated more data than ever before; this data continues to grow at an exponential rate. As constituents became aware of the data's existence, they fought for access to it. Consumer rights activists argued information is explicitly linked with political authority and control, giving way to the Freedom of Information Act (FOIA) declaring any person has the right to request access to federal agency records or information (Currie 2016).

Sharing government data with citizens was initially a slow process due to the lack of channels for dissemination. Rather than letting the data go untouched and unseen by the majority of a population unless prompted for, the FOIA and technological openness serve to provide standardized formats, low costs and ease of access to government data. Citizen access to information is fueled in part by **"**openness as government disclosure and openness as open systems" providing value through entrepreneurship contributing to support of open data policies and the open data movement (Currie 2016). Newly available open data prompted new ways of thinking and ideas for approaching existing problems through a fact-based background.

Through accessibility and freedom of restrictions regarding use, the open data model for managing government-created data is a transformation from prior internal systems. In order to contribute to a city's data repository, city records are often digitized in order to provide a plentiful source of economic and institutional capital capable of propelling data-driven management and city planning goals in both the private and public sector (Currie 2016). Open source communities promote activity and content sharing on the end of the user and promote increased production of data and collection from the side of the publisher (Brovelli 2015). Open data produced by the government should be non-partisan in existence and removed from the context of its creation. Over time, accumulated data can be examined for identifying resulting effects in a community regarding new data mediators, policies or practices (Currie 2016).

2.2.2. Public Participatory GIS and E-governance

Open access and information sharing leads to civic engagement. Many e-government initiatives, which entail "delivery of information and services online through the internet and other digital means," are accessible from personal devices, as opposed to the previous process of physically going to a government entity for more information (Lu 2016). VGI and public

participatory GIS (PPGIS) provide an opportunity for heightened public involvement through open access tools and accessible/actionable information. VGI focuses on both active and passive collection of spatial information from citizens as sensors while PPGIS tends to inform government planning agencies through solely active public involvement (Lu 2016). The participatory component of VGI through complex modes of engagement can affect the types of data collected for a decision process (Brown 2013). The term PPGIS emerged in the mid-1990s for specific use cases where GIS was used with community intervention to aid insight towards policy decision-making; there was a shift from viewing GIS through only technical lens to a perspective involving institutions and greater society, all while emphasizing democracy (Brovelli 2015). In either case, individuals have the opportunity to create their own unique data that can further drive other processes.

Involving citizens more through open access and technology is bringing about new forms of governance. E-governance uses the internet a platform to bridge each individual to city hall, providing easily accessible opportunities to interact with the government. E-governance at its various stages (Figure 13) must be designed to encourage the public to become involve while not dissuading them through ease of operating and interacting. The stages progress from Stage 1 with one-way communication from the government to constituents with zero interactivity to Stage 5 with robust, two-way communication that is integrated into all key functions of the government and advances with the latest technology to best serve citizens.

		Stages of E-government				
		Administrative functions				Political functions
		Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
	Types of government	Information: dissemination/ catalogue	Two-way communication	Service and financial transaction	Vertical and horizontal integration	Political participation
Internal	Government to government	Agency filing requirements	Requests from local governments	Electronic funds transfers		N/A
	Government to public employees	Pay dates, holiday information	Requests for employment benefit statements	Electronic paychecks	One-stop job, grade, vacation time, retirement information, etc.	N/A
External	Government to individual–services	Description of medical benefits	Request and receive individual benefit information	Pay taxes online	All services and entitlements	N/A
	Government to individual–political	Dates of elections	Receive election forms	Receive election funds and disbursements	Register and vote: federal, state, and local (file)	Voting online
	Government to business–citizen	Regulations online	SEC filings	Pay taxes online, receive program funds (SB, etc.), agricultural allotments	All regulatory information on one site	Filing comments online
	Government to business–market-place	Posting request for proposals	Request clarification or specs	Online vouchers and payments	Marketplace for vendors	N/A
Technologies used		Basic Web technology, bulletin boards	Electronic data interchange, email	Electronic data interchange, electronic filing system, digital signature, interoperable technology, public key infrastructure	Integration of the technologies required for phase 1, 2, and 3.	Public key infrastructure, more sophisticated interface and interoperable technologies, chatrooms

Adopted from Hiller and Bélanger (2001).

Figure 13 Stages of E-government, *source:* Moon 2002

Within the context of this analysis, the MyLA311 system falls between Stages 2 and 3 of e-governance through the focus on customer service fulfilment for citizens and the city. Involving people in the production of civic data is a process of co-production through e-governance (Sangiambut 2016). Within this framework, a government no longer solely provides services but promotes market-based decision making as a manager of service providers (Hood 1995). In this regard, the functions of the government are shaped by how well they use technology and open source software/interfaces to interact with citizens. The design and user interface of the technology created for engagement decides the scale of how democratizing effect data production and use will have on a city (Sangiambut 2016).

Accessibility of interaction between both the user and the service leads to high levels of engagement for both parties. Open access e-government is no longer a one-way transaction of

"government-to-citizen"; rather it opens up a two-way "citizen-to-government-to-citizen" dialogue and exchange of ideas (Sieber and Johnson 2015). The ability of online interfaces allows the public to stay regularly informed through the ability to provide updates at a minimal cost. GIS technology must empower the public through its exploitation to effect policy and governance (Brovelli 2015). Data produced by individuals can be incorporated into management and aid decision-making for policy or other action within a community. Accessibility of e-government lessens the gap between constituents and the government by providing means for transparency, responsiveness and accountability (Lu 2016).

Citizen involvement in government processes is mutually beneficial, but there must be action on both ends for a dynamic understanding of the urban environment. The rescripting of citizen engagement should not place a burden solely on citizens to be active agents in claiming and monitoring their urban spaces (Sangiambut 2016). Participation translates into existing citizen participation from the real world to the digital world within the context of contributing to GIS data (Brown, 2013). As cities modernize and integrate more data sharing capabilities into their operations, individuals are able to extol individual emancipation from inefficiency (Sangiambut 2016). Coproduction of data as a byproduct of smart government initiatives allows cities to find opportunities for growth in understanding of technology, organizational management and how to approach challenges to better meet the needs of constituents (Schwester 2009).

More recently and relevant to the study area, Currie (2016) details Los Angeles's relationship with open data and how it seeks to "enable new administrative models and inspires new modes of civic involvement" on multiple fronts and mediums, such as the Facebook post in Figure 14 posted by L.A. City Councilwoman Monica Rodriguez's office. Through online

marketing citizens are more likely to engage with a civic entity, in this case a city councilwoman, and have it feel akin to normalized interaction outside the realm of the civic environment. Currie's research details how open data reshapes modes of administration and policy through a modernized, data-centric lens appealing to both citizens and the government. As a result, analysis and research has been performed on different types of existing 311 data, providing a baseline for how to utilize 311 data while acknowledging inherent limitations of the data, as discussed in a later chapter.

Figure 14 A Facebook social media post encouraging self-reporting by constituents

2.3. Analyzing 311 Service Requests

As descriptions of events and of spatiotemporal attributes, service request data is naturally tailored for use in GIS analyses revealing patterns relating to human nature. Spatial analyses of 311 service requests require a vast amount of data to provide a baseline where patterns of spatial distribution can be identified and avoid spatial autocorrelation (Mullen et al. 2014). Service request data can be seen as representative of the state of the "built environment,"

comprising of urban design, land use and the transportation system to encompass patterns of human activity within the physical environment (Gehl 2010). Changes to the built environment exist on a broad time spectrum, with some fleeting and others remaining for longer periods of time to shape the lives of people. The spatiotemporal fidelity through an address and the creation and service dates of 311 service requests types allows for in-depth analyses spanning a multitude of fields.

The human component of submitting a service request can be tied to a human-centric analysis of the environment. O'Brien (2016) finds that reporting of civic problems is distinguished behaviorally from reporting public issues arising from natural deterioration and people are found to specialize in one or the other. Investigating the type of person who self-reports community problems may identify that basing community intervention solely based on 311 data perpetuates inequalities simply because some communities may be partial to dealing with problems in their own way and not through self-reporting.

In a later publication on 311 data analysis, O'Brien (2017) finds the coproduction of data translates civic motivations into impacts. Many of these motivations are not themselves inherently civic, yet the submitted data platforms channel reports into positive outcomes for the community. Kontokosta (2017) finds that neighborhoods under reporting 311 data are often characterized by higher proportions of males, unmarried individuals and minorities, alongside higher unemployment and a smaller population of individuals fluent in English, providing further evidence for socioeconomic status and household characteristics have a non-trivial effect on the propensity to submit 311 service requests.

When analyzing 311 service requests of an area, it is important perform the analysis at a scale that will reveal useful patterns and new information at the local scale. Drawing on previous

research on citizen–government interaction, service delivery, and civic engagement, Minkoff (2015) focuses on how contacting propensity and condition both explain spatial variations in service request volume. Through an exploration at the census-tract-level variation in 311 service request volume within New York City, he finds using a larger administrative dataset at micro places allows one to estimate a more precise effect than audits at a smaller number of places would. The perspective of 311 occurring at an increasingly localized level is emphasized through White (2016) 's evaluation of 311 data as a measure of neighborhood-level realized demand for services and provider of information regarding relative intensity of how local neighborhoods utilize government services rather than 311 data as a measure of political participation and propensity for participation. Similar to the methods used in this analysis, she uses 311 data as a proxy for neighborhood-level propensity to participate and noted limitations. White (2016)'s research provides examples of how to best visualize the data both geographically and numerically while not committing an ecological fallacy.

2.4. Multivariate Clustering

Service request data sets comprising of more than a few thousand objects benefit from the assistance of algorithms for identification of patterns present in the data. Ester et al. (1996) defines a clustering algorithm as a process used for class identification to group a database's objects into subclasses with intentional meaning. The density-based notion of clustering lends weight to the idea of similar clusters comprising a neighborhood (Ester et al. 1996). Performing a multivariate clustering analysis of service request type frequencies located in a defined space, a census tract, provides the points needed for identifying distinct signatures in cluster types. Within ArcGIS Pro, a multivariate clustering analysis establishes natural clusters of feature characteristics based solely on feature attribute values within polygons.

Seeking to understand unique signatures within urban context at a localized scale, Wang et al. (2017) uses R to provide a multivariate cluster analysis of three cities' 311 data identifying spatial signatures of parts of the city that have similar patterns of requesting services. Sociodemographic patterns are identified within the clusters established from the multivariate clustering analysis of 311 service request type patterns' locational attributes through census data. The analysis produced a methodology for creating a low-cost decision support tool for urban stakeholders seeking information at the local level from city service data. The research warranted that 311 data analysis models based on regularly-updated open 311 data can have considerable potential regarding insight of city operations and neighborhood planning; it remains noteworthy that the structure of 311 reports correlated with socioeconomic quantities does not serve as an evidence of any causal relation, yet still provides insight to conditions present in the built environment (Wang et al. 2017).

Wang et al. (2017)'s multivariate clustering methodology was informed by previous analysis revealing the different structures or characteristics within cities. Bettencourt (2010) clusters cities into different classes of urban dynamics through analysis of data for gross metropolitan product, personal income, violent crime and patents for the construction of meaningful, science-based metrics for ranking and assessing local features. Spatiotemporal assessment of the characteristics of the cities leads to advancement in understanding the theory of urban evolution and promotes new tools for the formulation of improved urban policy. Kontokosta (2017) produces a proof-of-concept model of analyzing New York's 311 service requests to create a benchmark for and validate neighborhood resilience capacity. Kontokosta (2017) determines the average levels of community resiliency and activity at the neighborhood

scale through analysis of five years' worth of 311 requests for service in New York as a proxy for local activity.

