Enriching the Demographic Survey Sampling for the Los Angeles County Annual Homeless Count with Spatial Statistics

by

Katrina M. Kaiser

A Thesis Presented to the FACULTY OF THE USC DORNSIFE COLLEGE OF LETTERS, ARTS AND SCIENCES UNIVERSITY OF SOUTHERN CALIFORNIA In Partial Fulfillment of the Requirements for the Degree MASTER OF SCIENCE GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY

August 2021

Copyright © 2021

Katrina M. Kaiser

To my friends and neighbors who have experienced homelessness

Acknowledgements

First, I am grateful to my advisor Dr. Steven Fleming, my committee members Dr. An-Min Wu and Dr. Elisabeth Sedano, and my initial thesis instructor Dr. Jennifer Bernstein. I am also grateful to the rest of my professors at USC who taught the skills needed to put together this thesis, particularly Dr. Laura Loyola who oversaw the development of my initial prospectus in Fall 2019.

I am also very thankful for the homelessness research team at USC, particularly: Patricia St. Clair, who championed this project on my behalf for LAHSA and has provided much valuable insight on the history of USC's contract with LAHSA and how the count has changed over time; Laura Gascue, my work supervisor and the statistician who formerly performed the survey sampling; Stephanie Kwack, a senior research programmer and my cubicle neighbor who welcomed me inviting myself into the survey project; Gerry Young, a research programmer who assembled the harmonized analytical file for the homelessness data from multiple years and formats of previous work; Dr. Randall Kuhn of UCLA, a statistician who helped me refine the parameters of the statistical analyses of this thesis. I also thank the rest of the team, including director Dr. Benjamin Henwood and others, and LAHSA data management staff for the opportunity to collaborate on ways to innovate the Homeless Count.

Finally, thank you to my partner, close friends, and family for your support and encouragement during the past year and a half of thesis work. I appreciate my longtime roommate Farah Hussain, who provided insight from her knowledge as a social worker and therapist with clients experiencing homelessness in LA County.

iii

Table of Contents

Dedication	ii
Acknowledgements	iii
List of Tables	∕ii
List of Figures	iii
List of Abbreviations	X
Abstract	xi
Chapter 1 Introduction	1
1.1. Study Area	.2
1.2. Background Context	.4
1.2.1. History of LAHSA and HUD-mandated Point-in-Time counts of homelessness	.4
1.2.2. Current Demographic Survey Methodology	.7
1.3. Problem Statement	10
1.4. Impact Statement	13
1.5. Overview of Analysis	14
Chapter 2 Related Work	16
2.1. Literature Review	16
2.1.1. Spatial Distribution of Homelessness	16
2.1.2. Counting People Experiencing Homelessness	18
2.1.3. Other Spatial Cluster Detection Literature on Aggregated Counts	19
2.1.4. VGI and Homelessness	20
2.1.5. Perception and Subjectivity around Urban Homelessness	21
2.1.6. Sampling for Surveys of People Experiencing Homelessness	22
2.1.7. Other Spatial Sampling Literature	23
2.1.8. Area Characteristics Correlating with Homelessness	25

2.1.9. Other Area Deprivation Classification Literature	28
Chapter 3 Methods	29
3.1. Methods Overview	29
3.2. Data Acquisition	30
3.3. Data Cleaning	31
3.4. Analysis	33
3.4.1. Validation of existing "hot-spot planning" process	33
3.4.2. Determination of statistically significant spatial patterns	34
3.4.3. Identification of neighborhood characteristics correlating with homelessness .	38
3.4.4. Alternative Sampling Geography	39
Chapter 4 Results	40
4.1. Validation	40
4.1.1. Analysis of Change in PIT Counts	40
4.1.2. Analysis of Change in Hot Spot Designations	46
4.1.3. Visualization of Hot Spot Designations versus PIT Counts	50
4.1.4. Comparison of Hot Spot Planning Data Sources	55
4.2. Cluster Detection	62
4.2.1. Creating Spatial Weights Matrices	62
4.2.2. Local Moran's I (Cluster and Outlier Analysis)	66
4.2.3. Getis-Ord Gi* (Hot Spot Analysis)	73
4.3. Statistical Tests Against Neighborhood Attributes	80
4.3.1. Chi-Square Tests	80
4.3.2. Correlation Coefficients	80
4.4. Alternative Geography from Neighborhood Attributes	82
Chapter 5 Discussion	87

5.1. Main Conclusions	87
5.1.1. Identifying Hot Spots	87
5.1.2. Alternative Geographies	88
5.2. Limitations	90
5.3. Feasibility of Implementing Suggestions in Sampling Workflow	90
5.3.1. Hot Spot Strata	90
5.3.2. Geographic Strata	91
5.4. Future Research Directions	92
5.4.1. Current and Potential Plans for HC 2022 and Beyond	92
5.4.2. Additional GIS Analysis Directions	94
References	
Appendix A: ACS Classification Variables	
Appendix B: SCAG General Plan Land Use Code List	101

List of Tables

Table 1: Data Sources 30
Table 2: Proportions of census tracts where unsheltered individuals or CVRTMs are always,never, or sometimes found 2018-2020, by SPA45
Table 3: Proportions of census tracts with the same or different "hot-spot" stratum designation from 2018-2020 47
Table 4: Proportion of census tracts that are always, never, or sometimes an Individual orCVRTM "hot-spot", by SPA
Table 5: Counts and proportions of census tracts designated as "hot-spots" for individuals and CVRTMs based on input data and overall, by year
Table 6: Regressions of PIT Counts on Historical and Planning Session HS Flags
Table 8: Chi-Square Test Statistics for Neighborhood Characteristics vs. PIT Counts 80
Table 9: Correlation Coefficients of Neighborhood Characteristics vs. PIT Counts

List of Figures

Figure 1: Study Area
Figure 2: Map of prime, transitional, and marginal classified census tracts (Marr et al. 2009, 311)
Figure 3: Diagram of Spatial Cluster Detection Workflow
Figure 4: Change in Individuals found during PIT Count, 2018-2019
Figure 5: Change in Individuals found during PIT Count, 2019-2020
Figure 6: Change in CVRTMs found during PIT Count, 2018-2019
Figure 7: Change in CVRTMs found during PIT Count, 2019-2020
Figure 8: Individual "hot-spot" tracts and PIT counts in SPA 6, 2019
Figure 9: Individual "hot-spot" tracts and PIT counts in SPA 6, 2020
Figure 10: CVRTM "hot-spots" and PIT counts in SPA 3, 2019 53
Figure 11: CVRTM "hot-spots" and PIT counts in SPA 3, 2020
Figure 12: Comparison of Historical and Planning Session Individual "Hot-Spots" in SPA 6, 2019
Figure 13: Comparison of Historical and Planning Session Individual "Hot-Spots" in SPA 6, 2020
Figure 14: Comparison of Historical and Planning Session CVRTM "Hot-Spots" in SPA 3, 2019
Figure 15: Comparison of Historical and Planning Session CVRTM "Hot-Spots" in SPA 3, 2020
Figure 16: Incremental Spatial Autocorrelation of Individual PIT Counts, 2018
Figure 17: Incremental Spatial Autocorrelation of Individual PIT Counts, 2019
Figure 18: Incremental Spatial Autocorrelation of Individual PIT Counts, 2020 64
Figure 19: Incremental Spatial Autocorrelation of CVRTM PIT Counts, 2018
Figure 20: Incremental Spatial Autocorrelation of CVRTM PIT Counts, 2019
Figure 21: Incremental Spatial Autocorrelation of CVRTM PIT Counts, 2020

Figure 22: Cluster and Outlier Analysis results for individual PIT counts, 2018	7
Figure 23: (Cluster and Outlier Analysis results for individual PIT counts, 2019	8
Figure 24: Cluster and Outlier Analysis results for individual PIT counts, 2020	9
Figure 25: Cluster and Outlier Analysis results for CVRTM PIT counts, 2018	0
Figure 26: Cluster and Outlier Analysis results for CVRTM PIT counts, 20197	1
Figure 27: Cluster and Outlier Analysis results for CVRTM PIT counts, 202072	2
Figure 28: Getis-Ord Gi* Hot Spots for Individual PIT Counts, 2018	4
Figure 23: Getis-Ord Gi* Hot Spots for Individual PIT Counts, 20197	5
Figure 30: Getis-Ord Gi* Hot Spots for Individual PIT Counts, 2020	6
Figure 31: Getis-Ord Gi* Hot Spots for CVRTM PIT Counts, 2018	7
Figure 32: Getis-Ord Gi* Hot Spots for CVRTM PIT Counts, 2019	8
Figure 33: Getis-Ord Gi* Hot Spots for CVRTM PIT Counts, 2020	9
Figure 34: Classified census tracts based on attributes correlated with unsheltered individuals, 2019	3
Figure 35: Cluster Factor Distributions for Individual Factors, 2019	4
Figure 36: Classified census tracts based on attributes correlated with CVRTMs, 2019	5
Figure 37: Cluster Factor Distributions for CVRTM Factors, 2019	6

List of Abbreviations

ACS	American Community Survey
CD	Council District
CoC	Continuum of Care
CT	Census Tract
CVRTM	Cars, Vans, RVs, Tents, Makeshift Shelters
GIS	Geographic information system
HMIS	Homeless Management Information Systems
HUD	Department of Housing and Urban Development
LAHSA	Los Angeles Homeless Services Authority
LA-HOP	Los Angeles Homeless Outreach Portal
MAD	Median Average Deviation
SPA	Service Planning Area
USC	University of Southern California

Abstract

Each year, the Los Angeles Homeless Services Authority (LAHSA) conducts its Homeless Count, enumerating people who are experiencing homelessness in Los Angeles County. The count includes a Demographic Survey, where surveyors interview unsheltered people in a sample of census tracts in LA County. The survey data is a key tool for informing homelessness policy. The survey's current sampling methodology does not account for the spatial relationship between tracts but approaches the distribution of homelessness in a tabular way, using a "hot-spot planning process" that relies on administrative boundaries. LAHSA also uses administrative boundaries to sample tracts rather than accounting for the characteristics of the tracts where unsheltered people tend to live. This represents an opportunity for a spatial analysis approach to homelessness data that improves the stability of results, accounts for spatial variability in the data, and characterizes areas in ways that are relevant to the lived experience of unsheltered people. This thesis studies and compares the results of LAHSA's existing "hot-spot planning process" against "hot-spot" cluster detection statistics from spatial analysis. The thesis finds that spatial cluster detection tools identify additional areas for full inclusion in a survey sample. This thesis also identifies environmental and demographic characteristics correlated with homelessness and uses them to classify alternative geographies for stratification. Robust, representative sampling for the Homeless Count Demographic Survey is important to better understanding and serving this vulnerable, growing population. A spatial approach to homelessness data is a major enhancement that is novel for Los Angeles County and for homelessness policy overall.

Chapter 1 Introduction

On any given night, the Los Angeles Homeless Services Authority (LAHSA) estimates that over 48,000 people in Los Angeles County experiencing homelessness sleep without shelter. This estimate from January 2020 represents a 9% increase from the previous year. High-quality data on the characteristics and lives of unsheltered people, who are particularly visible and vulnerable, is urgently needed to support housing and healthcare solutions for them. This thesis proposes applying a spatial analysis lens to the intermediate data assembled as part of LAHSA's Annual Homeless Count to improve the conceptualization of data on unsheltered people. As part of the count, LAHSA conducts a Demographic Survey of a sample of unsheltered adults experiencing homelessness. To identify the areas to be sampled and surveyed, LAHSA conducts what it refers to as a "hot-spot planning process" to identify census tracts containing significantly more unsheltered people than other tracts in their Service Planning Area (SPA). SPAs are large administrative units for public health service provision in LA County. However, this "hot-spot" identification does not use spatial data concepts such as density, adjacency, or spatial autocorrelation and thus does not employ the statistical rigor or underlying spatial concepts that define a hot spot analysis in the spatial analysis domain. The process also relies on administrative boundaries that are not germane to how unsheltered people decide where to live. This thesis:

- evaluates the extent to which the existing "hot-spot planning process" describes spatial patterns among unsheltered people,
- identifies persistent hot spots with many unsheltered people in the context of their neighborhoods using spatial cluster detection techniques,
- identifies neighborhood attributes that characterize the places where unsheltered people live, and

• proposes alternative geographic boundaries to use in the survey sample stratification based on these neighborhood attributes.

This research was conducted with guidance from statisticians from USC's School of Public Policy and School of Social Work who have partnered with LAHSA to tabulate data from the Count since 2017. The research methods are designed to provide decision support for future Homeless Count Demographic Survey sampling decisions.

1.1. Study Area

Below is a color map of the study area, which is comprised of the Los Angeles County Continuum of Care (CoC). A "Continuum of Care" is an integrated system of care that guides and tracks the delivery of health care and services over time. The US Department of Housing and Urban Development (HUD) administers CoCs across the nation that specialize in assisting individuals experiencing homelessness. While LAHSA is an independent, joint powers authority vested by the LA County and City of LA governments, it operates in the same CoC as the LA County Department of Public Health. This excludes Long Beach, Pasadena, and Glendale, which have their own departments of public health and form their own CoCs.



Figure 1: Study Area

This map also indicates the boundaries for the eight different SPAs, which is what LAHSA uses as the main geographic boundary for the hot spot planning process. The smallest SPA by land area is the Metro SPA, and the largest SPA is the Antelope Valley SPA. The SPAs reflect the highly varied natural and urban landscape of LA County, where the most densely populated areas are in the center and the least densely populated areas are in the deserts and mountains. The boundaries of the 2,163 census tracts used for the LA Homeless Count are also visible on this map. The census tracts vary in both land area and in total population, and the smallest tracts are in the densely populated center of the county.

The study period used for this thesis is 2018-2020. In 2018, three fewer tracts were enumerated as part of the Homeless Count; these tracts are along the boundary with the other CoCs near Long Beach and La Cañada Flintridge (near Glendale).

1.2. Background Context

1.2.1. History of LAHSA and HUD-mandated Point-in-Time counts of homelessness

LAHSA, which orchestrates the Homeless Count, has been the lead homelessness agency in the Los Angeles CoC since 1993. In addition to the Homeless Count, LAHSA's main functions are to coordinate housing and services for families and individuals experiencing homelessness in the County.

This research is designed to supplement official data collection and policymaking around homelessness, so it uses terminology defined by HUD and other policymaking bodies as follows:

An *unsheltered* homeless person resides in:

• A place not meant for human habitation, such as cars, parks, sidewalks, abandoned buildings (on the street).

A sheltered homeless person resides in:

• An emergency shelter.

• An transitional housing or supportive housing for homeless persons who originally came from the streets or emergency shelters. (U.S. Department of Housing and Urban Development 2004, 4)

This manuscript generally uses the person-first phrasing "people experiencing homelessness" rather than "homeless people", except where there are direct quotations, since that is the

preference of the researchers from USC and LAHSA¹. It is the established terminology for many homelessness service providers as well, since it encourages speakers and writers to "consciously think of homelessness as an issue rather than a condition" of an individual (Hannon 2014). This research focuses on "unsheltered people" as a subset of people experiencing homelessness, as in the HUD definition above, while qualifying additional subsets of that group as needed.

