# The Space of Evictions: Determinants of Eviction in Alachua County

Renz Torres, URP 6272
2022-03-26

## Contents

## Introduction

Evictions, along with rising housing prices and non-renewals, are increasingly destabilizing tenants and the housing system nationwide. This report presents a spatial analysis of evictions in Alachua County over the past 20 years. I pose these questions: where do evictions occur in Alachua County, how do eviction-prone areas cluster or how do they act as outliers, and what are social and built environments of these areas? To these ends I will be using a variety of clustering techniques including Optimized Anselin Local Moran's I algorithm and Optimized Hotspot Analysis to determine clusters of block groups with high and low evictions, as well as outlier groups. I will also be using generalized linear regression and geographically weighted regression analyses to determine social and environmental characteristics of block groups of evictions.

## Data Sources

The ultimate source of the Alachua County Evictions 2002-2021 dataset is the Alachua County Clerk of Court. I used a pre-made web scraper directed at the case record search to acquire information about filings, including addresses. The original source is available here: https://www.alachuaclerk.org/court_records/index.cfm?section=login&r=901483. Acquired in January 2022, these data range from 2002-01-01 to 2021-12-31. The raw data were processed and filtered using R with the help of the *tidyverse* and *postmastr* libraries to create valid addresses. These addresses were geocoded manually using an address table created from County data (below) and manual searching for a few addresses.

The Alachua County Addresses dataset was retrieved in January 2022 from the Alachua County Property Appraiser's website (https://maps.acpafl.org/pages/gis-datasets-page). These data contain coordinates for every addressed location, down to unit number, in Alachua County. These data were used to geocode the Evictions dataset above. Addresses with many second-level units were grouped and summarized as a single centroid for geocoding. Certain addresses not available in the dataset, mostly demolished or renamed addresses, were manually geocoded using various historical references.

The Alachua County 2020 Block Groups dataset was retrieved in February 2022 from the US Census Bureau's website (https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.2020.html). Originally scoped for the entire state of Florida, I limited it to Alachua County. I downloaded datasets using the *tidycensus* R library from the US Census Bureau: Table P2 from the 2020 Census, and many variables (summarized in the *Data Description*) from the ACS 5-year estimates from 2016-2020. These data can also be referenced on the Census Bureau's data portal, https://data.census.gov.

I reprojected geographical data in NAD 1983 Florida GDL Albers, units in meters (WKID: 3086).

## Data Description

The Evictions dataset maps the location of the address indicated in the defendant's Residence. Only residences in Alachua County were considered. Originally a database with multiple tables, the scraper flattens the database into a single summarized table. The specifications are listed below:

| **Field** | **Description** |
|---|---|
| Case Number | The original case ID as recorded in the Clerk's database |
| Filed | Date the plaintiff filed the eviction |
| Status | Whether the case is ongoing or proceedings have ended. "99 CLOSED" and "CLOSED" for our purposes are the same |

| Field | Description |
|---|---|
| Summary | Concatenates the case number, type of case, the name of the presiding, and plaintiffs vs defendants. |
| Action | Summarizes the charges against the defendant |
| Residence | The address of residence as recorded of the defendant(s), if known |
| Total Costs | Amount required to be paid to the court |
| Total Due | Amount that still needs to be paid (to my understanding) |
| pm.* | Objects from the *postmastr* library used to format and standardize the addresses in `Residence` |
| Type | The format of the match between pm.address and the geocoding table.<br>• Standard: pm.address was exactly matched in the database<br>• Ordinal: pm.address was matched with the street number lacking its ordinal suffix, e.g. "8" vs "8th"<br>• Nodir_standard: matched but lacking a directional prefix on the street name<br>• Nodir_ordinal: matched but lacking a directional prefix and an ordinal |
| MEAN_X and MEAN_Y | The coordinates as matched in the geocoding table |

There are additional tables that have been unflattened from this table, including list of plaintiffs, list of defendants, and the docket. Currently they are not being included in the analysis. The total number of records is about 32 thousand; about 90% of cases were successfully geocoded.

Block Groups contains standard TIGER information about the underlying geometries of the block group – tract, county, and state, as well as acres of land and water, location of the centroid, and the perimeter and area of each block group. I summary joined Evictions to these count the number of evictions occurring in each block group then log transformed the quantity.

| Field | Description |
|---|---|
| GEOID | The Census' code used to identify block groups |
| Geographical Area Name | The long-form name of the block group in question |

Below is a description of the fields obtained and used from the P2 dataset. Descriptions of each variable are summarized below using the original census information.

| Code | Concept | Subcode | Label |
|---|---|---|---|
| P2 | HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE | P2_001N | Total (universe) |
| P2 | HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE | P2_002N | Hispanic or Latino |
| P2 | HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE | P2_003N | Not Hispanic or Latino |
| P2 | HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE | P2_004N | Population of one race |
| P2 | HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE | P2_005N | White alone |

| Code | Concept | Subcode | Label |
|---|---|---|---|
| P2 | HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE | P2_006N | Black or African American alone |
| P2 | HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE | P2_007N | American Indian and Alaska Native alone |
| P2 | HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE | P2_008N | Asian alone |
| P2 | HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE | P2_009N | Native Hawaiian and Other Pacific Islander alone |
| P2 | HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE | P2_010N | Some Other Race alone |
| P2 | HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE | P2_011N | Population of two or more races |

In the final dataset, only the proportions of each category were used as opposed to raw counts. *Total* was removed from the final dataset for analysis and each other variable was transformed and normalized into a proportion using *Total*.

Below is a description of the variables obtained from the ACS dataset. Similarly, any raw totals were suppressed from the final dataset and only the row-transformed proportions were used.

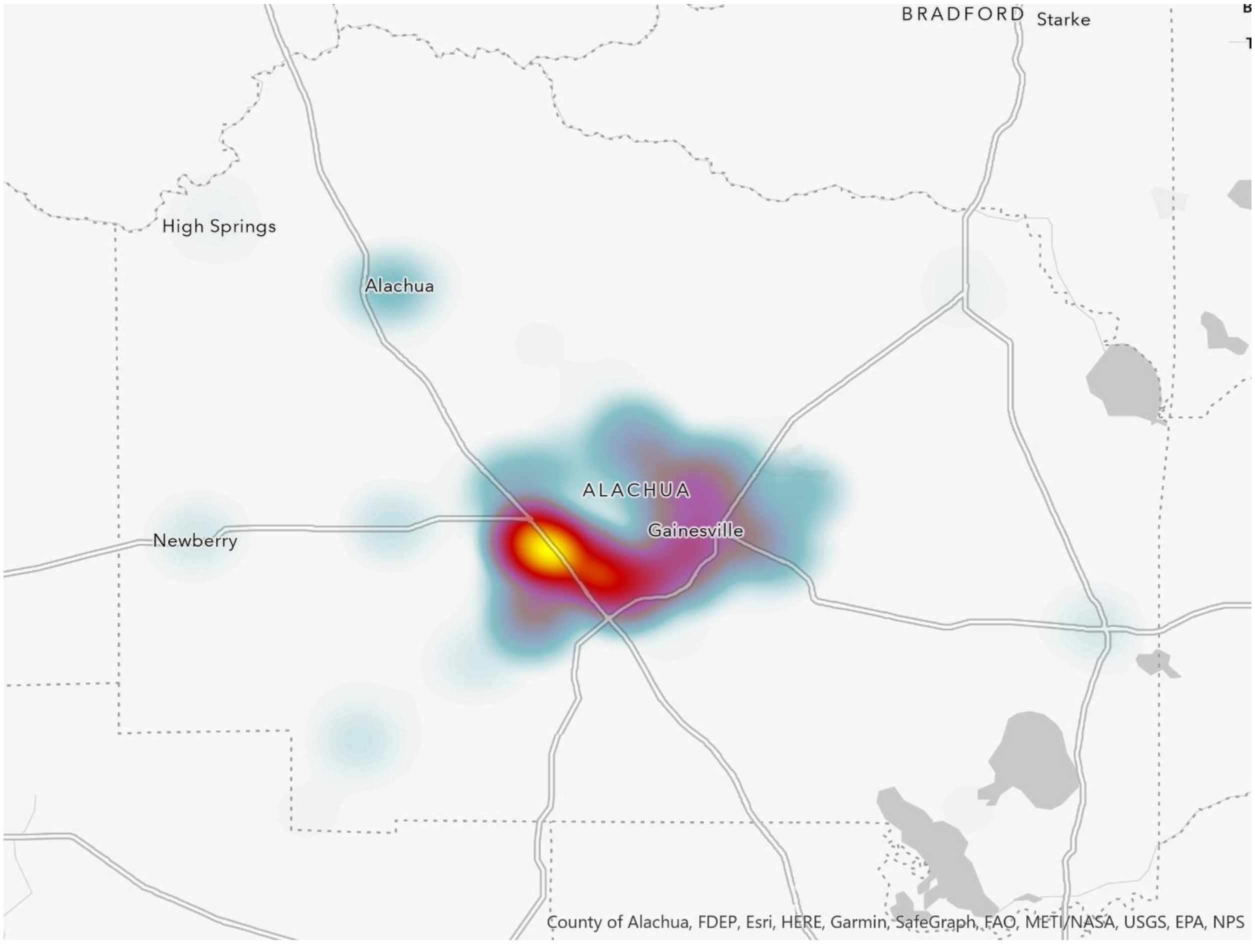
| Code | Concept | Subcode | Label |
|---|---|---|---|
| B19013 | MEDIAN HOUSEHOLD INCOME IN THE PAST 12 MONTHS (IN 2020 INFLATION-ADJUSTED DOLLARS) | B19013_001 | Median household income in the past 12 months (in 2020 inflation-adjusted dollars) |
| B25003 | TENURE | B25003_001 | Total (variable universe) |
| B25003 | TENURE | B25003_002 | Owner occupied |
| B25003 | TENURE | B25003_003 | Renter occupied |
| B11012 | HOUSEHOLDS BY TYPE | B11012_001 | Total (variable universe) |
| B11012 | HOUSEHOLDS BY TYPE | B11012_002 | Married-couple household |
| B11012 | HOUSEHOLDS BY TYPE | B11012_003 | Married-couple household: With children of the householder under 18 years |
| B11012 | HOUSEHOLDS BY TYPE | B11012_004 | Married-couple household: With no children of the householder under 18 years |
| B11012 | HOUSEHOLDS BY TYPE | B11012_005 | Cohabiting couple household |
| B11012 | HOUSEHOLDS BY TYPE | B11012_006 | Cohabiting couple household: With children of the householder under 18 years |
| B11012 | HOUSEHOLDS BY TYPE | B11012_007 | Cohabiting couple household: With no children of the householder under 18 years |
| B11012 | HOUSEHOLDS BY TYPE | B11012_008 | Female householder, no spouse or partner present |
| B11012 | HOUSEHOLDS BY TYPE | B11012_009 | Female householder, no spouse or partner present: Living alone |

| Code | Concept | Subcode | Label |
|---|---|---|---|
| B11012 | HOUSEHOLDS BY TYPE | B11012_010 | Female householder, no spouse or partner present: With children of the householder under 18 years |
| B11012 | HOUSEHOLDS BY TYPE | B11012_011 | Female householder, no spouse or partner present: With relatives, no children of the householder under 18 years |
| B11012 | HOUSEHOLDS BY TYPE | B11012_012 | Female householder, no spouse or partner present: With only nonrelatives present |
| B11012 | HOUSEHOLDS BY TYPE | B11012_013 | Male householder, no spouse or partner present |
| B11012 | HOUSEHOLDS BY TYPE | B11012_014 | Male householder, no spouse or partner present: Living alone |
| B11012 | HOUSEHOLDS BY TYPE | B11012_015 | Male householder, no spouse or partner present: With children of the householder under 18 years |
| B11012 | HOUSEHOLDS BY TYPE | B11012_016 | Male householder, no spouse or partner present: With relatives, no children of the householder under 18 years |
| B11012 | HOUSEHOLDS BY TYPE | B11012_017 | Male householder, no spouse or partner present: With only nonrelatives present |
| B25008 | TOTAL POPULATION IN OCCUPIED HOUSING UNITS BY TENURE | B25008_001 | Total |
| B25008 | TOTAL POPULATION IN OCCUPIED HOUSING UNITS BY TENURE | B25008_002 | Owner occupied |
| B25008 | TOTAL POPULATION IN OCCUPIED HOUSING UNITS BY TENURE | B25008_003 | Renter occupied |
| B25024 | UNITS IN STRUCTURE | B25024_001 | Total |
| B25024 | UNITS IN STRUCTURE | B25024_002 | 1, detached |
| B25024 | UNITS IN STRUCTURE | B25024_003 | 1, attached |
| B25024 | UNITS IN STRUCTURE | B25024_004 | 2 |
| B25024 | UNITS IN STRUCTURE | B25024_005 | 3 or 4 |
| B25024 | UNITS IN STRUCTURE | B25024_006 | 5 to 9 |
| B25024 | UNITS IN STRUCTURE | B25024_007 | 10 to 19 |
| B25024 | UNITS IN STRUCTURE | B25024_008 | 20 to 49 |
| B25024 | UNITS IN STRUCTURE | B25024_009 | 50 or more |
| B25024 | UNITS IN STRUCTURE | B25024_010 | Mobile home |
| B25024 | UNITS IN STRUCTURE | B25024_011 | Boat, RV, van, etc. |
| B25034 | YEAR STRUCTURE BUILT | B25034_001 | Total |
| B25034 | YEAR STRUCTURE BUILT | B25034_002 | Built 2014 or later |
| B25034 | YEAR STRUCTURE BUILT | B25034_003 | Built 2010 to 2013 |
| B25034 | YEAR STRUCTURE BUILT | B25034_004 | Built 2000 to 2009 |

