Using Local Geographic Features to Predict Changes in the Los Angeles Homeless Population

Environment and Planning B: Urban Analytics and City Science XX(X):2–28 © The Author(s) 2019 Reprints and permission: sagepub.co.uk/journalsPermissions.nav DOI: 10.1177/ToBeAssigned ijr.sagepub.com/

(\$SAGE

Michael R Lindstrom¹, Ryan Shijie Du¹, Xiang Yang Ng¹, Dominic Diaz¹², Margaret Koulikova¹, Matthew T Nero³⁴, Hannah C Ross⁵, Sanjay Shukla⁶, Andrea L Bertozzi¹⁷, and P Jeffrey Brantingham⁸

Abstract

Homelessness is a complex phenomenon with many factors, both societal and personal, playing a role. By combining annual homeless point-in-time estimates for Los Angeles, along with other publicly available data, we conduct a multimethod study that builds a series of models correlating the estimated homeless populations with localized features in the city. We first establish a descriptive link between regional features on the scale of census tracts, including median rent and number of coffee shops, and the homeless population, through correlation analysis and topic modelling. Then, we make a first step at quantitative predictions by combining our data into a shallow neural network to forecast homeless population changes. By inverting the network and looking at its structure, we identify regional features that may be significant in determining how the homeless population changes over time; notably, in trying to select one single feature as a predictor, both the street and vehicle homeless counts are more likely to decrease over a year if they began with an above average count and census tracts with zero counts may be more likely to remain that way rather than increasing.

Keywords

homelessness, point-in-time counts, multimethod study, machine learning, topic modelling, artifical neural networks, neural network inversion

Introduction

Homelessness is a serious social problem, one which is especially pronounced in large cities such as Los Angeles, with an estimated 36,300 people experiencing homelessness as of January 2019 (LAHSA, 2019). This number is even larger, at 58,936, for the entire Los Angeles County. In (Elliott and Krivo, 1991), the authors conducted a study across large municipal districts, with domiciled populations ranging from 50,000 to over 1 million, and identified lack of affordable housing units, high poverty rates, poor economic conditions, and limited mental health care facilities as potential factors contributing to homelessness. They also found that investing in residential beds and mental health care may decrease homelessness. In another study (Quigley, 1990), the author summarized many factors that correlate (positively or negatively) with the homeless population on a city-by-city basis. Some positive correlates included the domiciled population, poverty and unemployment rates, along with rent. Number of vacant units was negatively correlated with homeless populations. Some studies have case reports from homeless people stating that being unable to pay rent or property damage (Anderson and Christian, 2003) has lead to homelessness. Indeed (Glynn and Fox, 2019) finds that median rent is correlated with homelessness. The authors also find that Los Angeles is among the cities with the strongest correlations between homelessness and rental prices. But the problem of homelessness is more complicated than economics alone as many become homeless as a result of mental illness (Susser et al., 1993), post-traumatic stress syndrome (Taylor and Sharpe, 2008), domestic abuse, drug/alcohol abuse, child abuse/neglect (Herman et al., 1997) (Rew et al., 2001), among many other tragic events. There are yet still more complex causes for changes in the homeless population, sometimes owing to 'greyhound therapy', whereby homeless people are given a one way ticket on a bus from one city to another (Hyde, 1985). Even the temperature of a city (Metraux et al., 2016) has been found as a positive correlate for the homeless population. Given the myriad of complex factors at work, we limit our focus to identifying geographic features of Los Angeles (on the scale of census tracts) that can be used in predicting changes in the homeless population living on the streets and in vehicles. This is distinct from other studies that examine such patterns over the scale of cities. Also, different to many previous studies on homelessness to date, we will be looking at specific subsets of the homeless population: those living on the streets and those living in vehicles.

In the last week of January each year, the Los Angeles Homeless Services Authority (Cox et al., 2017) (lat, 2019b) estimates the number of homeless people living in each census tract and makes

Corresponding author:

Michael Lindstrom, Department of Mathematics, University of California Los Angeles, USA Email: mikel@math.ucla.edu

¹Department of Mathematics, University of California, Los Angeles

²Center for Applied Mathematics, Cornell University, Ithaca, New York

³Department of Physics & Astronomy, University of California, Los Angeles

⁴School for Independent Learners, Los Altos, California

⁵Department of Statistics, University of California, Los Angeles

⁶Department of Psychology, University of California, Los Angeles

⁷Department of Mechanical and Aerospace Engineering, University of California, Los Angeles

⁸Department of Anthropology, University of California, Los Angeles

this data publicly available. In general, the problem of taking an accurate reading of the total homeless population is extremely difficult. Many authors believe the counts are underestimates (Phelan and Link, 1999) (Berry, 2007), even suggesting upwards of 70% may not be counted in some areas. However, the methodology of LAHSA does mitigate other potential problems by limiting the count to a narrow time window, thus reducing double-counting. In our study, we interpret the counts as proxies for the true homeless population, which should generally increase/decrease, respectively, if the true homeless population increases/decreases. Accurate modelling of the homeless population counts has recently been considered in (Glynn and Fox, 2019), but such methodologies are beyond the scope of our study.

A deeper understanding of factors influencing the homeless population on a small scale, not just entire cities, is very important. Currently, complaints about homeless encampments are among the top reasons Los Angeles residents phone for city services (lat, 2019a). Public health is also a large concern regarding surging homeless populations (Smith, 2019). In March 2017, Measure H was adopted by the voters of Los Angeles County and brought into effect for the 2018-2019 Fiscal Year (Chou, 2018) (County,). Through a small tax increase of 0.25%, the county hopes to use these funds to alleviate homelessness through outreach programs, developing more shelters, helping people to find jobs, and preventing people in danger of becoming homeless from becoming homeless in the first place. Being able to pinpoint locations in the city where more resources are needed could be very beneficial for these interventions. And by better understanding the dynamics of the homeless population, interventions such as the aforementioned can be better evaluated.

Besides the homeless count estimates provided by the LAHSA, through public databases and web crawling, we obtained estimates for a myriad of other features of each census tract, such as the median cost of rent, median household income, number of coffee shops, number of bus stops, crime rate, etc. Our work begins by focusing upon the features of a census tract and how they relate to homelessness at a descriptive level; we then make some preliminary steps towards forecasting future homeless populations in a census tract.

We study a subset of Los Angeles city containing 898 census tracts with mean and median areas of 0.43 mi² and 0.29 mi², respectively. From the US Census Bureau, the mean domiciled population of these census tracts is approximately 4070, with a median of approximately 3900. Over the years considered in our study, each census tract had a mean of 11 homeless people living on the streets, 7 living in vehicles, and 6 living in shelters. The respective median values were 4, 4, and 0. The variances are very large, however, with respective standard deviations of 30, 11, and 31, showing that some census tracts have very large homeless populations.

We use multiple methods in our study. Figure 1 provides a summary of the research and how the techniques fit together. The remainder of the paper is organized as follows: we begin by describing our dataset. Then we explain the techniques used for the study. Following this, we present our findings, provide discussion, and conclude our work. A reader primarily interested in the results and with a knowledge of basic machine learning could move directly from here to the Results section.

Data

Table 1 summarizes the data we use in this study. See the Supplementary Material for full details.



Figure 1. The study consists of a descriptive component with correlation analysis and topic modelling and a prescriptive component with neural networks. First, we use correlation analysis to identify the factors most correlated or anticorrelated with the homeless counts, being able to compare our results with those of other studies on larger scales. We then use topic modelling via nonnegative matrix factorization to describe each census tract as a weighted mixture of different descriptors such as "high affluence". Through examining the relative presence of each topic geographically, we observe clear correlations and anticorrelations between certain topics and the homeless population. Then, to predict how the homeless population of a census tract will change over time, we use artificial neural networks to predict a ternary output, whether the population change will be in the lower, middle, or upper tertile (third of changes), roughly equating to decreasing, static, or increasing population counts, respectively. The most successful neural networks, when examined, allow us to identify combinations of features that may be more likely to yield a given change in the homeless population.