Since 2003, HUD has mandated that Continuums of Care report on the number of people experiencing homelessness every other year through a point-in-time (PIT) count (U.S. Department of Housing and Urban Development 2004). HUD publishes guidelines for two types of PIT counts of unsheltered people: services-based counts and night-of (also known as "public places" or "street") counts. Los Angeles County uses a night-of count and therefore this type is the focus of this thesis. HUD's guidelines for PIT counts of unsheltered people require that:

- CoCs report on the number of unsheltered people at least biennially,
- the count occurs during the last 10 days in January,
- either a complete census or an approved sampling and extrapolation method are used,
- only uninhabitable areas (e.g., deserts) or areas otherwise determined to have no unsheltered people (e.g., gated communities) may be excluded, and
- persons must not be double-counted, among other standards (U.S. Department of Housing and Urban Development 2014).

Since 2013, LAHSA has conducted its PIT count every year, which is beyond the HUD standard of every two years. LAHSA conducted its first count under the auspices of HUD's CoC mandate in 2005, though other counting projects occurred beforehand. LAHSA uses a complete

¹ While the phrases "unhoused" or "houseless" were also considered, these tend to be terms used in activism that is organizing for a right to housing and shelter for all people, rather than terms used by policymaking bodies or research literature.

census night-of count method to gather its raw count data, and supplements it with a samplingbased Demographic Survey in order to estimate characteristics of the unsheltered adult homeless population (USC 2020). Researchers describe the PIT methodology as follows:

The PIT Count is a visual-only tally of people experiencing unsheltered homelessness and the number of cars, vans, recreational vehicles (RVs), tents, and makeshift shelters assumed to be housing homeless individuals. LAHSA provides training to over 8,000 volunteers who canvas designated CTs [census tracts].

The HC20 PIT count was conducted by volunteers in all 2,163 CTs that make up the CoC, with special outreach teams counting LA metro stops, riverbeds, and other hard to reach places. In 2020, LAHSA also collected counts from Safe Parking program locations as well (USC 2020, 3)

Statistical contractors hired by LAHSA propose and implement refinements to the PIT count and associated surveys each year, but this thesis deploys spatial analysis techniques that the full-time researchers and staff have not had the time or resources to explore. LAHSA staff have contracted with staff from USC since the 2017 count. Previous statistical analysis contractors that have worked with LAHSA include Applied Research (LAHSA 2007) and University of North Carolina (LAHSA 2009; Agans et al. 2014). The team at USC includes a group of statisticians from the Schaeffer Center for Health Policy & Economics managed by Patricia St. Clair and a group of homelessness researchers from the Suzanne Dworak-Peck School of Social Work managed by Dr. Benjamin Henwood. This thesis author is a health policy research programmer at the Schaeffer Center working alongside colleagues who work on the Homeless Count, and most of our work experience involves analysis of tabular data in SAS, Stata, and Excel rather than spatial data in a GIS. LAHSA has a GIS systems analyst on staff who maintains geodatabases, creates reports, and produces web mapping applications for the public. Since LAHSA staff and contractors must focus on these tasks and on meeting the yearly deadlines for the Homeless Count, there is an opportunity for a student researcher to leverage data and software access to implement a spatial analysis approach to the homelessness data. This

approach is a novel contribution for LAHSA specifically and for official measures of homelessness generally.

1.2.2. Current Demographic Survey Methodology

The current Demographic Survey is a portion of LAHSA's yearly Homeless Count portfolio of studies (USC 2019; 2020). In addition to estimating characteristics of unsheltered adults, the Demographic Survey is also used to "determine the multiplier for the number of people living in the cars, vans, RVs, tents, and makeshift shelters (CVRTM) counted during the street PIT Count". The multiplier allows LAHSA to report an estimated count of all unsheltered people living in the county.

The sample stratification design from LAHSA's 2019 methodology is as follows:

A two-stage stratified random sample was used for the DS19. In stage one, all CTs [census tracts] were allocated to different strata and a random sample of CTs from each pre-defined stratum was selected. In the second stage, selected CTs were covered by the survey team. Interviewers conducted surveys with any homeless person they could locate in a CT who agreed to the survey. Respondents were assumed to be selected at random from the homeless population in the CT. CT sample selection probabilities were then used in the computation of analytic weights. (USC 2019, 3)

This thesis focuses on the stratification design. Currently, each census tract is allocated to one stratum, and the strata are defined by geography and "hot-spot status" as determined by LAHSA's "hot-spot planning process". The geographic classification system is defined by City Council Districts (CDs) inside the City of LA, and SPAs outside the City of LA. Both SPAs and CDs are used because the LA City Council has requested council district-level reports on the characteristics of unsheltered people.

Some census tracts that contain consistently high numbers of people living in homelessness are surveyed every year. This is based on prior institutional knowledge, not necessarily an empirical analysis of counts of unsheltered people over time. These include two census tracts in Skid Row and 10 census tracts in Venice are classified into their own communities and are fully covered by the sample. Thus, in the first stage of the stratification process, these census tracts are tagged for survey in addition to a random selection of other census tracts in each stratum. In 2020, tracts containing Family Solution Centers (FSCs), Youth Survey Sites, and safe parking locations were also fully included in the sample.

Within each geographic bounding area, "hot-spot status" is defined by the type of homelessness that is likely to be visible in each census tract. This information comes from two primary sources: the previous year's PIT count, and the "hot-spot planning sessions" conducted in each SPA during October and November. Using the PIT alone would bias the sample towards places where unsheltered people were found in January, so service provider representatives (including people with lived experience of homelessness) use these planning sessions to provide volunteered geographic information (VGI) on where they have seen unsheltered people living more recently. This involves placing dot stickers on a printed physical map and describing what they have seen in those locations; that information is then entered by a LAHSA GIS engineer in an ArcGIS geodatabase and aggregated to the census tract level. Planning sessions also collect information about the best time of day to find potential respondents. The historical PIT data can be thought of as nighttime location data, and the planning session data would add the daytime locations of unsheltered people. Planning session data is also supplemented with information from the homelessness management information system (HMIS) database.

A Median Average Deviation (MAD) formula is used to identify individual census tracts as "hot-spots". This is a variation on the Mean or Median Absolute Deviation statistic, which is a measure of the spread of the data around a measure of central tendency. It functions similarly to the standard deviation but is less susceptible to outliers. A tract is defined as a "hot-spot" "if the

estimated number for that CT in that category falls above the SPA median plus the average absolute median deviation for the SPA" (USC 2020, 6). This formula is applied to both the PIT count and planning sessions count, and a census tract becomes a "hot-spot" if this criterion is met based on either one of the counts. One quality issue with the "hot-spot planning session" data is that census tracts that have not been pointed out in a planning session are assigned values of zero unsheltered people, when these values should be treated as missing values for the purpose of finding median values. The MAD process is repeated for individuals living on the street, families, and people living in CVRTMs.

While census tracts may be identified as a "hot-spot" in multiple categories, a tract's final stratum was assigned based on the following prioritization rules:

a. CTs identified as family "hot-spots" were assigned to the family stratum (in 2020, families were part of the "full inclusion" stratum and youth were included as the next top stratum).

b. Any remaining CTs identified as containing a vehicle or encampment "hot-spot" were assigned to the CVRTM stratum.

c. Any remaining CTs identified as containing an individual "hot-spot" were assigned to the individual stratum.

d. All remaining CTs that were not identified as "hot-spots" were assigned to a nonhotspot stratum.

After multiplying the 23 geographic strata (15 CDs + 8 SPAs) by the 4 "hot-spot" strata,

inclusion). From the 92 strata, a random sample of tracts is selected. The number of tracts

a tract could be assigned to one of 92 strata defined (except for tracts designated for full

selected in each geographic bounding area (SPA/CD) is a function of the total number of

unsheltered people that need to be interviewed to reach a 5% significance criterion. Then those

tracts in each SPA/CD are allocated to each "hot-spot stratum" based on a Neyman proportional

allocation, which means that statisticians "sampled more CTs from SPAs with a larger number of

CTs or larger variability in CT homeless populations to account for heterogeneity in the homeless population within and across SPAs" (USC 2020, 8). This method of allocation ensures that the number of tracts selected in each geography is proportional to "the product of the number of CTs per geography and the geography [previous year's unsheltered population] standard deviation" (USC 2020, 8).

After tracts are selected, all people who appear to be experiencing homelessness are interviewed during the survey period, and this is assumed to be a random selection. Field work for the Demographic Survey begins in December and continues until the statistical significance count targets have been met. These targets are based on the previous year's total unsheltered population. In 2020, the survey covered 505 census tracts between December 5th, 2019 and February 29th, 2020.

1.3. Problem Statement

LAHSA's use of "hot-spot" when sampling for the Demographic Survey is different from that term's use in spatial science. In spatial science, hot spots are clusters of events or objects in space that are statistically significantly different from their expected distribution based on a random process. The technical term for the concept that spatial data from near locations are more similar than data from distant locations is "spatial autocorrelation". LAHSA's "hot-spots", like hot spots in spatial science, are areas with "higher concentrations of… homeless adults living in CVRTMs, or homeless adults living on the street" relative to the surrounding area (USC 2020, 6). Because the Demographic Survey samples tracts and not people, sampling tracts with higher numbers of unsheltered people is a way for researchers to increase survey efficiency. However, the method by which LAHSA "hot-spots" are chosen is not statistically meaningful. Population count data tends to follow a power law probability distribution, which means a small

portion of areas will contain most of the population. This distribution does not lend itself to a MAD, but it does suggest that there is an opportunity to identify statistically significant spatial clusters. While LAHSA identifies census tracts as "hot-spots" in the context of their SPA, this is a tabular approach to a tract's geography that does not reference that tract's relationship with neighboring tracts in the overall spatial pattern of people experiencing homelessness.

LAHSA's "hot spot" planning process also seeks to find different areas with different subpopulations of people experiencing homelessness living there (sample representativeness). Each area's subpopulation of adults experiencing different kinds of homelessness (streetdwelling individuals, families, CVRTMs) is assessed separately, and then tracts are assigned to a "hot-spot" stratum identifying a kind of homelessness based on the prioritization rules listed above. This stratification design is susceptible to the ecological fallacy, which is when researchers erroneously associate average characteristics observed for aggregated groups with individual members of a group. There is a difference between sampling people who live in cars and sampling areas with people who live in cars. Even if a surveyor goes to a vehicle hot spot, they are still asked to survey everyone they come across who appears to be homeless whether the respondent lives in a vehicle. The stratification design assigns each tract to one type of homelessness, so it is difficult to determine whether the stratification achieves representativeness such that people from dense encampments and lone individuals are both represented. This thesis focuses on the goal of sample representativeness in the hot spot planning process and decouples it from the goal of survey efficiency.

Since the experience of homelessness differs by type of homelessness and by neighborhood, sampling geographies that characterize the preferred environments of people experiencing homelessness will improve the representativeness of the Demographic Survey. The

current stratification of census tracts relies on covering administrative boundaries (SPAs, City Council districts, and census tracts) that are not germane to how unsheltered people may choose where to live. Since these boundaries are arbitrary with respect to the presence of homelessness, LAHSA's hot spot planning process is susceptible to the modifiable areal unit problem (MAUP). This means that patterns observed are constrained by these administrative boundaries, and a different configuration might produce a different pattern. Additionally, this approach is based on the geographic perspective of the service providers who need to provide population estimates to public officials, and not necessarily the perspective of the survey subjects. People experiencing homelessness are not likely to consider what council district they live in, but they may be more likely to consider whether they are near services, whether they can utilize unoccupied public spaces such as those near freeways, or whether it is safe to park or sleep without harassment from housed neighbors or police. Characterizing areas based on the relationship between the built environment and different types of homelessness allows researchers and outreach professionals to gain more insight into the needs of their clients. LAHSA has somewhat refined its process to incorporate these considerations when adding Safe Parking tracts to the survey sample, but a more comprehensive approach will improve the representativeness of the survey results.

Finally, the current stratification and sampling method does not address the variable size and density of census tracts and other administrative boundaries. Census tracts are common and simple to use for policymaking analyses, but the boundaries of census tracts assume roughly similar numbers of people living in houses. These houses are fixed in space, but people experiencing homelessness are not fixed to one census tract. According to USC team members, one pitfall of the current sampling process is that unsheltered people who may have been counted

in one census tract may be required to move across the street to another tract during a Sanitation Department sweep, where they cannot be surveyed because they are outside the sampled area. Census tracts also vary with respect to land area, but the tabular approach to tract-level data treats a small urban census tract no differently from an expansive tract in the desert. This phenomenon also decreases survey efficiency since tracts with more land area require more resources to cover.

1.4. Impact Statement

This thesis will contribute to the project of understanding and ameliorating homelessness locally in LA County and regionally in Southern California by supplementing the Los Angeles Homeless Count with additional spatial statistics. The Homeless Count gathers important data that is necessary to understanding where people experiencing homelessness are living, and how many of them are unsheltered. This lack of shelter increases the risks of danger and illness for an already vulnerable population, and this thesis will seek better information about where and how these people live. While the focus of this thesis is the sample selection, the intermediate analyses of spatial patterns in the Homeless Count results over time will also benefit the immediate research team at USC, staff at LAHSA, other homelessness service providers, and policymakers. Visualizing how "hot-spots" change or do not change each year and identifying which of those are persistent and statistically significant will aid audiences' understanding of how unsheltered people choose where to live. The results of the time series comparison may also help LAHSA identify additional neighborhoods that require more targeted services or interventions.

The Demographic Survey that LAHSA fields provides rich insights into the nature of different types of homelessness, and it is important to design the sample correctly so that the population can be understood accurately. Classifying built environments based on the

unsheltered people who live in them can guide service provision. For example, the adult unsheltered community is stratified into individuals, families, and CVRTMs. Each of these categories define different living situations, and the people in those different living situations have different service needs. Some areas, such as Hollywood, Skid Row, and Venice, have robust communities of both sheltered and unsheltered people experiencing homelessness. On the other hand, people living under a given freeway overpass may form a small local neighborhood that may or may not count as a "hot-spot". Finally, individuals setting up a tent in a remote area may not be part of a community of unsheltered people, but they may travel for services and ought to be identified. Policymakers and social service planners must not only understand the differences between categories of people experiencing homelessness, but they must also understand how unsheltered people group into communities to develop appropriate approaches.

1.5. Overview of Analysis

This thesis applies tools to identify hot spots in a statistically significant way, explores a way to leverage spatial analysis to create different sampling geographies, and proposes different stratification methods that better capture the different experiences of homelessness. Validation of LAHSA's existing "hot-spot planning process" involves comparing the expected "hot-spots" with PIT counts for unsheltered people each year. The next analysis investigates hot spot identification methods that account for spatial relationships. After that, this thesis finds neighborhood attributes that correlate with increased prevalence of people experiencing homelessness. New sampling geographies that account for these attributes are proposed and visualized alongside the PIT counts for that year.

The next sections of this thesis are as follows. Chapter Two is a review of related work from both academic research and policymaking bodies' current research on homelessness, as

well as a review of other literature on spatial survey sampling and cluster detection. Chapter Three describes the data acquisition, data cleaning, and the analysis workflow itself. Chapter Four is an explication of the results, and Chapter Five summarizes conclusions, makes recommendations for the next iteration of the Homeless Count Demographic Survey, and identifies opportunities for additional refinements and analyses.

Chapter 2 Related Work

The following literature review explores current research around the spatial distribution of homelessness, counting the unsheltered homeless, built environment characteristics of homeless communities, VGI and locating homelessness, spatial survey sampling, and spatial cluster detection.