In a multivariate cluster analysis, the parameters determine the course of the machine learning analysis. In Wang et al. (2017)'s analysis in R, the Initialization Method parameter uses Random seed locations option to frame the clustering algorithm as a sensitivity analysis to identify which features are always found within the same cluster through randomly selected seed features. One drawback of Random seed locations is that the algorithm incorporates heuristics and can return differentiating results each time the tool is ran, despite identical parameters and data. Two options available for clustering algorithms are K-means and K-medoids. The K-means function, described in Figure 15, uses machine learning and iterative refinement between classifying the characteristics making up a cluster and then computing the mean/centroid of all data points within the cluster type until there is little change within the makeup of the clusters (Trevino 2016). While K-medoids is more flexible in accommodating noise and outliers in the input features, K-means is faster and preferred for larger datasets, often bolstering the decision to use it in an analysis (Esri 2018).

$$ \text{objective function} \leftarrow J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 $$

Figure 15 K-Means Cluster Function (Sayad 2019)

The multivariate clustering tool often has the option to generate the optimal number of clusters, yet Wang et al. (2017) notes that four clusters is the best overall metric, as demonstrated through a comparison of results attained using both the Silhouette method and the Elbow

method. The Silhouette method is a commonly used method of interpretation and validation of consistency within clusters of data. It measures how similar an object is to its own cluster (internal relation) with other clusters (external relation). The Elbow method observes the percentage of variance explained as a function of the number of clusters using a pre-determined set of information. In summary, Wang et al. (2017)'s research found that within use of both Silhouette and Elbow methods, only four clusters were presented in analysis of three separate cities (New York, Chicago and Boston), as this number of clusters revealed detail about a city's sociodemographic profile while providing a reasonable trade-off between having too many clusters and an acceptable clustering quality.

2.5. Collective Efficacy

Collective efficacy relates to specific tasks taken to maintain public order. The idea of collective efficacy revolves around the engagement of community members through a basis of trust and a willingness to intervene for the common good to effectively create a safe and orderly environment. Collective efficacy can help stratify the social characteristics of a place of residence or work through inherent structural contexts (Sampson 1997). While much of the research surrounding collective efficacy is done within the context of crime, the underlying connection of the residents taking action to remedy physical ailments of their community prevails. Areas with similar attributes form clusters of individuals with common values as residents to maintain effective social controls with variation depending on the neighborhood (Sampson 1997). The collective actions of individuals inhabiting an area then enact change through a shared mindset.

The action of submitting a 311 service request is aligns to a citizen exercising informal social control to better influence the community and shape public order to a certain set of

standards. A service request is a task-specific construct built upon citizen engagement and shared

expectations (Morenoff 2001). In submitting a request to remedy a physical malady of the

environment, people become transformational agents heavily influencing environments around

them. Participation in the collection of data can bring about many benefits and provides an

opportunity for collaboration on achieving social, economic, political, scientific and

environmental goals (Lu 2016). In effect, it is their individual beliefs of personal efficacy that

motivate and guide their actions, further influencing those around them. This shared sense of

collective efficacy through collective influence lends citizens success in shaping and bearing

command on their surroundings (Bandura 2000).

Sometimes, people in a community can instead place more weight on the disorder present

in the built environment negatively influencing the actions of the people who reside in it rather

than focusing on the altruistic act of people bettering the built environment to make it more

preferable for living. While there is still an emphasis on betterment of the community as an end

goal, the "collaborative effort to maintain a certain standard of communal life" stems from

negative connotations (Wilson 1982). Through this perspective, small problems in the

environment can easily result in much larger problems occurring down the line. This reasoning

purely stems solely from the physical component of disorder: each service request included in

the analysis representing an attempt at self-policing a physical problem in a community.

If community members do not report a service request for an ongoing problem, it is

inferred that the problem is engrained within the fabric of the community and attempting to

remedy it would be considered feckless. However, if a service request is seen as an attempt to

raise the standard of living throughout the community if it is made with the intention of

remedying the problem and providing the requestee with a level of social control. Within the

scope of research by Wilson (1982), a lack of cohesion in a community reduces chances of any one person acting as a representative agent of the entire community. While this view of collective efficacy is more relaxed, the mindset does much more for simply (temporarily) remedying a problem, rather than attacking the problem at the root.

The success of the MyLA311 Program relies heavily upon citizen engagement. Citizens are the "coproducers" of public services through interactions with service agents to redirect city services. Rather than an agent presenting a finished product or service to the citizen, the agent and citizen produce the desired transformation together to effectively exert influence on both policy and maintenance of city programs (Whitaker 1980). The number of requests disseminating from an area can lead to revisions in the management, supervision, and training regarding the request type, therefore affecting the distribution of service allocations within the city. City management can then better predict the need for specific service requests in different locations and distribute the funds and resources appropriately.

In a city as large as Los Angeles, citizens take pride in the upkeep of their communities, referring to certain areas as distinct neighborhoods to gain a more geographically distinct identity. The city must actively respond to the changing tastes, circumstances and behaviors of constituents and their request for assistance in tackling new problems (Whitaker 1980). Through the data, it is evident that service requests created by city employed agents make up a sizable portion of the requests, it is clear that city services are not likely to be as helpful as they could and should be without the citizens themselves taking action. Collective efficacy, in all of its forms, is rooted in community action with long-lasting benefits, all of which stem from citizens taking the initiative towards bettering their surroundings.

Submitting a 311 service request provides a citizen with the opportunity to influence their surroundings and physical aesthetic of a location while exerting some sort of social control. Within this research, there is a focus on the aesthetic of a location due to the physical attributes that the chosen service requests impact. This strong "sense of place," or locational identity, combined with the demographics of an area shape its local signature (Handy 2002). Determining the reasoning behind civic engagement of service requests can be supported through the theories behind collective efficacy.

Chapter 3 Data and Methodology

This chapter describes the data and methods used to evaluate the relationship between MyLA311 service request data, census data, and the locations of service requests. First, this chapter describes the data used in the analysis. The first phase of this analysis involved identifying the types of service requests used and the resulting temporal scale. Next, the census data and its attributes are discussed, as well as other geographic boundaries used in the analysis. Finally, the methodology behind how and why clustering is used is discussed.

3.1. Methods Overview

The methodology for this thesis demonstrates the process of the steps taken for an analysis of the MyLA311 data (Figure 16). The data acquisition entails using publicly available data to form the basis for the analysis. Data preparation involved cleaning extraneous fields and removing irrelevant service requests in both Excel and ArcGIS Pro to ensure only relevant data was taken into consideration. This thesis uses ArcGIS Pro's machine learning geoprocessing Multivariate Clustering tool to determine clusters with similar features regarding 311 reporting frequencies. The tables produced from the multivariate clustering analysis tool were then brought into Excel for further analysis and visualizations through radar and bar charts.
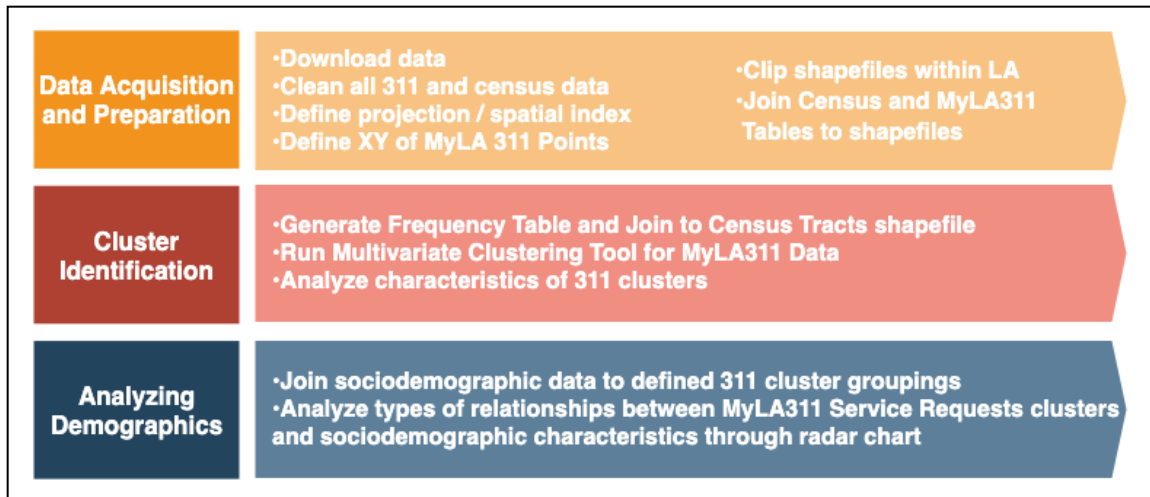
Figure 16 Overview of methodology process

3.2. Data and Processing

All of the data used in this analysis is publicly available, allowing the analysis to be reproducible. All data from MyLA311 used in the analysis is from 2018. While the source of the data is MyLA311, the data can be found and downloaded through Data.LACity.org and is provided by the City of Los Angeles Information Technology Agency in the category "A Well Run City". Shapefiles for the city boundaries and other administrative geographies can be found through the Los Angeles GeoHub. Finally, the 2017 5-Year Estimate American Community Survey (ACS) information can be found through the United States Census Bureau website.

Table 2 Data sets used in the analysis

Name	Spatial Unit	Source	Data Purpose
MyLA311 2018 Data	Latitude and longitude of points	Los Angeles Open Data Portal	Illegal dumping, graffiti removal, homeless encampments, dead animal removal, broken street lights
2017 5-Year Estimate ACS Data	Census Tract level for City of LA	U.S. Census Bureau	Determines demographic / socioeconomic characteristic traits in relation to service requests
TIGER/Line Census Tracts	Census Tract level	U.S. Census Bureau	Will be used to spatially join data. Offers visual context of data and boundaries

3.2.1. MyLA311

The MyLA311 data is comprised of service requests made during 2018 for the selected

service request types. As a proof of concept for determining clusters, the scope of the analysis

required only one year of service requests, yet the temporal and spatial component can be made

larger or more granular depending upon the goals of an analysis. The frequencies of the service

requests served as exploratory variables later on as the basis of the multivariate clustering.

MyLA311 data was cleaned of any personal identification information and extraneous

commentary or attachments and then reproduced by the City of Los Angeles My311 Data

Management team. The origin of the data stems from individual requests for service made to the

City of Los Angeles. When the service requests are geocoded into individual points on a map

through respective latitude and longitude, each point falls within a census tract polygon. Figure

17 is representative of the MyLA311 service request points at their base level on a map

symbolized by service request type.
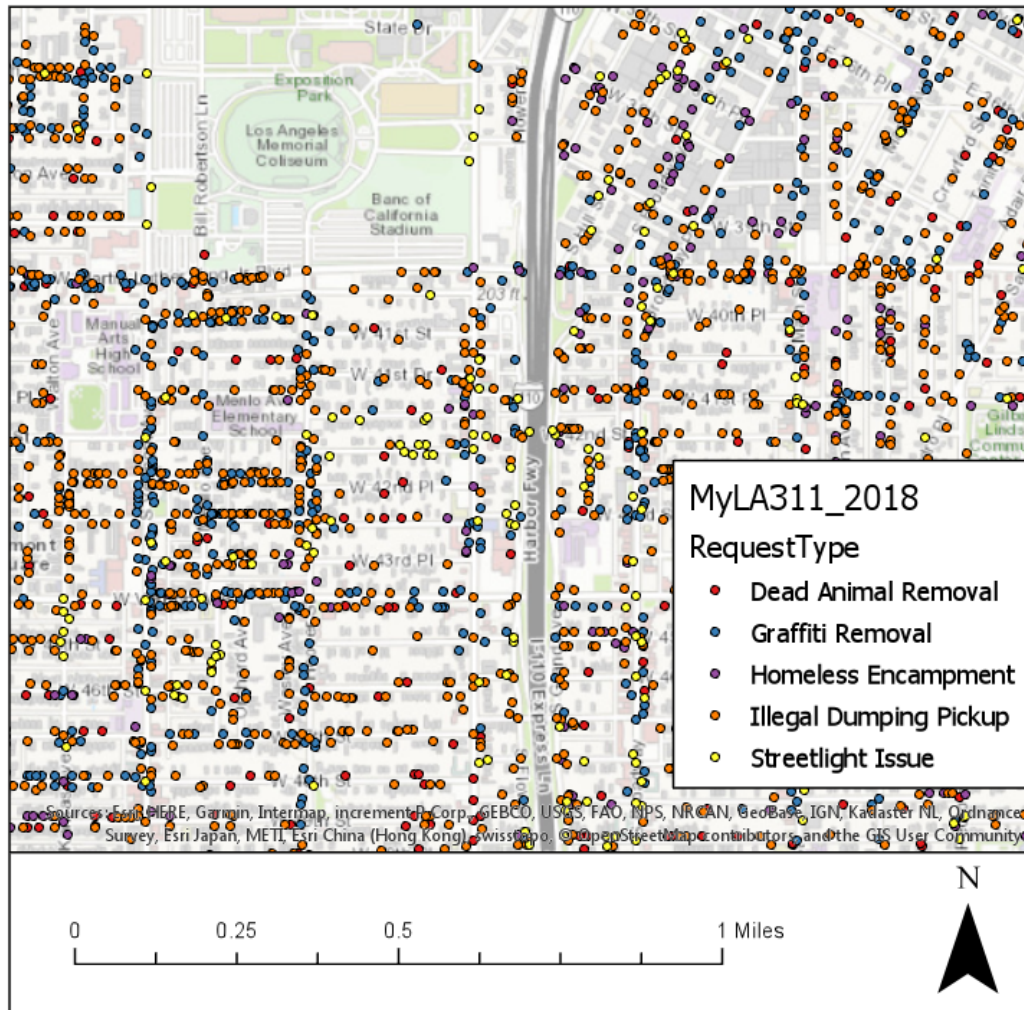
Figure 17 MyLA311 2018 data displayed as points at a large spatial scale in a section of South Los Angeles

