Dan Levine
Advanced Spatial Analytics
May 9, 2022

Finding spatial concentrations and discrimination in arrest warrants stemming from low-level citations

Motivation/background

The U.S. Constitution guarantees the rights against unreasonable search or seizure, making it illegal for police to detain individuals without specific cause, search vehicles or persons without probable cause, or enter a home without a search warrant. To enforce this right, courts have blocked evidence that is collected through an illegal search or following an illegal stop from being admitted against a defendant. In the parlance of the Supreme Court, all evidence that may be gathered following an impermissible stop is “fruit of the poisonous tree” and impermissibly tainted.[1]

However, this protection was severely weakened through a recent Supreme Court case, *Utah v. Strieff.*[2]

Police in South Salt Lake, Utah were surveilling a house they suspected was being used for drug dealing. A police officer monitoring the home decided to stop and search someone leaving the house, although without other specific suspicion, such a stop was unconstitutional, (as the prosecution later acknowledged). the officer stopped Edward Streiff, asked for his ID, looked him up and found there was an outstanding arrest warrant stemming from an earlier traffic violation. The officer arrested Streiff, conducted a search incident to arrest, and discovered methamphetamine and drug paraphernalia. Strieff was charged with drug possession, but in court argued the evidence should be discarded because his initial detention by the police unconstitutionally violated the Fourth Amendment. The question was appealed all the way to the Supreme Court, where the majority ruled that the evidence could be admitted in this case, even though the stop was illegal. In the court’s determination, discovering that Streiff had an outstanding warrant for a prior incident was an “extraordinary” condition, one so unusual and unique as to sever the initial stop from the following arrest and search.

Under the precedent set by this case, police officers would be able to make otherwise illegal stops—pulling over drivers or detaining pedestrians without any prior cause, for example. If they

[1] *Nardone v. United States* (1939) https://supreme.justia.com/cases/federal/us/308/338/
[2] *Utah v. Streiff.* 2015 https://www.supremecourt.gov/opinions/15pdf/14-1373_83i7.pdf

then discover that individual has an outstanding warrant, they can legally search the individual, despite the unconstitutionality of the initial stop.[3]

This would be a minor concern if the Supreme Court's reasoning were correct and a police officer encountering an individual with an outstanding warrant were extraordinary. However, there are many individuals in a similar situation to Strieff with warrants against them for prior, minor, often non-criminal offenses.

Municipal citations, for traffic violations, or minor infractions like loitering, or trivial matters like failure to mow a lawn or repair a sidewalk in accordance with local ordinance, can incur fines of several hundred dollars. These citations are not criminal, but to collect on outstanding balances, municipalities can have an arrest warrant issued against individuals who have failed to pay. The effect is a criminalization of poverty, a system through which individuals who are not able to come up with the money to pay off a citation wind up subject to arrest and, following *Streiff*, search following even an unconstitutional stop. [4]

The growing burden of municipal fines and warrants have been increasingly studied and reported in recent years. Indeed, the dissenting justices made use of several studies in refuting the majority's reasoning in *Streiff.* [5]

However, the spatial structure of where there are individuals with outstanding warrants and how this distribution matches with other spatial patterns has not been as explored.

This project uses unusually open data on individuals with outstanding warrants from Austin, Texas and applies several spatial analysis and geostatistics techniques to measure the spatial patterns of where the density and prevalence of warrants is greatest.

Data

Warrant records

Studies of the prevalence and frequency of warrants stemming from municipal citations are limited by the availability of data. Records are held by municipal counts and sheriff or police departments across several small jurisdictions. These courts have had little reason to make their records open or available digitally. Moreover, there are compelling privacy concerns that could shield identifying details about the individuals who have warrants against them.

---

[3] Kerr, Orin, and Amy Howe. 2016. "Opinion analysis: The exclusionary rule is weakened but it still lives." SCOTUSblog.
https://www.scotusblog.com/2016/06/opinion-analysis-the-exclusionary-rule-is-weakened-but-it-still-lives/

[4] Wesley Dozier & Daniel Kiel. 2021. Debt to Society: The Role of Fines & Fees Reform in Dismantling the Carceral State. U. Mich. Journal of Law Reform.
https://doi.org/10.36646/mjlr.54.4.debt

[5] *Utah v. Streiff.* 2015 https://www.supremecourt.gov/opinions/15pdf/14-1373_83i7.pdf

Austin, Texas' Municipal Court and police department offer unusually open data through their respective web portals. The police web site allows a search by date of birth and returns the name and demographic details about all matching individuals with warrant records.[6] The Municipal Court web site can be searched by name and data of birth together and returns details about each case, the outstanding balance, warrant status, and an abbreviated home address for each individual.[7]

For a project in Prof. Ravi Schroff's Data Driven Methods for Public Policy course, my colleague Branden DuPont built a web scraping program to collect these data. The script first searched the sheriff database, inputting every possible data of birth for individuals up to 110 years old and storing the returned names and details. Secondly, the names and dates of birth were entered to the Municipal Court database search and the corte record details were stored. The resulting dataset included approximately 58,000 unique case records of 36,000 unique individuals.

One limitation of the data is that records are only consistently available for individuals who have open warrants and open cases. Someone who is issued a citation and paid it off promptly will not have a warrant issued and so is not searchable in this database. Once an outstanding balance is paid off, the record is also (generally) removed.