2.1. Literature Review

2.1.1. Spatial Distribution of Homelessness

This topic refers to the broad practice of describing how one or more encampments or communities of people experiencing homelessness are situated in an urban area. A mix of journal papers, theses and dissertations, case studies, and reports discuss homelessness in Los Angeles specifically and on the West Coast more broadly. For LA County, an Esri case study produced in conjunction with LAHSA analyzes the 2017 PIT count data, including a choropleth map of homeless population density and an Optimized Hot Spot analysis of homeless population density for the whole county CoC (Esri n.d.). Following this example, this thesis will use density rather than counts and the Optimized Hot Spot tool to describe the spatial distribution of unsheltered people. Additionally, a GIST thesis by Krystle Harrell, a researcher in Portland, evaluates both aggregated PIT count data and self-collected data based on "reported campsites" submitted in 311 calls (Harrell 2019). She created a grid overlay with cells of roughly one city block to aggregate the campsites, and then leveraged the Kernel Density, Hot Spot Analysis (Getis-Ord Gi*), and Cluster and Outlier Analysis (Anselin Local Moran's I) to explore spatial patterns. The Hot Spot Analysis was used to identify areas for additional neighborhood-level and sub-neighborhood level study. Her thesis is an instructive example since it compares multiple different cluster identification statistics on homelessness data and discusses the intermediate steps taken to identify optimal search radii and cell size.

This thesis refers to similarly situated student researchers along the West Coast because there are not many peer-reviewed works about unsheltered people that deploy multiple spatial analyses. The West Coast context is also unique because California and Oregon, among other western states, reported the highest percentages of all people experiencing homelessness in unsheltered locations in 2020 (US Department of Housing and Urban Development 2021).GIST theses by Shaw (2018) in Orange County and Harrell (2019) in Portland collect supplemental data in a riverbed and in parks. Since publicly available PIT count data is aggregated by census tract, additional observation can provide higher-resolution information useful for verifying patterns found in public data or identifying where unsheltered people have moved between official counts (Shaw 2018; Harrell 2019). USC and LAHSA do not leverage aerial photography data in the way that Shaw does, but team members are working with a GIS application developer, Akido Labs, to develop a tool like Esri's Collector for both the PIT count and the survey.

Outside of the West Coast context, other research has identified how the mobility and instability of the population experiencing homelessness affects their spatial distribution patterns in a way that is distinct from the housed experience. Researchers in St. Louis, Missouri analyzed geocoded addresses from people who had experienced homelessness and found that "sleep locations of homeless adults were much more concentrated in the urban core at baseline than were their previous housed and follow-up locations" using mean center and standard deviation ellipses, and that "These core areas had higher poverty, unemployment, and rent-to-income ratios and lower median incomes" (Alexander-Eitzman, Pollio, and North 2013, 679). This result suggests that unsheltered people tend to move towards more urban, higher-poverty areas during their time without permanent shelter, which informs this thesis' use of socioeconomic data. Additionally, research from Osaka using spatial auto-regression also indicates that "the number of homeless people in a census block is significantly influenced by the number of homeless in neighboring

census blocks" (Iwata and Karato 2011, 45). While Los Angeles County is more polycentric than St. Louis and Osaka, not to mention warmer for unsheltered people, these results reinforce this thesis's assumption that people experiencing homelessness will tend to group together to access services and that clustering is likely to happen in neighborhoods with specific characteristics.

2.1.2. Counting People Experiencing Homelessness

This topic area is like the topic area described above, but it focuses on the process by which US civic agencies have collected and reported homelessness numbers and characteristics for HUD. For the purposes of this thesis, the PIT count is the closest thing there is to ground truth. Since people who are not seen during a PIT count cannot be reported or analyzed, it is instructive to see how other researchers estimate potentially missing people who go unobserved by enumerators during a census. For example, in addition to the aerial photography example above, researchers from University of North Carolina developed a "hidden homeless" count for LAHSA in 2009 and 2011 in addition to the street, shelter, and youth counts that are presently executed. (Agans et al. 2014). However, estimates were unstable with a large standard error, and LAHSA has since discarded this estimate. Additionally, a 2013 report by researchers at University of California at San Francisco and University of California, Berkeley focuses on factors contributing to undercounting youth experiencing homelessness, including difficulties such as: people who do not want to be seen from the street, people who specifically avoid large encampments of adults for safety reasons, people who seek less exposed areas during the cold of January, people in more rural areas with lower street density, and "homeless sweeps" happening the night before the count in order to reduce the official numbers in an area (Auerswald et al. 2013).

Together, these snapshots of the homeless count literature clarify that even a census-style PIT count is not a perfect measure of unsheltered homelessness. This must be kept in mind when using a survey sample to make inferences about the broader unsheltered population: sampling

should endeavor to reach hard-to-count populations if possible. This strategy may vary depending on the geography of the study area. For example, San Francisco County designs its homelessness count demographic survey sample by canvassing the whole area and interviewing every third person (Applied Survey Research 2019). It is a much smaller county, so costs to survey it are lower than in an expansive county like Los Angeles; the county can sample people rather than tracts. As discussed in Chapter 1, the most recent LAHSA survey sample fully includes all census tracts where youth or families experiencing homelessness are seen. While this thesis will not focus on hidden homelessness, youth, or families because they are sampled or counted differently, the behaviors identified in the research from UC San Francisco and UC Berkeley underscore the importance of sampling less obvious areas, even if those areas do not hold a high number of potential survey subjects.

2.1.3. Other Spatial Cluster Detection Literature on Aggregated Counts

Most research systematically comparing spatial cluster detection methods tends to focus on point-level data and not counts aggregated to bounding geographies. However, spatial health data tends to come in aggregated counts for patient privacy reasons, so analyses in that area inform this thesis. Using different types of injuries in one fire district available as both points and counts aggregated to census tracts, one paper compares a Bernoulli-distribution-based cluster detection method on point data with a Poisson-distribution-based cluster detection method on tracts (Warden 2008). The paper finds that fewer clusters are found when analyzing point data rather than when analyzing aggregated counts, although there is overlap in the results from the two methods. This result suggests that some clusters identified from aggregated counts are artifacts of the MAUP. A different paper carries out a comparison of different types of Kulldorff's spatial scan statistic on drug crimes aggregated and normalized as a rate over census tracts (Quick and Law 2013). The tool used to calculate the spatial scan statistic (SaTScan) is like the Density Based Clustering tool in

ArcGIS Pro, which identifies clusters of point features within surrounding noise based on their spatial distribution. However, there is not an analogous tool that can be used on counts aggregated within boundaries. These papers do not directly inform this thesis, but they suggest that LAHSA's future plan to collect higher resolution spatial data on homelessness could provide more information for targeting outreach efforts.

2.1.4. VGI and Homelessness

This topic area is relevant because VGI is collected for the planning sessions that go into identifying "hot spots". VGI is "digital spatial data that... are created by citizens who... gather and disseminate their observations and geographic knowledge" (Elwood 2008). Elwood indicates that VGI is most valuable when it comes from the research subjects themselves, particularly where they are from marginalized or under-represented populations. Curtis et al. (2018) analyze spatial video geonarratives from people living in Skid Row to identify specific blocks, buildings, or street corners where drug activity is prevalent (Curtis et al. 2018). The authors used kernel density estimation to visualize the frequency of keywords about drug use, police activity, or other incidents recorded during the study. Another paper used workshops to ask youth experiencing homelessness in Portland to identify their activity spaces on paper maps (Townley et al. 2016). As described in Chapter 1, LAHSA-led "hot spot planning sessions" gather information from service providers and housed community members in an area as well as from people who are experiencing or previously experienced homelessness. While the service providers at the planning sessions are there in a formal data creation capacity, they are only reporting what they have seen after the fact—it is not a realtime recording of precise locations. The papers described above inform this thesis's approach to validating the data from the planning sessions against the data from the PIT count.

2.1.5. Perception and Subjectivity around Urban Homelessness

This topic focuses on the way that the boundaries of areas where people experiencing homelessness are living are socially constructed and in a process of constant negotiation. This phenomenon is seen in sweeps and in the gentrification of neighborhoods (where gentrification refers to the processes by which property values and rents increase, not necessarily processes of displacement). One local paper is an ethnographic study by UCLA public policy researchers analyzes the cultural geography of the changing boundary between two adjacent neighborhoods downtown: Skid Row and Gallery Row (Collins and Loukaitou-Sideris 2016). Skid Row has a century-long history as a neighborhood for low-wage transient labor and for people experiencing homelessness, the latter group appearing particularly after the de-industrialization of the local economy in the 1970s. Skid Row has a legal boundary on Main Street as part of the LA Planning Commission "containment plan", but it does not match up with any census tract boundaries. On the other hand, Gallery Row emerged after "neighborhood revitalization" programs designed to bring artists to the area drew higher-income residents and business patrons, per Richard Florida's theory of the "creative class". The study found that the gentrification of Gallery Row brought increased attention to the plight of Skid Row, and residents of Skid Row in turn used that opportunity to reclaim political power. At the same time, the boundary between the two neighborhoods hardened since there was increased policing and business improvement district activity in the Gallery Row area. This research indicates how neighborhoods are defined by subjective or fluid boundaries that may not correspond with administrative boundaries.

Other researchers using 311 or complaint data on homelessness have found that some people are more likely than others to see and report a tent, for example, and the data will be biased towards areas where those people are (Goldfischer 2019; Harrell 2019). Specifically, a paper evaluating 311 data regarding complaints about visible homelessness in New York City suggests that gentrifying

neighborhoods saw an increase in the amount of anti-homeless 311 calls and an increase in the amount of enforcement (Goldfischer 2019). People experiencing homelessness may become more visible when their surroundings are redeveloped, so socioeconomic indicators from the ACS that suggest gentrification will be part of neighborhood stratification for this thesis. These papers also suggest that the types of providers present at each planning session or the assumptions of enumerators carrying out the PIT count will affect results, since demographic research is socio-politically embedded in the context of the observer.

In addition to the GIS analysis component, Goldfischer (2019) examines the language used in the homelessness field. The researcher notes that New York city officials and polices shifted from using the phrase "encampment" to "hot spot" as part of a rising tide of anti-homeless sentiment. "Hot spot" and "encampment" are also part of the terminology that LAHSA uses. Goldfischer also interrogates the visual signifiers of homelessness that spur 311 complaints, which are important to consider because the LAHSA PIT count and the selection of survey subjects during the Demographic Survey are both based on identifying people who appear to be experiencing homelessness. The extent to which these people or dwellings contrast with the aesthetic of the neighborhood around them may play a part. Together with the UCLA paper, these papers also illustrate how the concept of place is a subjective phenomenon, and this subjectiveness affects current understandings of homelessness. Observers with a different understanding of the boundaries of neighborhoods will arrive at different conclusions about which neighborhoods have "many" or "few" people experiencing homelessness.

2.1.6. Sampling for Surveys of People Experiencing Homelessness

Other research quantifies what information is lost or not when shrinking a survey sample, or when getting a convenience sample of people experiencing homelessness. Two studies that are twenty years apart, Koegel, Burnam, and Morton (1996) and Golinelli et al. (2015), "explore how

three progressively less inclusive sampling frames affect understandings of the size and characteristics of homeless populations" in Los Angeles (Koegel, Burnam, and Morton 1996, 378). Using baseline data that comprehensively enumerated two neighborhoods in LA, the earlier study compared the demographic and behavioral characteristics of samples to the baseline results by regressing each characteristic on fixed effects for sample frame, site, and subject gender. The researchers found little bias but saw that estimates of the population size were too small, and men were more likely to be undercounted than women when using service-based sampling frames. This research predates the current method of comprehensive enumeration during a PIT count, so the question of underestimation of population size is less relevant. A similar 2015 paper interviewed youth experiencing homelessness in LA County from 2008-2009 and found that a shelter-based sampling frame for a youth survey significantly biases the demographic and risky behavior estimates (Golinelli et al. 2015). HUD and other federal agencies have been funding research into the causes of homelessness and characteristics of people experiencing homelessness for several decades. The fact that there are conflicting results from different times on the effects of restricting a sampling frame to shelters only suggests that there is room for additional research with more nuanced sample stratification. In the current methodology, the sample size is driven by the previous year's PIT count, which is understood to be ground truth as the "baseline" count is in this research. However, LAHSA is starting to collect spatial data on known encampments.

2.1.7. Other Spatial Sampling Literature

The literature on spatial methods for survey sampling specifically suggests that it is necessary to consider spatial autocorrelation of results in some way. For example, researchers stratifying Worcester County, Massachusetts to execute a national health study used socioeconomic data to stratify a county in a similar fashion to this thesis's proposed stratification design. They created a simple two-way stratification design based on a hazard exposure index (population

density, average daily traffic, and density of pollution sources) crossed with an adaptive capacity/social character index (education level, median income, poverty level, linguistic isolation, and racial minority proportion), with both indices standardized to a score of one to five (Downs et al. 2010). These 25 combinations were reduced to 18 strata by dividing the City of Worcester into 5 strata and combining smaller towns in the rest of the county into strata that were 1) contiguous, 2) comparably scored, and 3) added up to an acceptable number of births per year based on being "within 10% of the county MOS [measure of size] divided by the number of strata" (Downs et al. 2010, 1321). This is like the way the geographic strata are created from SPAs and City Council Districts for the LAHSA Demographic Survey, but instead of using town and city boundaries alone the methodology includes an analysis of those areas' characteristics. Notably, the researchers in Massachusetts showed a map of the proposed strata to a community advisory board; the mapped strata needed to align with stakeholders' lived experience of Worcester County's socioeconomic geography to be approved. Additionally, other research finds that geographic clustering of perceptions may bias survey results. Brown, Wood, and Griffith (2017) collected data about perceptions of West Nile Virus eradication methods in Dallas and learned that dissatisfaction was clustered in low-socioeconomic-level neighborhoods. The researchers found that "analyses of [spatial autocorrelation] can help the geographic targeting of survey administration, while the qualitative components can deploy purposive sampling strategies informed by the [spatial autocorrelation] analysis and survey findings" (Brown, Wood, and Griffith 2017, 15). This conclusion suggests that LAHSA's Demographic Survey results will be more representative if potential "perception clusters" can be identified and used in stratification.

Finally, a brief review of literature from international public health researchers was conducted to find best practices for spatial sampling that transcended United States political boundaries. Researchers fielding a demographic, health, and air quality survey in Delhi state that "[t]he optimal sampling design seeks to capture maximum variability by the minimum possible sites, which involves an appropriate distance/spacing between sample sites so that spatial autocorrelation can be minimized/eliminated" (Kumar 2007, 583). To select households, the researchers stratified on average air pollution levels, proximity to roads, and proximity to industrial clusters. Similarly, this thesis uses the presence of freeways and land use to classify areas of Los Angeles County into strata. Additionally, public health researchers seeking to field a sample in a medium-sized city in Burkina Faso used principal components analysis and hierarchical ascendant classification of aerial photography and field data to design a five-class typology of the city's built environment (Kassié et al. 2017). After that, the survey design was not that different from the LAHSA Demographic Survey: researchers randomly sampled sub-spaces from each type of area, and then randomly sampled plots for surveyors to visit (before further sub-sampling households). The goal was to evaluate differences in health outcomes or healthcare access across classes. Adopting this approach for the LAHSA survey sample design would also allow the public to learn more about the nature of homelessness in physically and socioeconomically different areas.