Several steps needed to be completed for quality control of the point data. The downloaded data was comprised of service requests for only illegal dumping, graffiti removal, homeless encampments, dead animal removal, and broken streetlights. These service requests were chosen because of their visual characteristics, which prompt individuals to report them. The raw data contained many attribute fields, exemplified in Table 3, and had to be cleaned through filtering and removing attributes not essential to the analysis, which mainly used SR Number, RequestType and location features. Additionally, city employees completing requests on the job have the option to self-report the need for a service request when they are out in the field in order

to provide a record of addressing an issue. These service requests, along with any with a "cancelled" status, were filtered and then removed from the data set in Excel so that the analysis solely focused on reports made by citizens.

Table 3 MyLA311 Data Attributes

	Description	Attributes
Time	Descriptors of the important time points of requests	CreatedDate, UpdatedDate, ServiceDate, ClosedDate
Location	Descriptors related to the geo-location of the requests. Many of the descriptors provide redundant information about the locations of requests	AddressVerified, ApproximateAddress, Addresss, HouseNumber, Direction, StreetName, Suffix, ZipCode, Latitude, Longitude, Location, TBMPage, TBMColumn
Type	Semantic descriptors of the requests	RequestType, RequestSource, MobileOS, Anonymous
Agency	Descriptors indicating which agency handled the request	Owner, AssignTo
Other	Other varied descriptors relating to identification, status, or relationship to the administrative entitites	SRNumber, Status, ActionTaken, CD (Council District), CDMember, NC (Neighborhood Council, NCName, PolicePrecint, APC (Area Planning Commission)

Within this analysis, data for broken streetlight requests was comprised of two separate requests for single streetlight repairs and requests for streets with multiple broken streetlights. In order to ensure that the data was within the City of Los Angeles boundaries, data was queried to only include service requests with "Y" in the AddressVerified attribute, therefore ensuring that it was internally validated by the city with GIS data. This validation comes from geocoding the given address of a service request through the city's internal Thomas Brothers Map locator with all addresses within the city. While each service request attribute includes an address, the supplied latitude and longitude attributes generated by the city were used in order to avoid manually geocoding any of the addresses, which would have been a lengthy and time-consuming process.

3.2.2. American Community Survey Data

The census data stems from the 2017 5-Year Estimate ACS and contains both socioeconomic and demographic variables at a variety of spatial aggregations. The data provides insight regarding population diversity, education, income and employment at the scale of census

tracts. All of the data is freely available through the United States Census Bureau's American Fact Finder Download Center. The data was processed, as described below, to only include attributes needed in the analysis.

Following the methodology used by Wang et al. (2017), demographic features represent a variety in population diversity, education, and income and employment in this analysis. In the case of the data from the ACS, data was downloaded with 'Census tract' selected as the Geographic Type in the 'Geography Filter Options'. Additionally, the selected data contains information from the following categories: "Hispanic", "Non-Hispanic White", "Non-Hispanic African-American", "Non-Hispanic Asian", "High school degree", "College degree", "Graduate degree", "Uninsured ratio", "Unemployment ratio", "Poverty ratio", and mean for "Income (all)", "Income of No Family", "Income of Families" and "Income of Households." The datasets created by the U.S. Census Bureau in the American Community Survey used in the analysis are detailed in Table 4.

Table 4 American Community Survey Census Datasets

Census Dataset	GEO.display-label	Attribute Fields
ACS_17_5YR_DP05	HC03_VC93	Percent; HISPANIC OR LATINO AND RACE - Total population - Hispanic or Latino (of any race)
ACS_17_5YR_DP05	HC03_VC99	Percent; HISPANIC OR LATINO AND RACE - Total population - Not Hispanic or Latino - White alone
ACS_17_5YR_DP05	HC03_VC100	Percent; HISPANIC OR LATINO AND RACE - Total population - Not Hispanic or Latino - Black or African American alone
ACS_17_5YR_DP05	HC03_VC102	Percent; HISPANIC OR LATINO AND RACE - Total population - Not Hispanic or Latino - Asian alone
ACS_17_5YR_S1501	HC02_EST_VC11	Percent; Estimate; Population 25 years and over - High school graduate (includes equivalency)
ACS_17_5YR_S1501	HC02_EST_VC13	Percent; Estimate; Population 25 years and over - Associate's degree
ACS_17_5YR_S1501	HC02_EST_VC14	Percent; Estimate; Population 25 years and over - Bachelor's degree
ACS_17_5YR_S1501	HC02_EST_VC15	Percent; Estimate; Population 25 years and over - Graduate or professional degree
ACS_17_5YR_S1701	HC03_EST_VC01	Percent below poverty level; Estimate; Population for whom poverty status is determined
ACS_17_5YR_S1903	HC03_EST_VC02	Median income (dollars); Estimate; Households
ACS_17_5YR_S1903	HC03_EST_VC22	Median income (dollars); Estimate; FAMILIES - Families
ACS_17_5YR_S1903	HC03_EST_VC47	Median income (dollars); Estimate; NONFAMILY HOUSEHOLDS - Nonfamily households
ACS_17_5YR_S1902	HC03_EST_VC02	Mean income (dollars); Estimate; All households
ACS_17_5YR_S2301	HC04_EST_VC01	Unemployment rate; Estimate; Pop. 16 years and over
ACS_17_5YR_S2701	HC05_EST_VC01	Percent Uninsured; Estimate; Civilian noninstitutionalized population

In order for the data and analysis to accurately portray the demographics of Los Angeles, it is important to carefully address the concepts of race and ethnicity. Race and ethnicity are categorized separately in the census, and respondents may report any combination of race and ethnicity. For the 2017 ACS, the census allowed individuals to report race as one or more of the following categories: American Indian and Alaska Native, Asian, Black or African American, Native Hawaiian and Other Pacific Islander, White or some other race. Ethnicity was broken into two categories, "Hispanic or Latino" and "Not Hispanic or Latino."

For this analysis, the categories of race and ethnicity were combined. Four categories of race and ethnicity were identified:

1. "Hispanic," which included all ethnicity responses of "Hispanic or Latino" and any selection for race

2. "Non-Hispanic African-American," which included responses for ethnicity of "Not Hispanic or Latino" and race of "Black or African American"

3. "Non-Hispanic Asian," which included responses for ethnicity of "Not Hispanic or Latino" and race of "Asian"

4. "Non-Hispanic White," which included responses for ethnicity of "Not Hispanic or Latino" and race of "White"

This decision was influenced by the "Los Angeles County: Predominant Racial or Ethnic Group by Census Tract" map produced by Allen and Turner (2002). The categorizations chosen for Allen and Turner (2002)'s map accurately encapsulate the diverse demographics of Los Angeles. Allen and Turner (2002)'s perspective acknowledges the information and context lost within Los Angeles if only race is symbolized, instead of race alongside ethnicity. The Hispanic population within Los Angeles is steadily growing and necessary for providing context of the makeup of the city, as evidenced by the maps procured in Figure 2 in Chapter 1. A difference of this analysis was that it did not combine "American Indian or Alaska Native" or "Native Hawaiian or Other Pacific Islander" with the Asian identifying respondents because of a minute response from those two groups which would not have greatly impacted the outcome of the census tract clustering.

3.2.3. Contextual Boundary Data

Administrative polygon and line shapefiles were downloaded to provide geographic

constraints and context in the analysis. Both of sources used to acquire this data were from

websites serving as public platforms to freely explore, visualize, and download location-based

Open Data in GIS files. The boundary data was downloaded from the Los Angeles GeoHub. In

order to be published on the Los Angeles GeoHub, Los Angeles Information Technology

Agency require the data to have adequate metadata regarding the source and accurate or

precision. The shapefiles served the purpose of providing locational context for the city's extent.

The TIGER/Line shapefiles were downloaded from the U.S. Census website. Census tract

boundaries were downloaded for Los Angeles County and were clipped to only include census

tracts within City of Los Angeles boundaries. The census tract TIGER/Line shapefiles were the

main polygons used within the analysis, as both the MyLA311 data and sociodemographic data

are joined to the polygons. The same polygons were then used for the frequency analysis of

MyLA311 data points to determine the number of points and types of points in each census tract

polygon. These polygons displayed the cluster signature type in the final maps.

3.3. Data Aggregation

Performing a multivariate clustering analysis requires the points from individual service

requests to be counted within some sort of administrative boundary, in this case polygons

comprised at the scale of census tracts. Given the differing values present (i.e. frequency of

service request types, dollars, percentages), the data must be standardized prior to analysis.

3.3.1. Scale of Analysis

The decision to use census tracts as the level of spatial aggregation for the analysis stems

from a necessity rooted in the need for identifying patterns at the localized level. Zip codes,

council districts, neighborhoods, census tracts and census block groups were all considered as options for data aggregation, yet census tracts provide the best trade-off between spatial granularity of having a sufficient number of sub-areas within each city and containing a statistically significant sample of service requests (Wang et al. 2017).

Within the analysis, 1059 census tracts were analyzed. While Los Angeles data originally had 1170 census tracts, this included sliver polygons from the TIGER/Line census tract shapefile that may have occurred during the clipping process from slight misalignments of boundary lines downloaded from the LA GeoHub. Located adjacent to normal polygons within the Los Angeles city boundaries, the sliver polygons are long, elongated areas which do not represent an entity and must be deleted. The misalignment of boundary lines and creation of sliver polygons during the clipping process did not affect the outcome of the analysis. The sliver polygons were identified through sorting the frequency analysis by count and individually checking the location of census tract ID's with zero service requests present. Excess census tracts were then removed and the spatial join of points within the census tract was finalized to only account for points within census tracts of city boundaries.

3.3.2. Data Normalization

American Community Survey (ACS) data required normalization for the comparison of different features. In creating a radar chart, the standardization process was necessary for a comparison within a population based upon features with different values and units. The units of attribute values differed based upon the measured feature, such as income in dollars versus a percentage of people in a population belonging to a category. ArcGIS Pro automatically standardized data by z-score to allow for proper comparison across unit types through the multivariate clustering geoprocessing tool and creation of the box plots in Chapter 4. The

equations in Table 5 were used to create standardized radar charts in Excel. Within the radar

chart, the data was grouped by attribute type in order to better visually identify differences in

patterns in the resulting chart.

Table 5 Standardizing in Excel

Excel Functions for Standardizing Data	
AVERAGE	Returns the average of its arguments
STDEV.P	Calculates standard deviation based on the entire population
STANDARDIZE	Returns a normalized value

3.4. Data Analysis

In order to perform the spatial analysis, all spatial data had to be cleaned within ArcGIS

and Excel for quality assurance (Table 6). This entailed removing data outside of the study area

and irrelevant data prior to performing spatial statistics through the frequency analysis tool and

using the multivariate clustering analysis tool. Tables created from ArcGIS analyses were

exported to Excel and then joined through V-Lookup functions. Pivot Tables and charts were

used to then visualize the final results.