Geocoding

The Municipal Court records include an abbreviated home address for nearly every individual. Specifically, the data include

- a street name (but not the specifying street type; e.g. Jay Street would be represented as 'Jay'),
- city,
- state, and
- ZIP code, in most instances as a 9-digit code (ZIP plus 4)

Thus the exact location (e.g. specific house) cannot be determined—perhaps for privacy reasons. But the location can be approximated within a reasonable small area: Many streets in Austin are fairly short, so the street name alone defines a small area. Alternatively, the ZIP code + 4 defines a limited number of houses, generally one segment and side of the street in a postal delivery route.[8] This is, in reality, a linear feature, but the homes within a particular ZIP + 4 segment could be represented as an area.

Though the data provided some geographic indicators, I still needed a way to use this information for geocoding. The structure of the data did not allow for accurate matches with the U.S. Census Geocoder.[9] The USPS does not offer a ZIP + 4 lookup. Some private providers

---

[6] https://www.austintexas.gov/police/warrants/warrantsearch.cfm
[7] https://www.austintexas.gov/AmcPublicInquiry/pubportal.aspx
[8] https://faq.usps.com/s/article/ZIP-Code-The-Basics
[9] https://www.census.gov/programs-surveys/geography/technical-documentation/complete-technical-documentation/census-geocoder.html

offer crosswalk files or search tools to get locations of ZIP + 4, but these services are aimed at the direct mail advertising industry and are pricey. I was able to use ArcMap geocoder tool[10] to search addresses and achieved good matches. Most locations matched to an individual street segment.

I manually reviewed addresses that matched to other (coarser) locations and kept only those that had correctly matched to a point of interest in Austin. I then filtered the data to only those locations within Travis County (the county Austin sits within).

Removing records without locations or with locations outside of Travis County pared the database to 19,207 unique individuals.

Geographic scope

Travis County, rather than Austin exclusively, was chosen as the area of analysis for a few reasons.

Austin has grown irregularly through annexation of county land, such that its borders are serpentine and change over time (see Figure 1). Areas just inside and just outside city limits appear to be quite similar.

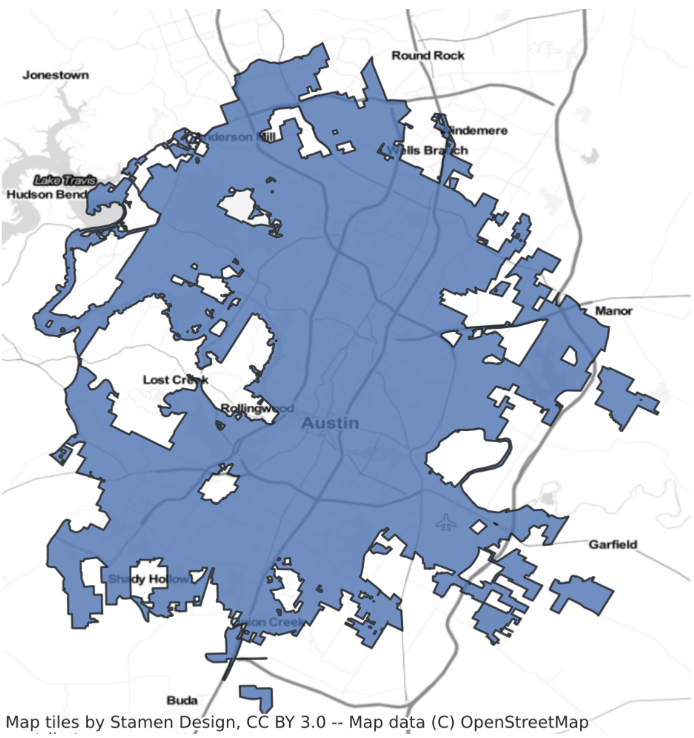


FIGURE 1. Austin city boundary.

---

10 https://desktop.arcgis.com/en/arcmap/latest/manage-data/geocoding/working-with-arcgis-online-geocoding-service.htm

The boundary *does* have legal significance for the issues being studied. Municipal citations are, by definition, only issued for violations that occur within Austin's city limits. But exploration of the data reveal numerous records of individuals with home addresses outside of Austin (most elsewhere in Texas, but also including locations from around the world). Analysis also finds that traffic citations are, by far, the most common initial citation type leading to warrants. Regardless of home location, drivers who violate traffic laws in Austin will be issued citations there and, if unpaid, these citations can be elevated to warrants.

This analysis thus assumes that drivers from anywhere in Travis County are likely to drive through Austin and could be issued warrants for unpaid citations accrued in the city.

Because warrant data covers exclusively violations within Austin, and lacks records of violations recorded by suburban jurisdictions, the counts and prevalence rates computed can only be considered an underestimate of the true number of individuals with outstanding warrants.

Census

Several demographic measures were taken from the U.S. Census American Community Survey, 5-year estimates from 2019 (the most recent year reliably available), and the Census Tract level. Specific values studies include: total population, White, Black, Asian, and Hispanic or Latino population, median income, population with income below the poverty level, households paying more than 50 percent of income on rent, households receiving SNAP benefits.

Police stops

Austin Police Department publishes data on all traffic stops.[11] Data from 2019, the most recent year available, were used to indicate the frequency of traffic stops within different areas. These are detailed data, but for this analysis, each recorded stop was simply counted with equal weight.

Methods

Group by Census Tract and compare prevalence to demographics

I spatially joined warrant locations to the Census Tracts the points sit within. Dividing the count of individuals with warrants by the total population of the Census Tracts yields a warrant prevalence.