Table 1. We collected publicly available data on several different regional features within Los Angeles. These characteristics include homeless population counts, physical features, housing costs, reported crimes, and economic and demographic data. Data came from different geographic zones: zip codes, census tracts, and PUMA zones. Some data was annual and some data was averaged over 5 years. In general, some approximation and processing was necessary in order to prepare data for use in our models and some features such as coffee shops were treated as static. The table lists of all variables considered in this study with their shorthand and meaning. The homeless populations, static features, and dynamic features are separated by horizontal rules. Years indicates years for which we have data.

totVehiclePeople2015-2017homeless population living in vehiclestotSheltPeople2015-2017homeless population living in shelterstotStreetPeople2015-2017homeless population living on streetsCoffeenumber of coffee shopsRestaurantCountnumber of restaurantsGenPopulation2015estimate of domiciled population	Shorthand	Meaning
totSheltPeople2015-2017homeless population living in shelterstotStreetPeople2015-2017homeless population living on streetsCoffeenumber of coffee shopsRestaurantCountnumber of restaurantsGenPopulation2015estimate of domiciled population	totVehiclePeople2015-2017	homeless population living in vehicles
totStreetPeople2015-2017homeless population living on streetsCoffeenumber of coffee shopsRestaurantCountnumber of restaurantsGenPopulation2015estimate of domiciled population	totSheltPeople2015-2017	homeless population living in shelters
Coffeenumber of coffee shopsRestaurantCountnumber of restaurantsGenPopulation2015estimate of domiciled population	totStreetPeople2015-2017	homeless population living on streets
RestaurantCountnumber of restaurantsGenPopulation2015estimate of domiciled population	Coffee	number of coffee shops
GenPopulation2015 estimate of domiciled population	RestaurantCount	number of restaurants
	GenPopulation2015	estimate of domiciled population
Busstops number of bus stops	BusStops	number of bus stops
AffordableHousingUnits number of affordable housing units	AffordableHousingUnits	number of affordable housing units
LibraryCount number of libraries	LibraryCount	number of libraries
RalphsCount number of Ralphs grocery stores	RalphsCount	number of Ralphs grocery stores
TraderJoesCount number of Trader Joe's grocery stores	TraderJoesCount	number of Trader Joe's grocery stores
WholeFoodsCount number of Whole Foods grocery stores	WholeFoodsCount	number of Whole Foods grocery stores
ZRI2014-2017 Zillow Rent Index	ZRI2014-2017	Zillow Rent Index
ZHVI2014-2017 Zillow Home Value Index	ZHVI2014-2017	Zillow Home Value Index

Table 1 continued from previous page					
PumaVeteranPctofCivPop2014-2016	PUMA percentage of veterans living as civilians				
PumaNonVeteranPctofCivPop2014-2016	PUMA percentage of non-veterans				
PumaMedFamilyIncome2014-2016	PUMA median family income				
PumaMedHouseholdIncome2014-2016	PUMA median household income				
PumaPovertyRatio2014-2016	PUMA percentage living below federal poverty threshold				
PumaMedGrossRent2014-2016	PUMA median gross rent				
PumaMedValue2014-2016	PUMA median property value of a housing unit				
PumaMedGrossRentPctIncome2014-2016	PUMA median percentage of income going towards rent				
PumaMedMonthlyHousingCosts2014-2016	PUMA estimate of monthly housing costs including mortgages and rent				
PumaUnemploymentRate2014, 2015, 2016	PUMA percentage unemployed				
PumaNotLaborForceRate2014-2016	PUMA percentage not in labor force				
PumaPctIns2014-2016	PUMA percentage with medical insurance covered by employer or a union				
ParkingCitationCount2015-2017	number of parking citations				
CrimeCount2014-2017	number of reported crimes				
TotalHousingUnits2014-2016	5-year average of number of homes available				
TotalVacantUnits2014-2016	5-year average of how many vacant units are present				
UnemploymentRate2014-2016	5-year average ending in given year of unemployment rate				
NotLaborForceRate2014-2016	5-year average ending in given year of percentage not in labor force				
BelowPovertyRate2014-2016	5-year average ending in given year of percentage living below federal poverty				
MedRentAsPercentOfIncome2014-2016	5-year average ending in given year of percentage income going towards rent				
TotalPopulation2014-2016	5-year average of total general population				
MedHouseholdIncome2014-2016	5-year average ending in given year of median household income				
MedFamilyIncome2014-2016	5-year average ending in given year of median family income				
MedRent2014-2016	5-year average ending in given year of median rent				
MedValue2014-2016	5-year average ending in given year of median property value of a housing unit				
MedMonthlyHousingCosts2014-2016	5-year average ending in given year of monthly housing costs including mortgages and rent				
VeteranPctofCivPop2014-2016	5-year average ending in given year of percentage of veterans living as civilians				
NonVeteranPctofCivPop2014-2016	5-year average ending in given year of percentage of non-veterans living				
EmployerInsPct2014-2016	5-year average ending in given year of percentage with medical insurance covered by their employer or union				

Methods

Here we provide an overview of the techniques used in this research. In general we have a data matrix with n records and m features, $\mathcal{X} \in \mathbb{R}^{n \times m}$, whose entry in row i and column j, is the value of the j^{th} feature for record i (a given census tract). Notationally, for a matrix A we denote $A_{i:}$ to be its i^{th} row and $A_{:j}$ to be its j^{th} column.

Correlation Analysis

A correlation coefficient $r \in [-1, 1]$ is a measure of how linearly correlated two features X and Y are (Johnson and Kuby, 2011). In our work, we examine the correlations between each feature and each of: the homeless population in vehicles and the homeless population in the streets. Due to the lack of accurate,

publicly available data on homeless shelters (see remarks in the Supplementary Material), we opt to not model the sheltered homeless population and focus only on the street and vehicle homeless populations.

Topic Modelling

Topic modelling is a means of describing each record of a data set as a mixture of overarching themes. In our data, for example, we may describe the data of a census tract as being a mix of "high commerce" and "high affluence". We make use of nonnegative matrix factorization (NMF) to achieve this. NMF has similarities to Principal Component Analysis (Jolliffe and Cadima, 2016), but in topic modelling, the decomposition forces all entries to be nonnegative.

Given our data matrix \mathcal{X} with nonnegative entries, we select a desired number of topics k and seek matrices $\mathcal{W} \in \mathbb{R}^{n \times k}$ and $\mathcal{H} \in \mathbb{R}^{k \times m}$ with nonnegative entries so that $\mathcal{X} \approx \mathcal{WH}$. This approximate factorization amounts to writing the i^{th} record

$$\mathcal{X}_{i:} \approx \sum_{j=1}^k \mathcal{W}_{ij} \mathcal{H}_j$$

as a linear combination of k topics $\mathcal{H}_{1:}, ..., \mathcal{H}_{k:}$ weighted by $\mathcal{W}_{i1}, ..., \mathcal{W}_{ik}$. We refer to \mathcal{W} as the weights matrix and \mathcal{H} as the topic matrix where the rows represent each topic.

Artificial Neural Networks

An artificial neural network is a means of generating a nonlinear fit for a set of data. We refer the reader to (Bishop et al., 1995) for a thorough background. In trying to understand the homeless population, our neural network will receive an input of many features of a census tract (median income, number of coffee shops, etc.) and then predict whether the change in homeless population of that census tract will be in the lower, middle, or upper tertile.