Table 6 Software required for analysis

Software	Manufacturer	Function
ArcGIS Pro 2.3.1	*Esri*	Geoprocessing Functions • Spatial Join • Table analysis and management • Statistical Analysis • Selecting and Extracting data • Frequency Analysis • Multivariate Clustering
Excel 16.23	*Microsoft*	Data Manipulation and Analysis, Charts

3.4.1. Processing and Joining Data to Shapefiles

In its raw format, the MyLA311 and census data downloaded were not embedded into shapefiles. After downloading the chosen census variables, the data was opened in Excel to be properly formatted and removed of extraneous information. This process entailed selecting only necessary columns of data from the various census datasets downloaded for the analysis. Once all of the columns of data were selected, a V-Lookup function in excel ensured that each census tract number and its respective values are returned in the final, cleaned Excel sheet to be joined to the spatial data.

The attributes of the TIGER/Line shapefiles were properly formatted prior to joining them to census data. The GEOID field was a text field and needed to be a 'double' field in order to join to the ID field in the census Excel data, which was also a 'double' field. To do so, the attribute table of the census tract was opened and a field with a name 'GEOID10' and type set as 'Double" was added. The field calculator then populated 'GEOID10' with data from 'GEOID'. The census data was then dragged onto the map so it could be joined to the census tract shapefiles. The 'GEOID10' of each dataset was the basis for the matching.

The MyLA311 data was formatted in a way to easily allow GIS use. Once the data was dragged onto the map, it appeared in the Table of Contents. The data was then able to populate

the map by right clicking the table and selecting "display XY data." The longitude and latitude fields of the MyLA311 data corresponded to longitude and latitude fields of the join tool, respectively. After clicking OK, an event layer was created in the map; however, this was just a temporary layer that must then be made permanent by using the Copy Features tool and exporting the data to the map. A spatial join then added the census tract values to the MyLA311 points' attributes.

A frequency analysis was then run to read the attribute table, and it created a new table containing unique field values and the number of occurrences of each unique field value, thereby creating a count of each service request type within each census tract. This table was then exported and manipulated by a pivot table in Excel. The pivot table ensured that each census tract is in the GEOID10 column and the count of each service request type was in successive columns. It was then saved as a CSV file and joined to the original census tract shapefile through matching the GEOID10 values.

3.4.2. Multivariate Clustering Analysis

ArcGIS Pro has a multivariate clustering tool in its spatial statistics geoprocessing toolbox that uses unsupervised machine learning methods to determine natural clusters of features based solely on feature attribute values. The classification method is considered unsupervised because it does not require a set of preclassified features to train the methods used to find the clusters in the data. The tool works to maximize both within-group similarities and between-group differences while adjusting for every possible combination of the features to cluster. While the methodology of the multivariate clustering in this thesis follows the same parameters of Wang et al. (2017), the analysis was performed using tools available through ArcGIS Pro.

Multivariate clustering analysis was performed only with MyLA311 service request data frequencies within the census tract polygons. Sociodemographic data was not taken into consideration regarding the creation of clusters. The first input for the analysis involved selecting all variables necessary for the multivariate cluster analysis to run. The analysis fields used in this thesis are the counts of service requests for graffiti removal, homeless encampment, illegal dumping, dead animal removal and broken streetlights. The counts within these census tracts served to distinguish clusters of features from one another.

The additional parameters entailed an integer value representing the Number of Clusters to create, an Initialization Method and clustering algorithm. The analysis was limited to four clusters with the reasoning detailed in Section 2.4. The random seed locations option for the Initialization Method parameter was selected to frame the clustering and identify features found within the same cluster through randomly selected seed features. K-Means, explained in detail in Figure 15, was selected as the clustering algorithm and was discussed further in Chapter 2. If run multiple times, the multivariate clustering geoprocessing tool will not repeatedly create the same output due to the use of random seed locations; however, the created cluster groupings should generally be similar given the same parameters and inputs. The tool itself is exploratory for this reason.

3.4.3. Sociodemographic Analysis

In order to analyze the sociodemographic characteristics, the selected data from the ACS 2017 5-Year Estimates were then spatially joined to the new shapefile of census tracts with resulting cluster types from the multivariate cluster analysis. After running the tool, a new output layer shapefile was created detailing the cluster type of each census tract. The attribute table of the new output only included attributes included in the analysis fields parameter from the source

layer. For this reason, a spatial join was then run with intersect to add the name of the census GEOID10 to the resulting cluster type. This new attribute table was then exported to Excel to create the radar chart of sociodemographic values of the clusters. A VLOOKUP function then matched the census tract cluster GEOID10 (census tract ID) with the respective sociodemographic values from the ACS 2017 5-year estimate data set.

Within Excel, the sociodemographic values of each census tract were then separated by cluster type. The average of cluster values was determined to form a baseline for the characteristics of a cluster. The average and standard deviation of *all* census tracts observed in the analysis was also determined. The data was then standardized by the STANDARDIZE function in Excel, which calculated a normalized value (z-score) through =STANDARDIZE ([Cluster Average Value], [Average Value of All Census Tracts], [Standard Deviation of All Census Tracts]). These z-scores were then used for the radar chart in order to visually display the average characteristics present within each 311-determined census tract cluster at the same scale, thus allowing comparison of the socioeconomic characteristics of the 311 clusters. The same methodology was followed to produce radar charts of the service request types per cluster.

Each radar chart displays multivariate results in a two-dimensional graphic by representing different axes of each exploratory variable starting from the center and extending outward. Each color-line represents a different cluster type, and the line size drawn connecting the data to each axis represents the relative magnitude of the variable for the group. The chart was selected for visually displaying group-observations because the overlay allows visual comparison of the relative position of each group. Each characteristic was standardized to achieve an equal weight and to simplify the interpretation by plotting z-scores instead of raw values on each axis.

Chapter 4 Results

This chapter documents the results of the analysis. The Multivariate Clustering Analysis for the MyLA311 service request data determined the constructed 311 clusters present within each census tract and the correlating sociodemographic census data. All the exploratory sociodemographic variables were observed against the clustering metrics to identify statistically significant correlations. The resulting clusters create urban signatures from 311 frequencies. Sociodemographic values present in the census tracts of clusters characterize and inform the features of the local community.

This chapter is broken into several sections to present the results of the analysis. Section 4.1 provides a visualization and a breakdown of the of MyLA311 service request data. Section 4.2 describes the results of the MyLA311 multivariate cluster analysis and 4.3 identifies the sociodemographic characteristics most present in the clusters. 4.4 summarizes the results of the analysis within the geographic context of Los Angeles.

4.1. Analysis of Service Requests

Prior to analyzing the patterns present in the MyLA311 data, it was necessary to analyze the makeup of the data itself within the context of Los Angeles census tracts. Through the frequency analysis, the counts of each service type and sum of total service requests per tract were identified. The purpose of the map in Figure 18 is to display where the greatest intensity of 311 calls for service are disseminating from in Los Angeles. The Hollywood Hills, north Valley and Harbor communities generally have the lowest number of requests. The highest intensity of MyLA311 service requests are found in the central Valley, Downtown and Westside communities.
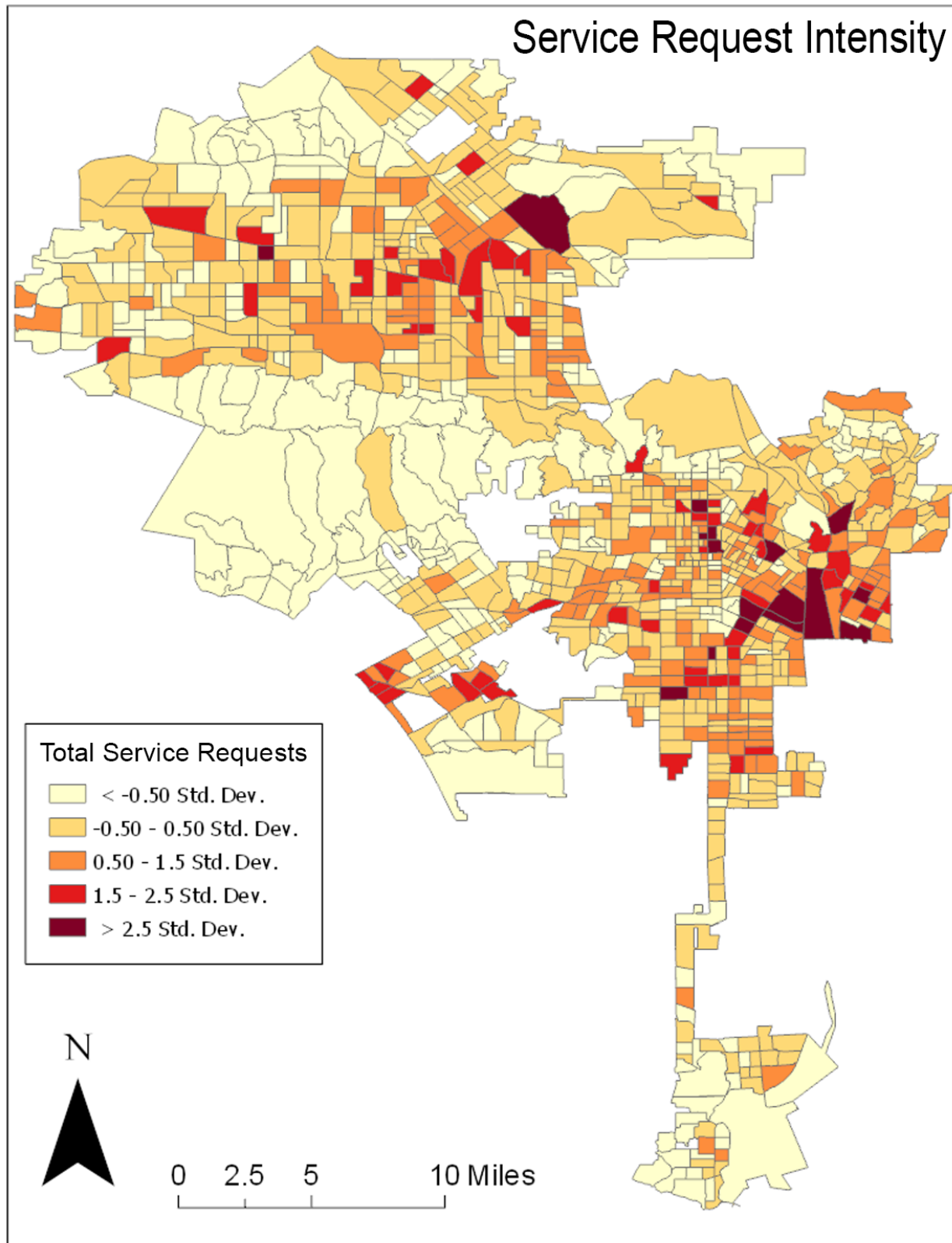
Figure 18 Map detailing the amount of 311 requests for service normalized by the standard deviation of each census tract

The overall distribution of service request type and source is detailed in Figure 19. The majority of all service requests are through the MyLA311 mobile application, phone calls to the 311 hotline or through the MyLA311 website. Figure 19 serves the purpose of showing how much data stems from the MyLA311 mobile application, mainly due to its ease of use in reporting a service request. An additional benefit of the mobile application is the ability to add a photo to document the problem. The multiple mediums available for submitting a service request are able to appeal to a variety of audiences with varying technological literacy, all while lowering barriers to accessing a platform or service used for service requests.

Graffiti removal was the most popular service request, followed by illegal dumping pickups, homeless encampments, dead animal removal, and finally streetlight issues. Observation of the service request sources revealed the preferences of individuals in regard to reporting certain service requests. For example, it is more likely a person would prefer to use the mobile application, followed by the website and phone call to report graffiti removal or homeless encampments, but would most likely use a phone to report illegal dumping or dead animal removal before considering reporting through the website or mobile application. Finally, it was revealed that service requests for streetlights, the smallest group of requests, is evenly spread across the website, phone, or mobile application as a medium for reporting problems within the city.