With exploratory data analysis, I sought to identify correlations between counts or prevalence rates of individuals with warrants and other economic or demographic indicators. I visually inspected bivariate scatterplots relating each possible explanatory factor to the warrant prevalence and choropleth maps of each variable. For variables that seemed to show a fit, I computed a linear, spatially-lagged regression.

---

[11] https://data.austintexas.gov/Public-Safety/2019-Racial-Profiling-RP-Arrests/m4cc-q8pr

Somewhat surprisingly, few demographic measures were seen to correlate closely with warrant concentrations.

The strongest demographic indicators of warrant prevalence were area income and Hispanic or Latino population. Spatially-lagged linear regression of the log of warrant prevalence rate by tract median income, the best-fitting model, showed an R^2 score of 0.45 (see Figure 2).

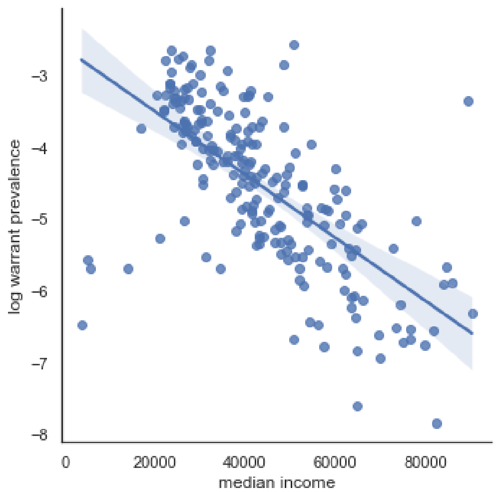


FIGURE 2: log-warrant prevalence ~ median income, at Census Tract level

The total number count of Latino or Hispanic residents in a Census Tract also correlated linearly with the prevalence of warrants. The spatially lagged regression showed an R^2 of 0.44.

Counts or proportions of other racial and ethnic groups were seen to be uncorrelated to warrant concentration.

Regression to find relationships

Exploratory analysis identified some relationships between features in the data, including a positive correlation between area Hispanic or Latino population in the count of individuals with outstanding warrants, and a negative relationship between area median income and warrant prevalence.

I used multivariate linear regression models to test the overall degree to which the area demographic factors explain the difference in the number of warrants and to determine which factors may have a significant correlation.

Parallel coordinate plots, created with GeoDa, helped to identify relationships between combinations of features. I also used GeoDa’s linear regression tool to build various models and also test the effect of adding spatial lag terms to correct or account for spatial dependency. I began with an array of demographic variables and stepwise removed the variables that did not have significant explanatory power.

The best-fitting linear model, with an R^2 value of 0.70, shows that the total Hispanic or Latino population and the total population receiving SNAP benefits are both positively associated with the number of outstanding warrants, controlling for the total population and the number of police stops in the Tract.

The coefficients on this model are small, but demonstrate a detectable discrimination against Hispanic and Latino residents: holding the population and rate of police stops constant and controlling for spatial effects, for each additional 33 Hispanic or Latino residents in a Census Tract there is one additional person with an outstanding warrant. Additionally, every 7 additional households receiving SNAP benefits imply one additional individual with an outstanding warrant.

Bin by area and map density

To measure the density of individuals with outstanding warrants and the variation across space, I fit point locations into a regular grid.

I overlaid a regular hexagonal grid on the study area, using the H3 grid specification.[12] The H3 data model was developed by Uber primarily to locate drivers and riders, but its regular, nesting hexagonal bins are ideal for this task, as well. Compared with rectangular bins, hexagons have much less variation in the distance between the edge end center and unlike a rectangular lattice, the center points of each hexagonal cell are equidistant to the center points of all its neighbors. Overlaying bins of H3 resolution 7, clipped to the boundaries of Travis County, produced a hexagonal lattice with 449 bins.

Locations of warrants were spatially joined to the hex bins and the count of individuals with warrants was connected to the hex bins as an attribute of each bin. Because the bins are equal in area, this count is also a relative density.

Area median income was spatially interpolated from the overlapping Census Tract to the hex bin. The two geometries do not perfectly align, so the hex bins inherited the median income attribute from the Tract with which it had the most overlapping area. The creation of hex bins and spatial interpolation was done using the Python `tobler` package.[13]

Check for spatial autocorrelation

Spatial autocorrelation of the density of individuals with warrants would indicate spatial clustering and indicate an underlying spatial process or confounding such that neighboring locations have similar densities of warrants. Spatial autocorrelation is expected in this case, as adjoining areas are often similar economically and demographically and would face similar

[12] https://eng.uber.com/h3/
[13] https://pysal.org/tobler/

policing conditions. Spatial autocorrelation was tested using Moran's and Geary's *c*, using both an queen-adjacency neighborhood (i.e., the six immediately adjoining hex bins) and an inverse distance measure with a threshold spanning the entire study area.

Measures of spatial autocorrelation were, in fact, remarkably small: with an adjacency neighborhood, *I* was 0.21 (on a scale of -1 to 1 where any positive value indicates some positive autocorrelation and higher values indicate more spatial autocorrelation), *c* was 0.77 (where values below 1 show positive spatial autocorrelation and values closer to 0 show stronger spatial autocorrelation.) Using distance weights, *I* was just 0.10 and *c* 0.96, almost negligible amounts.

TABLE 1: Measures of spatial autocorrelation

weight	Queen adjacency	Inverse distance
Moran's *I*	0.21	0.10
Geary's *c*	0.77	0.96

Identify clusters and hotspots

The overall measure of spatial autocorrelation across the county-wide study area may obscure unique, localized effects or non-stationary patterns. Local 'hot spots' or 'cold spots' can be especially revealing; in this case high-low areas would be places with a high concentration of individuals with outstanding warrants, while adjacent neighborhoods have low numbers. Low-high areas are the reverse, low density of warrants amid higher density.