More precisely, we allow an input $x^{(0)} \in \mathbb{R}^m$ where there are *m* features for $x^{(0)}$. We then take a linear combination of these features and add a bias term, which can intuitively be thought of as giving different weights of importance to different features as part of a decision-making process. There is an additional step of applying a nonlinear function to this result; this is necessary to prevent the entire model from being a linear regression and also to prevent numerical overflow errors. Mathematically, the input is transformed according to

$$z^{(1)} = W^{(1)}x^{(0)} + b^{(1)}$$
⁽¹⁾

for a (weight) matrix of coefficients $W^{(1)}$ and (bias) vector $b^{(1)}$, and then passed through a nonlinear function h to yield

$$x^{(1)} = h(z^{(1)}). (2)$$

This process can be repeated until after L iterations we arrive at $x^{(L)} \in \mathbb{R}^{m'}$. We find the best results occur with a shallow neural network with L = 1 and with a soft-max nonlinearity. When trying to classify the outcome of the population change, m' = 3, and we seek outputs as close to $(1, 0, 0)^T$, $(0, 1, 0)^T$, and $(0, 0, 1)^T$, as possible to indicate the respective possibilities of a first, second, and third tertile change. The network will classify the tertile of change based on the largest component of $z^{(1)}$. In a sense this ternary classification is a generalization of a logistic regression where the decision is no longer binary.

Neural Network Inversion and Structure

A neural network is a nonlinear mapping from the *m*-dimensional feature space \mathbb{R}^m to an *m'*-dimensional output space, $\mathbb{R}^{m'}$. Often an inverse does not exist. However, a neural network can be inverted in the sense of a constrained optimization problem (Lu et al., 1999). More colloquially, there could be many sets of features that yield the same neural network prediction and so the notion of determining which input yields a given output is badly defined. However, by adding more restrictions to the input, there is a way to ascribe a unique input to a given output. In this work, we look for the smallest deviation from average, with the fewest parameters changed from their average value, that results in a given prediction.

Techniques of deep Taylor series (Montavon et al., 2017) and layer-wise relevance propagation (Bach et al., 2015) are also often used for deep neural networks to gain insights into what a trained network finds to be most important in making a classification decision. We are also able to gain similar insights, but due to the simplicity of our fitting, this machinery is not necessary. Our classification task is nonlinear, but because the pre-activated states of the output neurons are linear with respect to the input features and the network will classify a population change based on the largest component of $z^{(1)} \in \mathbb{R}^3$ from (1), there is a well-defined feature distribution in parameter space that steers a trained neural network towards a given prediction. For i = 1, 2, 3, denoting the tertile of predicted change,

$$\nabla z_i^{(1)} = W_{i}^{(1)} \tag{3}$$

denotes this direction: for $\alpha > 0$ sufficiently large, an input $\alpha W_{i:}^{(1)}$ will result in a tertile classification *i*. If a particular relation among the features is "strong" enough, such as the the number of homeless present, along with other key signatures, we can know with certainty what the neural network will predict.

Results

Strongly Correlated/Anticorrelated Features

The correlation results are tabulated in the Supplementary Material. We correlate the 2017 populations with the static features and with the dynamic features of 2016. Our interpretation is that the overall statistics of the 2016 year should correlate with the 2017 population at the start of the year. We find the street and vehicle homeless populations are both correlated with metrics of disadvantage such as poverty rates, and anticorrelated with metrics of affluence like higher incomes. The street population also has a correlation with commerce (coffee shops and restaurants). There is a strong correlation with the local crime rate for both populations. This correlation does not mean the homeless are causing the increased crime - it only suggests an association.

Putting our results into context, many patterns are consistent with the correlations summarized in (Elliott and Krivo, 1991). We also find the homeless population is correlated with increases in measures of disadvantage, but in our analysis the street homeless population is correlated with number of affordable housing units. Comparing our results with (Quigley, 1990), there are again many similarities, however, the vacancy rate is positively correlated with the street homeless population and the rent is negatively correlated, opposite to the study cited. This could stem from the differences between a large-scale city-by-city analysis and a local analysis as we are doing here. Heuristically this seems plausible. Taking vacant units as an example, people in a city with a large number of vacant units relative to other cities may find it easier to find a dwelling; but, on the scale of census tracts, homeless individuals may be driven

out of areas with a higher housed population. The causal mechanism could also run the other way: the housed population may choose not to live in an area with a high homeless population. It is worth noting that the study cited collected total homeless counts, not just street homeless. There is also the concept of "ecological fallacy" (Piantadosi et al., 1988), which cautions against making inferences about individual characteristics from statistics of a large group; it may not be appropriate to assume the features of Los Angeles' homeless population will be consistent with the overall trend observed among many cities.

Topics and the Homeless Population

Features of the census tracts do not exist in isolation and in general a census tract will be a mixture of different groups of characteristics. Using all static features and the time dynamic features of 2016, we generate 3 topics illustrated in figure 2. Figure 3 places the topics and homeless populations on a map. We also compute the correlations between the homeless street counts, homeless vehicle counts, and each of the three topics. See Supplementary Material.

Neural Network Predictions

We wish to predict whether the change in population at a census tract will be in the first, second, or third tertile. As seen in the Supplementary Material, a first, second, and third tertile change correspond roughly to a decreased, unchanged, and increased homeless population in an area, respectively. This division also ensures that the datasets are well-balanced, having an equal portion of each possible result. We denote '-' for a change in the first tertile; '0' for a change in the second tertile; and '+' for a change in the third tertile. The technical details of how the network is set up are found in the Supplementary Material.

We utilize the static and dynamic features of table 1 (except for the PUMA fields) to predict the population changes. To predict the change in population for the year Y to Y+1, we use: the static features, the dynamic features of years Y-1 and Y, and the street, vehicle, and sheltered populations at the start of year Y. To avoid simply following a trend, we *do not include the homeless populations in year Y-1*. Confusion matrices that illustrate the breakdown of what the networks predict along with the true observations are found in the Supplementary Material. The predictive ability of the neural networks can be found in table 2. Understanding the decision-making process of the neural network is studied in the next section.



Figure 2. Topics formed using the static features and all 2016 dynamic features. As some features vary considerably in magnitude over census tracts, we normalize the features to their percentile-rank from 0 to 1. We perform NMF with k = 3 topics. The lengths of each bar represent the strength of a given feature. We identify topic 1 as "high affluence", topic 2 as "concentrated disadvantage", and topic 3 as "high commerce." These labels are not perfect descriptors. For example, within disadvantage, there are a modest number of restaurants/coffee shops. Also, these topics are not isolated and a given census tract could be described by combining these topics in different amounts. The correlations between the topics themselves are also quite insightful (see Supplementary Material), giving some understanding into the layout of Los Angeles. Unsurprisingly, affluence and disadvantage are anticorrelated, suggesting that areas of affluence and disadvantage, suggesting that a region can be either high commerce or highly disadvantaged, but not both.



Figure 3. Topics and homeless population counts over Los Angeles. The colours are as follows: red represents the upper tertile of values; yellow represents the middle tertile; and blue represents the lower tertile. Top row: heat map of 2017 street and vehicle population densities. Bottom row: heat map of topic weightings from \mathcal{W} by location; topics 1, 2, and 3 are left, middle, and right, respectively. Topics are generated from the static features and 2016 dynamic values. See Supplementary Material for correlation values. It appears that both the numbers of homeless people living on the streets and in vehicles are correlated with disadvantage, anticorrelated with affluence, and uncorrelated with commerce. In this regard, the topic modelling differs from the correlation analysis of individual features and the street homeless population: in the correlation analysis, there was a positive correlation with individual proxies for commerce. This may also stem from topic 2 including a small number of restaurants and coffee shops, providing resources for the street homeless, without their needing to frequent more affluent coffee shops and restaurants as may be found with topic 3.