Service Request Distribution

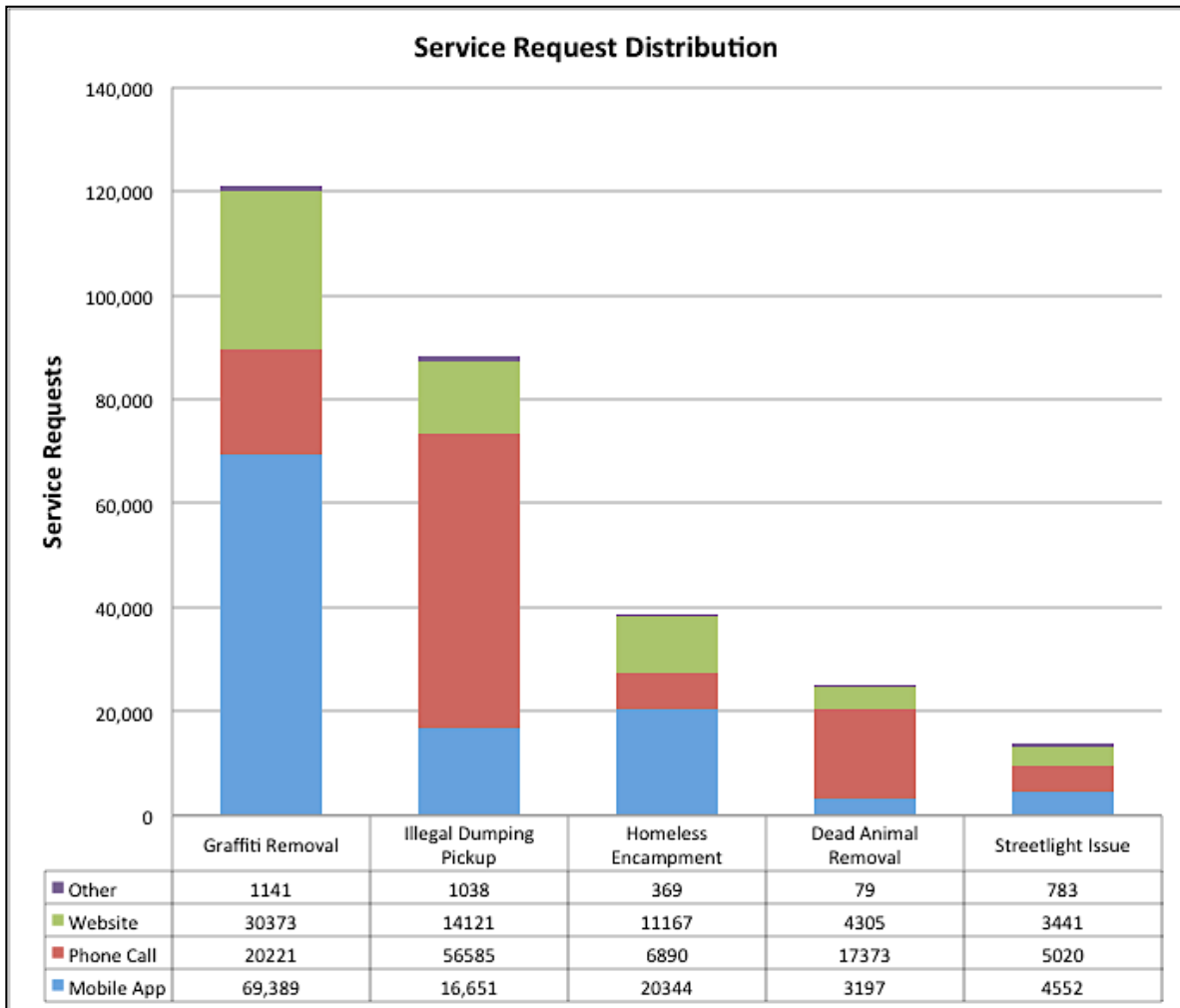|	Graffiti Removal	Illegal Dumping Pickup	Homeless Encampment	Dead Animal Removal	Streetlight Issue
Other	1141	1038	369	79	783
Website	30373	14121	11167	4305	3441
Phone Call	20221	56585	6890	17373	5020
Mobile App	69,389	16,651	20344	3197	4552

Figure 19 Chart detailing the sources of MyLA 2018 service request data and breaking down the types of the service request types analyzed in multivariate clustering

4.2. Service Request Clusters from MyLA311 Data

This multivariate clustering is done solely on the basis of 311 service request frequency from within each census tract–sociodemographic values from census data are not taken into account. Figure 20 is the visualization of service request clustering frequency regarding graffiti removal, homeless encampments, illegal dumping, broken streetlights and dead animal removal. The radar chart in Figure 22 details the average service request types in the resulting clusters based upon the z-score.
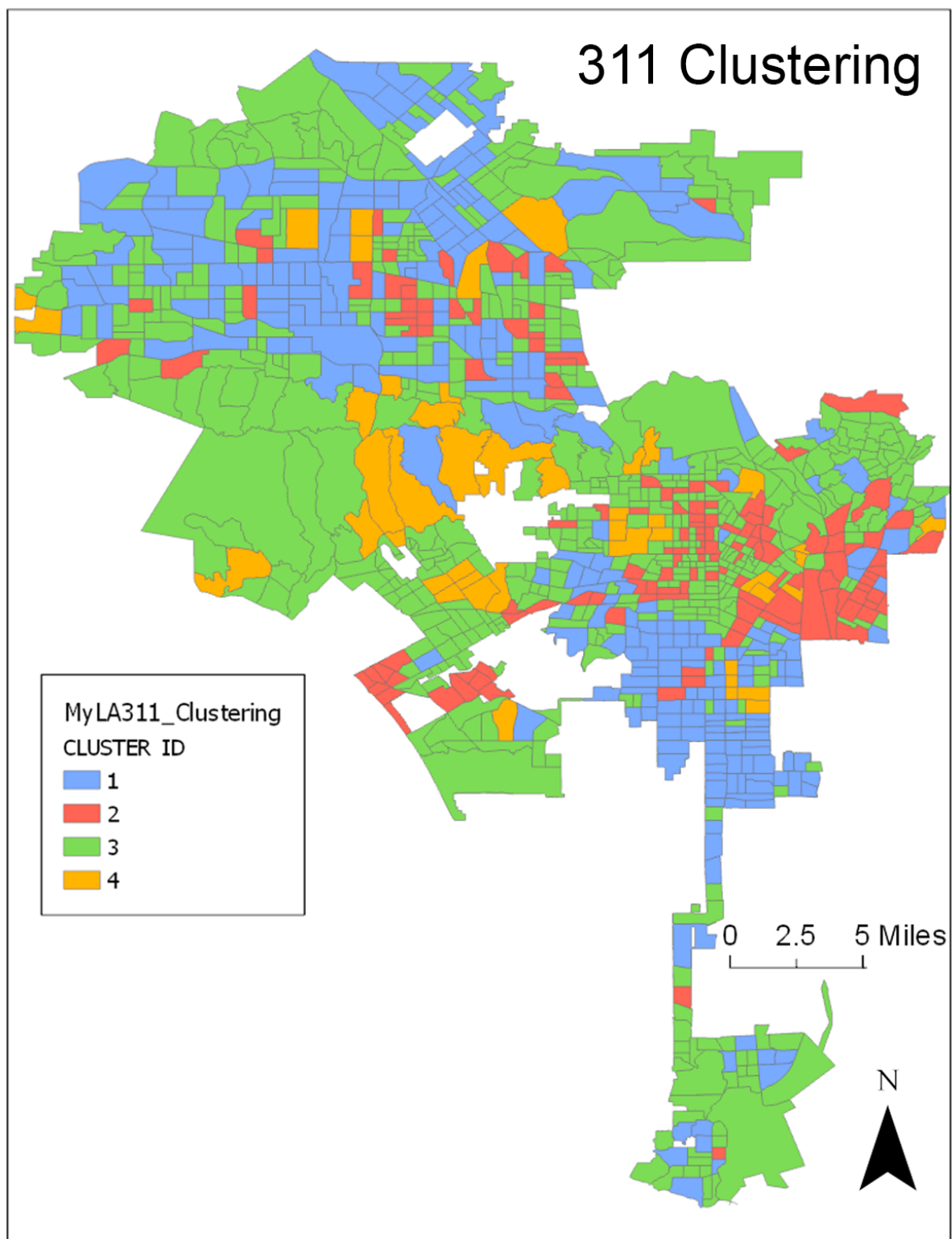
Figure 20 Multivariate Clustering of 311 Data

Resulting charts from the analysis provided depth to the distribution of the MyLA311 data in regard to the clusters. The standardization of the frequency of the different service request types discussed in Chapter 3 is necessary given the disparities within amounts reported, as exemplified earlier in the chart from Figure 19. Visually and statistically identifying clear differences in signatures of the clusters created from service requests are possible given the five variables present in the multivariate clustering analysis. After the multivariate analysis of 311 service requests, Figure 20 displays four distinct groups with differing signatures. Resulting tables and analysis of the data in Excel produced Figure 21 and Figure 22. Figure 21 details the average service requests per cluster. Figure 22 serves the purpose of visually comparing the z-scores of the different cluster signatures.
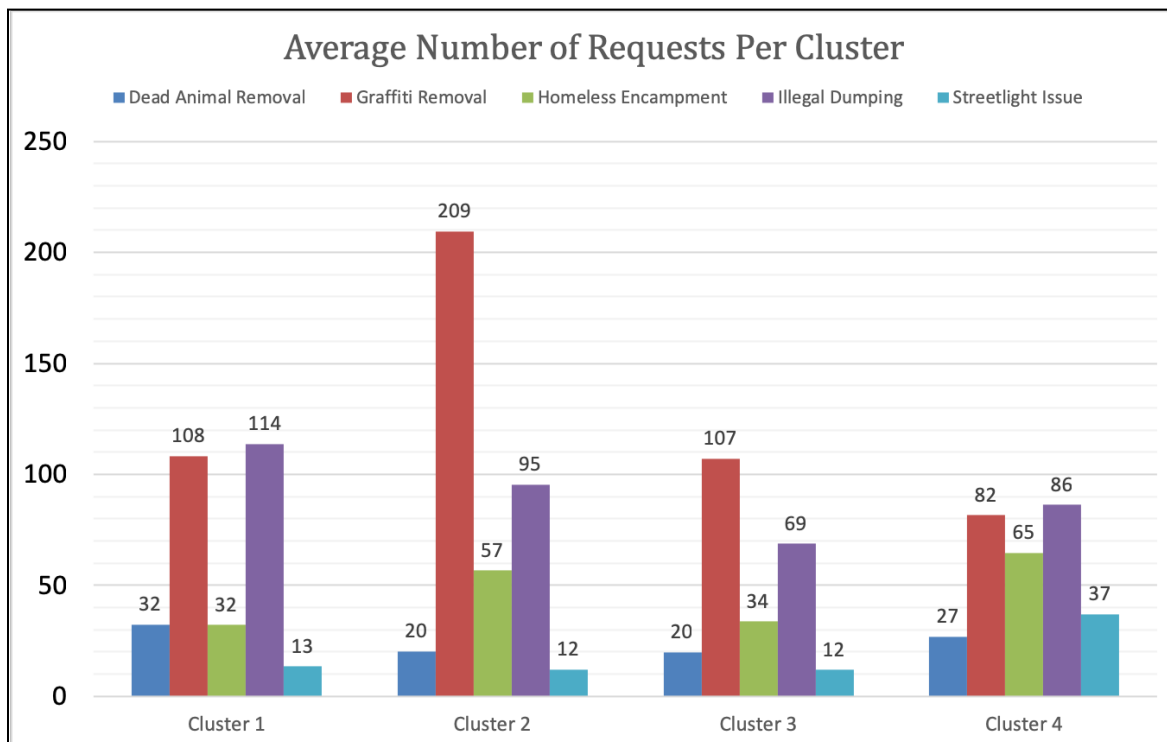