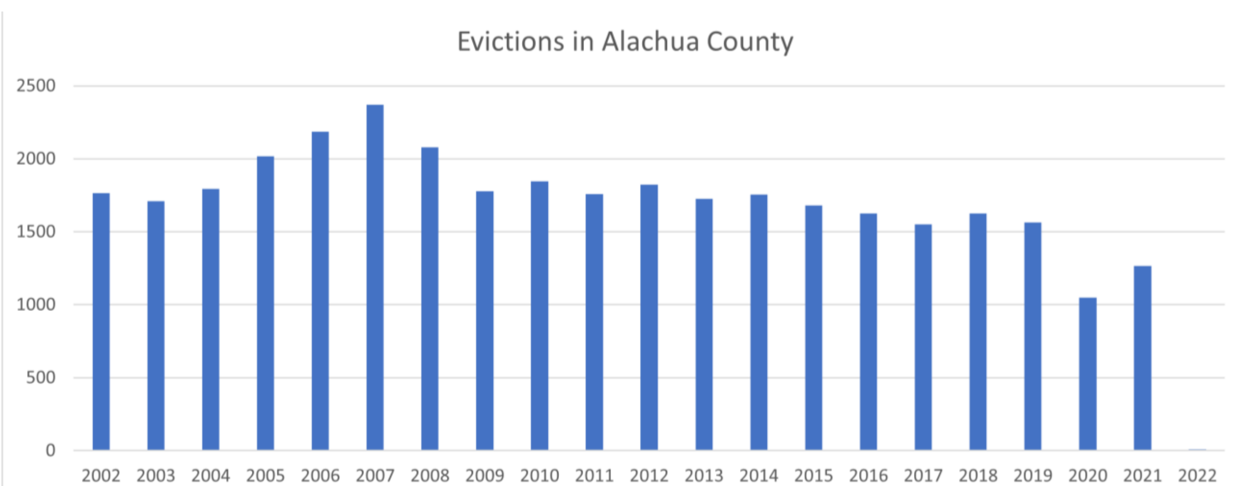
| Code | Concept | Subcode | Label |
|---|---|---|---|
| B25034 | YEAR STRUCTURE BUILT | B25034_005 | Built 1990 to 1999 |
| B25034 | YEAR STRUCTURE BUILT | B25034_006 | Built 1980 to 1989 |
| B25034 | YEAR STRUCTURE BUILT | B25034_007 | Built 1970 to 1979 |
| B25034 | YEAR STRUCTURE BUILT | B25034_008 | Built 1960 to 1969 |
| B25034 | YEAR STRUCTURE BUILT | B25034_009 | Built 1950 to 1959 |
| B25034 | YEAR STRUCTURE BUILT | B25034_010 | Built 1940 to 1949 |
| B25034 | YEAR STRUCTURE BUILT | B25034_011 | Built 1939 or earlier |
| B25035 | MEDIAN YEAR STRUCTURE BUILT | B25035_001 | Median year structure built |
| C17002 | RATIO OF INCOME TO POVERTY LEVEL IN THE PAST 12 MONTHS | C17002_001 | Total |
| C17002 | RATIO OF INCOME TO POVERTY LEVEL IN THE PAST 12 MONTHS | C17002_002 | Under .50 |
| C17002 | RATIO OF INCOME TO POVERTY LEVEL IN THE PAST 12 MONTHS | C17002_003 | .50 to .99 |
| C17002 | RATIO OF INCOME TO POVERTY LEVEL IN THE PAST 12 MONTHS | C17002_004 | 1.00 to 1.24 |
| C17002 | RATIO OF INCOME TO POVERTY LEVEL IN THE PAST 12 MONTHS | C17002_005 | 1.25 to 1.49 |
| C17002 | RATIO OF INCOME TO POVERTY LEVEL IN THE PAST 12 MONTHS | C17002_006 | 1.50 to 1.84 |
| C17002 | RATIO OF INCOME TO POVERTY LEVEL IN THE PAST 12 MONTHS | C17002_007 | 1.85 to 1.99 |
| C17002 | RATIO OF INCOME TO POVERTY LEVEL IN THE PAST 12 MONTHS | C17002_008 | 2.00 and over |
| B19057 | PUBLIC ASSISTANCE INCOME IN THE PAST 12 MONTHS FOR HOUSEHOLDS | B19057_001 | Total |
| B19057 | PUBLIC ASSISTANCE INCOME IN THE PAST 12 MONTHS FOR HOUSEHOLDS | B19057_002 | With public assistance income |
| B19057 | PUBLIC ASSISTANCE INCOME IN THE PAST 12 MONTHS FOR HOUSEHOLDS | B19057_003 | No public assistance income |

# Attribute Summary

Below is a heat map of all geocodable evictions in Alachua County.




Below is a temporal histogram of evictions in Alachua County.



There have been 34970 eviction filings between 01/01/2002 and 12/31/2021. Of these, 31770 were successfully geocoded. The median eviction in terms of time occurred in quarter 4 of 2009. The maximum evictions in a year, 2369 took place on 2007, and the minimum, 1050, took place on 2020.

The table below summarizes descriptions of centers for the attribute fields of Block Groups under study for P2.

| Field | Mean | Std. Deviation | Median | Min-Max |
|---|---|---|---|---|
| Hispanic or Latino proportion | 0.117 | 0.049 | 0.110 | 0.022-0.261 |
| Not Hispanic or Latino prop. | 0.883 | 0.049 | 0.890 | 0.739-0.977 |
| White alone prop. | 0.575 | 0.184 | 0.607 | 0.030-0.862 |
| Black or African American prop. | 0.193 | 0.1889 | 0.122 | 0.015-0.883 |
| American Indian or Alaska Native prop. | 0.002 | 0.002 | 0.002 | 0-0.011 |
| Asian prop. | 0.059 | 0.058 | 0.037 | 0-0.340 |
| Native Hawai'ian or Pacific Islander prop. | 0.000 | 0.001 | 0 | 0-0.006 |
| Some other race prop. | 0.006 | 0.004 | 0.006 | 0-0.022 |
| Two or more races prop. | 0.048 | 0.013 | 0.048 | 0.012-0.083 |
| Number of eviction cases | 201.1 | 288.0 | 82 | 1-1629 |

The same information for ACS variables:

| Field | Mean | St. Deviation | Median | Min-Max |
|---|---|---|---|---|
| Median household income in the past 12 months (in 2020 inflation-adjusted dollars) | 52335.68 | 29616.07 | 45985 | 2499 - 148621. |
| Owner occupied_prop | 0.536037 | 0.339437 | 0.633447 | 0 - 1. |
| Renter occupied_prop | 0.463963 | 0.339437 | 0.366553 | 0 - 1. |
| Married-couple household_prop | 0.347864 | 0.21324 | 0.351867 | 0.-0.848 |
| Married-couple household:With children of the householder under 18 years_prop | 0.117542 | 0.10119 | 0.099467 | 0.-0.457 |
| Married-couple household:With no children of the householder under 18 years_prop | 0.230322 | 0.143315 | 0.233553 | 0.-0.582 |

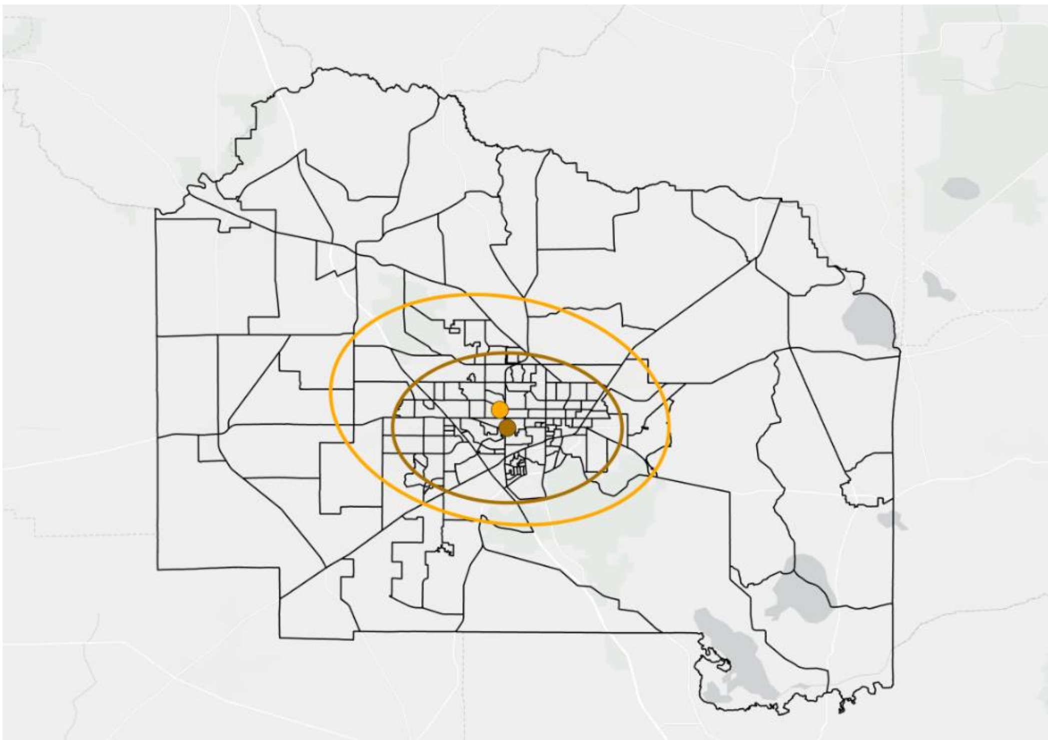
| | | | | |
|---|---|---|---|---|
| Cohabiting couple household_prop | 0.071889 | 0.083008 | 0.049087 | 0.-0.552 |
| Cohabiting couple household:With children of the householder under 18 years_prop | 0.011387 | 0.031391 | 0 | 0.-0.334 |
| Cohabiting couple household:With no children of the householder under 18 years_prop | 0.060502 | 0.076723 | 0.040949 | 0.-0.524 |
| Female householder, no spouse or partner present_prop | 0.335165 | 0.148419 | 0.340109 | 0.029-0.723 |
| Female householder, no spouse or partner present:Living alone_prop | 0.204658 | 0.118344 | 0.188097 | 0.-0.625 |
| Female householder, no spouse or partner present:With children of the householder under 18 years_prop | 0.043295 | 0.065401 | 0.017245 | 0.-0.362 |
| Female householder, no spouse or partner present:With relatives, no children of the householder under 18 years_prop | 0.049709 | 0.055956 | 0.033998 | 0.-0.272 |
| Female householder, no spouse or partner present:With only nonrelatives present_prop | 0.037503 | 0.070239 | 0 | 0.-0.367 |
| Male householder, no spouse or partner present_prop | 0.245083 | 0.136702 | 0.223596 | 0.-0.688 |
| Male householder, no spouse or partner present:Living alone_prop | 0.178178 | 0.111467 | 0.163163 | 0.-0.688 |
| Male householder, no spouse or partner present:With children of the householder under 18 years_prop | 0.008464 | 0.024048 | 0 | 0.-0.172 |
| Male householder, no spouse or partner present:With relatives, no children of the householder under 18 years_prop | 0.014314 | 0.025008 | 0 | 0.-0.138 |
| Male householder, no spouse or partner present:With only nonrelatives present_prop | 0.044127 | 0.087696 | 0 | 0.-0.526 |
| 1, detached_prop | 0.535125 | 0.357074 | 0.626853 | 0 - 1. |
| 1, attached_prop | 0.031074 | 0.050102 | 0 | 0.-0.257 |

| | | | | |
|---|---|---|---|---|
| 2_prop | 0.026001 | 0.058178 | 0 | 0.-0.344 |
| 3 or 4_prop | 0.053406 | 0.074569 | 0.008511 | 0.-0.33 |
| 5 to 9_prop | 0.07442 | 0.112811 | 0 | 0.-0.5 |
| 10 to 19_prop | 0.088307 | 0.138049 | 0 | 0.-0.642 |
| 20 to 49_prop | 0.058143 | 0.116978 | 0 | 0.-0.645 |
| 50 or more_prop | 0.048746 | 0.109412 | 0 | 0.-0.932 |
| Mobile home_prop | 0.084004 | 0.155029 | 0 | 0.-0.84 |
| Boat, RV, van, etc._prop | 0.000774 | 0.005613 | 0 | 0.-0.062 |
| Built 2014 or later_prop | 0.035248 | 0.064884 | 0.010089 | 0.-0.512 |
| Built 2010 to 2013_prop | 0.023901 | 0.041999 | 0 | 0.-0.249 |
| Built 2000 to 2009_prop | 0.154587 | 0.14727 | 0.12257 | 0.-0.859 |
| Built 1990 to 1999_prop | 0.181881 | 0.15156 | 0.154573 | 0.-0.731 |
| Built 1980 to 1989_prop | 0.184001 | 0.139151 | 0.169099 | 0.-0.796 |
| Built 1970 to 1979_prop | 0.182743 | 0.152648 | 0.153884 | 0.-0.788 |
| Built 1960 to 1969_prop | 0.109094 | 0.107101 | 0.073782 | 0.-0.375 |
| Built 1950 to 1959_prop | 0.069713 | 0.111656 | 0.02774 | 0.-0.671 |
| Built 1940 to 1949_prop | 0.024603 | 0.050592 | 0 | 0.-0.291 |
| Built 1939 or earlier_prop | 0.034229 | 0.090797 | 0 | 0.-0.72 |
| Under .50_prop | 0.140958 | 0.177929 | 0.074451 | 0.-0.832 |
| .50 to .99_prop | 0.095662 | 0.105136 | 0.058407 | 0.-0.475 |
| 1.00 to 1.24_prop | 0.049343 | 0.069507 | 0.028355 | 0.-0.376 |
| 1.25 to 1.49_prop | 0.045799 | 0.062814 | 0.023222 | 0.-0.353 |
| 1.50 to 1.84_prop | 0.066351 | 0.070391 | 0.047745 | 0.-0.452 |
| 1.85 to 1.99_prop | 0.025131 | 0.04021 | 0.008256 | 0.-0.232 |
| 2.00 and over_prop | 0.576757 | 0.257658 | 0.582606 | 0.-0.983 |
| With public assistance income_prop | 0.022467 | 0.047306 | 0 | 0.-0.411 |
| No public assistance income_prop | 0.977533 | 0.047306 | 1 | 0.589- 1 |
| Under 1.00_prop | 0.23662 | 0.218883 | 0.173212 | 0.-0.87 |

# Spatial Distribution Summary

## Mean Centers and Directional Distributions

The below map describes the unweighted (orange) and weighted (brown) mean centers and directional distributions using Block Groups and *Number of eviction cases* as the weight field.