2.1.8. Area Characteristics Correlating with Homelessness

Several papers seek to identify built-environment characteristics that are predictors for estimating the presence of unsheltered people experiencing homelessness, and these inform this thesis's approach to sample stratification. For example, the GIST thesis about homelessness in Portland used chi-square tests to find a "significant positive relationship… between campsites and the MUR [mixed-use-residential] areas across the city", as well as a relationship with proximity to support services or transit (Harrell 2019, 46). These relationships do not have uniform strength across her study area – "there is no one singular factor able to explain campsite spatial preference as the top 10 neighborhoods for campsite density exhibited a different distribution of the variables amongst the neighborhoods" (Harrell 2019, 46). In Osaka, a 2008 paper used spatial auto-regression
to evaluate the different factors that correlate with the population density of tent-based or streetbased people experiencing homelessness in Osaka. It found that "the availability of employment, public medical care and food" has a significant effect that differed for street-based and tent-based people, since the former are more mobile and less likely to live in a neighborhood with other unsheltered people (Suzuki 2008, 1023). This underscores the importance of separating out these populations for survey sample design. A thesis analyzing homeless count data in San Diego County used both principal components analysis and geographically weighted regression to create a heatstroke risk index value for census tracts and determine where there was correlation between heat vulnerability and homeless population density, respectively (Baker 2019). Together, these papers suggest that a multivariate causal inference tool characterizing how the relationship between homelessness and different environmental factors varies across space adds value to existing methodologies.

Other articles more explicitly classify different built environments based on how they shape the experience of homelessness. The primary reference paper that informs this thesis is a k-means clustering analysis of census data identifying neighborhoods of "marginal, transitional, or prime" space in Los Angeles County (Marr, DeVerteuil, and Snow 2009). These terms refer to lower, more moderate, and higher socioeconomic indicator scores based on census data. The researchers use this typology to stratify shelters for a survey sample to determine if there are similar survival strategies employed or demographic characteristics observed in common within or across people living in shelters in those spaces. Below is a copy of the map of classified tracts from Marr et al. (2009):



Source: US Census 2000



Figure 2: Map of prime, transitional, and marginal classified census tracts (Marr et al. 2009, 311) This paper suggests that people living in homelessness shelters in prime neighborhoods are more likely to live and survive independently, whereas people living in shelters in a marginal neighborhood are more likely to be part of a homeless community with higher utilization of support services. Based on this research, this thesis uses k-means clustering to classify built environments that correspond to different experiences of unsheltered homelessness. In a similar exploration of homelessness survival strategies, researchers in Ohio evaluated abandoned houses for their attractiveness to unsheltered people (Kaplan et al. 2019). The authors similarly find four distinct categories of living environments that correspond to different modes of living. These categories are applicable to considering different homelessness "hot spots" found in Los Angeles County. For example, it is likely that people living on the sidewalk in Hollywood have different priorities than people who live near a more remote freeway off-ramp in Lancaster. Together these papers illustrate how people experiencing different types of homelessness in Los Angeles have different patterns of movement or survival behaviors depending on the characteristics of their neighborhood.

2.1.9. Other Area Deprivation Classification Literature

Previous researchers have created measures such as an Area Deprivation Index or Social Vulnerability Index to identify communities at a disadvantage when measuring public health or resilience to natural disasters. For example, following researchers in Europe, Australia, and New Zealand, Singh (2003) constructed a "composite census-based socioeconomic index" designed for monitoring population health inequalities in the United States from 17 indicators (Singh 2003, 1137). While Singh used a factor analysis to weight the relative importance of these indicators, the geographically weighted regression and k-means clustering methods described above lend themselves better to analysis in a GIS. Still, this thesis uses some of the indicators identified by Singh, since his work has become the foundation for indices used by US policymakers and social science researchers.

Chapter 3 Methods

This chapter describes the data and methods used to analyze the yearly homelessness data and its spatial characteristics. The first topic is data acquisition, followed by data cleaning, and then the analysis workflow itself. The analysis is comprised of validation of LAHSA's MAD "hot-spots", spatial cluster detection in the homelessness data, identification of relevant neighborhood characteristics, and demonstration of alternative stratification geographies.

3.1. Methods Overview

This research entailed using a combination of publicly available data and specially acquired data. Data cleaning used both R and ArcGIS Pro. Data processing combined year-tract level homelessness data with ACS data from each prior year and split that table into three files (one for each study year). Those files were brought into a geodatabase and merged with information from land use and freeway data. Most analyses used ArcGIS Pro as well, except for the chi-square statistics. First, the analysis visualizes LAHSA-identified "hot-spots" year-over-year to validate their stability and predictiveness. Next, two different spatial science hot spot or cluster identification tools are applied to the homelessness data. These tools account for the variable size of census tracts and the spatial relationships between them. Then, this thesis quantifies the statistical relationship between different socioeconomic and built environment characteristics with the presence of different types of unsheltered people. Finally, this analysis demonstrates one method to classify census tracts based on the relevant characteristics found in the previous section. These empirically drawn geographies are designed to be more relevant and representative geographies than administrative boundaries, and they can be used in future sample stratification.

29

3.2. Data Acquisition

The data required for this thesis is largely publicly available, but some data acquisition required direct coordination with LAHSA through the other homelessness researchers at USC. Below is a table of data used in this research.

Data	Source	Data Host	Spatial Scale/Unit	Extent	Timeframe
Hot Spot Planning Data	LAHSA	USC	Census tract polygons	CoC	2018-2020
ACS Data	Census Bureau	Census Bureau	Census tract polygons	County	2017-2019
SPA Boundaries	LA County Department of Public Health	LA County Geohub	SPA polygons	County	Static
Land Use	SCAG	SCAG Geohub	Parcel polygons	County	2016
Freeways	LA County Department of Regional Planning	LA County Geohub	Polylines	County	Static

Table 1: Data Sources

The first item in this table is "Hot Spot Planning Data". The intermediate sampling design data that identifies hot spots has been acquired through USC team members based on data previously received LAHSA. LAHSA has furnished tract-level stratification information to USC in different formats in different years. A staff research programmer (Gerry Young) and the senior data advisor (Patricia St. Clair) at USC have provided a standardized version for this thesis, since the team needed to create one to make their own determinations for "hot-spot planning" for 2021. Each observation in this file is a census tract-year, and variables include the "hot-spot" flags, the inputs for creating those flags, and information about whether tracts were sampled or surveyed. This file also includes the PIT count results for each tract, so there is no need to acquire and join separate PIT count data. The second item in the table is ACS data. The socioeconomic data used for the census tract classification analysis comes from the Census Bureau's American Community Survey (ACS) 5-year summary file tables. The ACS data is available through the Census Bureau API, and a list of specific tables and descriptions of the variables created from those tables are in Appendix A.

The fourth and fifth items, land use and freeway data, also constitute components to the classification analysis. The land use data is maintained by the Southern California Association of Governments (SCAG) on their open data portal and is available as a downloadable geodatabase feature class. SCAG evaluates the land use codes from all five member counties and harmonizes them so that the classification system is the same. The land use codes are available in Appendix B. The freeway data is publicly available on the LA County Geohub, from which public data can be added to the project map without being downloaded into the project geodatabase.

The third item in the table is SPA boundary data. Administrative boundary data is also hosted on the LA County Geohub. The boundaries that LAHSA uses for reporting are: Service Planning Areas (used by the Department of Public Health), city council boundaries, city boundaries, and county supervisor districts. However, this thesis focuses on the use of SPA boundaries. The hot spot planning data includes attribute columns identifying these administrative units, but it only covers tracts within the CoC (excluding Long Beach, Glendale, and Pasadena). The hosted data covers the entirety of LA County, which is better for continuity in cartography and labeling.

3.3. Data Cleaning

Initial processing of homelessness and census data was completed in R using the dplyr and tidycensus packages. First, extraneous columns (such as those dealing with the Youth Count) are removed from the homelessness data, a join field is created, then the file is be split into three tables (one for each year from 2018-2020). The cleaning program downloads the relevant ACS variables,

creates indicator variables, and joins the ACS data from the year before with the corresponding year of homelessness data.

The land use and freeway data processing begin in ArcGIS Pro before that data is merged with the homelessness and ACS data in R. In the land use data, each parcel in the county is categorized according to the countywide General Plan. The numeric order of these aggregated codes generally goes from most populated (residential land use) to least populated (open space). The code dictionary is available in Appendix B. Since this feature class is available as a downloadable Esri geodatabase, the ArcGIS Pro Summarize Within tool aggregates this information to the census tract level. This tool summarizes the land area per tract within each land use code group, and then and selects the most dominant land use. The Summarize Attributes tool identifies the modal land use for a census tract based on the maximum land area sum for all land uses observed, then the Add Join tool is used to keep only the modal land use.

The freeway data processing creates a binary flag identifying whether a freeway intersects with a census tract. The Select By Location tool identifies census tracts that intersect with a freeway, then those tracts are output to a new layer and joined back to the main analytical files to create the flag variable. This tract-level table is output as a comma-separated values file.

To bring the land use and freeway data together with the homelessness and ACS data, an R program joins the land use and freeway data with each year of the other data and outputs each table as a comma-separated values file. Finally, an ArcGIS ModelBuilder model imports these output files into ArcGIS Pro and joins them with the census tract boundary shapefile (restricted to LA County). All data and base maps are projected using the NAD83 California State Plane Zone V projected coordinate system, except for the inset map of California in Figure 1 which is projected using NAD83 California Teale Albers.

32

3.4. Analysis

3.4.1. Validation of existing "hot-spot planning" process

The first aim of this thesis is to validate the extent to which LAHSA's existing "hot-spot planning process" (using the Median Average Deviation method) aligns with spatial patterns found during the PIT count. Despite the issues with visibility raised in the literature, the unsheltered PIT count is the closest that researchers can come to observing the "ground truth". These exploratory maps will help the homelessness research team at USC determine whether the previous year's PIT count is reliable when sampling for a survey conducted nearly a year later. They also help evaluate the extent to which that reliability differs for the different categories of unsheltered people. For instance, individuals on the street may have more flexibility to travel than people in tents or makeshift shelters, or people in vehicles may be less susceptible to "sweeps" by law enforcement that displace them to another area. Separating analyses for each category of homelessness accounts for these differences in mobility or visibility, which will improve data reliability. All these visualizations use the statistics and flags created by LAHSA to describe how communities of unsheltered people may be evolving over time.

- A set of countywide maps visualizes the raw change in the PIT counts for unsheltered individuals and CVRTM. USC research partners requested these maps to explore which areas have the most change between years.
- 2. A set of tables evaluates the stability of "hot-spots" identified by LAHSA over time, for each type of homelessness. These tables display the proportion of tracts in each SPA that change their "hot-spot" status during the study period. These identify which SPA to use as an example in the subsequent maps, since the countywide map is too small-scale to illustrate the census-tract level data in detail.

- 3. A set of maps visualizes the difference between the LAHSA-identified hot spots and final PIT counts for the 2019-2020 surveys, for each stratum
- 4. A set of maps and regression models compares hot spots identified by LAHSA for each year's demographic survey through the MAD process from historical data and from planning sessions. These models use a Poisson distribution link function since that is appropriate for non-normally distributed count data. For each category of homelessness, the regression model is:

Log(PIT Count) = Intercept +Year + SPA + Historical HS Flag + Planning Session HS Flag + (Historical HS Flag x Planning Session HS Flag)

3.4.2. Determination of statistically significant spatial patterns

Next, this research applies spatial analysis cluster detection techniques to the homelessness data to describe where distinct communities of unsheltered people in LA County form. As mentioned above, the way that LAHSA uses the MAD statistic to define "hot spots" is different than the way that term is used in spatial science. This investigation operationalizes spatial thinking concepts including adjacency, density, and spatial autocorrelation to add robustness to the "hot-spot planning" process. For example, one simple way for researchers to incorporate the concept of homeless population density into their analyses would be to normalize the aggregated counts to rates of people per square mile. This analysis is the first in the Esri tool demonstration with the 2017 homelessness data described in the literature review above. However, this thesis focuses on spatial statistics tools that incorporate all the concepts listed above.

Unlike the existing methodology which uses the tracts in each SPA as reference for whether a tract is a "hot-spot", these analyses will use the whole county for reference. This is because a single SPA does not always contain enough tracts for any single tract to achieve statistical significance. In general, the analyses use a p-value of 0.05 or smaller to identify statistical significance, meaning that clusters of people are estimated to not be the result of a random process 95% of the time. There is no False Discovery Rate correction applied. This emphasis on statistical significance means that fewer hot spots are identified than in LAHSA's methodology, but researchers will gain more kinds of information about the spatial distribution of homelessness.

This section tests the following spatial cluster detection statistics: Global Moran's I, Local Moran's I, and Getis-Ord Gi*. These statistics are like those used in the GIST thesis by Harrell, who was researching homelessness in Portland. Together, these cluster identification tools illustrate alternatives to LAHSA's current MAD-based hot spot planning process that incorporate Tobler's first law of geography: that near areas are more related to each other than distant areas. The tools analyze the counts of unsheltered people in each tract in the context of surrounding tracts, and they account for the variable sizes of census tracts.

- Formally test for spatial autocorrelation in the PIT count data for each year and type of homelessness with the Incremental Spatial Autocorrelation tool. This tool implements the Global Moran's I statistic, which "measures spatial autocorrelation based on both feature locations and feature values simultaneously. Given a set of features and an associated attribute, it evaluates whether the pattern expressed is clustered, dispersed, or random" (Esri n.d.). This tool assists with identifying an appropriate neighborhood distance for the next clustering analyses by finding the distance where spatial autocorrelation is strongest.
- 2. Create a Spatial Weights Matrix based on this distance and on a minimum number of neighbors. Because census tracts are not consistently sized in LA County, a custom spatial weights matrix will allow for smaller tracts to have an appropriate number of neighbors while larger, less densely populated tracts can still have at least two neighbors.
- 3. Use the Cluster and Outlier Analysis tool (Local Moran's I) to identify high-count clusters, low-count clusters, and spatially significant outliers (tracts with high counts of unsheltered

35

people surrounded tracts with low counts, and vice versa), and visualize them alongside SPA boundaries. With this statistic, "[a] positive value for I indicates that a feature has neighboring features with similarly high or low attribute values; this feature is part of a cluster. A negative value for I indicates that a feature has neighboring features with dissimilar values; this feature is an outlier. In either instance, the p-value for the feature must be small enough for the cluster or outlier to be considered statistically significant" (Esri n.d.).

4. Use the Hot Spot Analysis Tool (Getis-Ord Gi*) to identify hot and cold spots and visualize them alongside SPA boundaries. With this statistic, "[t]o be a statistically significant hot spot, a feature will have a high value and be surrounded by other features with high values as well. The local sum for a feature and its neighbors is compared proportionally to the sum of all features; when the local sum is very different from the expected local sum, and when that difference is too large to be the result of random chance, a statistically significant zscore results" (Esri n.d.).

Below is a graphical representation of the workflow, assembled with ModelBuilder:



Figure 3: Diagram of Spatial Cluster Detection Workflow

3.4.3. Identification of neighborhood characteristics correlating with homelessness

This section identifies and visualizes which environmental and socioeconomic attributes correlate with increased numbers of unsheltered people. The hypothesis is that certain characteristics will define neighborhoods that are more amenable to individuals living outside, and other characteristics may define neighborhoods that are more amenable to people living in CVRTMs. LAHSA has modified its sampling methodology to include this concept to an extent when including census tracts with Safe Parking lots, but the analyses in this section will explore it more systematically.