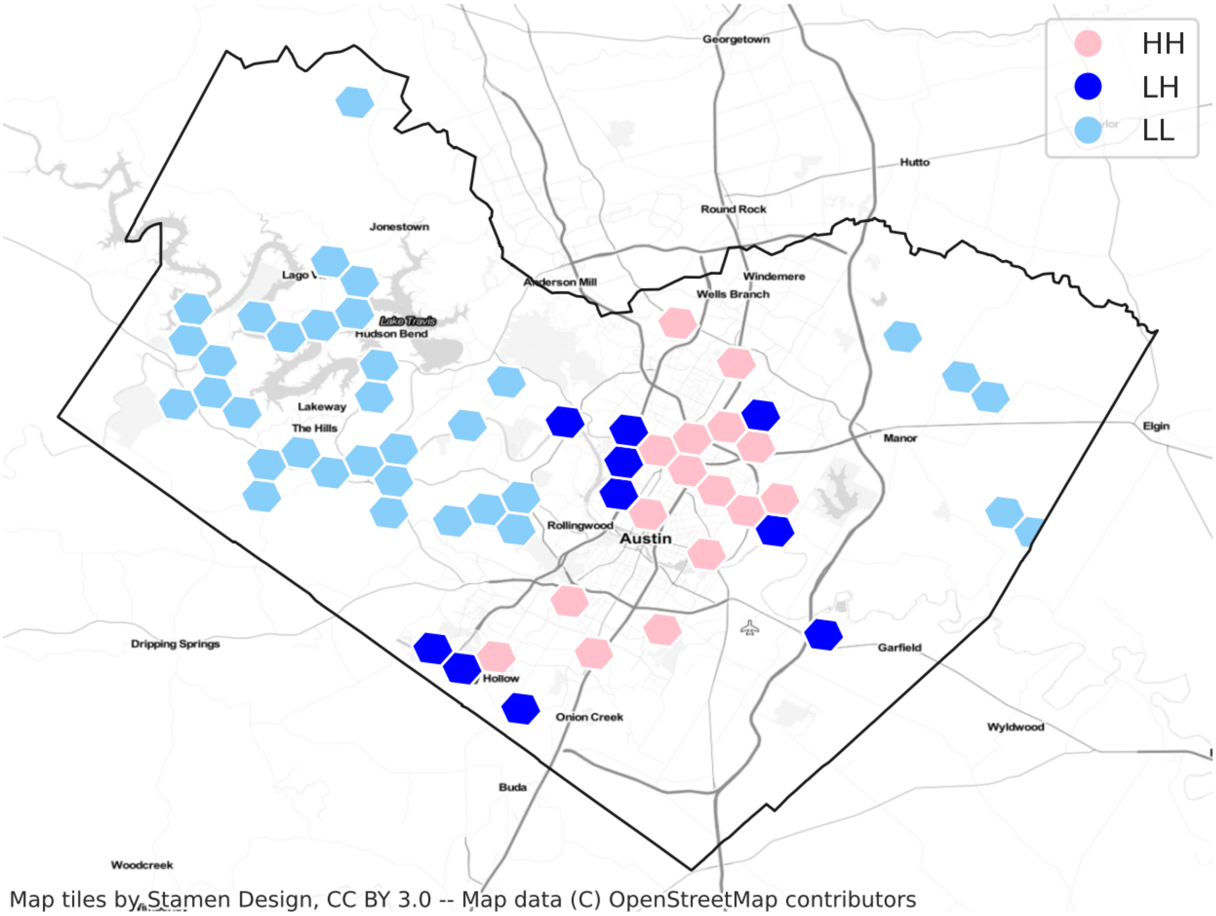
Local indicators of spatial association were computed for the count (equal to density) of individuals with outstanding warrants across hex bins, again using both queen adjacency and inverse-distance weights.[14] This used the Python `esda` package.[15]

When considering only immediate neighbors, the analysis revealed no significant high-low clusters (the 'hot spots' described above), but did identify many clusters of low rates across Austin's affluent western suburbs and clusters of high rates northeast of downtown. These are areas where low or high prevalences are unexpectedly clustered to a statistically significant degree. Moreover, the analysis finds several spots just east of downtown and in the southern reach of the county with significantly lower concentrations of warrants than their neighbors (see Figure 3).

Recomputing using the wider, inverse-distance weight uncovers a different pattern. When considering each area's warrant density relative to all other areas' (weighted by distance), large areas of eastern Travis County are shown to have significantly lower warrant density, while the

[14] Anselin, L. 1995. Local Indicators of Spatial Association—LISA. Geographical Analysis. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x
[15] Rey, S.J. and L. Anselin. 2007. PySAL: A Python Library of Spatial Analytical Methods. Review of Regional Studies https://pysal.org/esda/

more urban core of the county is a high-density cluster. More interesting are the numerous high- and low-value spots, which each could garner location-specific investigation (see Figure 3).