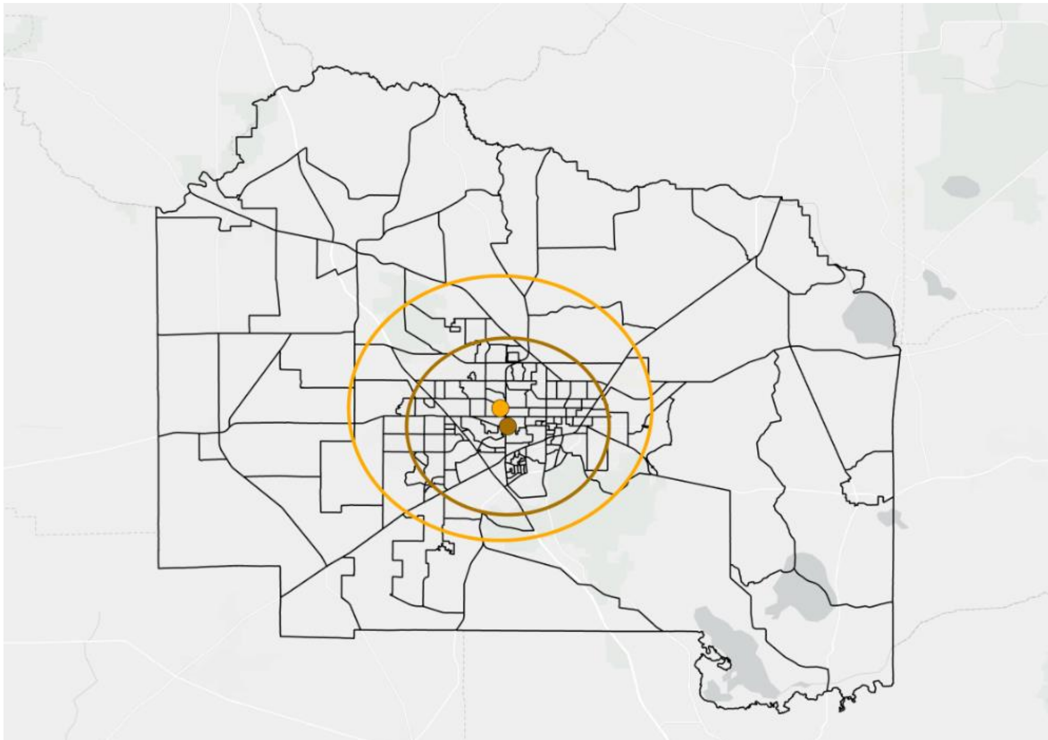


The weighted directional distribution is notably much smaller in area than the unweighted case, which may indicate that evictions are more generally concentrated at the block group level than the block group geometries themselves. The mean centers for unweighted and weighted cases are roughly in the same area. The following table summarizes the locations and areas of each statistic in the NAD83 Florida Albers projection:

| | |
|---|---|
| Unweighted mean center: location | 556874.136047, 629867.656715 |
| Weighted mean center: location | 557497.603185, 628228.658820 |
| Unweighted directional distribution: area | 4.326 * 10^3 sq. km |
| Unweighted directional distribution: perimeter | 7.490 * 10^2 km |
| Weighted directional distribution: area | 1.931 * 10^3 sq. km |
| Weighted directional distribution: perimeter | 5.003 * 10^2 km |

## Standard Distance

The below map describes the unweighted (orange) and weighted (brown) mean centers and *standard distances* using Block Groups and *Number of eviction cases* as the weight field.

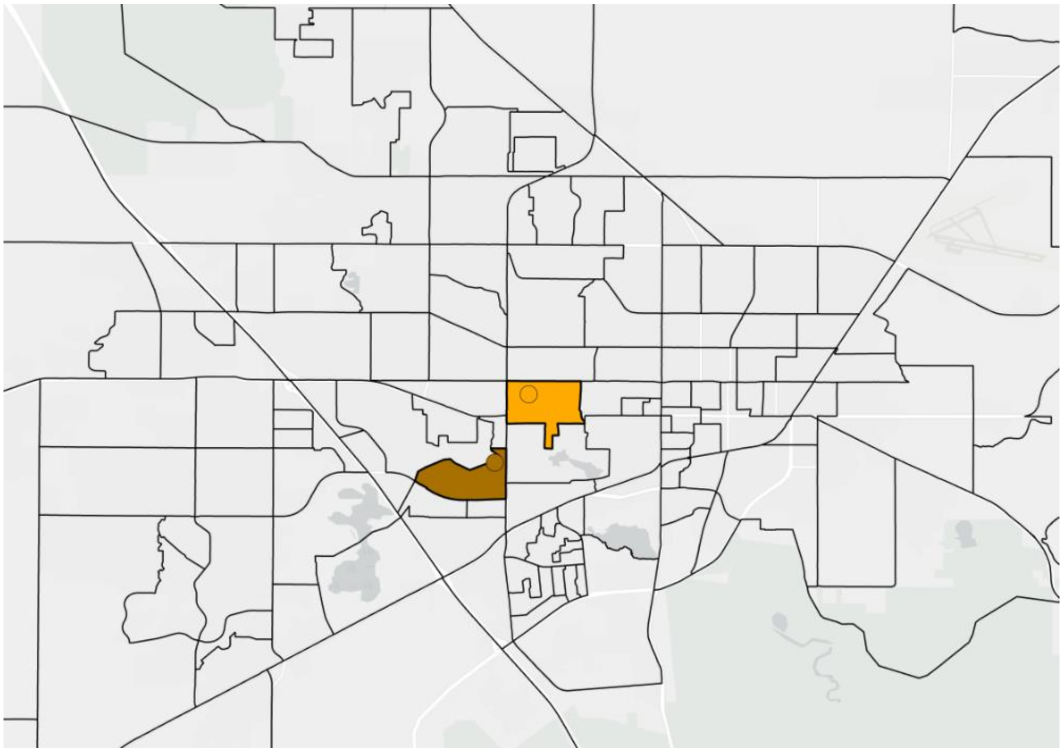


In general, we also find that the weighted standard distance is smaller than the unweighted standard distance, indicating a much tighter distribution of eviction cases in the county than the block groups alone.

| | |
|---|---|
| Unweighted standard distance | 1.20 * 10^2 km |
| Weighted standard distance | 8.00 * 10^1 km |

## Median Center and Median Feature

The map below describes the unweighted (orange) and weighted (brown) median centers and median features using Block Groups and *Number of eviction cases* as the weight field.



While both median centers are no more than a half mile away from each other, the slight southwest position of the weighted median center and feature may mean that evictions find their center in a further southwest quadrant than what is predicted by the block groups alone.

| | |
|---|---|
| Unweighted median center: location | 556874.136047, 629867.656715 |
| Weighted median center: location | 557497.603185, 628228.658820 |
| Unweighted median feature | GEOID 120010010002 |
| Weighted median feature | GEOID 120010015221 |

## Global Clustering Summaries

I completed an Average Nearest Neighbor analysis upon Block Groups. The analysis indicated that the block groups are more dispersed than would be expected at random, with significance below 5%. The full results are summarized below.

| | |
|---|---|
| Observed mean distance | 2106.4 m |
| Expected mean distance | 1908.2 m |
| Nearest neighbor ratio | 1.104 |
| *z*-score | +2.514 (dispersed) |
| *p*-value | 0.011 |

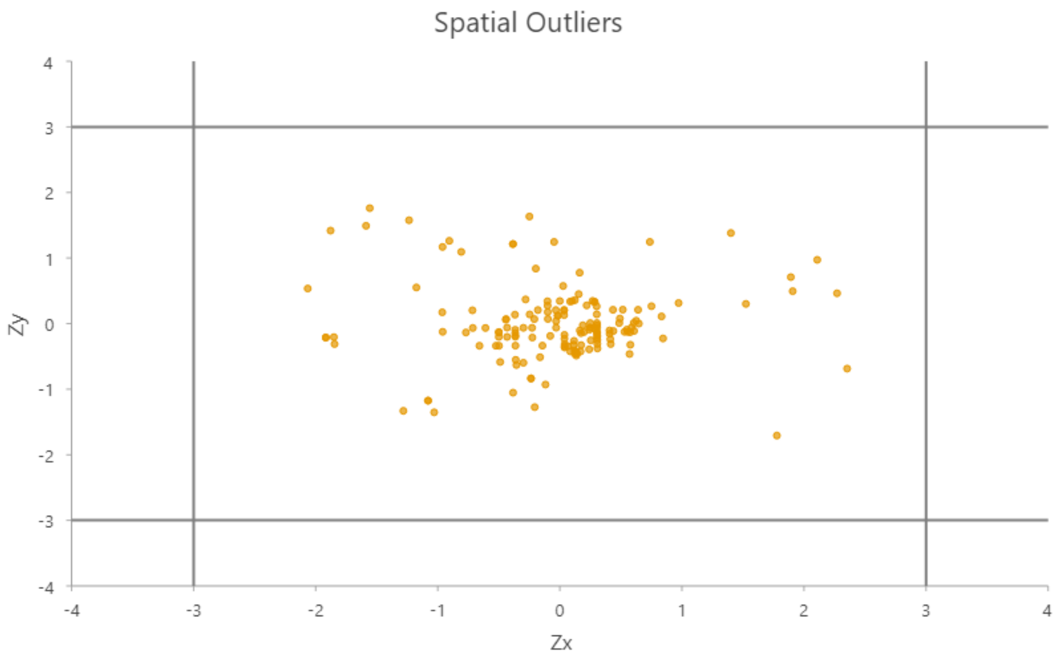
I also completed a Spatial Autocorrelation analysis using Global Moran's I. The analysis indicated that *Number of eviction cases* with block groups was much more significantly clustered than would be expected at random, with significance far below 1%. The full results are summarized below.

| | |
|---|---|
| Moran's index | 0.115 |
| Expected index | -0.0063 |
| Variance | 0.001 |
| *z*-score | +5.286 (clustered) |
| *p*-value | 0.000 |
| | |

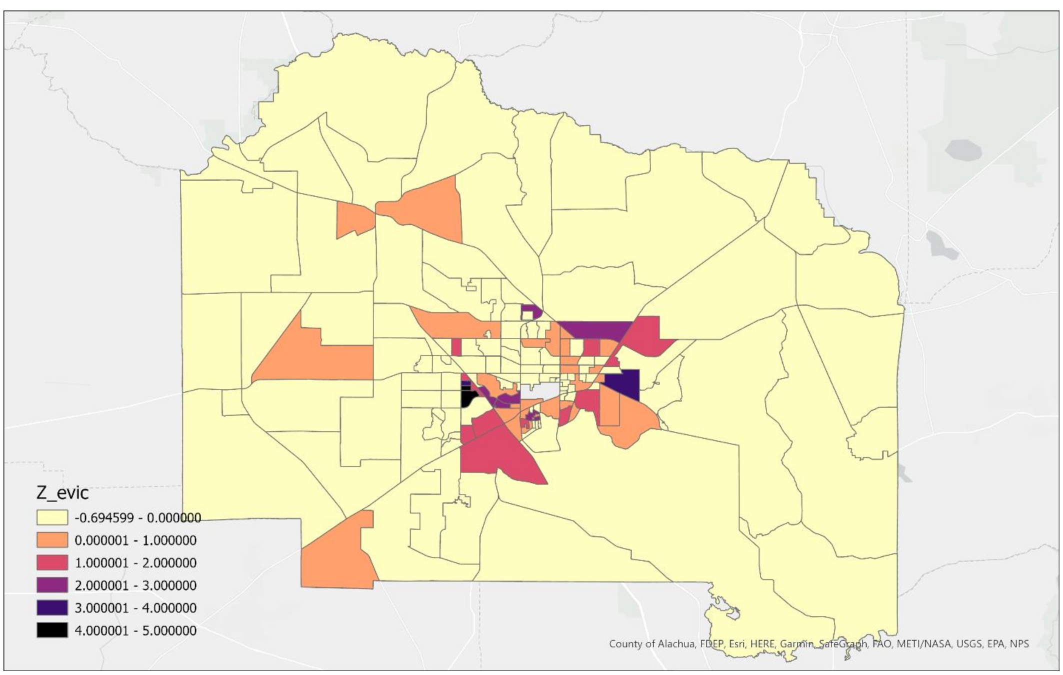
The Average Nearest Neighbor and Spatial Autocorrelation analyses are at odds with each other, but given the results of the centrality measures, we can safely ignore the average nearest neighbor results as an artifact of the design of block groups and the population distribution of Alachua County.