The socioeconomic attributes tested are based on the social vulnerability index work from Singh and the LA County neighborhood classification paper by Marr et al. These come from ACS data and include educational attainment indicators, income and poverty indicator, and housing characteristics. These are described in Appendix A. In addition to those attributes, this section also evaluates the relationship between the presence of unsheltered people and county land use data (descriptions of land use codes are in Appendix B) or freeway data. The tests use two different statistics: the chi-square statistic, which is the statistic that the Harrell thesis about Portland homelessness uses, and the correlation coefficient. The chi-square test is used for categorical variables. This section compares binary (0/1) flags for the presence of either any unsheltered individuals or any CVRTMs against the SCAG General Plan Code or a binary flag for whether a freeway runs through that census tract. Pearson's correlation coefficient is used for continuous numeric variables (with the caveat that these variables are not normally distributed). This section compares PIT counts of unsheltered individuals or CVRTMs against the variables from the ACS. The statistically significant characteristics from that analysis are then used in the next analysis.

3.4.4. Alternative Sampling Geography

Once this thesis identifies the neighborhood characteristics that have a statistically significant relationship to the presence of homelessness, it suggests a way to incorporate them into the stratification methodology using a geographic aggregation via k-means clustering analysis. This technique is based on the research by Marr et al. (2009) that classified LA County census tracts into Prime, Transitional, and Marginal spaces for an ethnographic study of people living in homeless shelters. These terms refer to higher, intermediate, and lower socioeconomic index values. This section applies the Multivariate Classification tool to the data about unsheltered individuals and people living in CVRTMs. Maps visualize the empirically determined geography alongside the SPA boundaries and PIT counts for the relevant year. This geography can be used as a stratification level for the sample design.

Chapter 4 Results

This chapter visualizes the outcomes of the "hot-spot planning" validation analyses, the clusterdetection analyses, the search for neighborhood characteristics relevant to homelessness, and the resultant creation of alternative sampling geographies.

4.1. Validation

4.1.1. Analysis of Change in PIT Counts

The first set of results quantifies how and where counts of unsheltered people are changing over time. These exploratory data analyses were among the first maps requested by the research team at USC. The following figures visualize census tracts where there was a sizable increase or decrease in the number of people (or CVRTMs) counted from year-to-year. For each map pair (Individuals or CVRTMs), the first map represents change from 2018-2019 and the second map represents change from 2019-2020. These maps use raw differences between counts aggregated to the census tract level, with graduated symbols in orange representing increases and symbols in blue representing decreases.



Figure 4: Change in Individuals found during PIT Count, 2018-2019



Figure 5: Change in Individuals found during PIT Count, 2019-2020



Figure 6: Change in CVRTMs found during PIT Count, 2018-2019



Figure 7: Change in CVRTMs found during PIT Count, 2019-2020

Even in small-scale countywide maps it is possible to see some changes over time. For example, in Figure 5, the map of changes for individuals from 2019-2020, a larger decrease in Santa Monica is offset by an increase in Venice Beach. In Figure 6, the map of changes for CVRTMs from 2018-2019, on the west side of the map in Malibu medium size decreases in tracts along Topanga Canyon are visible alongside smaller increases in tracts along the coast. This illustrates how reliance on historical data without accounting for patterns in neighboring tracts may lead to sampling tracts from where unsheltered people have left.

This change is not evenly distributed around the county. Based on conversations with the research team at USC, natural disasters such as fires or anthropogenic changes such as sweeps near a new building may lead to people finding a new place to live. The following tables display the proportions of census tracts in each Service Planning Area (SPA) where at least one unsheltered individual (or CVRTM) is always found, never found, or sometimes found during the three-year study period.

Table 2: Proportions of c	ensus tracts where unshe	eltered individuals or CV	RTMs are always, never,
-	or sometimes found	2018-2020, by SPA	·

SPA	Always 1+ Individual in PIT	Never 1+ Individual in PIT	Sometimes 1+ Individual in PIT
1	28.6%	16.7%	54.8%
2	30.7%	19.6%	49.7%
3	31.8%	21.7%	46.2%
4	68.8%	5.3%	25.9%
5	59.0%	11.8%	29.2%
6	64.9%	2.2%	32.9%
7	45.0%	8.3%	46.4%
8	34.8%	15.8%	49.0%

SPA	Always 1+ CVRTM in PIT	Never 1+ CVRTM in PIT	Sometimes 1+ CVRTM in PIT
1	46.4%	6.0%	47.6%
2	48.0%	10.3%	41.7%
3	26.7%	17.5%	55.4%
4	75.0%	1.9%	23.1%
5	55.3%	9.9%	34.8%
6	81.1%	1.8%	17.1%
7	47.1%	4.8%	47.8%
8	41.7%	14.2%	43.7%

Table 2, continued

These tables are a high-level means of identifying where there is more stability or volatility in the PIT count results. SPA 4, which covers Central LA, has the highest proportion of census tracts where at least one unsheltered individual is found every year during the PIT count in this study period. SPA 6 (South LA) has the highest proportion of tracts where at least one CVRTM is found each year. These tracts are areas where homelessness and poverty are persistent. SPA 3 (San Gabriel Valley) has the highest proportion of tracts where no unsheltered individuals have been found during the study period, and the same is true for CVRTMs. This area includes wealthier suburbs and remote, mountainous areas in the Angeles National Forest.

4.1.2. Analysis of Change in Hot Spot Designations

This set of results quantifies changes in the "hot-spot" designations of census tracts over time. First, the USC research team created a categorical variable for their own investigation that identifies whether a census tract falls into the same or different strata across the study period. "Stratum" refers to whether a tract is identified as an Individual, CVRTM, Family, or Youth (in 2020) "hot-spot" or a non- "hot-spot" based on the historical PIT counts and the planning session input for all categories of unsheltered homelessness. The table below displays the proportions of tracts in each SPA assigned to the various categories of change.

SPA	Same 2018/19/20	Same 2018/19	Same 2018/20	Same 2019/20	Different Each Year
1	41.7%	25.0%	13.1%	7.1%	13.1%
2	41.5%	20.0%	13.3%	13.1%	12.2%
3	41.8%	30.1%	12.5%	8.4%	7.2%
4	45.6%	21.3%	12.2%	11.9%	9.1%
5	52.2%	23.6%	7.5%	12.4%	4.3%
6	32.9%	26.8%	12.7%	14.9%	12.7%
7	33.9%	27.7%	11.4%	13.1%	13.8%
8	44.9%	22.7%	8.9%	14.6%	8.9%

Table 3: Proportions of census tracts with the same or different "hot-spot" stratum designation from2018-2020

According to this table, roughly one-third to one-half of census tracts in each SPA are designated as part of the same stratum during the study period. For example, LAHSA may always designate some tracts as CVRTM "hot-spots", and others may always be designated as non "hotspots". However, this designation happens after statisticians apply the "hot-spot" prioritization rules to tracts that may have a significant number of multiple kinds of unsheltered people. Because the tract prioritization rules and final designation obscures situations where a tract has multiple kinds of unsheltered people, this thesis disaggregates analyses by type of homelessness. Based on conversations with the USC research team, this approach is consistent with their plans for survey sampling going forward.

Maps of smaller areas for the next set of validation analyses make it easier to see how changes in the counts and locations of unsheltered people over time affect the "hot-spot planning process" that is part of the Demographic Survey sample stratification. To determine where the map areas should cover, the following tables display the proportions of census tracts in each SPA that are always, never, or sometimes designated as an individual or CVRTM "hot-spot" based on LAHSA's Median Average Deviation formula.

SPA	Always Individual HS	Never Individual HS	Sometimes Individual HS
1	2.4%	52.4%	45.2%
2	3.8%	45.5%	50.7%
3	4.5%	51.0%	44.3%
4	5.6%	54.1%	40.3%
5	4.3%	56.5%	39.1%
6	3.1%	45.6%	51.3%
7	6.6%	42.9%	50.2%
8	5.7%	49.8%	44.1%

Table 4: Proportion of census tracts that are always, never, or sometimes an Individual or CVRTM "hot-spot", by SPA

SPA	Always CVRTM HS	Never CVRTM HS	Sometimes Individual HS
1	11.9%	47.6%	40.5%
2	12.0%	45.3%	42.7%
3	15.0%	41.5%	43.2%
4	17.5%	46.3%	36.3%
5	16.8%	48.4%	34.8%
6	14.5%	43.9%	41.7%
7	16.6%	41.2%	41.9%
8	14.2%	49.4%	36.0%

Based on the tables above, SPA 6 (South LA) has the largest proportion of census tracts where the Individual "hot-spot" status changes over the course of the study period, and SPA 3 (San Gabriel Valley) has the largest proportion where the CVRTM "hot-spot" status changes. In the context of LA County as a whole, the area covered by SPA 6 is characterized by a higher-density urban street grid with low-rise commercial and a higher proportion of multifamily residential buildings. Census tracts tend to be smaller in area. On the other hand, SPA 3 is characterized by lower-density suburban streets and a mix of housing types. One caveat is that SPA 3 geographically contains the city of Pasadena, which has no homelessness data here because the city maintains its own health department. SPA 3 also covers low-density tracts including open space in the San Gabriel Mountains. Focusing on the SPAs where there are the most changes makes it easier to identify places where the "hot-spot" designation does not correspond to the largest PIT count results on a map. Comparing two different areas of LA County also helps illustrate how differences in the built environment across areas mean that the unit of analysis (census tracts) is not standardized; a single tract is not necessarily comparable to all other tracts in its SPA. For the next analyses the maps for individuals will be of SPA 6 and the maps for CVRTMs will be of SPA 3.

Additionally, the next maps focus on 2019 and 2020 because of validity concerns about the 2018 "hot-spot" flags. The following table illustrates how tracts identified as individual or CVRTM "hot-spots" from historical or planning session data were not always identified as such in the overall flag. These rows are highlighted below. There were also significantly fewer hot spots identified in 2018, particularly for individuals.

Table 5: Counts and proportions of census tracts designated as "hot-spots" for individuals and CVRTMs based on input data and overall, by year

Year	Overall Individual HS	Historical Data Individual HS	Planning Session Individual HS	N Tracts	Percent of Tracts in Year
	N	N	N	<mark>1,479</mark>	<mark>68.5%</mark>
	N	N	Y	<mark>55</mark>	<mark>2.5%</mark>
	N	Y	N	<mark>230</mark>	<mark>10.6%</mark>
2018	Ν	Y	Y	76	3.5%
	Y	Ν	Y	136	6.3%
	Y	Y	Ν	140	6.5%
	Y	Y	Y	44	2.0%
	Ν	Ν	Ν	1,488	68.8%
2010	Y	Ν	Y	236	10.9%
2019	Y	Y	Ν	254	11.7%
	Y	Y	Y	185	8.6%
2020	Ν	Ν	Ν	1,432	66.2%
	Y	Ν	Y	262	12.1%
2020	Y	Y	Ν	267	12.3%
	Y	Y	Y	202	9.3%

Year	Overall CVRTM HS	Historical Data CVRTM HS	Planning Session CVRTM HS	N Tracts	Percent of Tracts in Year
	N	N	N	<mark>1,499</mark>	<mark>69.4%</mark>
	N	N	Y	<mark>3</mark>	<mark>0.1%</mark>
	N	Y	N	<mark>30</mark>	<mark>1.4%</mark>
2018	Ν	Y	Y	5	0.2%
	Y	Ν	Y	140	6.5%
	Y	Y	N	376	17.4%
	Y	Y	Y	107	5.0%
	Ν	Ν	Ν	1,502	69.4%
2010	Y	Ν	Y	181	8.4%
2019	Y	Y	Ν	314	14.5%
	Y	Y	Y	166	7.7%
	Ν	Ν	Ν	1,293	59.8%
2020	Y	Ν	Y	624	28.8%
	Y	Y	N	164	7.6%
	Y	Y	Y	82	3.8%

Table 5, Continued

The practical effect of this coding error is small; only 361 tracts and 38 tracts out of over 2,000 are affected for unsheltered individuals and CVRTMs, respectively. For the 2019 and 2020 counts, a USC quantitative analyst (Laura Gascue) engaged in an interactive quality control process with LAHSA analysts preparing the data so that derived variables and flags behaved as expected. However, for 2018 there was not the same level of dialogue between the data teams, and sampling was based on the flags as they arrived from LAHSA. This analysis is focused on the conceptual validity of the hot spot methodology, not validity issues that arise from analyst error. Still, the following maps visualizing the overlap between "hot-spot" flags from different data sources are restricted to 2019 and 2020 due to this issue.

4.1.3. Visualization of Hot Spot Designations versus PIT Counts

The following yearly maps visualize the "hot-spot" status for tracts containing unsheltered individuals or CVRTMs with the PIT count results for that year and type of homelessness.



Figure 8: Individual "hot-spot" tracts and PIT counts in SPA 6, 2019



Figure 9: Individual "hot-spot" tracts and PIT counts in SPA 6, 2020



Figure 10: CVRTM "hot-spots" and PIT counts in SPA 3, 2019



Figure 11: CVRTM "hot-spots" and PIT counts in SPA 3, 2020

In the maps above, the dots represent the density of unsheltered individuals (green) or CVRTMs (purple) found during the PIT count in each census tract, not their exact locations (which are not available). The shaded census tracts represent areas identified as "hot-spots" based on the previous year's historical data from January and on "planning session" results from the fall of that same previous year. The overlap between the "hot-spot" areas and the PIT counts appears to be better for CVRTMs (Figures 10 and 11) than it does for individuals. This may be because individuals who do not maintain a tent or other makeshift shelter are more mobile than other types of people experiencing homelessness. The planning sessions are meant to reduce the effect of the lag between the time that historical data was collected and the time the survey goes into the field. However, even with additional planning sessions in place, there is not alignment between where "hot-spots" are flagged and where unsheltered people are found. Additionally, the fact that tracts are identified as hot spots based on counts rather than the population density of unsheltered people both reduces the efficiency of fielding the survey and makes it difficult to use the "hot-spot" formula for other service planning needs outside the survey.

4.1.4. Comparison of Hot Spot Planning Data Sources

This analysis decomposes LAHSA's "hot-spot" flags into flags based on its input data sources and visualize those against the PIT count results. As explained above, when LAHSA analysts identify a census tract as a "hot-spot", it is because the count from input data falls above a Median Average Deviation cutoff. Input data includes historical data from the previous year's PIT count and planning session data, where service providers identify where they have seen unsheltered clients on a map. In 2020, additional outreach data was incorporated into the planning session "hot-spot" flag. A tract is a "hot-spot" overall for a given category of homelessness if it is a "hot-spot" based on one or more of the input data sources.

For both years and both types of homelessness, the minority of LAHSA-identified "hotspots" are designated as such based on both input data sources. This phenomenon is visible in Table 5 in section 4.1.2. Additionally, very parsimonious general linear models regressing PIT counts on the historical and planning session "hot-spot" flags quantify the relative extent to which these flags predict high counts of unsheltered individuals or CVRTMs in a census tract.