HH
LH
LL
Georgetown
Hutto
Round Rock
Jonestown
Anderson Mill
Windemere
Wells Branch
Lake Travis
Hudson Bend
Lakeway
The Hills
Manor
Elgin
Rollingwood
Austin
Dripping Springs
Garfield
Hollow
Onion Creek
Wyldwood
Buda
Woodcreek
Map tiles by Stamen Design, CC BY 3.0 -- Map data (C) OpenStreetMap contributors

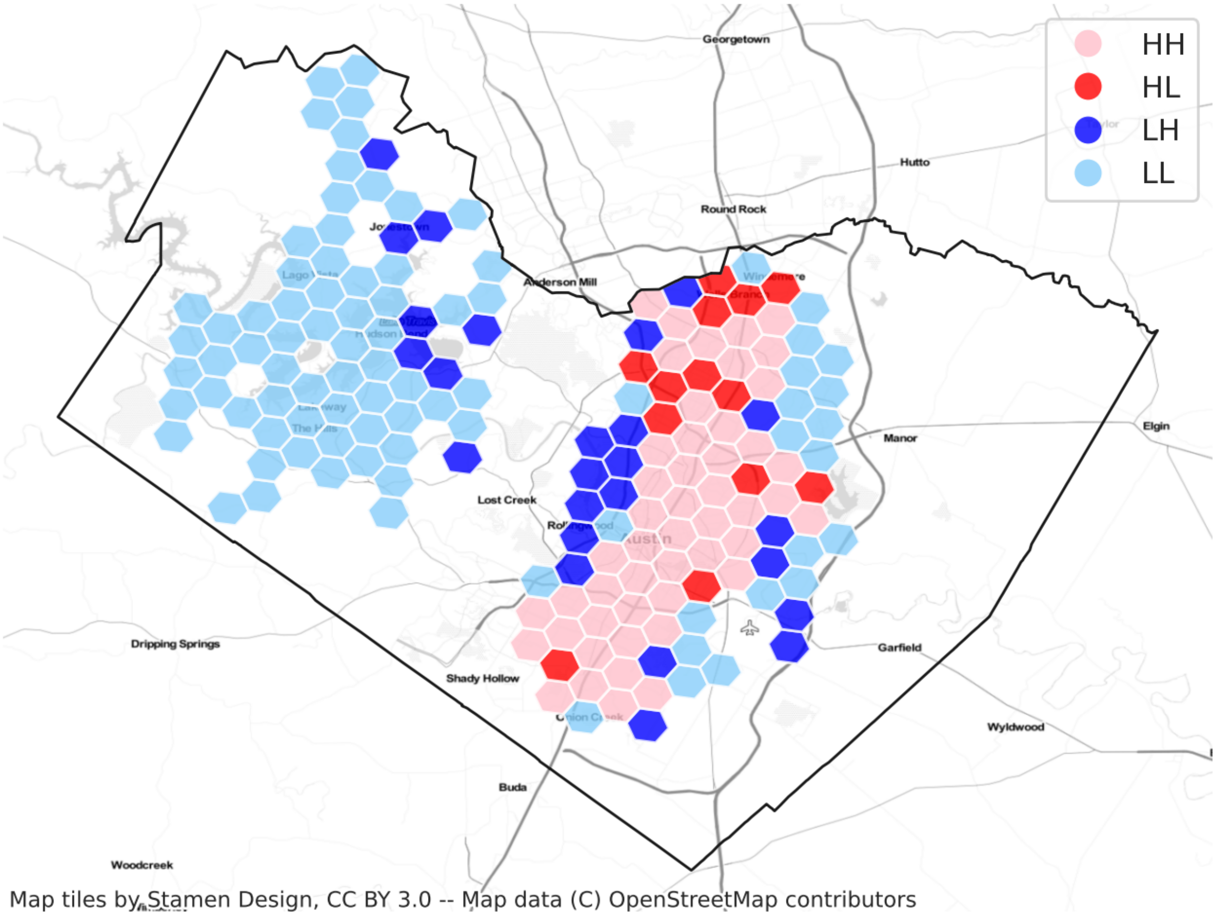


FIGURE 3. Spatial clusters and outliers in count and density of individuals with outstanding warrants. (top) computed considering only immediate neighbors. (bottom) computed considering all other areas, weighted by inverse distance. LL indicates areas of low density near other areas of low density. HH indicates areas of high density near other areas of high density. LH indicates spots of lower density than predicted from the values at neighbors, HL, the opposite. Mapped areas are significant at $p < 0.05$.

Evaluate patterns with semivariograms

I used semivariograms to assess the spatial structure of warrants and to detect whether the distribution of warrants varied between neighborhoods at different median income levels.

Semivariance is a geostatistical measure of the variance between samples as a function of the distance between them. The effective range of this function indicates the distance over which samples are instances of a common underlying process; beyond this range, samples are independent of one another.

I first tested semivariance of the count (also equaling density) of individuals with warrants in each hex grid cell, based on the distance between hex bin centroids. Semivariograms were computed and plotted with the Python `skgstat` package.[16]

This semivariogram reaches the sill over a very short range, less than 5 miles and within the first decile of inter-bin distances. Alternate models (spherical, exponential, Gaussian) and binning methods had little impact on the parameters. In all permutations, the range remained very short (and within the first decile of distiances) (see Figure 4).