## Outliers

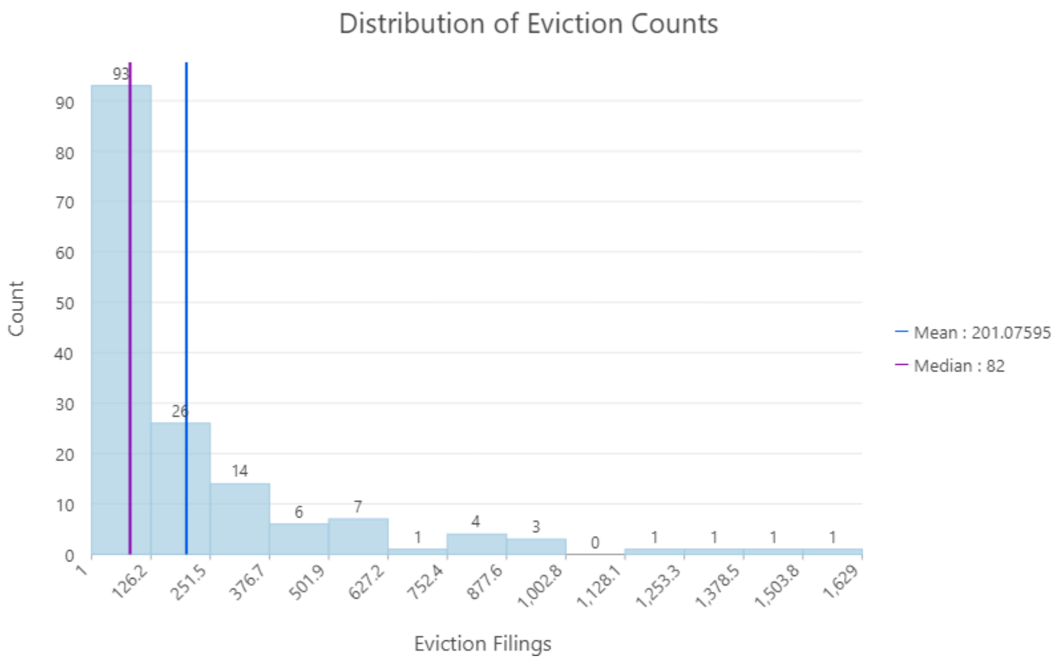
Using the Near Tool method described in lecture, I created a graph that displays the *z*-scores for all block group locations – no block groups fall outside the 3 standard distances indicated.



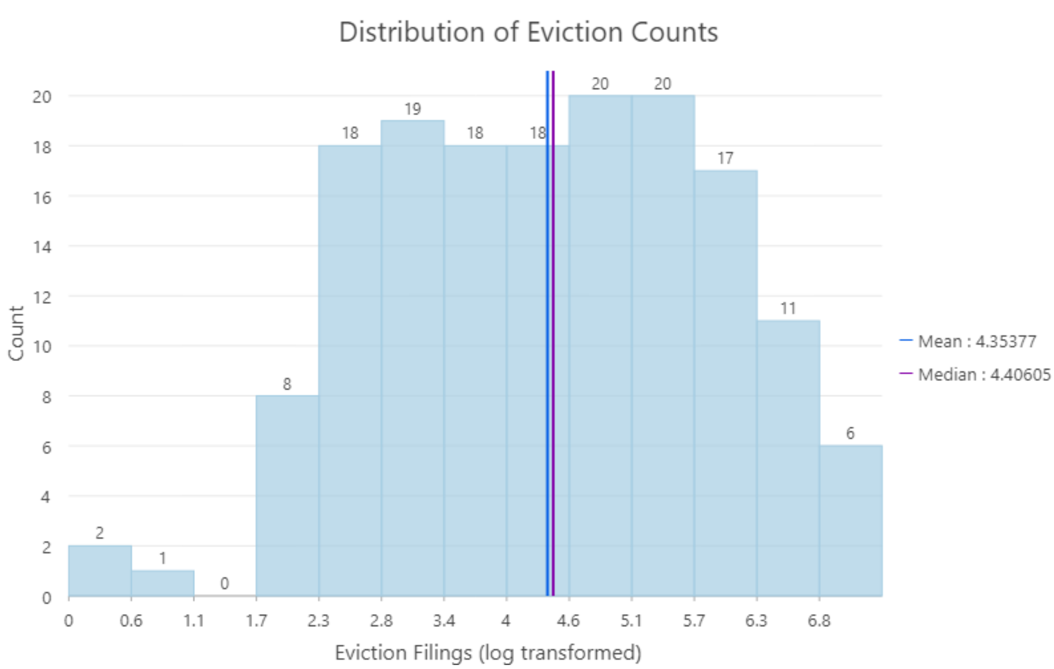

As for attribute outliers, the same *z*-score statistic found multiple block groups with evictions exceeding 3 standard deviations from the mean:



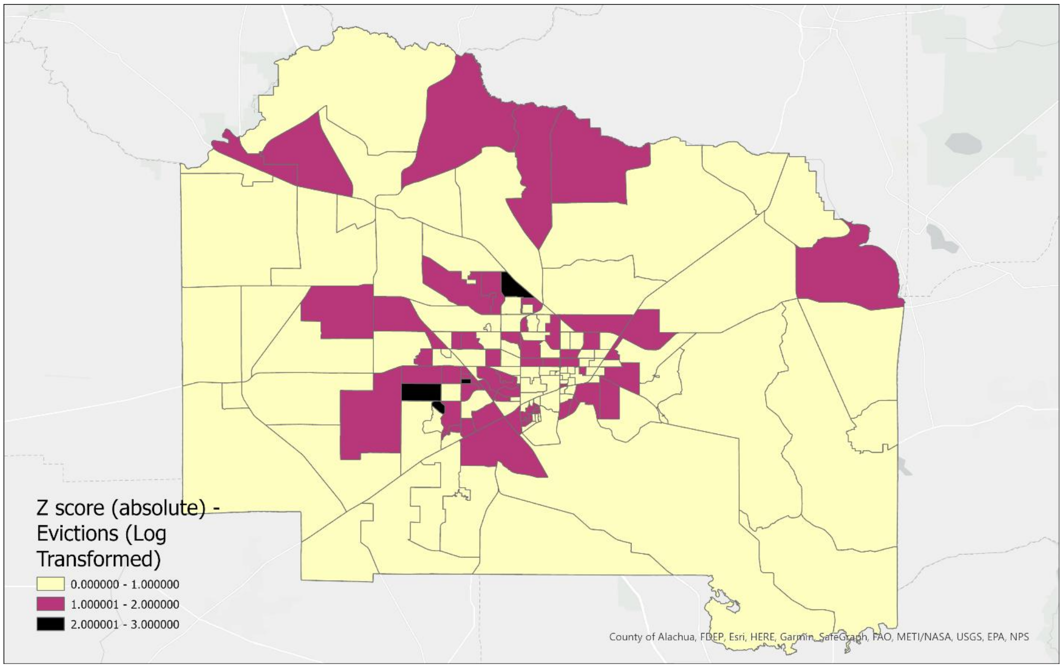

Although this could lead to rejection of analysis of certain block groups, we find that this field is *not* normally distributed, as seen below.




Indeed, with a logarithmic transformation, the data approach a normal distribution.




With this transformation, no block groups remain as outliers.

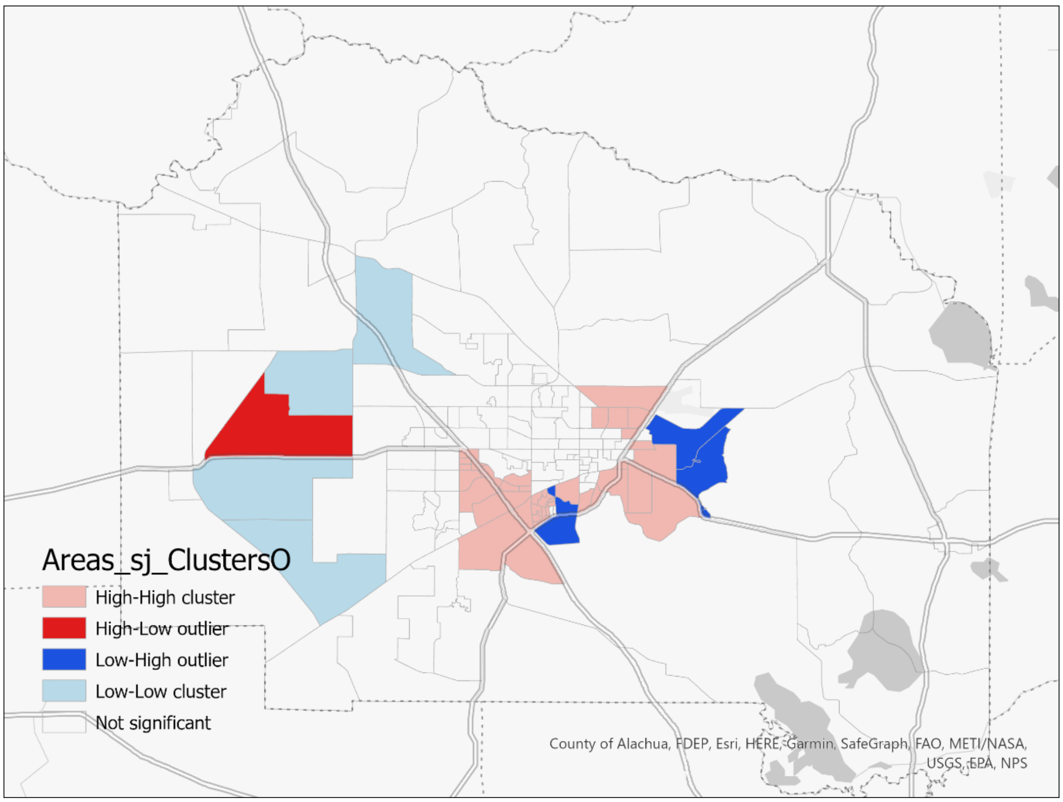



Certain locations do jump out with high Z scores – these include GEOID 120010022171, corresponding to Sparrow Condominiums and to the Holly Heights apartment complex which have many evictions as well as being surrounded by eviction-prone areas, GEOID 120010022082 and 120010022221, located south of Tioga which has no evictions despite many occurring around them, and 120010018032 in North Gainesville which has only a handful of evictions despite surrounding block groups having many.

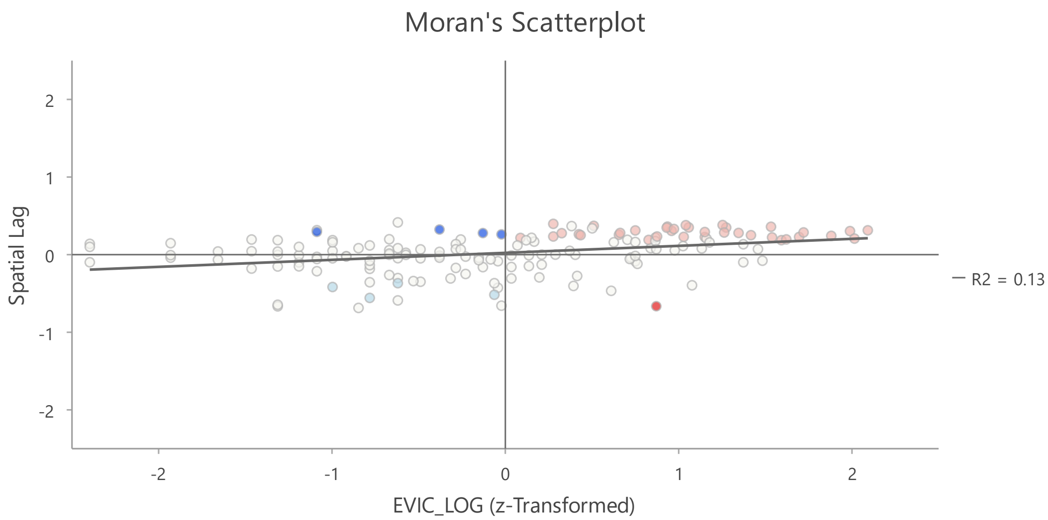
## Cluster, Outlier, and Hotspot Analyses

### Anselin Local Moran's I

I completed an Anselin Local Moran's I analysis to identify areas where evictions locally cluster and where eviction counts are local outliers. The analysis was completed using an Inverse Distance conceptualization, Euclidean distance, row standardization, FDR correction, and a default distance search threshold of 11.7 km was chosen by the algorithm.