Regression Term	Coefficient Estimate	Estimated Multiplier	Standard Error	t- Statistic	p-Value
(Intercept)	1.22	3.40	0.03	38.50	0.00
Year = 2020	0.01	1.01	0.01	1.03	0.30
SPA 2	-0.93	0.39	0.04	-25.77	2.05E- 146
SPA 3	-0.61	0.54	0.04	-16.94	2.43E-64
SPA 4	0.55	1.74	0.03	17.00	7.96E-65
SPA 5	0.49	1.64	0.04	14.06	7.14E-45
SPA 6	0.20	1.22	0.03	5.67	1.47E-08
SPA 7	-0.41	0.66	0.04	-11.29	1.49E-29
SPA 8	-0.62	0.54	0.04	-16.17	7.80E-59
Skid Row ("SPA 9")	4.73	112.84	0.05	98.25	0.00
Venice ("SPA 10")	1.54	4.65	0.06	24.26	4.74E- 130
Hollywood ("SPA 11")	0.80	2.22	0.06	12.80	1.68E-37
Individual Historical HS	1.11	3.05	0.02	62.09	0.00
Individual Planning Session HS	0.72	2.06	0.02	33.59	2.50E- 247
Individual Historical HS*Planning Session HS	0.03	1.03	0.03	1.08	0.28

Table 6: Regressions of PIT Counts on Historical and Planning Session HS Flags

Regression Term	Coefficient Estimate	Estimated Multiplier	Standard Error	t- Statistic	p-Value
(Intercept)	2.10	8.18	0.02	89.64	0.00
Year = 2020	-0.04	0.96	0.01	-2.72	6.57E-03
SPA 2	-1.05	0.35	0.02	-42.09	0.00
SPA 3	-1.94	0.14	0.03	-59.82	0.00
SPA 4	-0.18	0.84	0.02	-7.73	1.05E-14
SPA 5	-0.68	0.50	0.03	-23.49	5.49E- 122
SPA 6	-0.23	0.79	0.02	-9.47	2.91E-21
SPA 7	-1.26	0.28	0.03	-43.79	0.00
SPA 8	-0.84	0.43	0.03	-29.89	2.52E- 196
Skid Row ("SPA 9")	3.86	47.35	0.04	89.92	0.00
Venice ("SPA 10")	1.02	2.76	0.05	20.54	1.05E-93
Hollywood ("SPA 11")	-0.23	0.80	0.06	-3.63	2.78E-04
CVRTM Historical HS	1.35	3.87	0.02	74.76	0.00
CVRTM Planning Session HS	0.65	1.91	0.03	22.37	7.01E- 111
CVRTM Historical HS*Planning Session HS	-0.12	0.89	0.03	-3.69	2.23E-04

The models above suggest that the planning sessions can identify high-count tracts, but they do not identify tracts with counts as high as those found with the historical data. For both models, the coefficient estimate for the planning session flag term is smaller than that for the historical "hot-spot" term. For example, in the individual model the coefficient on planning session "hot-spot" (after exponentiating it) means that the expected multiplying factor for the count of unsheltered individuals for planning session "hot-spot" tracts is 2.06 times a baseline tract that is not a "hot-spot". The interaction term indicates that any separate effect for tracts identified with both methods is small or insignificant; most of the power comes from the historical "hot-spot" flag alone. These results are consistent with the practical understanding that tracts with at least one person, tent, or vehicle identified in planning sessions will be recognized as a "hot-spot" based on the MAD formula.

The models are not comprehensive nor meant to predict counts, so this analysis does not report adjusted R² nor other goodness-of-fit statistics. In addition to the flags of interest, these models only control for year and SPA. In 2019, the full inclusion communities of Skid Row and Venice were treated as their own SPA for "hot-spot planning", as was Hollywood. Still, these models can be used to aid researchers deciding whether to continue holding separate planning sessions to identify "hot-spots".



The following maps visualize the overlap between the historical "hot-spot" designations and the planning session "hot-spot" designations for the SPAs of interest for individuals and CVRTMs:

Figure 12: Comparison of Historical and Planning Session Individual "Hot-Spots" in SPA 6, 2019



Figure 13: Comparison of Historical and Planning Session Individual "Hot-Spots" in SPA 6, 2020



Figure 14: Comparison of Historical and Planning Session CVRTM "Hot-Spots" in SPA 3, 2019



Figure 15: Comparison of Historical and Planning Session CVRTM "Hot-Spots" in SPA 3, 2020 These maps show that it is useful to have both processes if possible (or some other method of closing the temporal gap between the historical data and the survey field period, such as using updated outreach data). On the other hand, it is valuable to understand how large the discrepancy is; providers participating in planning sessions have a daytime service-provision-oriented perception of where unsheltered people live rather than an understanding of sleeping locations. For example, it is
surprising that the large tract in the Angeles Forest (tract 9303.01) was identified as a "hot-spot" for CVRTMs during planning sessions for the 2019 survey when none were found in 2018 and only a couple of tents were found there in January 2019. No CVRTMs were found during the 2018 or 2020 PIT count. For 2020, there were also far fewer planning-session-only "hot-spots" for CVRTMs in SPA 3.

4.2. Cluster Detection

4.2.1. Creating Spatial Weights Matrices

First, this analysis iteratively implements the Moran's I statistic with different search radii for clustering considerations using the ArcGIS Pro Incremental Spatial Autocorrelation tool on each year of PIT count data. The tool identifies the first distance at which spatial autocorrelation in the data is strongest for unsheltered individuals and CVRTMs. For each run of the tool, the initial distance is set at 900 feet. This figure is based on the square root of the area of smallest census tract in LA County. There is no increment distance specified because no peak z-score could be identified when smaller increments were specified. Below are charts identifying the peak distances for each type of homelessness for each year. In 2018 and 2020, the peak distances for both types of homelessness were 4,178.83 feet and in 2019, the peak distance for both types of homelessness was 7,456.40 feet.

Spatial Autocorrelation by Distance



Figure 16: Incremental Spatial Autocorrelation of Individual PIT Counts, 2018

Spatial Autocorrelation by Distance



Figure 17: Incremental Spatial Autocorrelation of Individual PIT Counts, 2019

Spatial Autocorrelation by Distance



Figure 18: Incremental Spatial Autocorrelation of Individual PIT Counts, 2020

Spatial Autocorrelation by Distance



Figure 19: Incremental Spatial Autocorrelation of CVRTM PIT Counts, 2018

Spatial Autocorrelation by Distance



Figure 20: Incremental Spatial Autocorrelation of CVRTM PIT Counts, 2019

Spatial Autocorrelation by Distance



Figure 21: Incremental Spatial Autocorrelation of CVRTM PIT Counts, 2020

Next, the first peak distance is used to define the neighborhood around each census tract for the next analyses. For each year and type of homelessness, a spatial weights matrix file using this neighborhood definition is created. Because some census tracts have a large area and are surrounded by other large tracts, the spatial weights matrix defines the neighborhood as the tracts within this fixed distance or the two nearest neighboring tracts in the large-tract case.

4.2.2. Local Moran's I (Cluster and Outlier Analysis)

The following maps visualize statistically high-count and low-count clusters of census tracts based on the Moran's I test for spatial autocorrelation, as well as tracts that are high or low-count outliers surrounded by tracts with dissimilar densities of unsheltered people.



Figure 22: Cluster and Outlier Analysis results for individual PIT counts, 2018



Figure 23: (Cluster and Outlier Analysis results for individual PIT counts, 2019



Figure 24: Cluster and Outlier Analysis results for individual PIT counts, 2020



Figure 25: Cluster and Outlier Analysis results for CVRTM PIT counts, 2018



Figure 26: Cluster and Outlier Analysis results for CVRTM PIT counts, 2019



Figure 27: Cluster and Outlier Analysis results for CVRTM PIT counts, 2020

This analysis reveals high-count clusters of tracts (in light red) that homelessness researchers, service providers, and the public are already aware of, such as Hollywood, Venice, and Skid Row. In the current sampling methodology, tracts in these areas are always surveyed. However, it also reveals statistically significant high-count clusters and high-low outliers (in darker red) in Malibu, South LA, San Pedro and other areas where it would be important and efficient to survey. Meanwhile, tracts identified as low-low-count clusters (light blue) are less likely to have unsheltered people, and it may be efficient to sample a smaller proportion or skip them entirely during the survey. Dark blue tracts are "low-high outliers" that tend to be adjacent to higher count tracts, and these could be grouped together.

4.2.3. Getis-Ord Gi* (Hot Spot Analysis)

The following maps visualize statistically significant high-density clusters of census tracts based on the Getis-Ord Gi* statistical test. To be a statistically significant spatial "hot-spot", a highcount tract must be surrounded by other high-count tracts. This analysis does not identify highcount spatial outliers like the previous analysis.



Figure 28: Getis-Ord Gi* Hot Spots for Individual PIT Counts, 2018



Figure 29: Getis-Ord Gi* Hot Spots for Individual PIT Counts, 2019



Figure 30: Getis-Ord Gi* Hot Spots for Individual PIT Counts, 2020



Figure 31: Getis-Ord Gi* Hot Spots for CVRTM PIT Counts, 2018



Figure 32: Getis-Ord Gi* Hot Spots for CVRTM PIT Counts, 2019



Figure 33: Getis-Ord Gi* Hot Spots for CVRTM PIT Counts, 2020

The results of the spatial cluster detection analysis using the Getis-Ord Gi* statistic overlaps with the clusters identified from the local Moran's I analysis above, but the information presented here may be more intuitive for researchers identifying areas to fully include in the survey sample. For individuals and CVRTMs, there are persistent, statistically significant hot spots downtown, Venice/Santa Monica, and South LA. For individuals specifically, there are smaller hot spots in Hollywood, Lancaster, and near LAX airport. For CVRTMs specifically, there are hot spots in the East San Fernando Valley, Antelope Valley, and San Pedro.

4.3. Statistical Tests Against Neighborhood Attributes

4.3.1. Chi-Square Tests

The table below shows chi-square test statistics for the relationship between SCAG land use category or the presence of a freeway with the presence of different types of unsheltered homelessness. Statistically insignificant values are shown in *dark red italic* text.

Table 7: Chi-Square Test Statistics for Neighborhood Characteristics vs. PIT Counts

Neighborhood Attribute	2018		201	9	2020		
	Individuals	CVRTM	Individuals	CVRTM	Individuals	CVRTM	
Majority Land Use	28.05	62.37	46.11	51.19	66.48	58.56	
Contains Freeway	5.01	7.19	15.78	20.60	10.93	18.73	

Both categorical land use variables have a statistically significant relationship with both types of homelessness, except for the presence of a freeway for unsheltered individuals in 2018. Two-way frequency tables for the different land use categories versus the presence of different types of homelessness are available in Appendix F.

4.3.2. Correlation Coefficients

This table shows Pearson correlation coefficients between neighborhood attributes and counts of unsheltered individuals or CVRTMs based on census-tract level data. One caveat is that neither the homelessness counts nor the neighborhood attribute values are normally distributed. Significant correlation coefficients tend to be above 0.04 or below -0.04, and correlation coefficients above 0.2 or below -0.2 (in **bold** text) are both statistically and practically significant.

Neighborhood Attribute		Individuals			CVRTM		
		2019	2020	2018	2019	2020	
Percent of HH With Any Crowding		0.06	0.06	0.07	0.09	0.09	
Percent of HH With Moderate Crowding		-0.01	-0.02	0.03	0.04	0.04	
Percent of HH With Severe Crowding		0.11	0.13	0.09	0.12	0.11	
Percent of Population White Alone Non-Hispanic		-0.03	-0.02	-0.07	-0.07	-0.06	
Total Population		-0.01	-0.01	-0.03	-0.05	-0.03	
Percent of Adults Age 25+ Without Grade 9 Education		0.05	0.04	0.1	0.1	0.1	
Percent of Adults Age 25+ With a HS Diploma		-0.06	-0.06	-0.11	-0.12	-0.12	
Miles of Freeway		0.02	0.01	0.05	0.08	0.05	
Median Home Value		0.02	0.02	-0.1	-0.09	-0.06	
Median Rent		-0.12	-0.14	-0.16	-0.17	-0.17	
Percent of Housing Units Built Before 1980		-0.08	-0.07	-0.02	-0.03	-0.02	
Percent of HH Renting		0.21	0.22	0.11	0.16	0.16	
Homeowner Vacancy Rate		0.03	0.01	0.01	0.02	0.01	
Renter Vacancy Rate		0.07	0.08	0.03	0.04	0.05	
Median Family Income		-0.1	-0.12	-0.14	-0.16	-0.16	
Percent of Families Below Federal Poverty Line		0.19	0.14	0.22	0.23	0.17	
Percent of HH Without Access to a Vehicle		0.35	0.34	0.27	0.3	0.3	
Percent of Renter HH Paying >35% of Income on Rent		0.01	-0.01	0.04	0.07	0.03	
Percent of HH Without Phone Access		0.22	0.14	0.22	0.2	0.15	
Percent of Population Below 150% of Federal Poverty Line	0.19	0.2	0.2	0.19	0.23	0.22	
Percent of Civilian Employed Adults in White Collar Occupations	-0.05	-0.04	-0.04	-0.12	-0.12	-0.11	
Civilian Unemployment Rate		0.15	0.14	0.14	0.15	0.19	
Percent of HH Moved In Before 1990		-0.17	-0.18	-0.06	-0.1	-0.11	
Percent of HH Moved In 1990-1999		-0.18	-0.18	-0.07	-0.12	-0.1	
Percent of HH Moved In 2000-2009		-0.11	-0.13	-0.07	-0.03	-0.04	
Percent of HH Moved In 2010-2014		0.15	0.15	-0.02	0.12	0.13	
Percent of HH Moved In 2015-Onward		0.23	0.21	0.09	0.11	0.1	

Table 8: Correlation Coefficients of Neighborhood Characteristics vs. PIT Counts

The variables chosen for this analysis are partially supply factors, i.e. "people that could be homeless in the future", but they are mostly demand factors illustrating how welcoming or unwelcoming an area might be for unsheltered people. They are based on area deprivation index literature (Singh 2001) and the analysis by Marr et al. (2009). Variables related to poverty, income, and educational attainment are mostly significantly correlated with the presence of both kinds of homelessness. Variables indicating serious poverty such as lack of access to a vehicle or phone are more strongly correlated. Variables explicitly related to the built environment, such as the proportion of older buildings or miles of freeway, are correlated to some types of homelessness but not others. The proportion of residents who moved to a tract in 2015 or later is also significantly correlated with homelessness, particularly with unsheltered individuals. Finally, a variable for the racial mix of a tract is correlated for CVRTMs and not unsheltered individuals. The k-means analysis reference paper by Marr et al. (2009) incorporates the percent of white non-Hispanic residents in a tract, so that is the indicator used here. The race of residents is relevant because it is a factor in access to political resources that could be used to eject unsheltered people from an area.

4.4. Alternative Geography from Neighborhood Attributes

The following maps visualize how census tracts can be clustered into Prime, Transitional, and Marginal spaces (following Marr et al. 2009) based on the relevant socioeconomic and environmental attributes identified above. These designations are primarily based on how these geographies align with the map from the reference paper and how the underlying cluster attribute averages line up. The relative values of these attributes do not change much from year to year, and as seen in Table 7 above, their statistical relationships to the presence of homelessness do not change much from year to year. Based on feedback from the USC research team, the maps and charts below focus on the intermediate year 2019. In addition to null data for Pasadena and Long Beach, low-population tracts with invalid ACS data such as near the airport or in desert areas were not always classified by the k-means clustering analysis.