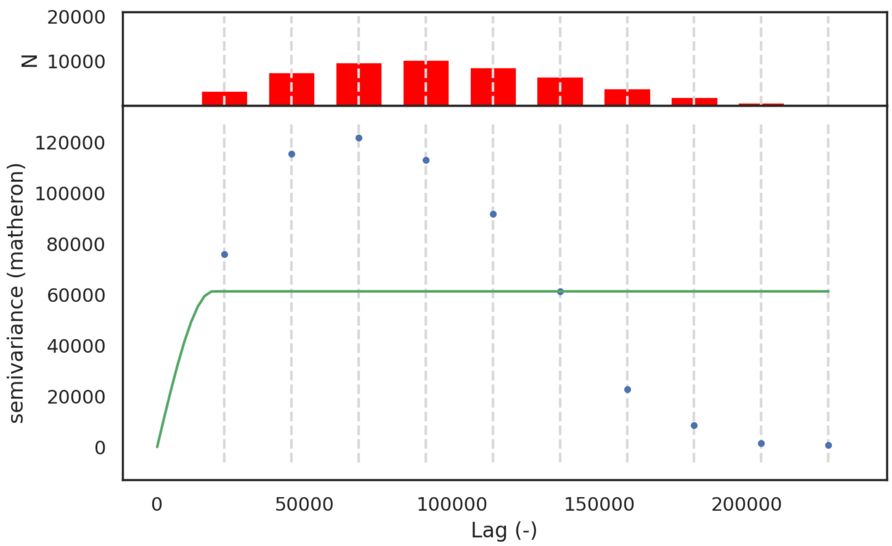


FIGURE 4: Semivariogram. x axis measures the distance between pairs of areas (measured in feet) Top figure is a histogram of the number of pairs of areas at each distance lag (red). Bottom figure shows the mean variance at each distance lag (blue) and the fitted function (green).

This short range shows that the patterns defining warrant assignment are more locally specific in nature and that the causes at one location are similar only to those of the immediate neighbors.

Comparing semivariance to detect differing warrant assignment processes at different income levels

Next, I divided the areas into three categories by income quantiles, yielding three equally-numbered sets of bins which sit within, respectively, the lowest-third median income, middle third median income, and highest-third median income Census tracts. I separately computed semivariograms for each of these subsets.

The lowest income group shows a much higher semivariance and the narrowest range. Comparatively, the middle- and top-third income groups show much lower semivariance overall.

[16] https://scikit-gstat.readthedocs.io/en/latest/userguide/variogram.html

The middle income group has a slightly wider range than the bottom income group. The range for the highest income is much greater than the others. (see Figure 5)

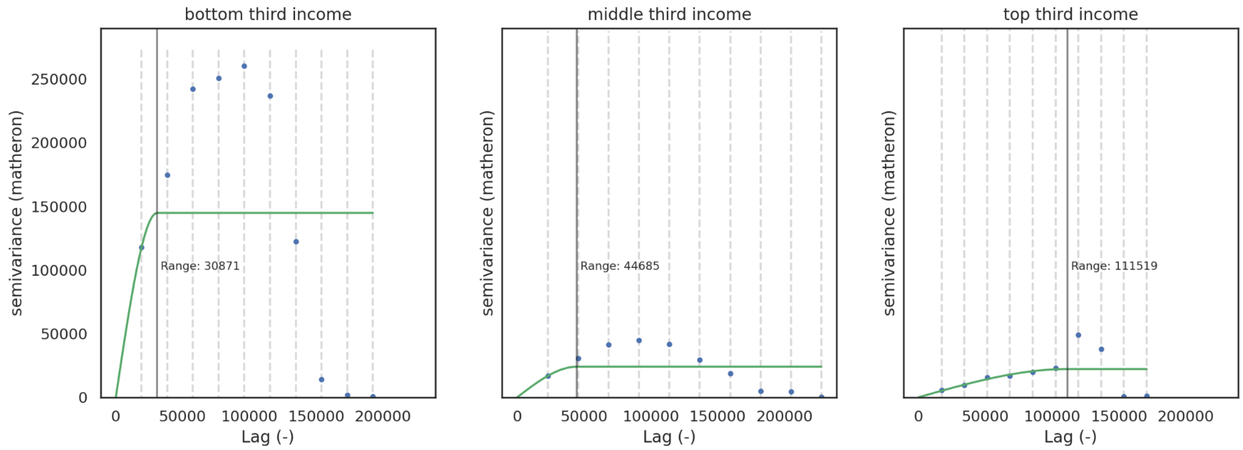


FIGURE 5: Semivariograms fit for areas within Census Tracts with the bottom third, middle third, and top third quantiles of median income.

This suggests that among these lower income areas, there is a higher variation in warrant distribution, that is, the dissimilarity in the allocation of warrants between neighbors or between neighboring areas is greater. In contrast, the higher income groups show more similarity across greater distance, suggesting a more similar underlying mechanism of warrant assignment.

Using Census Tracts as areas to compare prevalence

Using the raw count or density of individuals with warrants fails to account for the overall population of each area. What appear to be concentrations of warrants may simply be concentrations of population.

To test whether the spatial structures appeared the same when accounting for total population, I repeated the same semivariance analysis, but using Census Tract areas rather than the regular hex grid. By using Census Tract geographies, it was possible to use warrant prevalence (the number of individuals with outstanding warrants divided total population) as a measure.

However, Census Tract areas are unequally sized: they are drawn as areas with similar populations (approximately 4,000 people each), but this yields many more, smaller, closer-together Tracts in the denser urban core, and fewer, more-distantly-spaced Tracts in lower-density suburban areas. Adding total population thus corrects one possible bias in the data but introduces another, as the distance between Tracts is, essentially, inversely related to their density. Nonetheless, binning by Tracts is necessary to compare prevalence rates and Tract areas are not capricious spatial units, but ones drawn to generally align to meaningful geographic and political divisions.[17]

[17] https://www.census.gov/programs-surveys/geography/about/glossary.html#par_textimage_13

As in the preceding analysis, Census Tracts were split into three income quantiles and semivariograms were computed across each group, this time with warrant prevalence used as the variable measure. These semivariograms reveal a different set of spatial patterns.