Prepared using sagej.cls

Population	P-	P0	P+	R-	R0	R+	Accuracy
Street	64.6%	54.0%	40.6%	52.9%	70.2%	34.9%	53.1%
Vehicle	64.5%	53.6%	42.9%	57.6%	61.5%	40.9%	53.4%

Table 2. Predictive ability of the trained neural networks averaged over 1000 trials. Since our dataset is limited in size, 898 census tracts per year, as a means of data augmentation, for the street and vehicle homeless populations separately, we consolidated the homeless population changes of 2015-2016 and 2016-2017 into a single dataset. Thus the 4 datasets (2 populations and 2 annual changes) are turned into 2 datasets (2 populations) each with 1796 records to train and test with. The suffix '-' denotes a first tertile, '0' indicates a second tertile, and '+' indicates a third tertile change. 'P' denotes a neural network's precision (its correctness given its prediction) and 'R' denotes a neural network's recall (its correctness given the ground truth). The accuracy is the overall correctness (number of correct predictions out of all tests). Because the dataset is balanced, the accuracy and recall should be $33.\overline{3}\%$ if the network had no predictive power and relied upon chance alone. The network does have predictive powers, although there is greater difficulty in accurately predicting and recalling '+' events.

Understanding the Neural Networks

To understand what features the neural networks determine to be most significant in predicting population changes and to understand what circumstances yield a given prediction, we select the 3 top performing neural networks for both street and vehicle homeless predictions. We label these trained neural networks A, B, and C, for the street homeless, each with accuracy 59%; and we label them D, E, and F, for the vehicle homeless, with accuracies of 60%, 59%, and 59%, respectively. These neural networks perform significantly better than the average values: see table 2.

Due to the stochastic nature of training neural networks and selecting their training data, two trained networks may be equally effective at predicting outcomes but have different structure. This is especially true in our datasets where there is a large uncertainty in the features. However, commonalities found between neural networks may illustrate important patterns. Despite structural differences, the networks tend to yield the same predictions: of the 1796 (after augmentation) possible predictions, A, B, and C all predicted the same outcomes for 1171 of them; and D, E, and F all predicted the same outcomes for 1266 of them.

Inversion of the Neural Network We would like to gain insight into the "thought processes" of the neural network. We now ask the following question: if all features are at their mean value, how can we select the fewest number of features and vary them by the least total amount to force the networks to make a given prediction? This is addressed by a minimal ℓ^1 -norm (Boyd and Vandenberghe, 2004) with a constraint (the prediction) - see the Supplementary Material. In figure 4, we provide a high level summary of the results.



Figure 4. A summary of the smallest changes that result in the top neural networks making a given prediction. All features that are not explicitly listed are understood to be at the mean value. Note that all of neural networks D, E, and F predict that if all features are at the mean value, the vehicle population will increase. This is not the case for neural networks A, B, and C. A common pattern appears to be that an above average homeless count one year will decrease into the next. And a significantly below average count will not change. Significantly below average counts in year Y likely indicate populations of zero (since the populations can't be negative). If a region has a homeless population of zero, it may not be a viable place for a homeless person to live making changes in the population over that region unlikely.

Key Feature Signatures to Yield a Given Prediction The feature signatures that steer a neural network towards a given classification can be computed for networks A-F. In figure 5, we focus solely on the street population. The vehicle feature signatures are found in the Supplementary Material. The bars are the corresponding weights in the matrix W described in equation (3). The more similar a census tract's features are to these weights and the larger those feature values are, the more likely the neural networks are to predict a given population change.

Note that a feature such as "Below Poverty Rate Y" above may appear with the same sign in different cases that yield different predictions. In such cases, those features are not particularly significant; this merely indicates that signature is present in multiple conditions. An analogy could be made with directions: Northeast and North by Northeast are different directions (feature distributions) that ultimately lead to different destinations (predictions). But both directions do have a component of "North" in them (like "Below Poverty Rate Y" having the same sign in all predictions in figure 5).



Figure 5. Top: dominant distribution of features (as z-scores) so that a change in street count in the lower tertile of changes is predicted with neural networks A-C. Middle: dominant distribution of features (as z-scores) so that a change in street count in the middle tertile of changes is predicted with neural networks A-C. Bottom: dominant distribution of features (as z-scores) so that a change in street count in the upper tertile of changes is predicted with neural networks A-C. Bottom: dominant distribution of features (as z-scores) so that a change in street count in the upper tertile of changes is predicted with neural networks A-C. The features that appear have been culled so that all three neural networks have the same sign in their respective components and where the magnitude is at least 0.1.

Isolating the Most Significant Features in the Neural Networks Here we try to come up with interpretations for the behaviour of the neural networks after training. Due to each direction involving many pieces, it is hard to isolate one or two features that are most significant. It is interesting to note, however, that consistent with the smallest ℓ^1 -perturbations, the more above average a population is, be it street or vehicle, the more it may tend to a decrease in the population: among the 452 records where the street population in year Y was above the mean, the population change fell into the '-' classification in 295 of them the next year. And among the 588 records where the vehicle population in year Y was above the mean, the population change was '-' in 378 of them the next year. Using a χ^2 -test in assuming that one-third of the 452 and 588 records for street and vehicle homeless populations that began with an above average population should result in a '-' change, the deviations are significant at the $p = 5 \times 10^{-47}$ and $p = 4 \times 10^{-57}$ levels, respectively.

The '+' cases are more difficult. By looking at individual patterns present among the top neural networks, we do find that the sheltered homeless population and the number of shelters may be significant. The most promising results, which are not significant, are that among the 168 records with an above average sheltered homeless population, the street homeless population rose in 64 of them (significance is p = 0.2). And among the 112 records with an above average shelter count, the vehicle homeless population rose in 46 of them (significance is p = 0.08). Given the weakness of the significances, random chance cannot be ruled out to describe these predictors.

We speculated that if a given homeless population is significantly below average in a census tract, it may be more likely there is no change going into the next year. We test this idea in a similar spirit as above. There were 226 records that had no change in the street homeless population from year Y to Y + 1 (a fraction $f_s = 0.13$) and among the 374 records with a zero homeless street count in year Y, 179 had no change from year Y to Y + 1. A χ^2 -test with expected fraction of records having zero change is f_s , the result is significant at the level $p = 5 \times 10^{-94}$. Likewise, 258 records had no change in the vehicle homeless count between years (a fraction $f_v = 0.14$). What we observe is that among the 542 census tracts with a zero homeless vehicle count in year Y, 243 had no change from year Y to Y + 1. The significance is $p = 6 \times 10^{-91}$. It does appear that a census tract with zero homeless population is less likely to change.

Discussion

Speaking very broadly, the correlation analysis and topic modelling were of a more qualitative nature, illustrating trends and patterns within the homeless population at large, without taking into account temporal variations. Both the correlation analysis and topic modelling illustrate that homeless populations are more likely to be found in regions of concentrated disadvantage and less likely to be found in areas of higher affluence.

The neural network studies aimed to make quantitative predictions about the changes in the homeless population. In this way, we can see how features may be linked with a given change in the homeless population. The street and vehicle populations were both likely to decrease the following year if the respective populations were above average at the start of the year. As pure speculation, this may stem from social interventions offering assistance; the highly correlated increased crime driving homeless people away; the additional influence of law enforcementdispersing the homeless; or perhaps the homeless populations decrease due to crowding effects. Census tracts that begin with a population of zero for a given homeless population are indeed more likely to remain at a population of zero. We hypothesize this may be due to limited resources in an area keeping the homeless population near zero.