Figure 21 Average Service Requests per 311 Cluster

Cluster 1 had a high concentration of dead animal removals, followed by illegal dumping. Cluster 2 had a high concentration of graffiti removal requests and homeless encampment

reports. Cluster 3 had no service request with a high concentration and was generally characterized by a lack of reporting any type of service request. Finally, Cluster 4 had a high concentration of reporting streetlight issues.
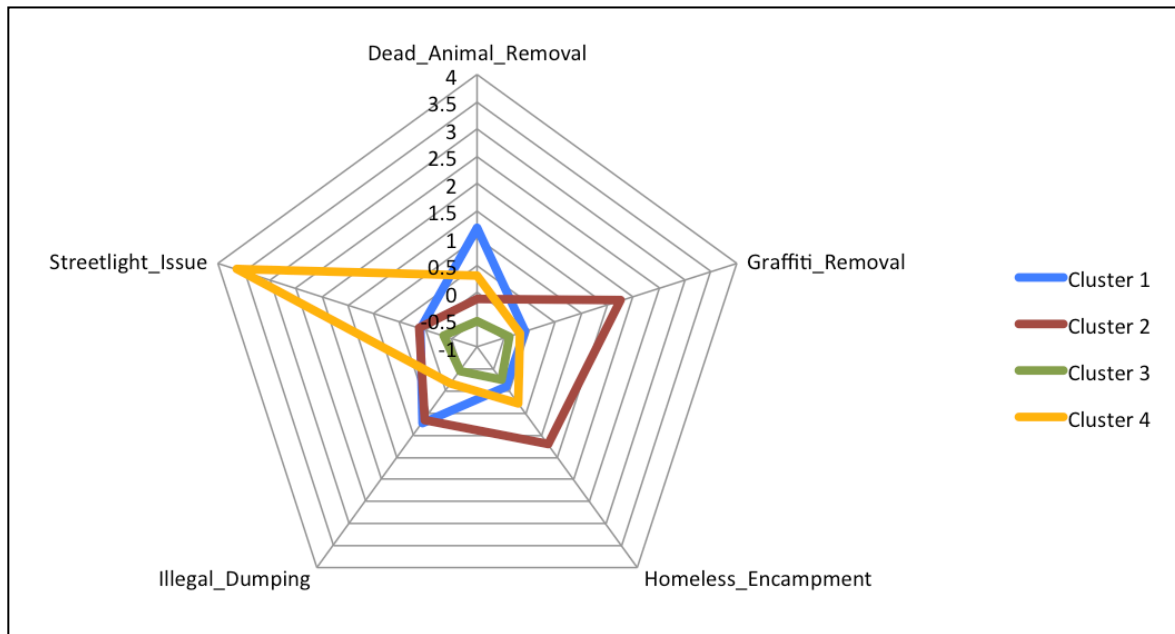
Figure 22 Radar chart of 311 service request cluster characteristics

The breakdown of the service requests is exemplified in Figure 23 and Table 7, both of which are provided as results of the multivariate clustering analysis within ArcGIS Pro. Within Table 7, the R-squared value for broken streetlights had the highest value at .625225, followed by dead animal removal at .528803, graffiti removal at .51247, illegal dumping at .31403, and finally homeless encampments at .241207. The R-squared value indicates the most effective variable for establishing clusters. It reflects the amount of variation in the original data retained after the clustering process; the larger the R-squared value is for a given variable, the better that variable is at discriminating among selected features for indication of a cluster type.
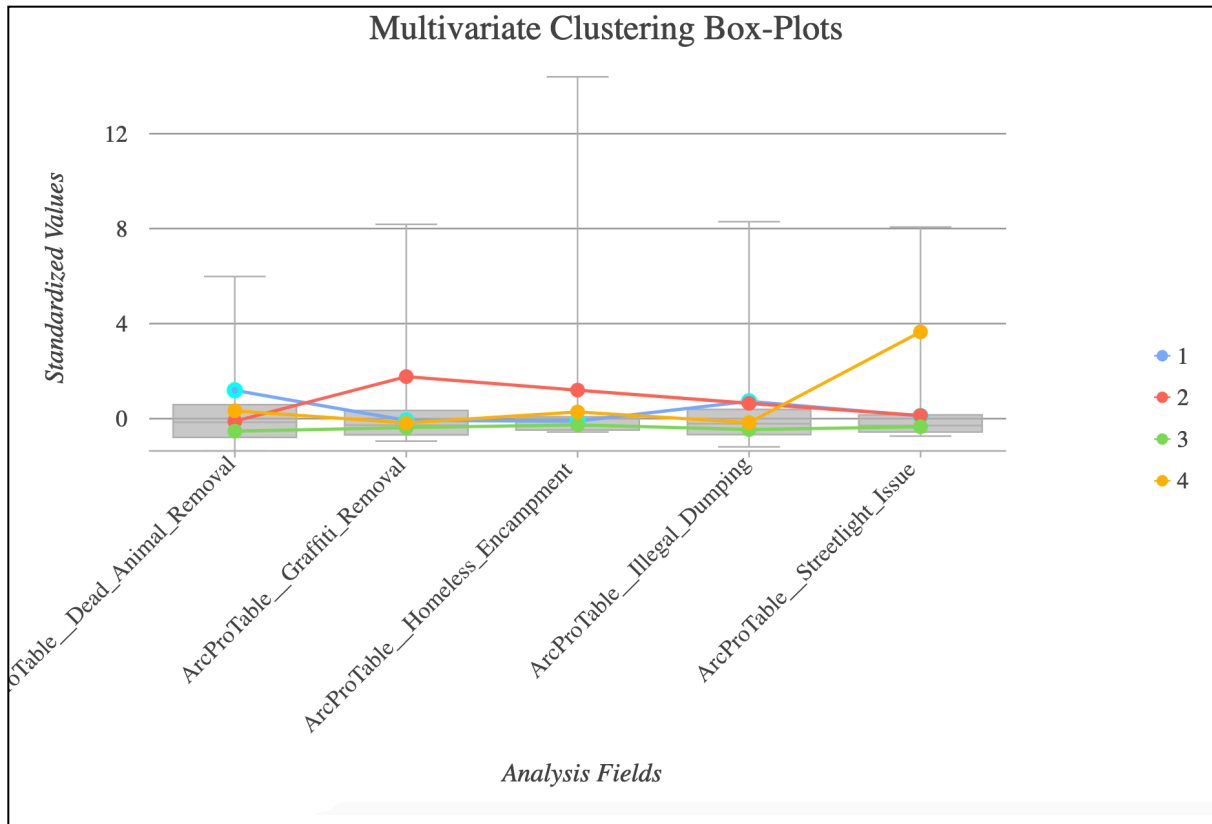
Figure 23 Multivariate Clustering Box Plots – 311 Data

Table 7 Multivariate Clustering Table Results – 311 Data

Variable	Mean	Std. Dev.	Min	Max	R Squared
Streetlight Issue	13.2597	17.9428	0	158	0.6252
Dead Animal Removal	23.7347	17.4104	0	128	0.5288
Graffiti Removal	118.2351	124.4432	0	1137	0.5125
Illegal Dumping	85.3352	71.4499	0	678	0.3140
Homeless Encampment	37.0038	65.3535	0	978	0.2412

4.2.1. 311 Clusters and Resulting Neighborhoods

Multivariate cluster analysis of 311 service requests in census tracts produced the resulting clusters and signatures. Communities in Cluster 1 seemed to be most concentrated within South Los Angeles and the Valley. Communities in Cluster 2 were generally located around the Downtown, East Los Angeles area and Venice neighborhoods with some adjacent clusters located in the Valley. It is worth noting that the dark colored census tracts in Figure 18

showed the areas submitting the most 311 service requests, and these same census tracts also comprised Cluster 2. Problems in Cluster 2 communities are not simply reported, but *heavily* reported. Out of all cluster profiles, Cluster 2 exemplified the most extreme frequency of reports for graffiti removal and homeless encampments, with illegal dumping as the third highest priority. Many of the communities in that fell in Cluster 2 were experiencing homelessness as a wicked problem, heavily affecting the quality of life for residents according to service request data.

Communities in Cluster 3 generally had the lowest reporting rates out of all cluster profiles. While each service request type remained below the standard deviation for frequency of reports, the frequency for reporting homeless encampments was the highest. Generally, these areas either simply did not experience many problems, or less likely, did not report many of the problems and dealt with them in other ways. Regardless of the reasoning behind the reporting patterns of 311 data for the neighborhoods in Cluster 3, insight to how the built environment functioned was still gained from observing the signature's low service request intensity.

Finally, Cluster 4 was comprised of the fewest census tracts. This cluster profile was unexpected given its very high frequency of reporting broken streetlights, something that characterized no other cluster profile. There was no discernible pattern of the pockets of census tracts belonging to Cluster 4, save for several census tracts grouped together in the Hollywood Hills area. No census tracts fell into Cluster 4 down the Harbor Freeway and in the San Pedro area.

4.3. Sociodemographic Attributes of Resulting 311 Clusters

The clusters created from the multivariate clustering of MyLA311 data provided insight regarding the sociodemographic features of communities that submitted service requests. There

were some surprises found, mainly regarding the spatial distribution of 311 data clusters. Since this study only looked at a portion of the possible 311 service requests from 2018, performing this analysis with more types of service requests would provide additional insight to the problems experienced by different communities in Los Angeles. Ultimately, the information revealed by the sociodemographic clustering must be taken in conjunction with the information revealed about a census tract's 311 urban signature to provide the most detailed information about a location at the local level.

4.3.1. Sociodemographic Cluster Characteristics

In order to define the urban signature of the resulting 311 clusters, the sociodemographic data from census tracts must be taken into account. General insight was provided from the maps procured for the study area in Chapter 1 from the same data. The multivariate clustering of MyLA311 data proves beneficial in visualizing and determining the multiple sociodemographic characteristics and service requests that make up an urban signature, as exemplified by the radar chart in Figure 24 and corresponding z-score values in Table 8.
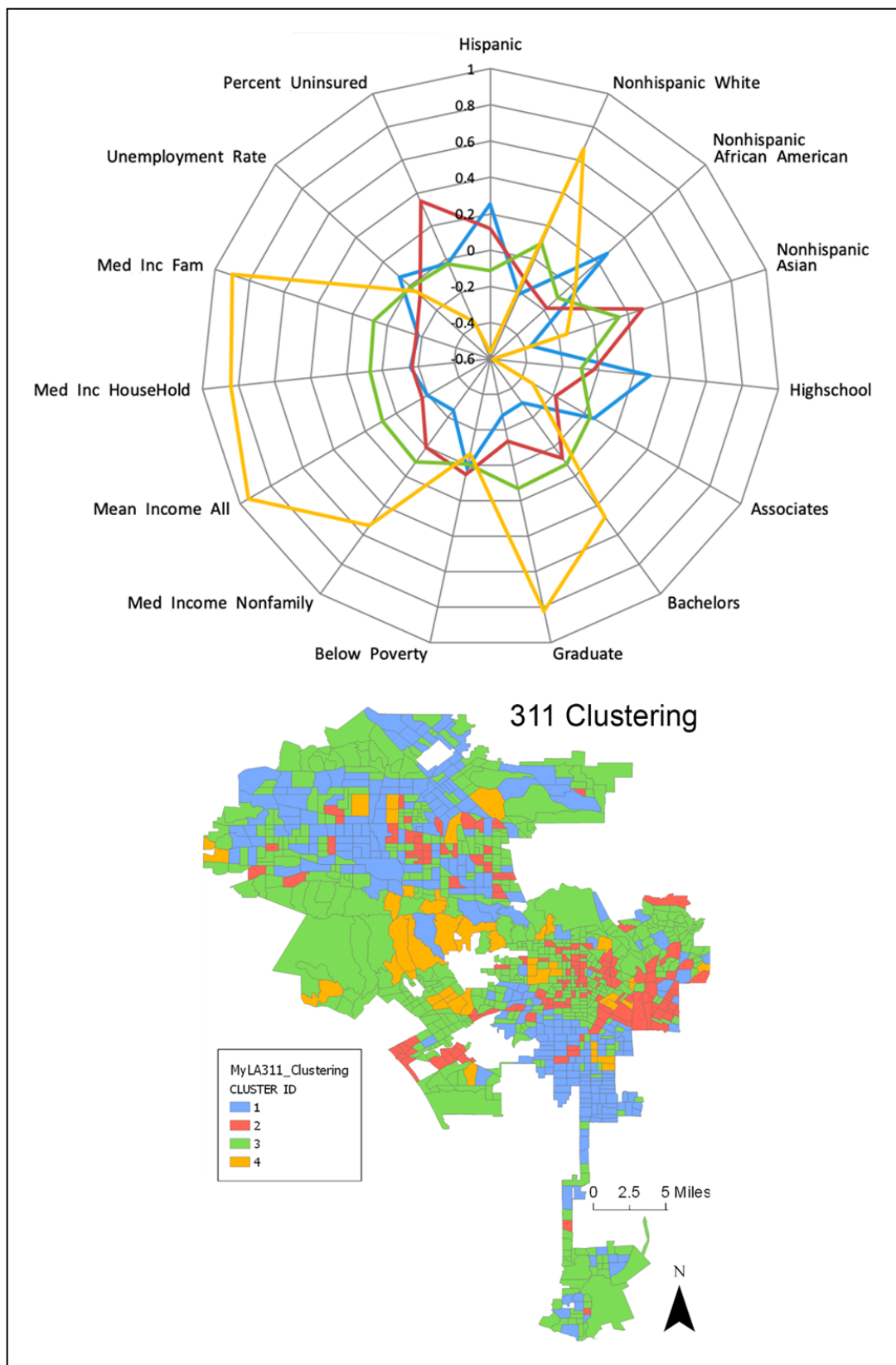
Figure 24 Radar Chart of Sociodemographic data from 311 Clustering