As before, the variance is much higher within the lowest income group, again suggesting more variation in the assignment of warrants across low-income neighborhoods. The high-income group has the lowest variance, the middle-income group sits evenly between the other two (see FIgure 6a).

The distribution of warrant prevalence rates within each income group reinforces this pattern. The lowest-income group shows nearly equal counts of Tracts with prevalence between 0 and 0.175; the middle-income group has a greater concentration of tracts with rates near zero; and at the highest income, nearly all tracts have a rate near zero (see Figure 6b).

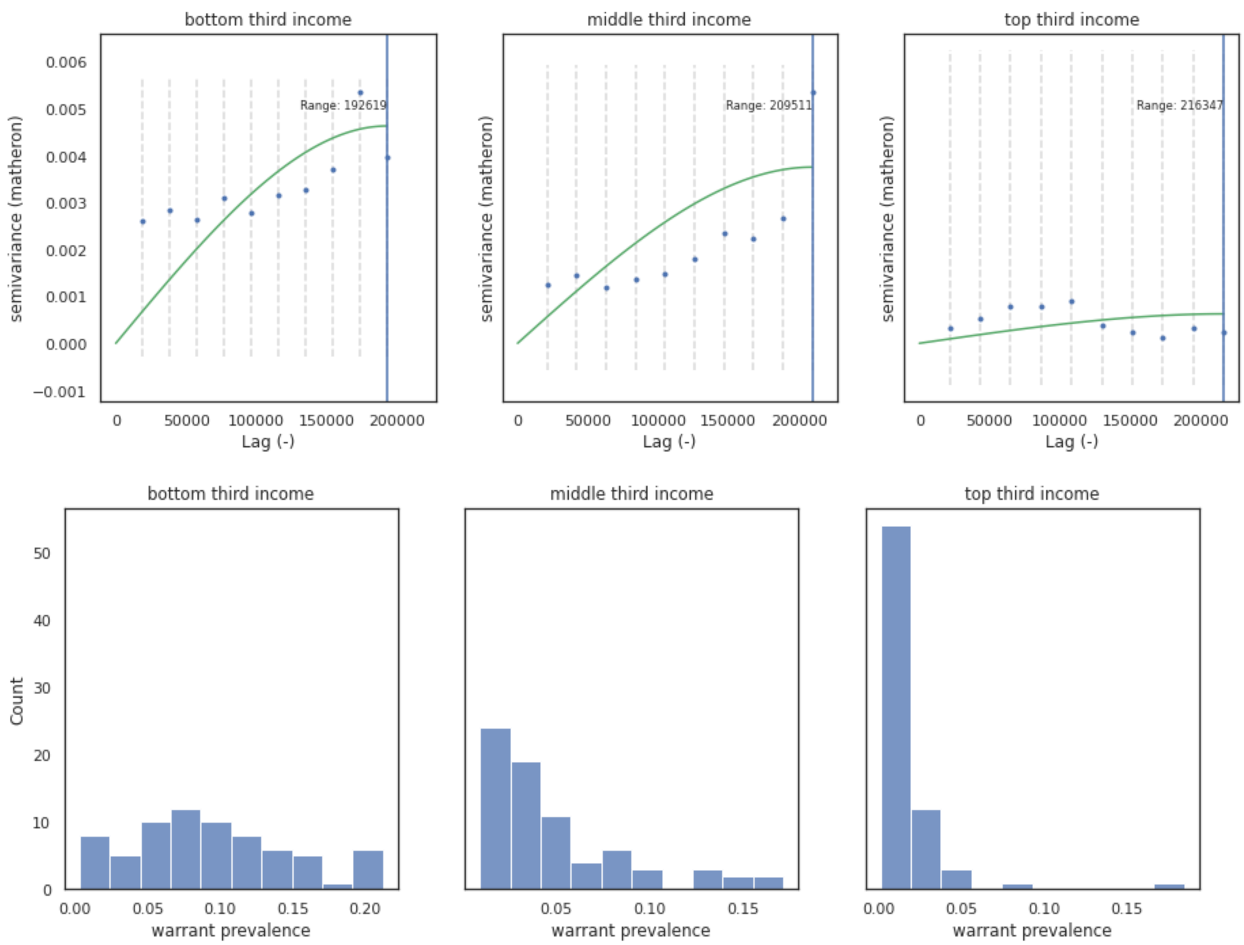


FIGURE 6: a. (top) semivariograms fit to variance in warrant prevalence across Census Tracts. b. (bottom) histograms displaying the distribution of warrant prevalence across Census Tracts. Each plot is computed for Tracts in the bottom, middle, and top income quantiles.

Whereas the semivariance across hex bins showed a very steep curve and short effective range (especially for the lowest-income set), the Tract, warrant prevalence semivariograms each show variance increasing continuously with distance and register effective range only at the outer limit of the study area (see Figure 6a). This is true for each income group. This indicates that Tracts are spatially related to one another across any distance; there is no point at which the variation in warrant prevalence is independent of an underlying spatial process. It may also indicate that once accounting for base population the pattern of warrant assignment is similar across a wide area and not as locally-driven.

Geographically Weighted Regression

Since the bulk of outstanding warrants are seen to come from traffic tickets, then the variation in the prevalence of outstanding warrants may be explained by differing amount of police enforcement or issuance of traffic tickets.

I used regression models to test whether police stop volume explains warrant density.