Conclusion

Our focus has been on studying homelessness on small geographic scales. We found the homeless population on a small scale has distinct correlates from similar analyses done comparing between cities. We found that different groups within the homeless, those living in vehicles and those living on the streets, have regions of high and low total populations that can be matched qualitatively to a given type of neighourhood. Both populations are found in areas of concentrated disadvantage.

We also found a degree of population change can be predicted with neural networks by taking into account localized features, even with a lot of noise and uncertainty in the data. Through examining the neural networks, we have found the street and vehicle homeless populations may be more likely to decrease from one year to the next if they begin at an above average count and that regions that begin with zero homeless may be more likely to remain so. The use of "may" in all these conditions is important as neural networks are also capable of inaccurate predictions. The patterns observed could also reflect the presence of hidden variables that are not directly documented in our datasets.

Our methodology does have predictive power, but the problem is obviously very complicated. There are a range of predictive powers, with predictions of population increases being the weakest. More data, more accurate data, or different methods may be required for deeper understanding. Our predictive study has found associations between features and changes in the homeless populations, but we have not identified causal mechanisms. Identifying causal mechanisms is a much harder problem and is left as future work. If there are causal mechanisms, they are likely indirect. For organizations who work directly with the homeless, these findings could be useful to flag regions in the city where more resources or attention will be required with higher or lower probability. However, we believe this work is really just a first step towards a more thorough study to understand the factors that lead to homelessness, to identify the most important features, and to arrive at viable solutions.

Supplementary Material

Data

The Los Angeles Homeless Services Authority (LAHSA) provides data on counts of homeless populations by census tract, including people living in vehicles, living in shelters, and living on the street. Locations and counts of physical features were obtained through public Los Angeles city and county databases, The Yellow Pages, and the Google Places API. Features include number of homeless shelters and services, restaurants, coffee shops, bus stops, grocery stores, libraries, and affordable housing units. The real estate website Zillow provided measurements of rent and housing costs via the Zillow Rent Index (ZRI) and Zillow Home Value Index (ZHVI). Data on reported crimes and parking citations were obtained through public Los Angeles city databases.

The American Community Survey (ACS) was another source of economic and demographic data, including median household income, median rent, unemployment rates, total population, total housing units, percentage of population who served in the military, and other variables. We were interested in two forms of ACS data released by the United States Census Bureau: 5-year estimates, which are based on

survey responses averaged over five years prior to their release year, and Public Use Microdata Samples, which include actual survey responses geographically coded to a Public Use Microdata Area (PUMA) from a given year. PUMA zones are much larger than census tracts: a PUMA contains at least 100,000 domiciled individuals, whereas in our dataset for Los Angeles County, each census tract has no more than 12,000 domiciled individuals. Assigning estimates from a PUMA to all census tracts within that PUMA runs the risk of flattening variations between census tracts within PUMAs, but the PUMA data come from only one year. The ACS 5-year estimates allow for data at the geographic level of a census tract, but averaging over time runs the risk of obscuring changes from year to year. In this sense, ACS PUMA estimates maintain temporal resolution but lack spatial resolution, while the ACS 5-year estimates maintain spatial resolution but lack temporal resolution. Some variables inherently require more temporal or spatial specificity than others. Both the ACS PUMA and ACS 5-year estimates are imperfect measures for economic and demographic characteristics of census tracts, so our data may contain varying degrees of redundancy between the PUMA estimates and the 5-year estimates.

Since many of the physical features, such as restaurant locations, did not have reliable start and end dates, we assume them to be static variables over the time period 2015-2017. For some databases, location data was missing for a significant proportion of entries, and some data sources were incomplete or inconsistent. For other datasets, values were interpolated from zip codes to census tracts. We approximate the centers of census tracts using their centroids and calculate distances between census tracts or a point and a census tract as the Euclidean distance between the point and the centroid of the census tract. Some variables come from Los Angeles city databases. We limited analyses using these variables to census tracts set of census tracts for which we had data over all variables in order to make use of greatest number of variables.

In addition to these issues with explanatory data we collected, past work in social sciences and statistics literature suggest some uncertainty with point-in-time counts of homeless populations themselves, particularly the undercounting of unsheltered homeless populations (Phelan and Link, 1999). LAHSA also frequently issues corrections or updates based on methodology changes from year to year (Los Angeles Homeless Services Authority, 2018).

We used various techniques, assumptions, and extrapolations to process the data for use in our models. We outline these methods below.

- *Census tracts*: All data were organized by geographic regions called census tracts, as defined by the 2010 U.S. Census and shapefiles provided by Los Angeles County (Los Angeles County, 2012). Distances between census tracts were estimated as distances between centroids of tracts.
- Homeless population counts: Data on homeless populations in 2015 (Los Angeles Homeless Services Authority, 2015), 2016 (Los Angeles Homeless Services Authority, 2017) (Los Angeles Homeless Services Authority, 2017) was provided by the LAHSA. In particular, we calculated counts of people living in the street, people living in vehicles, people living in shelters, and the total homeless population for each census tract for each year from 2015-2017. The 2015 counts were provided according to census tracts defined by the U.S. Census Bureau in the year 2000. In order to include homeless population counts from 2015 in our analysis, it was necessary to convert these values to 2010 census tracts. Assuming uniform distribution of homeless populations within each tract, we calculated population count estimates based on the proportion of land area

shared between the 2000 tract and the 2010 tract, according to census tract relationship files (U.S. Census Bureau, 2017a).

- Zillow indices: Data from the Zillow Rent Index (ZRI) and Zillow Home Value Index (ZHVI) were calculated from Zillow zip code data (Zillow, 2017). Monthly values were averaged over each year from 2014-2017, and the census tract data were calculated by locally averaging over all zip codes with centroids falling within a chosen radius of the census tract centroid.
- *Parking citations*: Data on parking citations come from a Los Angeles city database, accessed through DataLA (Los Angeles Department of Transportation, 2017). Between 5 and 25 percent of the citations from each year from 2015-2017 did not have a precise location available. By assuming the citations with locations were a representative sample of all citations in a given year, we estimated the yearly count of citations in each census tract with an appropriate scaling factor to approximate totals over the entire year.
- ACS 5-year estimates: Data from the American Community Survey 5-year estimates were accessed through the American Factfinder portal. The 5-year estimates are released every year but contain data from the previous four years (U.S. Census Bureau, 2017d). The difference between two 5-year estimates would therefore come from the first year of the earlier estimate and the last year of the later estimate. For example, the difference between the 2011-2015 ACS estimates and the 2012-2016 ACS estimates comes from 2011 and 2016. We collected data from the 2010-2014, 2011-2015, and 2012-2016 ACS 5-year estimates. In some entries, the ACS uses a '-' or '+', respectively, at the end of a median estimate to indicate that the median is in the respectively lowest or highest interval of an open-ended distribution (U.S. Census Bureau, 2017d). In such cases, we used the median estimate given. For most analyses, we restricted our analysis to only the census tracts for which an estimate was available across all categories. Variables from the ACS 5-year estimates include: total housing units, total vacant units, the unemployment rate, the percentage of people not in the labor force, the percentage of people living below the federal poverty threshold, median rent as a percentage of income, total population, median household income, median family income, median rent, median home value, median monthly housing costs, the percentage of civilians who are veterans, and the percentage of people receiving medical insurance covered by their employer or union. Further information can be found in the American Community Survey Subject Definitions from the U.S. Census Bureau.
- ACS PUMA estimates: Data from the American Community Survey 1-year Public Use Microdata Samples (PUMS) were accessed through the American Factfinder portal. We used the PUMS housing and population responses and weights to calculate weighted means (U.S. Census Bureau, 2017c) and medians of particular categories for each Public Use Microdata Area for each year from 2014-2016. PUMAs include approximately 100,000 people and are generally much larger than a census tract (U.S. Census Bureau, 2017e). By examining PUMA shapefiles to determine whether the centroid of a census tract falls within a PUMA, we assigned data to each census tract from its corresponding PUMA. Variables from the PUMA estimates include: percentage of civilians who are veterans, median family income, median household income, percentage of people living below the federal poverty threshold, median gross rent, median property value, median gross rent as a percentage of income, median monthly housing costs, unemployment rate, the percentage of people not in the labor force, and the percentage of people receiving health insurance through a

current or former employer or union (U.S. Census Bureau, 2017b). Further information can be found in the American Community Survey PUMS ReadMe from the U.S. Census Bureau.