Table 8 Z-Scores of Cluster Characteristics

Cluster	Hispanic	NonHispanic White	NonHispanic African American	NonHispanic Asian	Highschool	Associates	Bachelors	Graduate	Below Poverty	Med Income Nonfamily	Mean Income All	Med Inc HouseHold	Med Inc Fam	Unemployment Rate	Percent Uninsured
1	0.2513	-0.2099	0.2659	-0.3651	0.2885	0.0593	-0.3018	-0.2820	0.0178	-0.2493	-0.2002	-0.1548	-0.1776	0.0750	-0.0238
2	0.1176	-0.1237	-0.1807	0.2833	-0.0243	-0.1860	0.0753	-0.1345	0.0531	0.0033	-0.1656	-0.1636	-0.1685	-0.0777	0.3499
3	-0.1124	0.0992	-0.0999	0.1457	-0.0961	0.0394	0.1196	0.1291	-0.0087	0.1037	0.0881	0.0712	0.0793	-0.0093	-0.0288
4	-0.5607	0.6670	0.0397	-0.1562	-0.5852	-0.3288	0.4788	0.8232	-0.0629	0.5361	0.9435	0.8438	0.9019	-0.0454	-0.3766

The breakdown of racial composition of resulting clusters seemingly mirrors the clustering of groups identified by Wang et al. (2017) in the analysis. Cluster 1 had a high disposition of Hispanic and Non-Hispanic African American residents, with barely any Non-Hispanic white residents. Cluster 2 had mainly Non-Hispanic Asian residents, closely followed by Hispanic residents. Cluster 3 had the most equal distribution of requests from residents of different races and ethnicities. Cluster 4 was comprised of mainly Non-Hispanic white communities, with rarely any Hispanic communities.

The sociodemographic information regarding the highest level of education attained relates to the financial information, namely because of the correlation between education and income dependent on employment. Cluster 1 had the highest amount of those with only a high school education. While Cluster 1 also had the highest number of individuals with an associate degree, it also had the lowest response rate regarding a bachelor's degree or any graduate education. Cluster 2 followed Cluster 1 with second highest attainment of only a high school education, and had fewer associate degrees, yet more bachelor's and graduate education. Cluster 3 was the second most overall educated group, falling only behind Cluster 4. Able to separate itself from the other clusters, Cluster 4 had noticeably more four-year college and graduate school levels of education than all other groups. Three trends were identifiable in regard to financial situations. Cluster 1 and Cluster 2 generally had lower income, while Cluster 3 was

regarded as the "middle class" of the four clusters. Communities belonging to Cluster 4 were noticeably wealthier in all economic categories.

Percent uninsured, unemployment rate and below poverty were the remaining variables without different types. Overall, there was no major difference in the percentage below poverty and unemployment rate determined through the MyLA311 reporting clusters. The most noticeable difference is visible in the percentage of uninsured individuals. Cluster 2 had the highest rate of uninsured individuals, while Cluster 1 and Cluster 3 were generally around the same percentage of uninsured respondents. Cluster 4 had the lowest percentage of individuals without insurance; this is linked to the relation of income to insurance, discussed in Chapter 1.

Chapter 5 Conclusions and Discussion

This chapter concludes this thesis and provides a discussion about the objectives of the research through a discussion of the results, their significance, limitations and future research. This research analyzed the results of multivariate clustering of MyLA311 data for service requests and the sociodemographic characteristics present in the resulting clusters. Resulting clusters' sociodemographic variables were then analyzed through correlation statistics.

5.1. Significance of Findings

The analysis confirms that MyLA311, Los Angeles's 311 service request data, can be used to identify urban signatures within City of Los Angeles. Ultimately, this data reveals geospatial patterns relating to citizens who made service requests and their built environment. The findings and use of the MyLA311 data as VGI for geospatial analysis reveals new information at the local scale. It is important to utilize the separate-and-combine approach for understanding the features present in all of the data due to the variant patterns in subsets of the census tracts (Minkoff 2015). This approach addresses the variations in patterns over each subset and works best with the subsets' respective analysis independent of each other.

This research also contributes to the field of analysis using VGI from city service requests. The analysis provides another example where 311 service request data can be used alongside sociodemographic data for a multivariate cluster analysis revealing spatial patterns at a granular scale within an urban built environment specific to Los Angeles in comparison with the analysis of Wang et al. (2017). This thesis provides an alternative methodology for the process through Esri and Excel software instead of through R. The clusters formed appeared to be successful in identifying neighborhood level clusters of differences in reporting frequency and of sociodemographic characteristics.

5.1.1. Proof of Concept

The results of this study with MyLA311 data are meaningful for identification of urban signatures at the local scale within a city. This analysis was not intended to produce a model of 311 service requests and the community profiles but to provide a proof of concept for using MyLA311 service request data in analysis of the 311 service requests in describing urban activity and characteristics, allowing the data to be taken into consideration for future models. A stakeholder could examine the areas of interest in the resulting maps to gain a better understanding of the human context alongside concerns presented through 311 data. The analysis benefits using MyLA311 data is the capability of identifying the correlations between service requests and sociodemographic characteristics at a finer scale.

5.1.2. Implications for Los Angeles

The results of this analysis provide stakeholders with clearer pictures about the constituents and needs of an area. For example, a city council district can observe the different clusters in neighborhoods and the differences present within those boundaries. Given the granular level of a census tract, the patterns observed could show a concentrated image of community makeup and offer insight as to which problems are being reported the most by citizens. In theory, a council district's local governing body can then tackle those problems at the source and make a concentrated effort to alleviate them at the root cause. While this research solely focused on MyLA311 data collected in 2018, the data is available from 2016 to the present. This larger temporal scope of data can be analyzed to detect changes in the needs of a community and its constituents over time.

5.2. Study Limitations and Future Research

While this study successfully used MyLA311 data to identify urban signatures within City of Los Angeles, there were several observed limitations in this study. This section addresses these limitations and what work can be done in the future to improve upon these results.

5.2.1. Limitations

In creating a methodology for the multivariate cluster analysis in ArcGIS, there were minor problems relating to sliver polygons and ensuring only relevant information is taken into account. The main limitation of this study relates to the source of the data in how MyLA311 data includes only *reported* service requests, not *every* occurrence in the communities within City of Los Angeles. Thus, the multivariate clustering results were only based upon reported problems. As a result, the frequencies may not be 100% representative for the area. As a result, it must be accepted that the data serve as a proxy for the actual occurrences of service requests in the area. While this does not detract from the success of the methodology towards defining an urban signature, it can influence the clustering results.

An additional caveat stemming from the 311 service requests is the reporting patterns of individuals. Some individuals may be considered power users of reporting service requests and report at a high rate, whereas other individuals may rarely report or not at all. Kontokosta (2017) claims that disparity in reporting stems from a variety of problems, but mainly a lack of interest of interaction with the government, varying levels of trust regarding government agencies, or a lack of accessibility through technology

Moreover, only few types of service requests were available through the LA Open Data Portal, while many exist within the restricted backend of MyLA311's server. This extends to the politics of only specific MyLA311 service request type data's availability through the Open Data

Portal. While the city may tout its commitment to transparency and publicly available data, there is a need for more demonstrations of this commitment. In relation to this analysis, incorporating more service request types would assist in creating signature clusters based off of additional, physically prominent service request types.

Concurrently, the bias of citizen reporting can prejudice the frequency of specific reports, as the data for reports are based on what an individual has a propensity to report: for example, Individual A may report all problems that they see, whereas Individual B has a penchant for only reporting homeless encampments. While VGI through civic service requests is an asset for transforming civic activism into big data, it often begs the question regarding quality of data. The publicly available MyLA311 data does not include if a resident who lives near the service request location submitted it, or if it was submitted by someone traveling through the area, perhaps for work or leisure, who felt compelled enough to submit a service report. While it is impossible to arrive at a conclusion for who exactly is submitting the service request, the locational component of the service request provides insight into where people feel the need to fix physical problems.

At the time of publishing, the full list of 311 service request types is not available on MyLA311. As the city undergoes more advancements in technological capabilities and data management, it is expected that more data will be made available to the public. This in turn will lead to more opportunities for inclusion within a spatial analysis using City of Los Angeles proprietary data.

5.2.2. Future Research

In the future, it would be beneficial for the same analysis to be done with additional service requests in order to provide a more detailed picture of what problems citizens are

experiencing, as well as compare clustering from service requests in previous years. Given the vast options for service requests, if *all* types of service requests were available to the public on the Open Data LA website, it would then make sense to make smaller groups for multivariate clustering of service requests that are thematically similar, such as only refuse or investigation service requests listed in Table 1, so as not to provide an overload of information ultimately not revealing granular details in the cluster results. It would be also beneficial to analyze multiple years of 311 data to observe the changes over time within the census tracts.

This thesis serves as a baseline to provide a methodology to observe and analyze clusters of MyLA311 service requests. Future research direction can follow the existing research from other cities that uses the same type of 311 data to perform predictive analytics within a city. Lu (2016) utilizes a comparison of relative request share for each channel, a spatial hot spot analysis, and a regression model to compare 311 service type usages alongside sociodemographic variables. Performing these types of analysis with the 311 data for Los Angeles would prove beneficial to multiple stakeholders and offer additional proof for the modernization of data-driven city services and technology. Zha (2014) provides a methodology for using current 311 service request patterns for a prediction of future 311 service requests, providing a city with insight regarding budgeting of resources for future services. Should this analysis be performed with the MyLA311 Los Angeles data, it could provide a basis for preemptive problem solving in the city to saving time, money and other resources.

5.3. Conclusion

In conclusion, multivariate clustering of Los Angeles MyLA311 service requests produces clusters reminiscent of an urban signature to provide more insight as to the needs and problems facing certain communities in Los Angeles. When coupled with the sociodemographic

attributes in the census data, such service request clusters provided a more detailed characterization of the local communities. Identifying which census tracts or neighborhoods experience a low total amount of service requests could be indicative of a need to advertise the services more to the community, or indicative of less existing problems available to report. This sort of spatial analysis provides new context and understanding of an area and how the community interacts with the built environment based on observed knowledge of the 311 service requests at the local scale and the correlating neighborhood characteristics.

The analysis provides quantitative understanding of the built environment and makeup of the city through data relating to the communities inhabiting a space. By utilizing the citizen produced data of MyLA311, urban planners and policy-makers have the opportunity to approach civic issues with renewed insight and data to support any proposals or plans. Replication of this research within Los Angeles should address all available service requests rather than the five determined in this analysis as a proof of concept. Finally, by providing more insight to the makeup of communities and better understanding the problems residents face, Los Angeles policy makers and local government institutions can utilize data-driven analytics to best address problems while taking the human component into consideration.

References

Allen, James P. and Eugene Turner. 2002. *Changing Faces, Changing Places: Mapping Southern Californians.* California State University, Northridge. Center for Geographical Studies

Antoniou, V., & Skopeliti, A. 2015. "Measures and Indicators of VGI Quality: An Overview." *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2, 345.