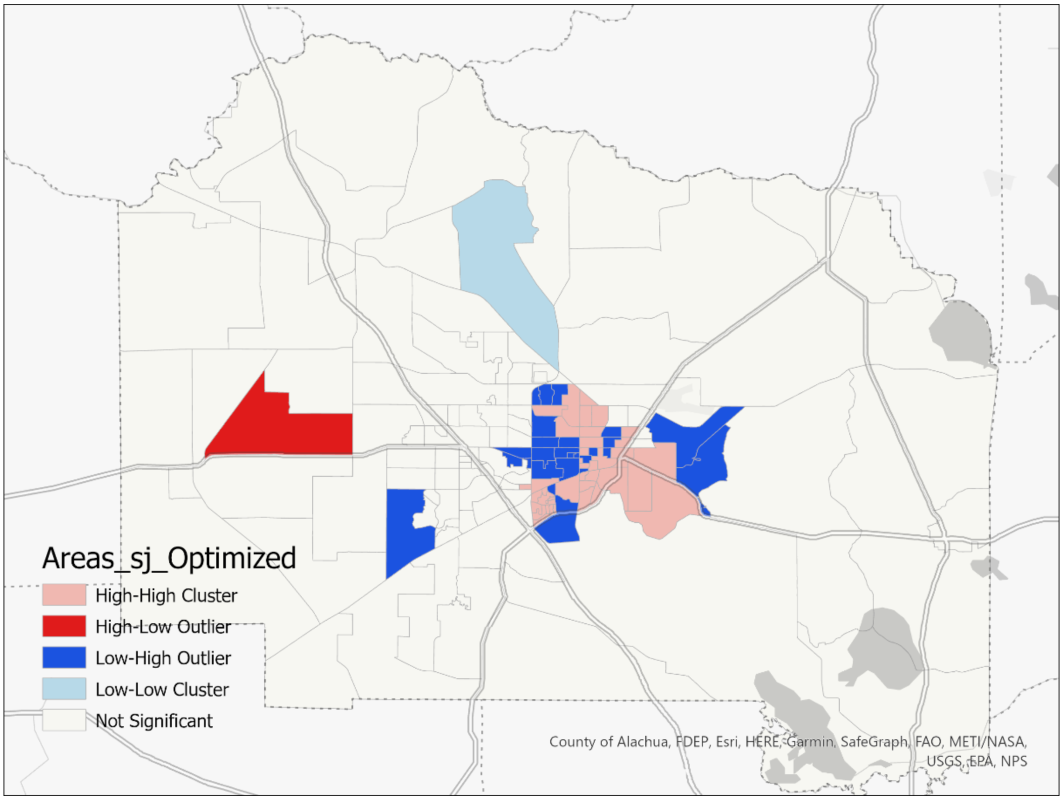

The analysis, completed on log-transformed eviction counts, found a high-high cluster of evictions across 36 block groups in southwest, south, and east Gainesville, outliers around 34th Street and Williston, and rural East Gainesville. Analysis also found a band of low clusters along US 41 and CR 235, with a high-low outlier around northeast Newberry-Tioga.



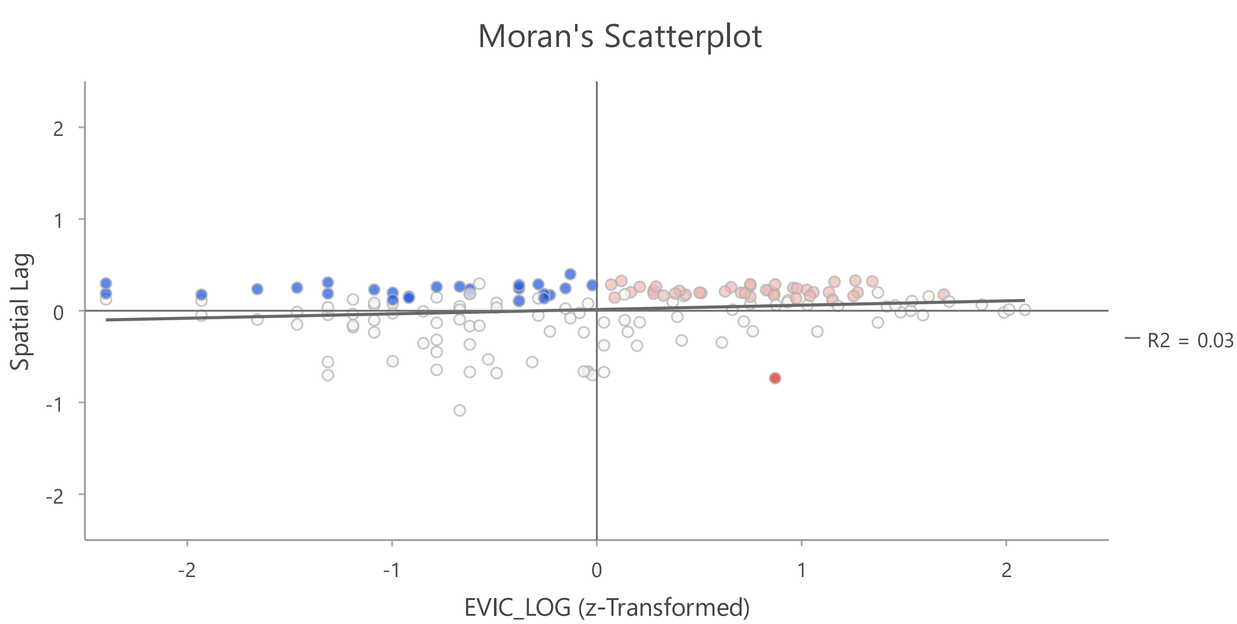

Examining the spatial lag shows the clustering of high z-transformed counts in the top-right quadrant. While there might be a rough pattern of spatial dependence occurring (that, specifically here, a block group's proximity to other eviction-prone block groups affects evictions), the small R^2 value does not provide any evidence to reject the assumption of spatial independence.

## Optimized Outlier Analysis

Next, I completed an optimized outlier analysis on the log-transformed eviction counts. I used 999 permutations and FDR correction as parameters. The algorithm identified two locational outliers in the dataset. No optimal distance for clustering was found, so a distance band of 9.1 km was derived from the average distance to the nearest 8 neighbors.

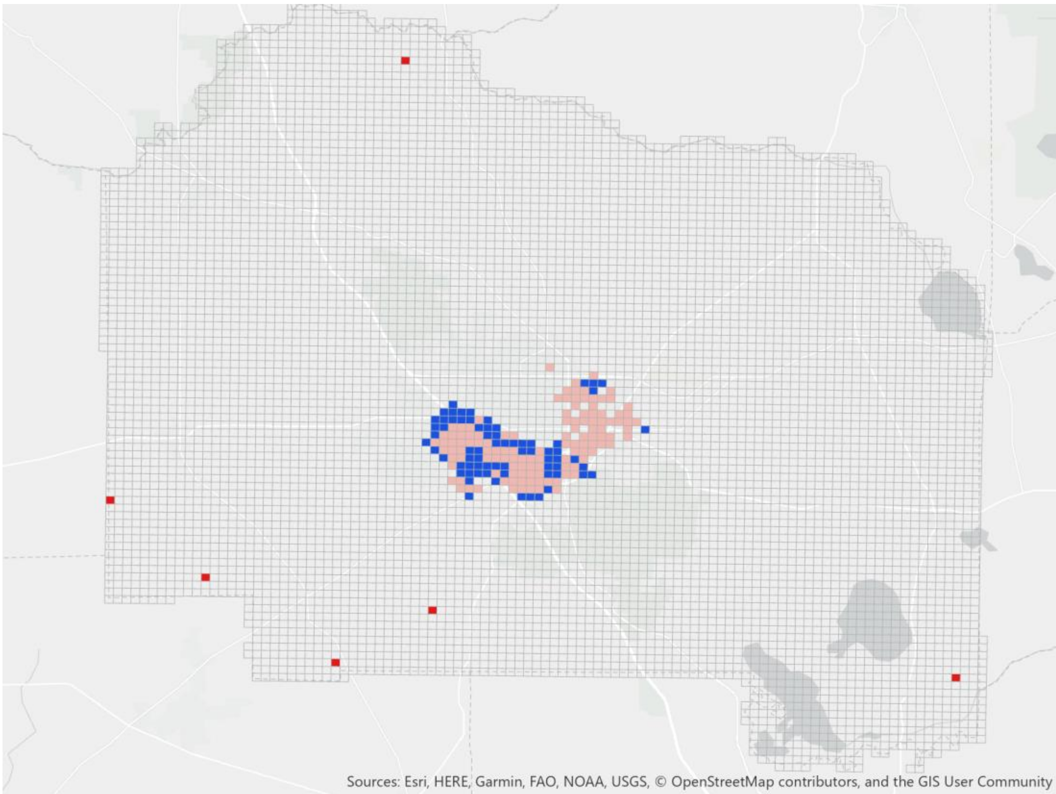



A high-high cluster of 40 block groups with high evictions was found mostly in east and central Gainesville, with a band of 25 low-high outliers corresponding to UF-related areas and to rural east Gainesville. A single high-low outlier block group was found in Newberry-Tioga, a low-high outlier in between Gainesville and Archer, and a mostly rural low-low cluster in the northern part of the county.

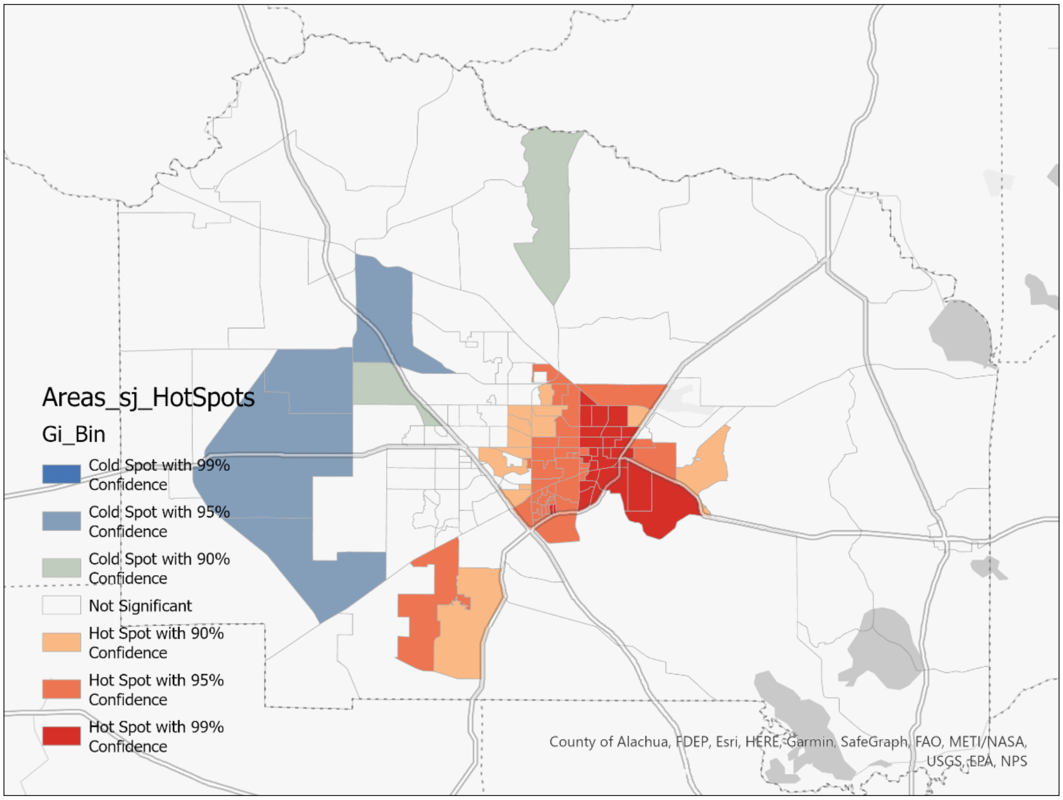


Z values against spatial lag cluster similarly to the non-optimized Local Moran's I, with most high-high values in the top right quadrant and low-high outliers in the top left. The small $R^2$ here does not provide evidence to refute spatial independence.

Conducting the same analysis using fishnet aggregation polygons on the original eviction data (hexagons oddly enough produced no significant outcomes), we find a high-high cluster of evictions stretching from southwest Gainesville through to east Gainesville, with low-high outliers mostly on the outer edges of those areas and in the southwest part of southwest Gainesville. There are also high-low outliers sprinkled across the rural parts of the county.