- Affordable housing units: Data on affordable housing units managed under the Los Angeles Housing and Community Investment Department come from a Los Angeles city database (Los Angeles Housing and Communicty Investment Department, 2017), accessed through DataLA. This includes locations listed from 2003 through December 2017. Some locations were listed as under construction or did not list the number of units available; these were discarded. The remaining units were counted as being within a census tract based on latitude/longitude coordinates. If not already available, these were calculated from an address using Google Geocoding API (Google, 2017a).
- *Bus stops*: Data on Los Angeles County Metro bus stop locations come from a Los Angeles county database, accessed through the Metro Developer website (Los Angeles County Metropolitan Transportation Authority, 2017). This includes physical bus stop locations from an earlier version of the bus stop data, listed June 17, 2017. These stops were counted as being within a census tract based on latitude/longitude coordinates.
- *Coffee shops*: Data on coffee shops within Los Angeles County come from calls to the Google Places API as of October 2017 (Google, 2017b). These stops were counted as being within a census tract based on whether or not latitude/longitude coordinates were contained within a census tract. Although the Google Places API is not a comprehensive database, these counts provide an estimate of the count of coffee shops within each census tract.
- 2015 domiciled population: Data on housed population estimates by census tract as of 2015 come from the Los Angeles County Department of Mental Health, accessed through the County of Los Angeles Open Data portal (County of Los Angeles Department of Mental Health, 2017).
- *Reported crimes*: Data on yearly reported crime counts by census tract come from the Los Angeles Police Department, accessed through DataLA (Los Angeles Police Department, 2017). A small percentage of these (< 0.5%) were missing latitude/longitude location coordinates. If address information was available, then latitude/longitude coordinates were calculated using the Google Geocoding API(Google, 2017a); otherwise, those crime records with missing locations were discarded. Reported crime counts were calculated for each year from 2014-2017.
- *Restaurants*: Data on restaurant counts by census tract come from the Los Angeles Office of Finance, accessed through DataLA in October 2017 (Los Angeles Office of Finance, 2017). This represents active restaurants which have not notified the Office of Finance of closure. Some restaurants listed start dates after the date that the data was accessed; these were discarded. If not already available, latitude/longitude coordinates were calculated from an address using Google Geocoding API (Google, 2017a).
- *Homeless shelters*: Data on counts of homeless shelters and services come from the County of Los Angeles Location Management System, accessed through the County of Los Angeles Open Data portal (Los Angeles County Open Data, 2017).
- *Libraries*: Data on counts and locations of public libraries come from a Los Angeles city database, accessed through DataLA (Los Angeles Public Libraries, 2017).
- Grocery stores: Data on counts and locations of grocery stores come from webscraping The Yellow Pages (DexYP, 2017). While not a comprehensive database, these counts provide an estimate of the grocery stores near or within each census tract. Since we were interested in differences between

grocery store chains as well, we searched for locations of different chains, including Ralphs, Trader Joe's, and Whole Foods.

We remark that Google Places API data were not cached: they were used only to generate counts within each census tract.

When necessary, latitude/longitude data was converted to the California State Plane Coordinate System using the MATLAB function "sp_proj.m," found on Mathworks (Stevens, 2010).

In order to use the census tracts with the most data possible, we restricted most analyses to tracts with centroids falling within the boundaries of the City of Los Angeles(Los Angeles County Department of Public Works, 2016) and with valid ACS values.

Sheltered Homeless Population

The sheltered homeless population is also counted by the LAHSA. This population includes those living in shelters, temporary housing, and having hotel/motel vouchers. To study this population effectively, we require accurate information on homeless shelters in an area. The data we found publicly available yielded a correlation near 0 when comparing the counted LAHSA homeless population and the number of available shelters in an area. Through corresponding with LAHSA it seems the available public data on shelters is out of date, including shelters that are no longer in operation. Also, certain types of sheltered homeless, such as those with hotel/motel vouchers, are not included in the LAHSA reports.

Correlation Analysis

Tables 3 and 4 summarize the correlation coefficients for the 2017 homeless population with the static and 2016 dynamic features.

For brevity, the focuses in the correlation analysis are the 2017 homeless populations counts. But we remark that while the number of features with strong enough correlations to be considered is mostly consistent from year-to-year for each population individually, the 2016 vehicle population appears to be an exception, having only 6 features that meet the selection criteria. See table 5. We speculate this stems from the counting methodology, as others have also noted some anomalies in the imputed vehicle population over the years (Flaming and Burns, 2017).

Topic Modelling

Choice of k = 3 Topics The NMF algorithm entails minimizing the Frobenius error (the square root of the sum of the squares of all entries) $||\mathcal{X} - \mathcal{WH}||_F$ over all $\mathcal{W}, \mathcal{H} \ge 0$.

We plot the Frobenius error in using k topics to approximate \mathcal{X} in figure 6. Most of the decrease in error occurs by k = 3 topics and this justifies our choice of topic number.

Topic and Population Correlations

Neural Networks

Tertile Ranges Table 7 lists the population change ranges that result in a given tertile for the street and vehicle homeless populations.

Confusion Matrices The confusion matrices for the trained neural networks are given in figures 7-8.

Feature	Correlation with Street Population
Coffee	0.344203
RestaurantCount	0.410193
AffordableHousingUnits	0.238052
ParkingCitationCount2016	0.492895
CrimeCount2016	0.474599
TotalVacantUnits2016	0.128503
MedRent2016	-0.16607
PumaMedGrossRent2016	-0.15568
MedMonthlyHousingCosts2016	-0.17301
PumaMedMonthlyHousingCosts2016	-0.17635
PumaNotLaborForceRate2016	0.141521
BelowPovertyRate2016	0.230879
PumaPovertyRatio2016	0.224231
MedHouseholdIncome2016	-0.16625
PumaMedHouseholdIncome2016	-0.17167
MedFamilyIncome2016	-0.12757
PumaMedFamilyIncome2016	-0.14144
EmployerInsPct2016	-0.20926
PumaPctIns2016	-0.20953
VeteranPctofCivPop2016	-0.11872
PumaVeteranPctofCivPop2016	-0.16143

Table 3. The correlations for the 2017 street homeless populations that are either above 0.1 or below -0.1. The bar separates the static features from the dynamic features. Like the vehicle homeless population in table 4, the street homeless population is correlated with metrics of disadvantage, such as the number living below the poverty line; and negatively correlated with measires of affluence like median income. There appears to be a correlation with commerce.

Neural Network Structure We ran our neural network in *TensorFlow*. We found best results by using all of the available data, minus the PUMA values. The difference was a marginal 1% decrease in overall accuracy by including PUMA data, but we opted to remove the PUMA data.

We had 47 input features mapping to 3 outputs, with a soft-max activation applied to the linear transformation to produce the final activations. The softmax activation maps $z \in \mathbb{R}^d$ to $a(z) \in \mathbb{R}^d$ via

$$a_i = \exp(z_i) / \sum_j \exp(z_j)$$

We did not use any hidden layers. The objective function was cross-entropy. For each neural network 70% of data were used for training and 30% were used for validation. The data normalization in subtracting the mean and dividing by the standard deviation was done only after a training set had been selected, thereby preventing testing data from influencing the training process. The minibatch sizes were 100.