Figure 34: Classified census tracts based on attributes correlated with unsheltered individuals, 2019



Figure 35: Cluster Factor Distributions for Individual Factors, 2019



Figure 36: Classified census tracts based on attributes correlated with CVRTMs, 2019



Figure 37: Cluster Factor Distributions for CVRTM Factors, 2019

In these figures, the light green tracts and lines represent Prime areas, the medium bluegreen tracts represent Transitional areas, and the pink tracts represent Marginal areas. This method of classifying geography (k-means clustering) is designed to increase the representativeness of the sample population. Homelessness is represented as green or purple dots, corresponding with Individuals and CVRTMs as represented in Section 4.1. While there are more unsheltered people in Marginal areas, they are distributed across each of the different classes of tracts. The unsheltered people found in different classes of tracts are likely to have different lifestyle characteristics and survival strategies that policymakers and outreach workers may be interested in. These people may also have different demographic characteristics. Depending on the needs of the surveyors, these geographies can supplement or replace the existing SPA- and CD-boundary-based geographic strata.

Chapter 5 Discussion

5.1. Main Conclusions

This thesis aims to identify ways for statisticians conducting the LA Homeless Count to leverage spatial thinking for two goals: selecting a more representative sample for the Demographic Survey and gaining insight about the unsheltered population. While the analyses presented here illustrate the value of adding spatial analysis to the existing methodologies, further work is necessary to achieve the first aim and measure whether changing the tract stratification would achieve a more representative, diverse sample of tracts and respondents. Both the spatial cluster detection analyses and the neighborhood characteristics analyses achieve the second aim of new insights. This thesis identified new areas with significant densities of unsheltered people and found socioeconomic indicators that are correlated with the presence of unsheltered people.

5.1.1. Identifying Hot Spots

First, the "validation" analyses illustrated that movement of unsheltered people during the year means that "hot-spot" identification based on the historical PIT count data may miss new clusters of people or CVRTMs that emerge afterwards. Depending on the neighborhood and type of homelessness, unsheltered people living in a high-count tract dispersed to surrounding tracts, or, on the other hand, previously dispersed unsheltered people agglomerated into encampments. The "hot-spot planning sessions" potentially help with this temporal gap, but it is not clear why there is so little overlap with historical data as well as a smaller relationship with the final PIT counts for a year. LAHSA and USC's existing plans to incorporate outreach data into "hot-spot" planning are a step in the right direction.

This thesis approaches the homelessness data with hot spot identification techniques from spatial statistics in a way that accounts for spatial autocorrelation, so census tracts (and the people living there) are analyzed in the context of others near them. Unlike the existing tabular approach which uses other tracts in a SPA for reference, this hot spot analysis uses a fixed distance band to describe the spatial relationship between tracts, which corrects this phenomenon. The "cluster detection" analyses suggest that statisticians should add additional neighborhoods beyond Hollywood, Santa Monica, and Venice to the areas that are always sampled and surveyed for the Demographic Survey. Both the Moran's I and Getis-Ord Gi* statistical tests identified the following areas as significant clusters of hot spots: Malibu, San Pedro, South LA near Inglewood, the east San Fernando Valley area, and areas in and around Lancaster.

This thesis adds value to the existing "hot-spot planning" process by expanding researchers' understanding of where to find and survey unsheltered people, and these maps can also help when targeting service provision. As noted in other related literature, the presence of clustering or hot spots of unsheltered people is directly related to their visibility in communities. Increased visibility can make people vulnerable to harassment from housed residents, police, or politicians. However, increased visibility can also catalyze better outreach (whether from official providers or activist organizations), increased financial resources, and more political capital for unsheltered people in specific areas.

5.1.2. Alternative Geographies

The correlation test analysis found that socioeconomic status indicators were broadly correlated with the presence of unsheltered individuals and CVRTMs in all years of the study period. These variables are meant to proxy the effect of disinvestment in a neighborhood (and its housed residents). Some poverty-related variables that the USC research team was interested in, including the proportion of rent-burdened households, turned out to not be significantly correlated in five out of six cases. Variables indicating extreme poverty, such as lack of access to a phone or vehicle, or severe residential overcrowding, were both statistically and practically significantly correlated. Demographic statistics used as indicators for political power over presence of homelessness were also significantly related to homelessness, including the percent of white non-Hispanic residents and the percent of more recent arrivals to an area. It is possible that the strength of the correlation differs across different types of homelessness because CVRTMs (particularly tents and makeshift shelters) are more visible to housed residents than individuals. Housed residents may be more likely to advocate against encampments of multiple CVRTMs.

Overall, understanding the survival strategies of different people based on their different environments not only helps sampling but also helps agencies target their service provision accordingly. Even though homelessness was correlated with lower socioeconomic indicators, homelessness appears all over the county in Prime, Transitional, and Marginal neighborhoods. Sampling based on neighborhood characteristics, and therefore on these survival strategies, is important because people who fall into homelessness have different histories and needs. For instance, people who become unsheltered may try to stay in their original community, or they may join an encampment, or they may avoid specific neighborhoods. The way that people evaluate these options will differ based on their socioeconomic and demographic considerations (e.g., single women may avoid Skid Row for safety, LGBTQ+ youth may gravitate towards Hollywood for community and access to services).

The relevant characteristics do not appear to change much from year-to-year because this thesis uses five-year average ACS summary data and static built environment data on land use and freeways. It is possible to simplify this retrospective analysis of correlated factors and evaluate all years with one statistical test. However, if one-year data were available at the census tract level, a yearly analysis might pick up phenomena like new development, higher-income residents, or other

89

indicators of gentrification more clearly. Additionally, a 2020 ruling by Federal Judge David O. Carter requiring the offer of shelter to unsheltered people near freeways may diminish the extent to which freeways are related to homelessness in the future (Cuniff 2020). Still, identifying transit yards, airports, or other less residential areas where unsheltered people might seek refuge with more static land use data will help with targeting services.

5.2. Limitations

There are some statistical considerations that could impact the validity of the neighborhood characteristics and area classification analyses. First, the chi-square and Pearson's correlation coefficient tests compare tract attributes with counts of unsheltered people, not densities of unsheltered people per areal unit. This approach is consistent with the chi-square tests used in the reference paper by Harrell (2019) but using densities would normalize the scale of the counts and ensure that tracts were more comparable. Additionally, there are multiple redundant socioeconomic indicators that are all used in the k-means classification analysis, so it is likely that the classification model is over-specified. This is also consistent with the number of variables used in the reference paper by Marr et al. (2009), but an additional step to reduce the dimensionality of the data would create a more valid model. Finally, the k-means classification algorithm treats land-use codes as ordinal numbers when they are categorical codes. While the numeric codes do roughly correspond to decreasing levels of residents, a better approach would use variables evaluating what proportion of a tract's land area was residential, commercial, industrial, or open space.

5.3. Feasibility of Implementing Suggestions in Sampling Workflow

5.3.1. Hot Spot Strata

Either the hot spot designations based on the Local Moran's I statistic or the Getis-Ord Gi* statistic could replace the existing methodology for "hot-spot planning" based on historical PIT

count data. For the 2020 Count, tracts were assigned to the following types of "hot-spot" strata: Full Inclusion tracts, Hot Spot tracts (split further into Youth, CVRTM, or individual strata), and nonhot-spots. Based on the Local Moran's I Cluster and Outlier analysis, the high-count clusters and outliers for either type of homelessness could be added to Full Inclusion stratum. For the purposes of surveying, low-count outliers could be grouped with adjacent high-count cluster areas since unsheltered people may move about the broader area between data collection time and survey fielding time. Meanwhile low-count clusters could be excluded from the sample entirely, for efficiency. Based on the Getis-Ord Gi* hot spots, which may be easier to interpret, all hot spots for either type of homelessness could be added to the Full Inclusion stratum. In either case, tracts that don't have statistically significant counts of either kind of unsheltered people could form the nonhot-spot stratum. Since the new non-hot-spot stratum would have more tracts, this thesis recommends using the "Optimal" allocation in SAS's SURVEYSELECT procedure instead of Neyman allocation and using the tract land area as the survey cost parameter. Finally, this process should be repeated for each category of homelessness, so a tract is not assigned to one "hot-spot" stratum but may instead be included in one or more of the individual, vehicle, and tent and makeshift shelter samples. For future analysis in collaboration with USC researchers, this thesis author will test this potential sampling strategy and compare it with current strategies.

5.3.2. Geographic Strata

The scheme classifying tracts into Prime, Transitional, or Marginal spaces for CVRTMs or individuals could interact with the existing geographic strata in the survey sampling process rather than replace them. The current use of SPA and CD in sampling is based on requests for demographic information reports by city and county officials. Using SPA and CD for stratification allows for margins of error to be computed for statistics in these reports, and policymakers would likely want to retain that information.

91

This thesis' analysis is a proof of concept that incorporates neighborhood characteristics that are specifically correlated with different types of homelessness, but it's possible to simplify the process. Alternative classification that changes for each type of homelessness or from year-to-year may be difficult to explain. Since many of the relevant variables are socioeconomic and demographic indicators that are relevant to multiple policy issues, using a pre-existing classifier like the California Healthy Places Index or a different social vulnerability or area deprivation index may also work.

5.4. Future Research Directions

5.4.1. Current and Potential Plans for HC 2022 and Beyond

Since research for this thesis began in the fall of 2019, the policy landscape around homelessness in LA County has shifted so that more resources are available for improved research and outreach. The most obvious change is that recovery funds meant to fight the economic and public health fallout of the ongoing COVID-19 global pandemic have been used to rapidly provide temporary shelter and services. In California, Project Roomkey and Project Homekey respectively rented and purchased surplus hotel rooms to bring unsheltered people indoors. This hotel room strategy lowered people's risk of contracting respiratory illness compared to the risk they would have faced in a congregated shelter setting. Successive Federal stimulus packages have allowed LA and other large West Coast metropolitan areas to creatively tackle the longstanding issue of people living without shelter. The City of LA is planning to spend more than \$1 billion in the next fiscal year on homelessness, a sevenfold increase over the budget from five years ago (Oreskes and Zahniser 2021). New temporary "bridge" shelters and permanent supportive housing developments will also benefit from this money.

The new policy environment represents a major opportunity to bring in spatial data and spatial thinking. Until now, researchers working on the count have had to focus on meeting federal

deadlines for data collection while navigating the increasing politicization of homelessness at the city and Council District level. This thesis is not an indictment of LAHSA's and USC's current good work. However, if housed residents and policymakers continue to entertain assumptions about unsheltered people that are not based in data, unsheltered people will continue to die on the street. The fact that LAHSA has only made slow progress towards incorporating spatial data until recently is an argument for directing more resources towards analytical innovation, not fewer resources.

Looking ahead, the USC research team is currently planning to break out vehicles (CVR) and tents/makeshift shelters (TM) in their analyses. This will help with the representativeness of the Demographic Survey sample because the CVR and TM populations have very different levels of mobility and visibility. When investigating the different types of unsheltered people alongside the USC research team, Dr. Randall Kuhn of UCLA found that the demographic survey tends to slightly over-count street-dwelling individuals and under-count vans and RVs compared to the proportions found in the PIT count. The maps in this thesis of raw change over time and of clustering detected in the PIT counts can also be reproduced with the CVRTM data broken out in this fashion as well.

Additionally, LAHSA is currently working with Akido Labs to build a spatial database and app for homelessness data collection. One app is currently in use for monitoring and triaging COVID-19 in encampments, and another app is in development for use during the next Annual Homeless Count (Smith 2021). A prototype based on the app used for the City of Long Beach's count will be piloted in July and August of 2021, and this thesis' author is the point of contact between LAHSA and local community organizations that will carry out the pilot in Koreatown, Los Angeles. It is possible that this point data will need to be de-identified, aggregated, or censored in some way before it is publicly available on the LA City Geohub. This type of detailed spatial data collection will allow for more spatial thinking about homelessness in LA County. Since the USC homelessness research team did not have to work on a Homeless Count for 2021 due to the pandemic, the team is currently evaluating ways to predict or interpolate the 2021 count based on historical and outreach data from the Homelessness Management Information System (HMIS) and calls to the LA Homeless Outreach Portal (LA-HOP). This data contains more spatial information and has more temporal granularity, which will improve the timeliness of researchers' understandings of where major encampments are. This may continue to supplement or entirely replace "hot-spot planning session" data, since it also includes the daytime locations of unsheltered people. As spatial information about known encampments accumulates, tracts including them could be treated as "full inclusion" survey tracts or they could form a stratification class for survey sampling.

5.4.2. Additional GIS Analysis Directions

Other options for refining the hot spot identification process may provide even more visibility for the spatial patterns of unsheltered people's living situations. For example, the ArcGIS Pro Areal Interpolation tool, which is based on kriging, can be used to smooth out counts from irregularly shaped census tracts into identically shaped and sized areas like a hexagon grid. Using a regular grid would make it easier to quickly evaluate both relative counts and density, and these grid cells could be used as sampling units rather than census tracts. Also, the temporal nature of the historical data makes it possible to perform Emerging Hot Spot Analysis, which evaluates new versus persistent hot spots based on the Getis-Ord Gi* used in this thesis. This analysis would be particularly meaningful with access to more temporally granular data, such as monthly or quarterly information from HMIS or LA-HOP. More frequent data deliveries could also beget more creative data visualizations such as 3D space-time maps or animations.

There are also additional ways to evaluate which neighborhood characteristics are relevant to the presence of unsheltered people. One variable that this thesis did not include was the presence of riverbeds, even though there is literature about encampments near riverbeds in Southern California. Rivers could be incorporated into the existing analyses in the same manner that freeways are currently included. This thesis evaluates the statistical relevance of each neighborhood characteristic individually, but geographically weighted regression (and other generalized linear modeling in development) would evaluate the relative relevance of each variable together. It is likely that there is spatial heterogeneity among the geographic areas, and that some built environment characteristics may be more impactful than others with respect to the presence of homelessness in different areas. Such a modeling exercise could also inform future prediction efforts.

An even more robust way to incorporate neighborhood characteristics into prediction would use EBK regression prediction. Socioeconomic and built environment attribute data would be predictors, and historical PIT count data would be the dependent variable. Such a model might help predict where homelessness would grow ahead of time, and the analysis could encourage policymakers to direct resources to areas where there may be people at risk of falling into homelessness. In Stata or other statistical analysis software, spatial auto-regression is also an option.

Finally, the Build Balanced Zones tool available in ArcGIS Pro could create sampling areas that are larger than a census tract in neighborhoods with small census tracts. As mentioned in the first chapter, people who have been "swept" out of a tract cannot be surveyed until they return to the tract since they are out of the sampled area. As a result, research team members were interested in combining tracts in dense areas with small tracts to alleviate this issue. The Build Balanced Zones tool can use a genetic algorithm to create spatially contiguous zones in the study area based on balancing an attribute target sum (e.g., number of unsheltered people) with other relevant neighborhood characteristics. Since LAHSA uses Esri software to manage its spatial data, either an

95

in-house spatial data analyst or an outside researcher could explore any of these specific tools in future work.