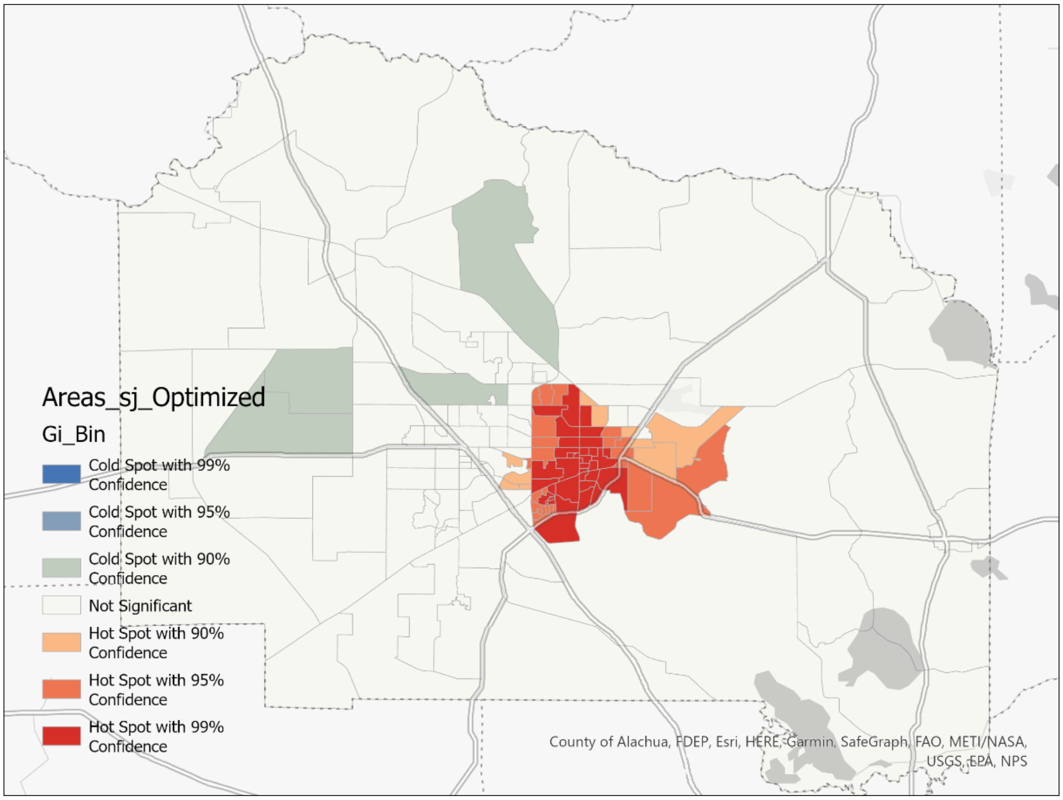


## Getis-Ord Gi* Hotspot Analysis



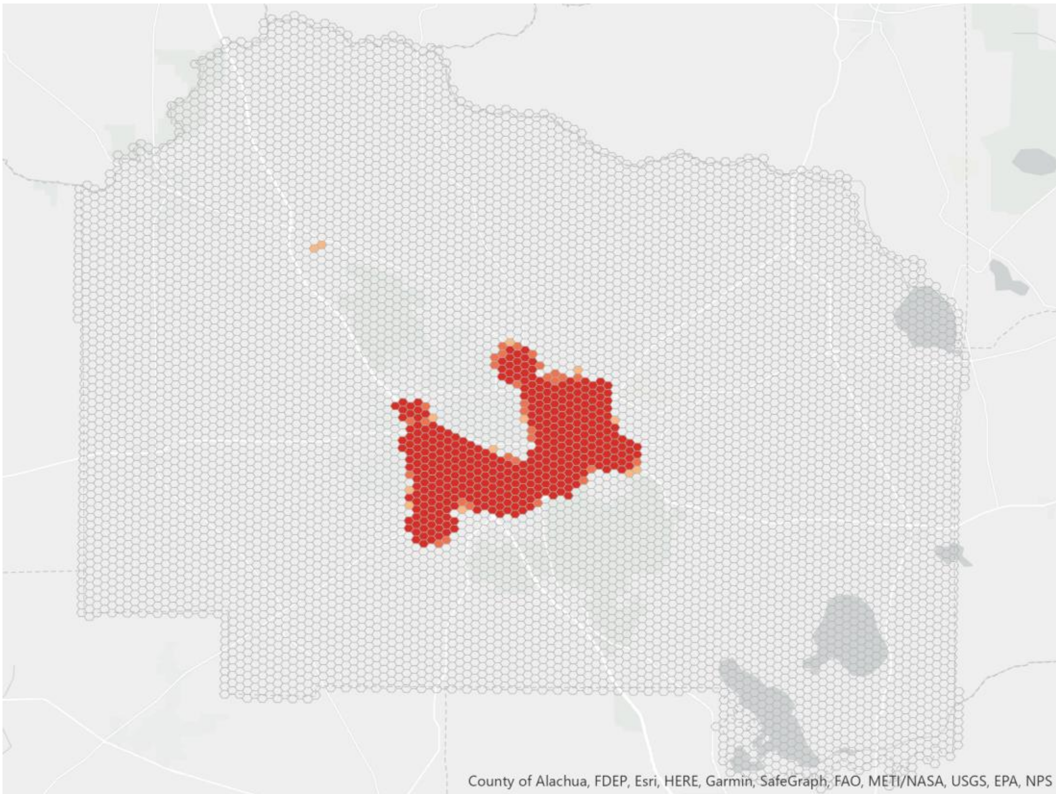

Hotspot analysis returned a cold band along the US 41-CR 235 corridor from Archer to Alachua via Newberry. It also found a random weak hotspot in southern Alachua County, and a confident hot spot in East Gainesville. It also found a weak cold spot in northern Alachua County.

## Optimized Getis-Ord Gi* Hotspot Analysis




Compared to the unoptimized hotspot analysis, only Gainesville east of 34th Street lights up as a significant hotspot, with isolated cold spots in the west and north parts of the county.

Carrying out a hotspot analysis using hexagons instead of block groups as the summarizing geometry, we come up with the following:



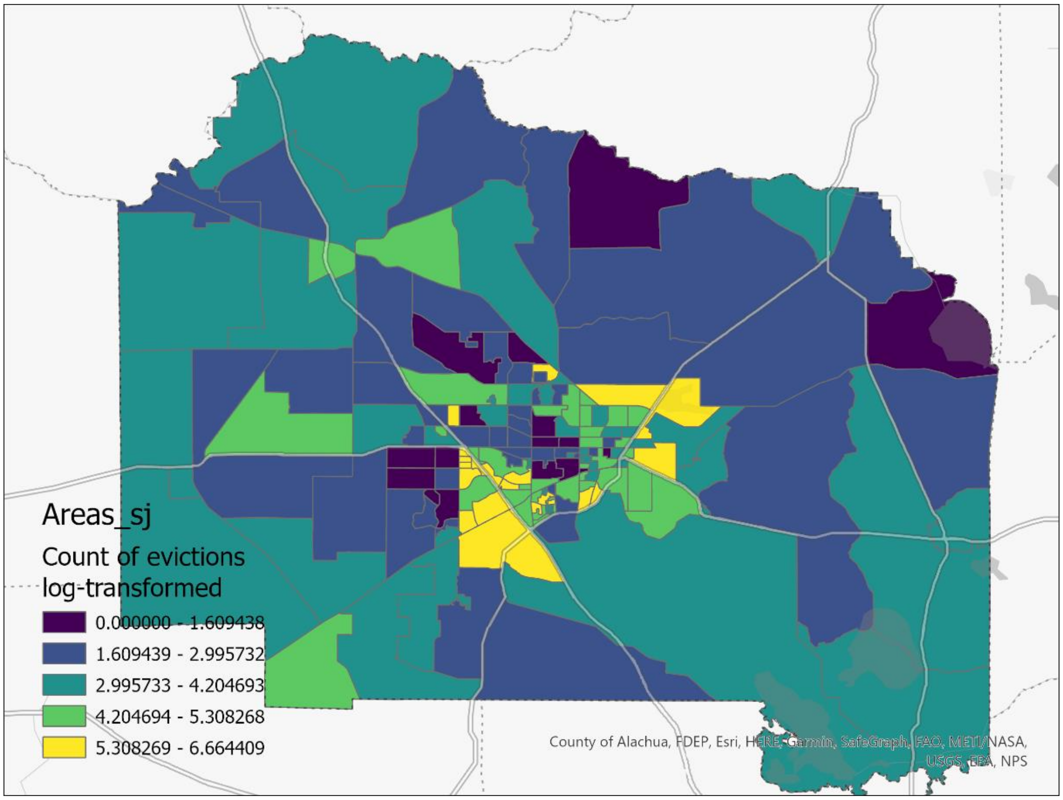

This finds that most of Gainesville except the northwest neighborhoods, and a small area of the City of Alachua lights up as a hotspot.

## Discussion

Between Anselin LMI and its optimized equivalent, the only areas consensed upon were a portion of east Gainesville surrounding Hawthorne Road. Other areas, for example a triangle of block groups in Southwest Gainesville, disappeared in the optimized LMI, as well as new low outliers appearing across Gainesville north and south of UF. The rural band of low-low clusters was not preserved between the two analyses. The optimized LMI, when performed using a fishnet grid on the original eviction data (not log-transformed), ends with a similar band across southwest and east Gainesville of high-high clustering, except with much more in the way of cold-hot outliers.

The two hotspot analyses consensed upon central and east Gainesville as a hotspot for evictions, albeit to various significances. A rural cold band that appeared in the unoptimized hotspot analysis disappeared in the optimized analysis. However, the hexagon grid version of the optimized hotspot analysis did return a similar hotspot area to that of the optimized LMI using a fishnet grid.

The degree to which east and central Gainesville reappears as areas of interest across the analyses is surprising, in that this area doesn't have the highest evictions in the county. On inspection of a simple visualization of log-transformed eviction counts, one would expect that the southwest cluster more appear as a high-high cluster or a hotspot, but that wasn't the case.




There *are*, however, fourth quantile areas clustered around east Gainesville that appear frequently. Perhaps that more cold bands/low-low outliers don't appear in rural parts of the county is because of the distance weighting algorithms used that may assign lower numbers of neighbors to larger and rural block groups.

As for which analysis. Anselin's LMI or Hotspot, provides the better treatment of the data, we have to turn to the objects of analyses for each. Anselin's LMI generally considers whether a feature is locally different from its neighborhood, whereas Hotspot Analysis analyzes whether a neighborhood is different from the global. This difference causes, for example, a hotspot in Hotspot Analysis to be a low-high outlier feature in Anselin's LMI. In general I prefer the results of LMI to Hotspot Analysis to describing the behavior of block groups, as the size of neighborhoods is generally small and the block groups are variable in size, being big with large-distance neighborhoods in rural areas and small with small-distance neighborhoods in other cases. Anselin's LMI also is more sensitive to local differences in features, an important factor in determining whether an area is actually a high risk area for evictions.

## Regression Analyses

To understand not only the clustering pattern but also possible determinants of evictions, I completed non-spatial and spatial regression analyses on my dataset using variables from the Census and American Community Survey.

I pulled variables on the following topics:

- Race and ethnicity
- Family and household type
- Income as a proportion of the poverty level
- Householder marital status
- Age of housing

- Type of housing structure
- Public assistance information

In other analyses, race and family type (householder status and presence of children) were previously identified as social determinants of eviction. Poverty is used as a stand in for otherwise incomplete data on income. I wanted to determine built environment determinants of eviction so used type and age of housing structure as stand-ins. Anecdotally, it seems that much of the low-income housing stock in Gainesville is composed of 1980s constructed 8-plex housing. Moreover, single-family detached housing seems to predominate in wealthier, stabler neighborhoods.

In the following analysis, I used a log-transformed eviction count as my dependent variable, transformed by: $y' = \ln(y+1)$ where y' is the transformed and y is the original eviction counts, to deduce determinants for evictions. Of the almost 60 original variables, I chose a subset of 29 that represented key racial, ethnic, householder, income, and environmental characteristics.

Criteria for passing models include:

- R^2 >= 0.50
- p value for coefficients <= 0.05
- VIF <= 7.5
- JB p value >= 0.1
- Spatial Autocorrelation p > 0.10

In a previous run of exploratory regression models, block groups representing UF were found to be outliers. These block groups were removed in all following analyses with the loss of accounting for five evictions in a small sliver of non-UF property.