Feature	Correlation with Vehicle Population
CrimeCount2016	0.190963
ZRI2016	-0.12929
MedRent2016	-0.19354
PumaMedGrossRent2016	-0.13641
ZHVI2016	-0.12742
MedValue2016	-0.14653
PumaMedValue2016	-0.10825
MedMonthlyHousingCosts2016	-0.2015
PumaMedMonthlyHousingCosts2016	-0.1454
UnemploymentRate2016	0.136703
PumaUnemploymentRate2016	0.112522
PumaNotLaborForceRate2016	0.131614
MedRentAsPercentOfIncome2016	0.101068
BelowPovertyRate2016	0.179729
PumaPovertyRatio2016	0.176377
MedHouseholdIncome2016	-0.19993
PumaMedHouseholdIncome2016	-0.1396
MedFamilyIncome2016	-0.20052
PumaMedFamilyIncome2016	-0.13499
EmployerInsPct2016	-0.16388
PumaPctIns2016	-0.10522
VeteranPctofCivPop2016	-0.11436

Table 4. The correlations for the 2017 vehicle homeless populations that are either above 0.1 or below -0.1. Like the street homeless population in table 3, there are positive correlations with measures of disadvantage and negative correlations with measures of affluence. There does not seem to be a correlation with commerce.

Feature	Correlation with Vehicle Population
CrimeCount2015	0.165237
MedMonthlyHousingCosts2015	-0.11971
BelowPovertyRate2015	0.115518
MedHouseholdIncome2015	-0.12218
MedFamilyIncome2015	-0.11742
EmployerInsPct2015	-0.11687

Table 5. The correlations for the 2016 vehicle homeless population that are either above 0.1 or below -0.1. There are very few features that meet the selection criteria.

In figure 9, we plot the cross-entropy loss against the number of epochs used. This was necessary to avoid overfitting and ideally reduce the number of epochs needed. There seems to be minimal overfitting and the benefits of choosing a large number of epochs is minimal; we chose to use 2000 epochs.



Figure 6. How the Frobenius error $||\mathcal{X} - \mathcal{WH}||_F$ changes with the number of topics k. While the error does decrease beyond k = 3, the biggest gains occur by k = 3 topics.

	Vehicle	Street	Affluence	Disadvantage	Commerce
Vehicle	1	0.257626	-0.16499	0.210166	-0.0738
Street	0.257626	1	-0.21014	0.186901	-0.01551
Affluence	-0.16499	-0.21014	1	-0.59118	-0.10787
Disadvantage	0.210166	0.186901	-0.59118	1	-0.64976
Commerce	-0.0738	-0.01551	-0.10787	-0.64976	1

	Ta	ıb	le	6.	٦	The	corre	latio	nsl	between	the	home	less	popul	latic	on d	count	s a	nd	topic	weight	ts.
--	----	----	----	----	---	-----	-------	-------	-----	---------	-----	------	------	-------	-------	------	-------	-----	----	-------	--------	-----

]	Prediction	1
		_	0	+
vec	_	17.5%	3.3%	6.4%
ser	0	5.9%	24.2%	14.7%
Ő	+	9.6%	7.0%	11.3%

Figure 7. A confusion matrix for the ternary classification neural network for the street homeless. The standard deviations of all computed percentages are less than 2% based on 1000 trials.

Population	Tertile	Range
street	1st ('-')	$(-\infty, -1.5]$
	2nd ('0')	(-1.5, 2)
	3rd ('+')	$[2,\infty)$
vehicle	1st ('-')	$(-\infty,-1]$
	2nd ('0')	(-1, 3.5)
	3rd ('+')	$[3.5,\infty)$

Table 7. The population changes for a given tertile. The second tertile overlaps with 0, so it can be thought of as representing a static population.

		I	Prediction	1
_		_	0	+
vec	_	19.4%	3.6%	7.1%
ser	0	5.5%	20.6%	12.4%
0p	+	8.7%	9.3%	13.5%

Figure 8. A confusion matrix for the ternary classification neural network for the vehicle homeless. The standard deviations of all computed percentages are less than 2% based on 1000 trials.



Figure 9. The training and testing losses as the street/vehicle neural networks are trained. Each point was averaged over 50 trials. We find the testing losses are best for a 20,000 epochs for the street population and 20,000 for the vehicle population. The improvement from 2,000 to 20,000 is negligible compared to the computational time required for more epochs.

Mathematics of Inverting the Neural Network For a vector x, the ℓ^1 norm is defined by $||x||_1 = \sum_i |x_i|$ where i runs over the components of x.

Denoting $W^{(1)}$ and $z^{(1)}$ from the Methods section by W and z, we remark that from a softmax activation, i^* is the index that maximizes Wx + b if and only if that corresponding node has the greatest activation in h(z). Thus, to activate output node i^* the most, we require that

$$W_{i:x} + b_i \le W_{i^*:x} + b_{i^*}, \quad \forall i \in \{1, 2, 3\} \setminus \{i^*\} \implies (W_{i^*:} - W_{i:})x + (b_{i^*} - b_i) \ge 0, \quad \forall i \in \{1, 2, 3\} \setminus \{i^*\}.$$

By defining $\mathcal{Q}^{(i^*)}$ to be the matrix formed by concatenating the row vectors $(W_{i^*:} - W_{i:}), i \in \{1, 2, 3\} \setminus \{i^*\}$ and $c^{(i^*)}$ to be the column vector formed by concatenating the scalars $(b_{i^*} - b_i), i \in \{1, 2, 3\} \setminus \{i^*\}$, we formulate this as:

$$x = \arg\min_{x} ||x||_{1}, \quad \text{s.t. } \mathcal{Q}^{(i^{*})}x + c^{(i^{*})} \ge 0$$
(4)

to find the smallest and most sparse deviation from x = 0 that yields a prediction i^* . We solve (4) using the Python *minimize* routine.

Sparse Inverses We plot the smallest, sparse perturbations that yield a given prediction of '+', '0', and '-', for the top neural networks. For brevity, we have omitted features where none of the top neural networks produced an output with magnitude of at least 0.1. These results in a sense indicate the least possible change from being average a census tract could make and result in a given prediction.

Feature Signatures for a Vehicle Population Changes Figure 11 illustrates the feature signatures that lead to the neural networks towards various predictions for the vehicle homeless counts.

References

- (2019a). Political Los Angeles Times: Complaints about homeless camps in L.A. have surged. but are cleanups pointless?
- (2019b). Political: Scenes from the 2019 Greater Los Angeles homeless count.
- Anderson, I. and Christian, J. (2003). Causes of homelessness in the UK: A dynamic analysis. *Journal of Community & Applied Social Psychology*, 13(2):105–118.
- Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., and Samek, W. (2015). On pixelwise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS one*, 10(7):e0130140.
- Berry, B. (2007). A repeated observation approach for estimating the street homeless population. *Evaluation review*, 31(2):166–199.
- Bishop, C., Bishop, C. M., et al. (1995). *Neural networks for pattern recognition*. Oxford university press.

Boyd, S. and Vandenberghe, L. (2004). Convex optimization. Cambridge university press.

Chou, E. (2018). LA county to roll out \$402 million spending plan for Measure H. will it make a difference?

County, L. Measure h documents.



Figure 10. Dominant set of sparse features that result in a given prediction. A, B, and C are the top neural networks for the street homeless population; D, E, and F are the top neural networks for the vehicle homeless population. The '-', '0', and '+' represent predictions of change in the lower, middle, and upper third of values. A feature value of zero represents the mean value. Top row: inversions for street homeless count changes. Bottom row: inversions for vehicle homeless count changes. An all zeros feature vector results in a prediction of '+' for D, E, and F and hence no plot is shown.