References

- Agans, Robert P., Malcolm T. Jefferson, James M. Bowling, Donglin Zeng, Jenny Yang, and Mark Silverbush. 2014. "Enumerating the Hidden Homeless: Strategies to Estimate the Homeless Gone Missing from a Point-in-Time Count." *Journal of Official Statistics* 30 (2): 215–29. https://doi.org/10.2478/jos-2014-0014.
- Alexander-Eitzman, Ben, David E Pollio, and Carol S North. 2013. "The Neighborhood Context of Homelessness." *American Journal of Public Health* 103 (4): 679–85. https://doi.org/10.2105/AJPH.2012.
- Applied Survey Research. 2019. "San Francisco Homeless Count & Survey Comprehensive Report." San Francisco. https://hsh.sfgov.org/wpcontent/uploads/2020/01/2019HIRDReport SanFrancisco FinalDraft-1.pdf.
- Auerswald, Colette L, Jessica Lin, Laura Petry, and Shahera Hyatt. 2013. "Hidden in Plain Sight: An Assessment of Youth Inclusion in Point-in-Time Counts of California's Unsheltered Homeless Population." Sacramento. http://cahomelessyouth.ca.gov.
- Baker, Michael. 2019. "HEAT WAVES AND HOMELESSNESS: ANALYSIS OF SAN DIEGO AND RECOMMENDATIONS." UC San Diego. https://escholarship.org/uc/item/3s49k58k.
- Brown, Timothy T., Jennifer D. Wood, and Daniel A. Griffith. 2017. "Using Spatial Autocorrelation Analysis to Guide Mixed Methods Survey Sample Design Decisions." *Journal* of Mixed Methods Research 11 (3): 394–414. https://doi.org/10.1177/1558689815621438.
- 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): 401–27. https://doi.org/http://dx.doi.org.libproxy1.usc.edu/10.3828/tpr.2016.27.
- Cuniff, Meaghan. 2020. "Inside a Judge's Controversial Crusade to Solve Homelessness in L.A." *LA Magazine*, November 5, 2020. https://www.lamag.com/citythinkblog/inside-a-judgescontroversial-crusade-to-solve-homelessness-in-l-a/.
- Downs, Timothy J., Yelena Ogneva-Himmelberger, Onesky Aupont, Yangyang Wang, Ann Raj, Paula Zimmerman, Robert Goble, et al. 2010. "Vulnerability-Based Spatial Sampling Stratification for the National Children's Study, Worcester County, Massachusetts: Capturing Health-Relevant Environmental and Sociodemographic Variability." *Environmental Health Perspectives* 118 (9): 1318–25. https://doi.org/10.1289/ehp.0901315.
- Elwood, Sarah. 2008. "Volunteered Geographic Information: Future Research Directions Motivated by Critical, Participatory, and Feminist GIS." *GeoJournal* 72 (3–4): 173–83. https://doi.org/10.1007/s10708-008-9186-0.
- Esri. n.d. "Combating Homelessness in Los Angeles County—Analytics | Documentation." Accessed September 25, 2020a. https://desktop.arcgis.com/en/analytics/case-studies/la-countyhomelessness-1-overview.htm.
Documentation." Accessed February 13, 2021b. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm.

- Goldfischer, Eric. 2019. "From Encampments to Hotspots: The Changing Policing of Homelessness in New York City." *Housing Studies*. https://doi.org/10.1080/02673037.2019.1655532.
- Golinelli, Daniela, Joan S. Tucker, Gery W. Ryan, and Suzanne L. Wenzel. 2015. "Strategies for Obtaining Probability Samples of Homeless Youth." *Field Methods* 27 (2): 131–43. https://doi.org/10.1177/1525822X14547500.
- Hannon, Ryan. 2014. "Homeless People vs. People Experiencing Homelessness | Street Outreach Blog." Street Outreach Blog. December 22, 2014. https://www.streetoutreach.systems/homeless-people-vs-people-experiencing-homelessness/.
- Harrell, Krystle N. 2019. "Homelessness in Portland, Oregon: An Analysis of Homeless Campsite Spatial Patterns and Spatial Relationships." Portland State University. https://doi.org/10.15760/geogmaster.24.
- Iwata, Shinichiro, and Koji Karato. 2011. "Homeless Networks and Geographic Concentration: Evidence from Osaka City." *Papers in Regional Science* 90 (1): 27–46. https://doi.org/10.1111/j.1435-5957.2010.00306.x.
- Kaplan, Abram, Kim Diver, Karl Sandin, and Sarah Kafer Mill. 2019. "Homeless Interactions with the Built Environment: A Spatial Pattern Language of Abandoned Housing." Urban Science 3 (2): 65. https://doi.org/10.3390/urbansci3020065.
- Kassié, Daouda, Anna Roudot, Nadine Dessay, Jean Luc Piermay, Gérard Salem, and Florence Fournet. 2017. "Development of a Spatial Sampling Protocol Using GIS to Measure Health Disparities in Bobo-Dioulasso, Burkina Faso, a Medium-Sized African City." *International Journal of Health Geographics* 16 (1): 14. https://doi.org/10.1186/s12942-017-0087-7.
- Koegel, Paul, M. Audrey Burnam, and Sally C. Morton. 1996. "Enumerating Homeless People: Alternative Strategies and Their Consequences." *Evaluation Review* 20 (4): 378–403. https://doi.org/10.1177/0193841X9602000402.
- Kumar, Naresh. 2007. "Spatial Sampling Design for a Demographic and Health Survey." *Population Research and Policy Review* 26 (5–6): 581–99. https://doi.org/10.1007/s11113-007-9044-7.
- LAHSA. 2007. "2007 Greater Los Angeles Homeless Count Report." Los Angeles. https://documents.lahsa.org/planning/homelesscount/2007/HC07-full_report.pdf.

—. 2009. "2009 Greater Los Angeles Homeless Count Report." Los Angeles. http://documents.lahsa.org/planning/homelesscount/2009/HC09-fullreport.pdf.

- Marr, Matthew D., Geoff DeVerteuil, and David Snow. 2009. "Towards a Contextual Approach to the Place-Homeless Survival Nexus: An Exploratory Case Study of Los Angeles County." *Cities* 26 (6): 307–17. https://doi.org/10.1016/j.cities.2009.07.008.
- Oreskes, Benjamin, and David Zahniser. 2021. "Garcetti Plans Nearly \$1 Billion for L.A. Homeless Programs - Los Angeles Times." *Los Angeles Times*, 2021.

https://www.latimes.com/homeless-housing/story/2021-04-19/los-angeles-will-increase-budget-for-addressing-homelessness.

- Quick, Matthew, and Jane Law. 2013. "Exploring Hotspots of Drug Offences in Toronto: A Comparison of Four Local Spatial Cluster Detection Methods." *Canadian Journal of Criminology and Criminal Justice* 55 (2): 215–38. https://doi.org/10.3138/cjccj.2012.El3.
- Shaw, William Timothy. 2018. "THE EFFECT OF PUBLIC POLICY ON THE SPATIAL DISTRIBUTION OF ORANGE COUNTY'S HOMELESS POPULATION: A CASE STUDY IN THE LOWER SANTA ANA RIVER AREA." CSU Long Beach. https://doi.org/10839469.
- Singh, Gopal K. 2003. "Area Deprivation and Widening Inequalities in US Mortality, 1969-1998." *American Journal of Public Health* 93 (7): 1137–43. https://doi.org/10.2105/AJPH.93.7.1137.
- Smith, Doug. 2021. "This Year's Homeless Count Was Canceled. Should We Rethink It? Los Angeles Times." Los Angeles Times. 2021. https://www.latimes.com/california/story/2021-02-04/this-years-homeless-count-is-canceled-is-it-time-to-rethink-it.
- Suzuki, Wataru. 2008. "What Determines the Spatial Distribution of Homeless People in Japan?" *Applied Economics Letters* 15 (13): 1023–26. https://doi.org/10.1080/13504850600972394.
- Townley, Greg, L. Pearson, Josephine M. Lehrwyn, Nicole T. Prophet, and Mareike Trauernicht. 2016. "Utilizing Participatory Mapping and GIS to Examine the Activity Spaces of Homeless Youth." *American Journal of Community Psychology* 57 (3–4): 404–14. https://doi.org/10.1002/ajcp.12060.
- U.S. Department of Housing and Urban Development. 2004. "A Guide to Counting Unsheltered Homeless People." *Counting the Homeless: Unsheltered and Sheltered*. Office of Community Planning and Development. https://www.hudexchange.info/sites/onecpd/assets/File/Guide-for-Counting-Unsheltered-Homeless-Persons.pdf.

----. 2014. "Point-in-Time Count Methodology Guide." https://files.hudexchange.info/resources/documents/PIT-Count-Methodology-Guide.pdf.

- US Department of Housing and Urban Development. 2021. "The 2020 Annual Homeless Assessment Report to Congress."
- USC. 2020. "2020 Los Angeles Continuum of Care Homeless Count Methodology Report." Los Angeles.
- Warden, Craig R. 2008. "Comparison of Poisson and Bernoulli Spatial Cluster Analyses of Pediatric Injuries in a Fire District." *International Journal of Health Geographics* 7 (51): 17. https://doi.org/10.1186/1476-072X-7-51.

Variable Variable Name in R **Census Tables** Logic Median Family Income B19113 Median Family Inc Median HomeValue DP04 Median Home Value Median Rent Median Rent DP04 Percent of HH Without Phone No Phone B25043 Access Percent of Adults Age 25+ Without Pct below G9 edu B15003 Grade 9 Education Percent of Families Below Federal Pct Below PovLevel B17026 Poverty Line Percent of Housing Units Built Pct Bldg Pre1980 **DP04** Before 1980 Persons Per Room Percent of HH With Any Crowding Pct Crowd Any DP04 >1.0Percent of HH With Moderate 1<Persons Per Room Pct Crowd Moderate **DP04** Crowding <=1.5 Percent of HH With Severe Persons Per Room Pct Crowd Severe DP04 Crowding >1.5 Percent of Adults Age 25+ With a Pct HS edu B15003 HS Diploma Percent of HH Moved In 1990-Pct Movein 1990 99 **DP04** 1999 Percent of HH Moved In 2000-Pct Movein 2000 09 **DP04** 2009 Percent of HH Moved In 2010-Pct Movein 2010 14 **DP04** 2014 Percent of HH Moved In 2015-DP04 Pct Movein 2015 On Onward Pct Movein Pre 199 Percent of HH Moved In Before **DP04** 1990 0 Percent of HH Without Access to a Pct NoVehicle **DP04** Vehicle Percent of Population White Alone Pct Race WhiteNH B03003 Non-Hispanic Percent of Renter HH Paying Pct Rent GTE35 Pct **DP04** >35% of Income on Rent Inc Pct Renters DP04 Percent of HH Renting Management, Percent of Civilian Employed business, science, and Adults in White Collar Pct WhiteCollar C24060 arts occupations + Occupations Sales and office occupations Percent of Population Below 150% Pop below 150Pov C17002 of Federal Poverty Line Civilian Unemployment Rate Unemp Rate B23025 Vacancy Rate Home DP04 Homeowner Vacancy Rate owner Renter Vacancy Rate Vacancy Rate Renter **DP04**

Appendix A: ACS Classification Variables

Code	Land Use Description
1100	Residential
1110	Single Family Residential
1111	High-Density Single Family Residential (9 or more DUs/ac)
1112	Medium-Density Single Family Residential (3-8 DUs/ac)
1113	Low-Density Single Family Residential (2 or less DUs/ac)
1120	Multi-Family Residential
1121	Mixed Multi-Family Residential
1122	Duplexes, Triplexes and 2- or 3-Unit Condominiums and Townhouses
1123	Low-Rise Apartments, Condominiums, and Townhouses
1124	Medium-Rise Apartments and Condominiums
1125	High-Rise Apartments and Condominiums
1130	Mobile Homes and Trailer Parks
1131	Trailer Parks and Mobile Home Courts, High-Density
1132	Mobile Home Courts and Subdivisions, Low-Density
1140	Mixed Residential
1150	Rural Residential
1200	Commercial and Services
1210	General Office Use
1211	Low- and Medium-Rise Major Office Use
1212	High-Rise Major Office Use
1213	Skyscrapers
1220	Retail Stores and Commercial Services
1221	Regional Shopping Center
1222	Retail Centers (Non-Strip With Contiguous Interconnected Off-Street Parking)
1223	Retail Strip Development
1230	Other Commercial
1231	Commercial Storage
1232	Commercial Recreation
1233	Hotels and Motels
1240	Public Facilities
1241	Government Offices
1242	Police and Sheriff Stations
1243	Fire Stations
1244	Major Medical Health Care Facilities
1245	Religious Facilities
1246	Other Public Facilities
1247	Public Parking Facilities
1250	Special Use Facilities
1251	Correctional Facilities
1252	Special Care Facilities
1253	Other Special Use Facilities

Appendix B: SCAG General Plan Land Use Code List

1260	Educational Institutions
1261	Pre-Schools/Day Care Centers
1262	Elementary Schools
1263	Junior or Intermediate High Schools
1264	Senior High Schools
1265	Colleges and Universities
1266	Trade Schools and Professional Training Facilities
1270	Military Installations
1271	Base (Built-up Area)
1272	Vacant Area
1273	Air Field
1274	Former Base (Built-up Area)
1275	Former Base Vacant Area
1276	Former Base Air Field
1300	Industrial
1310	Light Industrial
1311	Manufacturing Assembly and Industrial Services
1312	Mation Picture and Television Studio Lots
1312	Packing Houses and Grain Elevators
1314	Research and Development
1320	Heavy Industrial
1320	Manufacturing
1321	Patroleum Refining and Processing
1322	Open Storage
1325	Major Metal Processing
1324	Chemical Processing
1320	Extraction
1331	Mineral Extraction Other Than Oil and Gas
1332	Mineral Extraction - Oil and Gas
1340	Wholesaling and Warehousing
1400	Transportation Communications and Utilities
1400	Transportation
1411	Airports
1412	Railroads
1413	Freeways and Major Roads
1413	Park-and-Ride Lots
1415	Bus Terminals and Vards
1415	Truck Terminals
1417	Harbor Facilities
1417	Harbor Facilities
1418	Navigation Aids
1/20	Communication Facilities
1420	Utility Engliting
1430	Fleetrical Power Eacilities
1431	Solid Waste Dispessel Easilities
1432	Solid Waste Disposal Facilities

1433	Liquid Waste Disposal Facilities
1434	Water Storage Facilities
1435	Natural Gas and Petroleum Facilities
1436	Water Transfer Facilities
1437	Improved Flood Waterways and Structures
1438	Mixed Utilities
1440	Maintenance Yards
1441	Bus Yards
1442	Rail Yards
1450	Mixed Transportation
1460	Mixed Transportation and Utility
1500	Mixed Commercial and Industrial
1600	Mixed Residential and Commercial
1610	Residential-Oriented Residential/Commercial Mixed Use
1620	Commercial-Oriented Residential/Commercial Mixed Use
1700	Under Construction
1800	Open Space and Recreation
1810	Golf Courses
1820	Local Parks and Recreation
1830	Regional Parks and Recreation
1840	Cemeteries
1850	Wildlife Preserves and Sanctuaries
1860	Specimen Gardens and Arboreta
1870	Beach Parks
1880	Other Open Space and Recreation
1900	Urban Vacant
2000	Agriculture
2100	Cropland and Improved Pasture Land
2110	Irrigated Cropland and Improved Pasture Land
2120	Non-Irrigated Cropland and Improved Pasture Land
2200	Orchards and Vineyards
2300	Nurseries
2400	Dairy, Intensive Livestock, and Associated Facilities
2500	Poultry Operations
2600	Other Agriculture
2700	Horse Ranches
3000	Vacant
3100	Vacant Undifferentiated
3200	Abandoned Orchards and Vineyards
3300	Vacant With Limited Improvements
3400	Beaches (Vacant)
4000	Water
4100	Water, Undifferentiated
4200	Harbor Water Facilities
4300	Marina Water Facilities

4400	Water Within a Military Installation
4500	Area of Inundation (High Water)
7777	Specific Plan
8888	Undevelopable or Protected Land
9999	Unknown