ER returned the following variable significance table:

| Category | Variable | Significant | % Neg | % Pos |
|---|---|---|---|---|
| HOUSE_TYP | F3_OR_4_PROP | 99.86 | 0 | 100 |
| HOUSE_TYP | F1__DETACHED_PROP | 99.67 | 100 | 0 |
| RACE | WHITE_ALONE_PROP | 99.21 | 100 | 0 |
| HOUSE_TYP | F5_TO_9_PROP | 97.47 | 0 | 100 |
| TENANCY | RENTER_OCCUPIED_PROP_X | 96.8 | 0.73 | 99.27 |
| HOUSE_TYP | F1__ATTACHED_PROP | 96.55 | 0 | 100 |
| FAMILY | FEMALE_HOUSEHOLDER__NO_SPOUSE_OR _PARTNER_PRESENT_PROP | 93.13 | 0.05 | 99.95 |
| RACE | BLACK_OR_AFRICAN_AMERICAN_ALONE_PROP | 92.61 | 6.87 | 93.13 |
| FAMILY | MARRIED_COUPLE_HOUSEHOLD_PROP | 89.44 | 97.45 | 2.55 |
| HOUSE_TYP | F10_TO_19_PROP | 80.75 | 0.84 | 99.16 |
| POVERTY | F2_00_AND_OVER_PROP | 78.5 | 97.42 | 2.58 |
| RACE | HISPANIC_OR_LATINO_PROP | 59.95 | 18.65 | 81.35 |
| POVERTY | UNDER_1_00 | 55.66 | 38.88 | 61.12 |
| HOUSE_TYP | F2_PROP | 51.97 | 7.66 | 92.34 |
| FAMILY | MALE_HOUSEHOLDER__NO_SPOUSE_OR _PARTNER_PRESENT_PROP | 48.67 | 35.5 | 64.5 |
| HOUSE_TYP | F20_TO_49_PROP | 43.69 | 42.57 | 57.43 |
| HOUSE_AGE | BUILT_1980_TO_1989_PROP | 18.92 | 10.21 | 89.79 |
| HOUSE_AGE | BUILT_1939_OR_EARLIER_PROP | 17.4 | 98.37 | 1.63 |
| HOUSE_AGE | BUILT_1970_TO_1979_PROP | 15.99 | 16.05 | 83.95 |

| | | | | |
|---|---|---|---|---|
| HOUSE_TYP | F50_OR_MORE_PROP | 14.8 | 30.06 | 69.94 |
| HOUSE_AGE | BUILT_2014_OR_LATER_PROP | 13.87 | 98.72 | 1.28 |
| HOUSE_AGE | BUILT_1990_TO_1999_PROP | 12.19 | 48.95 | 51.05 |
| HOUSE_AGE | BUILT_1950_TO_1959_PROP | 11.19 | 79.8 | 20.2 |
| FAMILY | COHABITING_COUPLE_HOUSEHOLD_PROP | 10.4 | 37.37 | 62.63 |
| HOUSE_AGE | BUILT_1960_TO_1969_PROP | 5.62 | 59.14 | 40.86 |
| HOUSE_AGE | BUILT_2000_TO_2009_PROP | 3.2 | 82.16 | 17.84 |
| HOUSE_AGE | BUILT_2010_TO_2013_PROP | 2.93 | 17.32 | 82.68 |
| PUBLIC_BEN | WITH_PUBLIC_ASSISTANCE_INCOME_PROP | 0.52 | 96.88 | 3.12 |
| HOUSE_AGE | BUILT_1940_TO_1949_PROP | 0.19 | 57.91 | 42.09 |

Of these, only two, WHITE_ALONE and BLACK_OR_AA, were determined to be collinear.

This table suggests broadly that characteristics related to proportion of building ages are not important in determining evictions. Householder characteristics were mostly found to not be significant except notably “Single female householder”. Only certain building size characteristics were found to be significant, specifically single-family, missing middle, and 5-8-unit housing. Race and ethnicity were found to be significant most of the time with mostly steady effects. Two poverty metrics, UNDER_1_00 and OVER_2_00, were found to be significant at least 50% of the time but the former has surprisingly ambivalent effect in determining evictions.

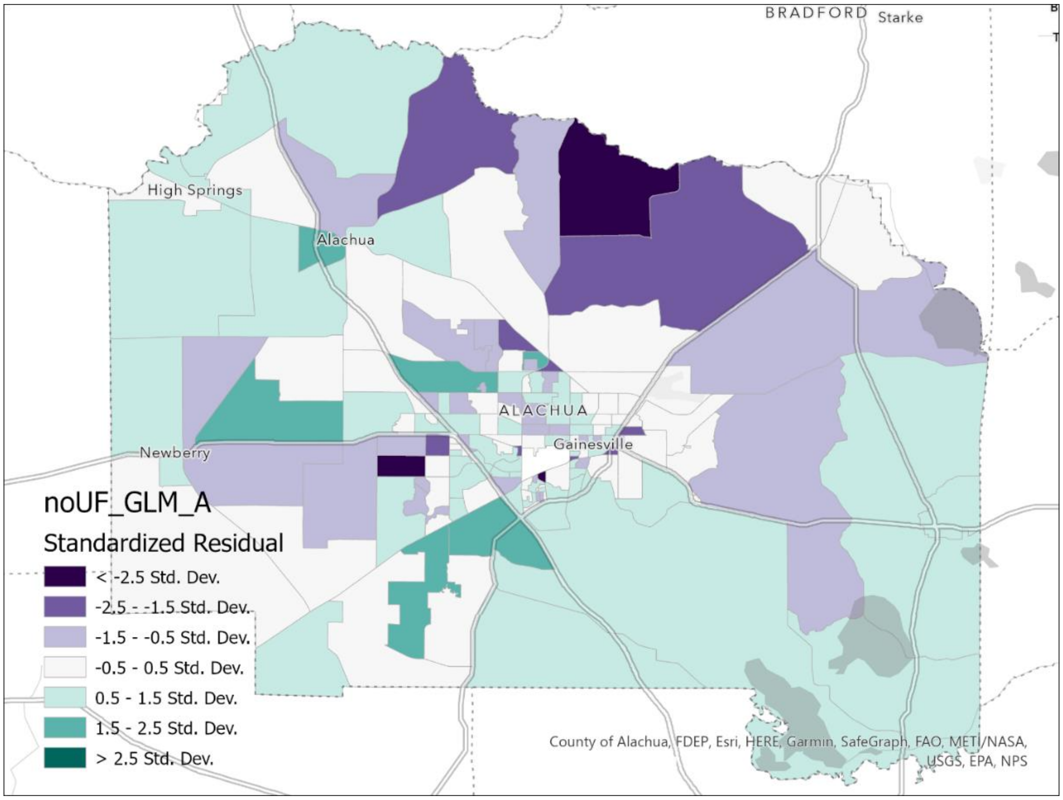
Exploratory regression returned with 14 models with less than 5 variables that passed all 5 minimum criteria. The top five (ranked by Adjusted R^2) are written below.

| Adjusted R^2 | Model parameters (-/+ correlation) |
|---|---|
| **0.68** | **(A) +HISPANIC_LATINO +BLACK_AA - SINGLE_FAMILY_DETACHED - BELOW_POVERTY_LINE** |
| 0.61 | +HISPANIC_LATINO+BLACK_AA+SINGLE_FAMILY_ATTACHED+5_9_UNIT |
| 0.61 | +HISPANIC_LATINO+BLACK_AA+5_9_UNIT+10_19_UNIT |
| **0.60** | **(B) +HISPANIC_LATINO+BLACK_AA+5_9_UNIT** |
| 0.59 | +BLACK_AA+SINGLE_FAMILY_ATTACHED+10_19_UNIT+BUILT_2010_2013 |

The two models marked A and B are of interest because of their high R^2 while balancing number of variables to avoid overfit.

## Model A

I completed a generalized linear regression on Model A to get more information on the underlying model. The map of residuals is listed below:




Three block groups were found to have residual outliers further than -2.5 standard deviations away (the model vastly underpredicted the number of evictions): one is an affluent single-family block southeast of Tioga, another is a small block group just southwest of Shands, and another mostly rural area southeast of Brooker.

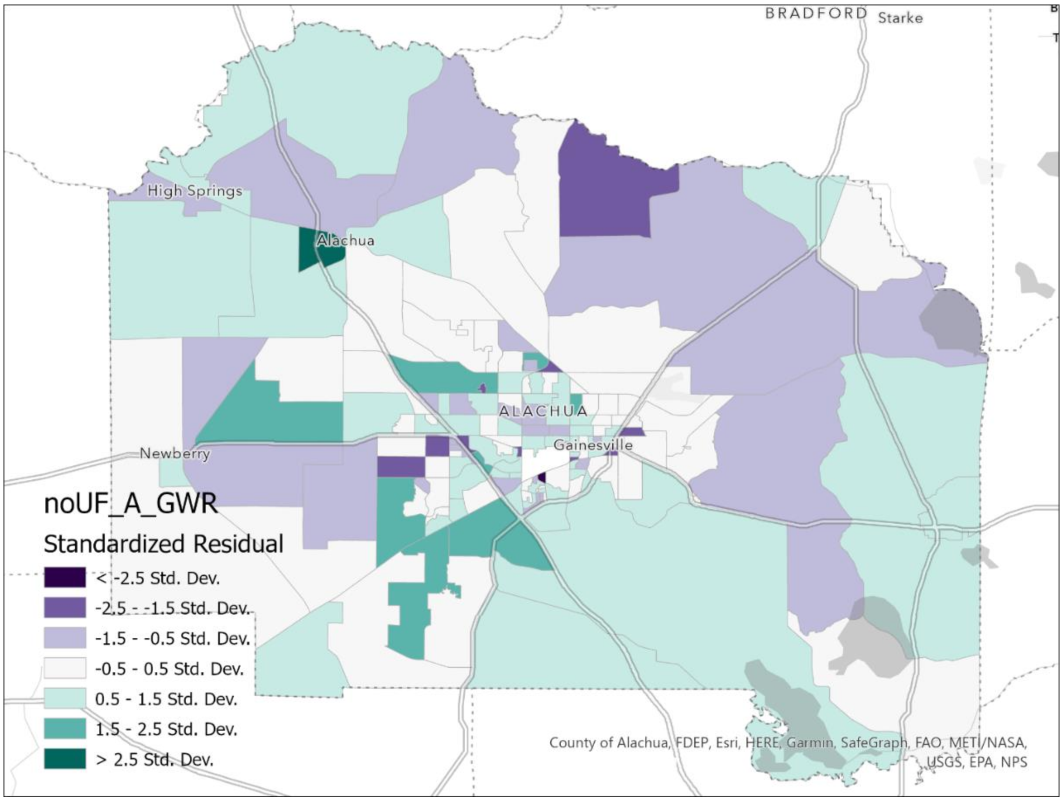
| | | | |
|---|---|---|---|
| Input Features | Areas_sj_noUF | Dependent Variable | EVIC_LOG |
| Number of Observations | 157 | Akaike's Information Criterion (AICc)[d] | 388.083337 |
| Multiple R-Squared | 0.691083 | Adjusted R-Squared[d] | 0.682954 |
| Joint F-Statistic | 85.010419 | Prob(>F), (4,152) degrees of freedom | 0.000000* |
| Joint Wald Statistic | 431.962993 | Prob(>chi-squared), (4) degrees of freedom | 0.000000* |
| Koenker (BP) Statistic | 1.002506 | Prob(>chi-squared), (4) degrees of freedom | 0.909416 |
| Jarque-Bera Statistic | 2.212085 | Prob(>chi-squared), (2) degrees of freedom | 0.330866 |

According to the results above, the global model performed relatively well with an Adjusted R^2 of 0.68 and a p value for the F test at << 0.001.

| **Variable** | **Coefficient**[a] | **StdError** | **t-Statistic** | **Probability**[b] | **VIF**[c] |
|---|---|---|---|---|---|
| Intercept | 3.504173 | 0.370817 | 9.449883 | 0.000000* | -------- |
| HISPANIC_OR_LATINO | 6.620830 | 1.875372 | 3.530408 | 0.000559* | 2.009764 |
| BLACK_OR_AFRICAN_AMERICAN | 4.536040 | 0.376304 | 12.054202 | 0.000000* | 1.204815 |
| F1__DETACHED_PROP | -2.283458 | 0.274107 | -8.330522 | 0.000000* | 2.209236 |
| UNDER_1_00 | -1.358148 | 0.386296 | -3.515820 | 0.000588* | 1.674455 |

GLR also predicts relatively high coefficients for HISPANIC_OR_LATINO and BLACK_AA. Of the former, every 10% change results in an increase of 1.93x. For BLACK_AA, every 10% change results in a increase in evictions of 1.57x. For single-family and poverty, the effect is less drastic, with a decrease of 1.25x and 1.14x respectively. All variables regardless were found to be significant at $p < 0.001$.

I also performed a geographically weighted regression using a bisquare conceptualization of weight and a golden search method. Golden search was not able to find an optimal number of neighbors through its algorithm, so it chose n_neighbors = 125.