Bandura, Albert. 2000. "Exercise of Human Agency Through Collective Efficacy." *Current Directions in Psychological Science* 9(3), 75-78.

Bégin, D., Devillers, R., & Roche, S. 2013. "Assessing Volunteered Geographic Information (VGI) Quality Based on Contributors' Mapping Behaviours." Paper presented at the Proceedings of the 8th International Symposium on Spatial Data Quality ISSDQ, 149-154.

Bengfort, Jacquelyn. 2019. "Cities Move 311 Systems to the Cloud and Improve Citizen Services." State Tech Magazine., last modified January 11, accessed January 15, 2019, https://statetechmagazine.com/article/2019/01/cities-move-311-systems-cloud-and-improve-citizen-services.

Bettencourt, Luís MA, José Lobo, Deborah Strumsky, and Geoffrey B. West. 2010. "Urban Scaling and its Deviations: Revealing the Structure of Wealth, Innovation and Crime Across Cities." *PloS One* 5 (11): e13541.

Brovelli, Maria Antonia, Marco Minghini, and Giorgio Zamboni. 2015. "Public Participation GIS: A FOSS Architecture Enabling Field-data Collection." *International Journal of Digital Earth* 8(5), 345-363.

Brown, Gregory, Maggi Kelly, and Debra Whitall. 2014. "Which 'public'? Sampling Effects in Public Participation GIS (PPGIS) and Volunteered Geographic Information (VGI) Systems for Public Lands Management." *Journal of Environmental Planning and Management* 57(2), 190-214.

Bundorf, M. Kate and Mark V. Pauly. 2006. "Is Health Insurance Affordable for the Uninsured?" *Journal of Health Economics* 25 (4): 650-673.

Chamard, Sharon. 2010. *Homeless Encampments*. US Department of Justice, Office of Community Oriented Policing Services Washington, DC.

Collins, Brady and Anastasia Loukaitou-Sideris. 2016. "Skid Row, Gallery Row and the Space in between: Cultural Revitalisation and its Impacts on Two Los Angeles Neighbourhoods." *Town Planning Review* 87 (4): 401-427.

Composto, Sarah, Jens Ingensand, Marion Nappez, Olivier Ertz, Daniel Rappo, Rémi Bovard, Ivo Widmer, and Stéphane Joost. 2016. "How to Recruit and Motivate Users to Utilize

VGI-systems." Paper, 19th AGILE Conference on Geographic Information Science, Helsinki, Finland.

Currie, Morgan. 2016. "The Data-Fication of Openness the Practices and Policies of Open Government Data in Los Angeles." PhD diss., UCLA.

Dear, Michael. 1992. "Understanding and Overcoming the NIMBY Syndrome." *Journal of the American Planning Association* 58 (3): 288-300.

Early, Dirk W. 2005. "An Empirical Investigation of the Determinants of Street Homelessness." *Journal of Housing Economics* 14(1), 27-47.

Esri. 2016. "Los Angeles Launched GeoHub." Accessed February 2, 2018. Retrieved from https://www.esri.com/esri-news/arcnews/spring16articles/los-angeles-launched-geohub

Esri. 2018. "How Multivariate Clustering works." Accessed December 21, 2018. Retrieved from https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/how-multivariate-clustering-works.htm

Ester, Martin, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." *Kdd* 96 (34), 226-231.

Federal Communications Commission. 1997. *Use of N11 Codes and Other Abbreviated Dialing Arrangements.* Washington, DC. Accessed on March 14,2019. https://transition.fcc.gov/Bureaus/Common_Carrier/Orders/1997/fcc97051.pdf

Flanagin, Andrew J. and Miriam J Metzger. 2008. "The Credibility of Volunteered Geographic Information." *GeoJournal* 72(3-4), 137-148.

Fonte, C. C., Lucy Bastin, G. Foody, T. Kellenberger, N. Kerle, P. Mooney, A-M Olteanu-Raimond, and L. See. 2015. "VGI Quality Control." *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2*, 317-324.

Gehl, J. 2010. *Cities for People*. Washington, DC: Island Press.

Goodchild, Michael F. 2007. "Citizens as Sensors: The World of Volunteered Geography." *GeoJournal* 69(4), 211-221.

Goodchild, Michael F. and Linna Li. 2012. "Assuring the Quality of Volunteered Geographic Information." *Spatial Statistics* 1, 110-120.

Handy, Susan L., Marlon G. Boarnet, Reid Ewing, and Richard E. Killingsworth. 2002. "How the Built Environment Affects Physical Activity: Views from Urban Planning." *American Journal of Preventive Medicine,* 23(2), 64-73.

Hood, Christopher. 1995. "The 'new public management' in the 1980s: Variations on a theme." *Accounting, Organizations and Society*, 20(2-3), 93-109.

Johnson, Steven. 2010. "What a Hundred Million Calls to 311 Reveal About New York." *Wired Magazine.*

Koenig, John F. 2001. "Spaces of Denial and Denial of Place: The Architectural Geography of Homelessness in Victoria, BC." Master's thesis, University of Victoria.

Kontokosta, Constantine E. and Awais Malik. 2018. "The Resilience to Emergencies and Disasters Index: Applying Big Data to Benchmark and Validate Neighborhood Resilience Capacity." *Sustainable Cities and Society, 36*, 272-285.

Kontokosta, Constantine, Boyeong Hong, and Kristi Korsberg. 2017. "Equity in 311 Reporting: Understanding Socio-Spatial Differentials in the Propensity to Complain." Presented at the Bloomberg Data For Good Exchange.

Los Angeles City Controller. 2019. Los Angeles City Controller Control Panel. Accessed December 12, 2018. https://controllerdata.lacity.org/

Los Angeles County Department of Public Health. 2017. *Recent Trends in Health Insurance Coverage in Los Angeles County*. Office of Health Assessment and Epidemiology. http://publichealth.lacounty.gov/docs/LaHealth_RecentTrendsInHealthInsuranceCoverage_yr2017.pdf

Lu, Qing and Peter A. Johnson. 2016. "Characterizing New Channels of Communication: A Case Study of Municipal 311 Requests in Edmonton, Canada." *Urban Planning, 1*(2), 18-31.

Lulka, David. 2013. "The Posthuman City: San Diego's Dead Animal Removal Program." *Urban Geography* 34(8), 1119-1143.

Matsumoto, Shigeru and Kenji Takeuchi. 2011. "The Effect of Community Characteristics on the Frequency of Illegal Dumping." *Environmental Economics and Policy Studies* 13(3), 177-193.

McCormack, Gavin R., Melanie Rock, Ann M. Toohey, and Danica Hignell. 2010. "Characteristics of Urban Parks Associated with Park use and Physical Activity: A Review of Qualitative Research." *Health & Place* 16(4), 712-726.

Megler, Veronika, David Banis, and Heejun Chang. 2014. "Spatial Analysis of Graffiti in San Francisco." *Applied Geography* 54, 63-73.

Minkoff, Scott L. 2016. "NYC 311: A Tract-Level Analysis of Citizen–Government Contacting in New York City." *Urban Affairs Review* 52(2), 211-246.

Moon, M. Jae. 2002. "The Evolution of E-government among Municipalities: Rhetoric Or Reality?" *Public Administration Review* 62 (4): 424-433.

Morenoff, Jeffrey D., Robert J. Sampson, and Stephen W. Raudenbush. 2001. "Neighborhood Inequality, Collective Efficacy, and the Spatial Dynamics of Urban Violence." *Criminology* 39(3), 517-558.

Mullen, William F., Steven P. Jackson, Arie Croitoru, Andrew Crooks, Anthony Stefanidis, and Peggy Agouris. 2015. "Assessing the Impact of Demographic Characteristics on Spatial Error in Volunteered Geographic Information Features." *GeoJournal*, 80 (4), 587–605.

Nam, Taewoo. 2012. "Modeling Municipal Service Integration: A Comparative Case Study of New York and Philadelphia 311 Systems." Dissertation, University at Albany, State University of New York.

Nam, Taewoo and Theresa Pardo. 2014. "The Changing Face of a City Government: A Case Study of Philly311." *Government Information Quarterly* 31. 10.1016/j.giq.2014.01.002.

Neis, Pascal, Dennis Zielstra, and Alexander Zipf. 2012. "The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011." *Future Internet* 4: 1– 21.

Novak, Kurt. 1995. "Mobile Mapping Technology for GIS Data Collection." *Photogrammetric Engineering and Remote Sensing* 61 (5): 493-501.

O'Brien, Daniel Tumminelli, Dietmar Offenhuber, Jessica Baldwin-Philippi, Melissa Sands, and Eric Gordon. 2017. "Uncharted Territoriality in Coproduction: The Motivations for 311 Reporting." *Journal of Public Administration Research and Theory* 27(2), 320-335.

O'Brien, Daniel Tumminelli. 2016. "Using Small Data to Interpret Big Data: 311 Reports as Individual Contributions to Informal Social Control in Urban Neighborhoods." *Social Science Research 59*: 83-96.

Office of Los Angeles Mayor Eric Garcetti. 2013. *Executive Directive No 3.* Retrieved January 30, 2019, from https://www.lamayor.org/sites/g/files/wph446/f/page/file/Executive-Directive-3-Open-Data.pdf?1426620075

Pain, Rachel, Robert MacFarlane, Keith Turner, and Sally Gill. 2006. "'When, Where, if, and but': Qualifying GIS and the Effect of Streetlighting on Crime and Fear." *Environment and Planning* A 38(11), 2055-2074.

Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." *Science* 277(5328), 918-924.

Sayad, Saed. 2019. "K-Means Clustering." Data Mining Map. Accessed March 18. https://www.saedsayad.com/data_mining_map.htm.

Schwester, Richard W., Tony Carrizales, and Marc Holzer. 2009. "An Examination of the Municipal 311 System." *International Journal of Organization Theory & Behavior* 12 (2): 218-236.

Scruggs, Carl. 2013. *Best practices in open data initiatives*. Office of Legislative Oversight.

Sieber, Renee E., and Peter A. Johnson. 2015. "Civic Open Data at a Crossroads: Dominant Models and Current Challenges." *Government Information Quarterly* 32(3), 308-315.

Tauberer, Joshua. 2014. *Open Government Data: The Book*. 2nd ed.

Tobler, Waldo R. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 46, no. sup1: 234-240.

Toregas, Costis. 2001. "The Politics of E-Gov: The Upcoming Struggle for Redefining Civic Engagement." *National Civic Review* 90 (3): 235-240.

Trevino, Andrea. 2016. "Introduction to K-Means Clustering." Oracle DataScience. Retrieved March 30, 2019, from https://www.datascience.com/blog/k-means-clustering.

Trinacria, Joe. 2018. "Pew Report: Philly Remains the Poorest of America's 10 Largest Cities." *Philadelphia Magazine*. April 4. https://www.phillymag.com/news/2018/04/06/pew-report-poverty/.

U.S. Census Bureau 2018. American Community Survey, 2017 American Community Survey 5-Year Estimates using American FactFinder. http://factfinder2.census.gov

Wang, Lingjing, Cheng Qian, Philipp Kats, Constantine Kontokosta, and Stanislav Sobolevsky. 2017. "Structure of 311 Service Requests as a Signature of Urban Location." *PloS One* 12(10): e0186314.

Wessel, Nate. 2016. "Assessing Predictors of Citizen Reports of Dead Animals in Cincinnati." Paper, University of Toronto.

Whitaker, Gordon P. 1980. "Coproduction: Citizen Participation in Service Delivery." *Public Administration Review* 240-246.

White, Ariel and Kris-Stella Trump. 2018. "The Promises and Pitfalls of 311 Data." *Urban Affairs Review* 54(4), 794-823.

Williams, C. H. 2007. "*The Built Environment and Physical Activity: What is the Relationship?*" Report, University of Indianapolis.

Wilson, James Q. and George L. Kelling. 1982. "Broken windows." *Atlantic Monthly* 249(3), 29-38.

Zha, Yilong Frank and Manuela Veloso. 2014. "Profiling and Prediction of Non-Emergency Calls in NYC." Semantic Cities: Beyond Open Data to Models, Standards and Reasoning: In Workshops at the AAAI-14 Conference on Artificial Intelligence.