Using the same grid of hexbins generated previously, I counted the number of police traffic stops within each bin.

I fit an OLS linear regression model and a Poisson regression model to the number of individuals with outstanding warrants within each bin indicated by the number of police stops. Each model showed a significant relationship, but not a strong fit (respectively, R^2 of 0.57 and pseudo-R^2 of 0.29).

These models fail to account for spatial relationships and spatial dependencies. For instance, the volume of traffic stops *nearby* a home location may be as important in explaining the number of warrants as the stops in the particular home region.

To take account of these possible spatial dependencies, I computed a geographically weighted regression between the two variables.[18] I used a fixed bandwidth distance, equal to the effective range determined by the semivariogram across all areas, computed above. The regression was modeled using the Python `mgwr` package.[19]

This model showed a somewhat worse overall fit than the simple linear model, with an overall pseudo-R^2 of 0.43. The benefit is the ability to examine in what places the model fits well. An examination of the local R^2 scores shows that the model fits much better in the northern part of the city, and more poorly in the downtown core and south (see Figure 7). This indicates that in the northern section, the prevalence of warrants is much more associated with the volume of police traffic stops, whereas in the core, the two factors are less related.

---

[18] Brunsdon, et al. 1998. Geographically weighted regression — modelling spatial non-stationarity. The Statisticion. https://doi.org/10.1111/1467-9884.00145
[19] Oshan, T, et al. 2019. mgwr: A Python Implementation of Multiscale Geographically Weighted Regression for Investigating Process Spatial Heterogeneity and Scale. International Journal of Geo-information. https://doi.org/10.3390/ijgi8060269

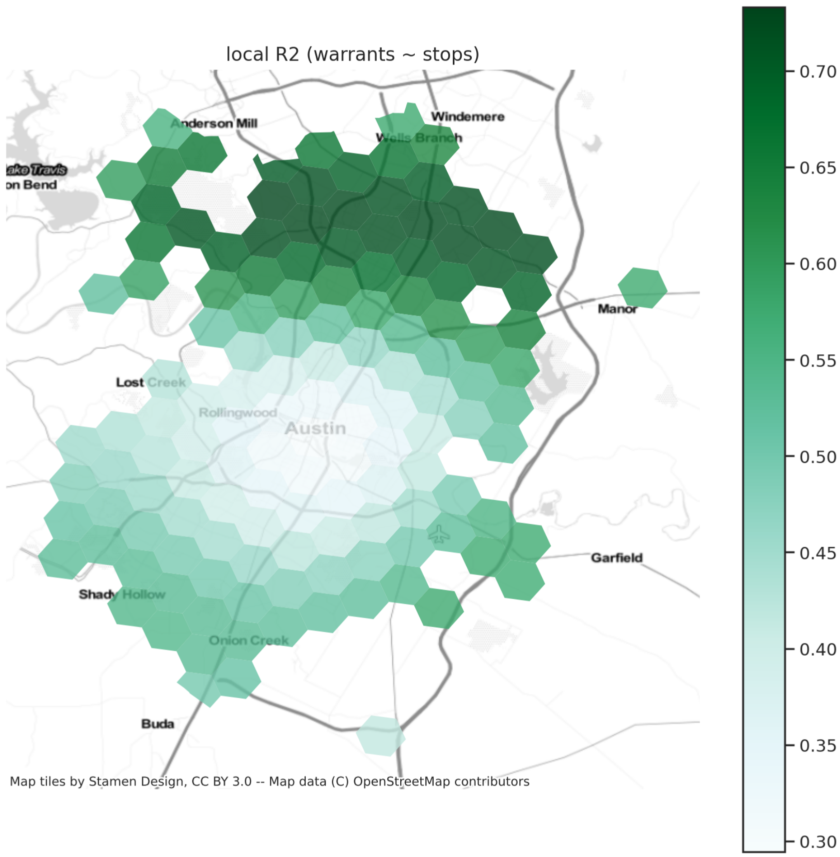


Figure 7: Local R^2 score from geographically weighted regression of the density of individuals with warrants ~ traffic stops.

Discussion

This analysis uses several spatial analysis techniques to evaluate the prevalence of outstanding arrest warrants stemming from non-criminal municipal citations. Further, these tools have explored geographic concentrations are a product of policing activity, residents' ability to pay citation fees, and other factors that may imply additional discrimination in the assignment of citations or the conditions of their escalation to warrants.

This study finds that there are significant local clusters with higher- and lower-densities of warrants. It determines that the pattern of warrant assignment is much more locally-varied among lower-income neighborhoods, suggesting additional, unique factors in these neighborhoods compared to relatively similar high-income areas. It finds that the Hispanic or Latino population in a neighborhood and the number of households receiving SNAP benefits are both significant indicators of a higher level of outstanding warrants, with these two factors alone explaining a high degree of the variation between areas and suggesting a discriminatory process against residents in these neighborhoods. Lastly, this study finds a differing relationship between the volume of police stops and the number of outstanding warrants across different sections of the city.

In applying various spatial analysis methods with varying parameters, this study also shows the variation and possible bias of analysis conducted at different scales. Local indicators of spatial autocorrelation are shown much differently using adjacency and inverse distance (an extension of the modifiable areal unit problem identified by Openshaw).[20] And semivariance appeared markedly different when measuring density across a regular lattice and prevalence across an irregular one.

[20] Openshaw S. 1977. Optimal Zoning Systems for Spatial Interaction Models. Environment and Planning A: Economy and Space. https://journals.sagepub.com/doi/10.1068/a090169