- Angeles Department of Mental Health County of Los (2017).County of los population census tract. city angeles estimated bv name, ethnicity, gender & age group: Cy 2015. https://data.lacounty.gov/Mental-Health/ County-of-Los-Angeles-Estimated-Population-by-Cens/rv2f-zsc7.
- Cox, R., Henwood, B. F., St. Clair, P., Ailshire, J., Lincoln, K., and Strong, D. (2017). 2017 los angeles continuum of care homeless count methodology report. *Los Angeles*.
- DexYP(2017). Yellow pages. https://www.yellowpages.com/.
- Elliott, M. and Krivo, L. J. (1991). Structural determinants of homelessness in the united states. Social problems, 38(1):113–131.
- Flaming, D. and Burns, P. (2017). Who counts? assessing accuracy of the homeless count.
- Glynn, C. and Fox, E. B. (2019). Dynamics of homelessness in urban America. *The Annals of Applied Statistics*, 13(1):573–605.
- Google (2017a). Google geocoding api. https://developers.google.com/maps/ documentation/geocoding/intro.
- Google (2017b). Google places api. https://developers.google.com/places/ web-service/intro.
- Herman, D. B., Susser, E. S., Struening, E. L., and Link, B. L. (1997). Adverse childhood experiences: are they risk factors for adult homelessness? *American journal of public health*, 87(2):249–255.



Figure 11. Top: dominant distribution of features (as z-scores) so that a change in vehicle count in the lower tertile of changes is predicted with neural networks D-F. Middle: dominant distribution of features (as z-scores) so that a change in vehicle count in the middle tertile of changes is predicted with neural networks D-F. Bottom: dominant distribution of features (as z-scores) so that a change in vehicle count in the upper tertile of changes is predicted with neural networks D-F. Bottom: dominant distribution of features (as z-scores) so that a change in vehicle count in the upper tertile of changes is predicted with neural networks D-F. The features that appear have been culled so that all three neural networks have the same sign in their respective components and where the magnitude is at least 0.1. Two of the tree neural networks suggested a below average vehicle count one year would result in a middle tertile population change the next year (not shown).

- Hyde, P. S. (1985). Homelessness in America: Public policy, public blame. *Psychosocial Rehabilitation Journal*, 8(4):21.
- Johnson, R. R. and Kuby, P. J. (2011). *Elementary statistics*. Cengage Learning.
- Jolliffe, I. T. and Cadima, J. (2016). Principal component analysis: a review and recent developments. *Phil. Trans. R. Soc. A*, 374(2065):20150202.
- LAHSA (2019). 2019 homeless count full report.
- Los Angeles County (2012). Census tracts (2010). https://egis3.lacounty.gov/ dataportal/2011/07/19/census-tracts-2010/.
- Los Angeles County Department of Public Works (2016). City boundaries and annexations/detachments. https://egis3.lacounty.gov/dataportal/2013/06/11/city-annexations/.
- Los Angeles County Metropolitan Transportation Authority (2017). Metro developer. https:// developer.metro.net/introduction/gis-data/download-gis-data/.
- Los Angeles County Open Data (2017). Homeless shelters and services. https://geohub. lacity.org/datasets/homeless-shelters-and-services.
- Los Angeles Department of Transportation (2017). Parking citations. https://data.lacity. org/A-Well-Run-City/Parking-Citations/wjz9-h9np.
- Los Angeles Homeless Services Authority (2015). 2015 homeless count results. https://www.lahsa.org/news?article=406-2015-homeless-count-results{&}ref=hc.
- Los Angeles Homeless Services Authority (2016). Homeless count 2016 result by census tract. https://www.lahsa.org/documents?id= 2187-homeless-count-2016-result-by-census-tract.xlsx.
- Los Angeles Homeless Services Authority (2017). Homeless count 2017 results by census tract. https://www.lahsa.org/documents?id= 1495-homeless-count-2017-results-by-census-tract.xlsx.
- Los Angeles Homeless Services Authority (2018). Statement on the 2017 youth count adjustment. https://www.lahsa.org/news?article= 379-statement-on-the-2017-youth-count-adjustment.
- Los Angeles Housing and Communicty Investment Department (2017).Hcidla affordable housing projects listing (2003)catalog and to present). https://data.lacity.org/A-Livable-and-Sustainable-City/ HCIDLA-Affordable-Housing-Projects-Catalog-And-Lis/u4mj-cwbz.
- Los Angeles Office of Finance (2017). Map of restaurants. https://data.lacity.org/ A-Prosperous-City/Map-of-Restaurants/ycz4-j47g.
- Los Angeles Police Department (2017). Crime Data from 2010 to Present. https://data.lacity. org/A-Safe-City/Crime-Data-from-2010-to-Present/y8tr-7khq.
- Los Angeles Public Libraries (2017). Library branches. https://data.lacity.org/ A-Livable-and-Sustainable-City/Library-Branches/a4nt-4qca/data.
- Lu, B.-L., Kita, H., and Nishikawa, Y. (1999). Inverting feedforward neural networks using linear and nonlinear programming. *IEEE Transactions on Neural networks*, 10(6):1271–1290.
- Metraux, S., Treglia, D., and O'Toole, T. P. (2016). Migration by veterans who received homeless services from the department of veterans affairs. *Military medicine*, 181(10):1212–1217.
- Montavon, G., Lapuschkin, S., Binder, A., Samek, W., and Müller, K.-R. (2017). Explaining nonlinear classification decisions with deep taylor decomposition. *Pattern Recognition*, 65:211–222.

- Phelan, J. C. and Link, B. G. (1999). Who are" the homeless"? reconsidering the stability and composition of the homeless population. *American Journal of Public Health*, 89(9):1334–1338.
- Piantadosi, S., Byar, D. P., and Green, S. B. (1988). The ecological fallacy. American journal of epidemiology, 127(5):893–904.
- Quigley, J. M. (1990). Does rent control cause homelessness? taking the claim seriously. *Journal of policy analysis and Management*, 9(1):89–93.
- Rew, L., Taylor-Seehafer, M., Thomas, N. Y., and Yockey, R. D. (2001). Correlates of resilience in homeless adolescents. *Journal of Nursing Scholarship*, 33(1):33–40.
- Smith, D. (2019). Skid row squalor draws health officials' scrutiny; county inspectors cite businesses for violations linked to trash-choked streets.

Stevens, A. (2010). Sp_proj. MATLAB Central File Exchange.

- Susser, E., Moore, R., and Link, B. (1993). Risk factors for homelessness. *Epidemiologic reviews*, 15(2):546–556.
- Taylor, K. M. and Sharpe, L. (2008). Trauma and post-traumatic stress disorder among homeless adults in Sydney. Australian and New Zealand Journal of Psychiatry, 42(3):206–213.
- U.S. Census Bureau (2017a). 2010 census tract relationship files.
- U.S. Census Bureau (2017b). 2016 acs pums data dictionary.
- U.S. Census Bureau (2017c). American community survey 2016 acs 1-year pums files.
- U.S. Census Bureau (2017d). American community survey multiyear accuracy of the data (5-year 2012-2016).
- U.S. Census Bureau (2017e). Pums technical documentation.

Zillow (2017). Data. https://www.zillow.com/research/data/.

Acknowledgements

ML acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) [funding reference number PDF-502-479-2017]. Cette recherche a été financée par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) [numéro de référence PDF-502-479-2017]. This work was supported by NSF grant DMS-1737770.

The authors would like to thank Marcus Roper and Stan Osher for help in facilitating this research, and Bao Wang and Guido Montufar for valuable discussions. The authors also thank Derek Yen for invaluable contributions to virtually all aspects of this research.