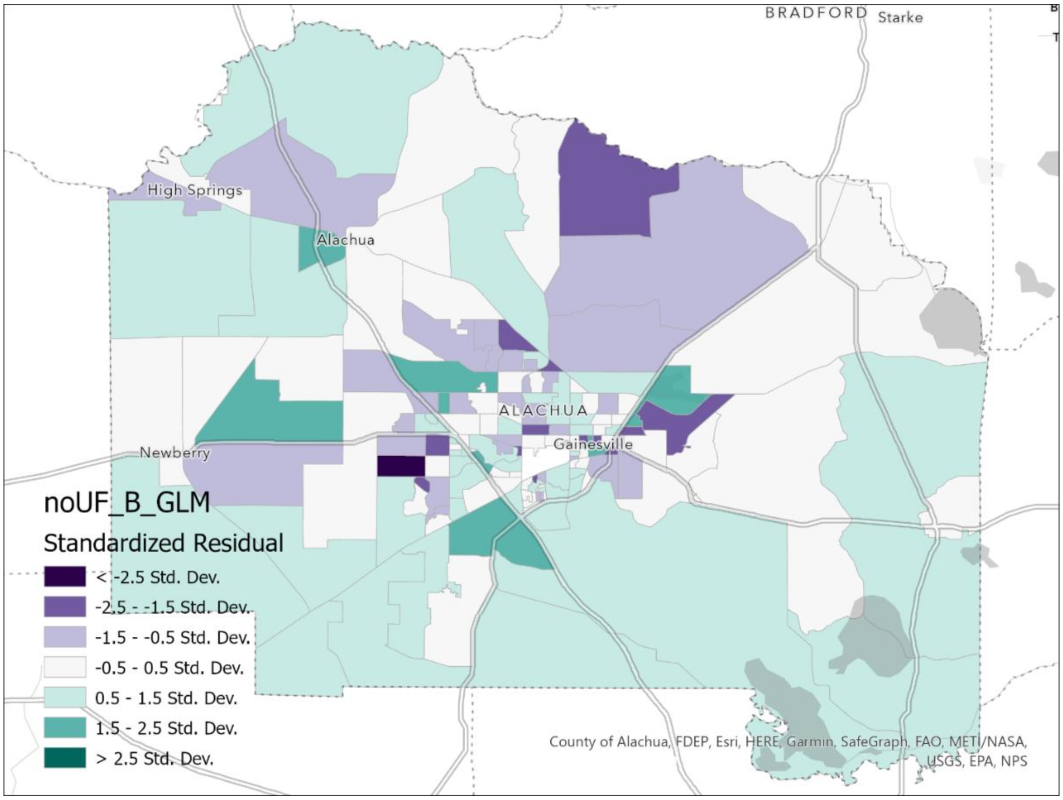

A small block group near Shands remains underpredicted compared to the global model, but two census blocks corresponding to Alachua were newly overpredicted (more evictions predicted than actual).

Overall the local model had a better fit to the data than the global model judging by the lower AICc and higher R^2 values.

| | |
|---|---|
| R2 | 0.7330 |
| AdjR2 | 0.7013 |
| AICc | 382.4004 |
| Sigma-Squared | 0.6186 |
| Sigma-Squared MLE | 0.5532 |
| Effective Degrees of Freedom | 140.4104 |

## Model B

The map of residuals for generalized linear regression is presented below:




The only deviant block group is an underprediction of an affluent area southeast of Tioga.

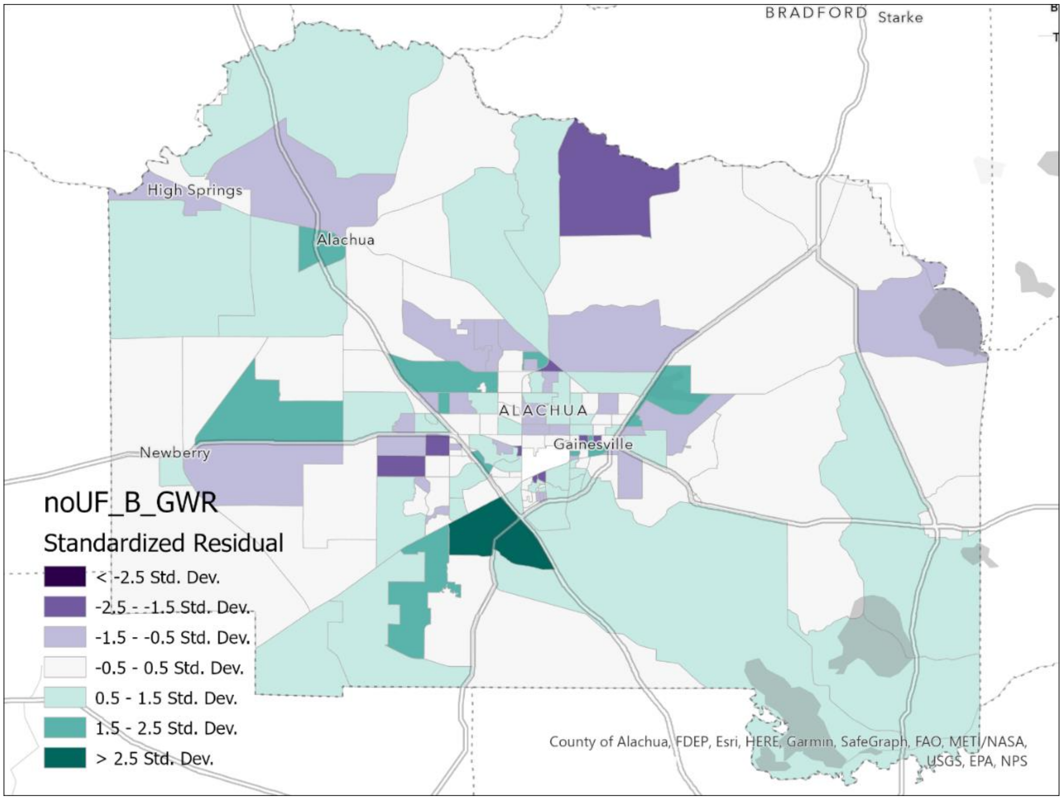
| Input Features | Areas_sj_noUF | Dependent Variable | EVIC_LOG |
|---|---|---|---|
| Number of Observations | 157 | Akaike's Information Criterion (AICc) | 423.279586 |
| Multiple R-Squared | 0.608092 | Adjusted R-Squared | 0.600408 |
| Joint F-Statistic | 79.132579 | Prob(>F), (3,153) degrees of freedom | 0.000000* |
| Joint Wald Statistic | 243.061715 | Prob(>chi-squared), (3) degrees of freedom | 0.000000* |
| Koenker (BP) Statistic | 1.070243 | Prob(>chi-squared), (3) degrees of freedom | 0.784262 |
| Jarque-Bera Statistic | 1.153970 | Prob(>chi-squared), (2) degrees of freedom | 0.561589 |

This global model B performed slightly worse than global Model A, coming in at only 0.61. The F statistic remains significant at $p << 0.001$.

| Variable | Coefficient | StdError | t-Statistic | Probability | VIF |
|---|---|---|---|---|---|
| Intercept | 1.295666 | 0.233554 | 5.547614 | 0.000000* | -------- |
| HISPANIC_OR_LATINO | 9.858470 | 1.845197 | 5.342774 | 0.000001* | 1.543693 |
| BLACK_OR_AFRICAN_AMERICAN | 4.486798 | 0.406884 | 11.027230 | 0.000000* | 1.117607 |
| F5_TO_9_PROP | 4.143371 | 0.869253 | 4.766587 | 0.000005* | 1.494599 |

This model also predicts huge impacts of HISPANIC_LATINO on evictions, with a 10% increase in HISPANIC_LATINO population being associated with a 2.69x increase in evictions. The same increase in BLACK_AA and 5_TO_9 is predicted to result in 1.57x and 1.51x increases respectively.

Using GWR, The local Model B underpredicts the same areas as local Model A: Alachua and the area southwest of Shands, but also a relatively mixed racial block group just south of I75 and Archer (composed of mobile home/multifamily tracts and single-family).




Golden search again could not find an optimal number of neighbors, so it chose n_neighbors = 118

| | |
|---|---|
| R2 | 0.6655 |
| AdjR2 | 0.6300 |
| AICc | 414.3864 |
| Sigma-Squared | 0.7663 |
| Sigma-Squared MLE | 0.6932 |
| Effective Degrees of Freedom | 142.0275 |

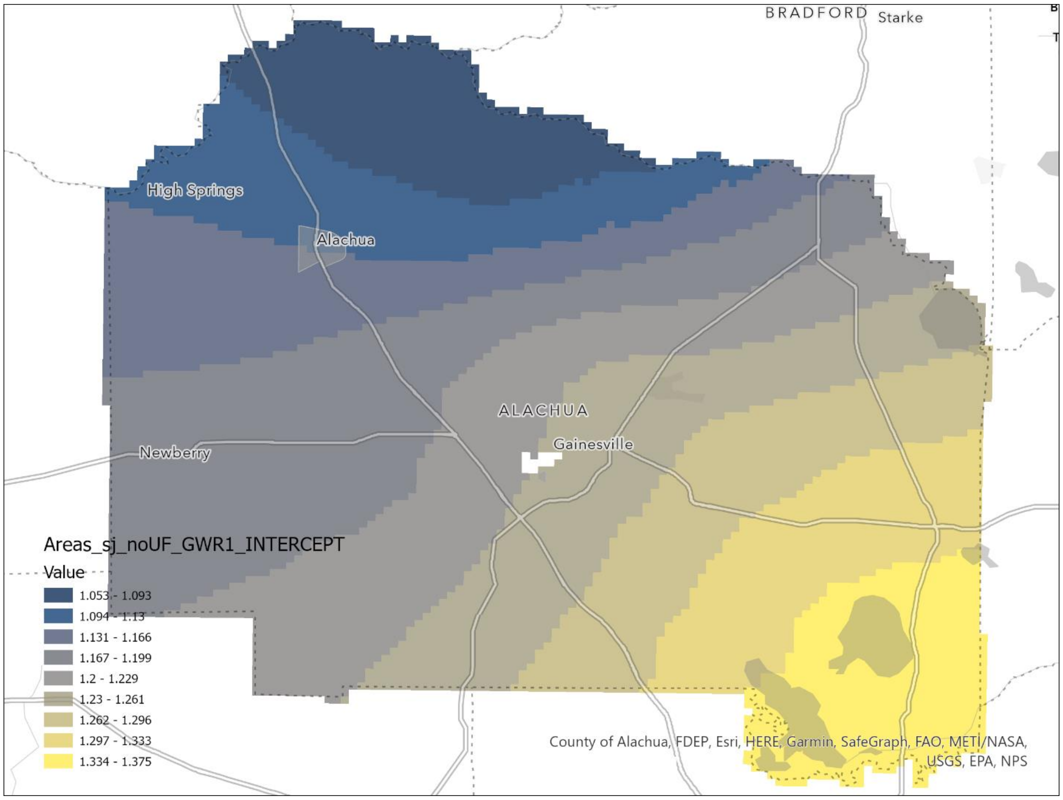
The local R^2 for Model B also is higher than the global Model B, and the AICc is lower in the local model than the global, indicating a better fit overall for the local model.

## Discussion

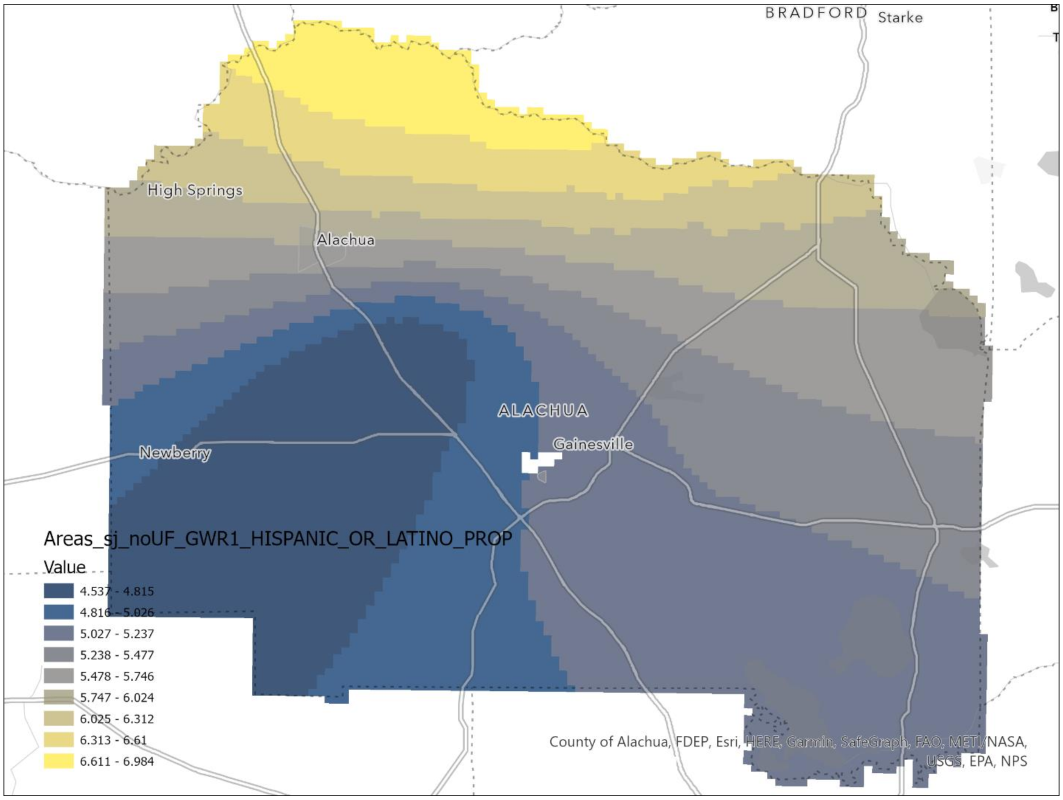
Of the models tested, local Model A performed the best with the lowest AICc and the highest R^2 values. Below I will discuss the implications of Model A.

Local models dictate different coefficients for the model $y' = C + \alpha\ \mathrm{HISPANIC\_LATINO} + \beta\ \mathrm{BLACK\_AA} + \gamma\ \mathrm{SINGLE\_FAMILY\_DETACHED} + \delta\ \mathrm{BELOW\_POVERTY\_LINE}$ based on space – each variable has a varying effect depending on location. To keep the underlying logic of the variables move from "characteristics of less privilege" to "characteristics of more privilege"), I used the model $y' = C +$

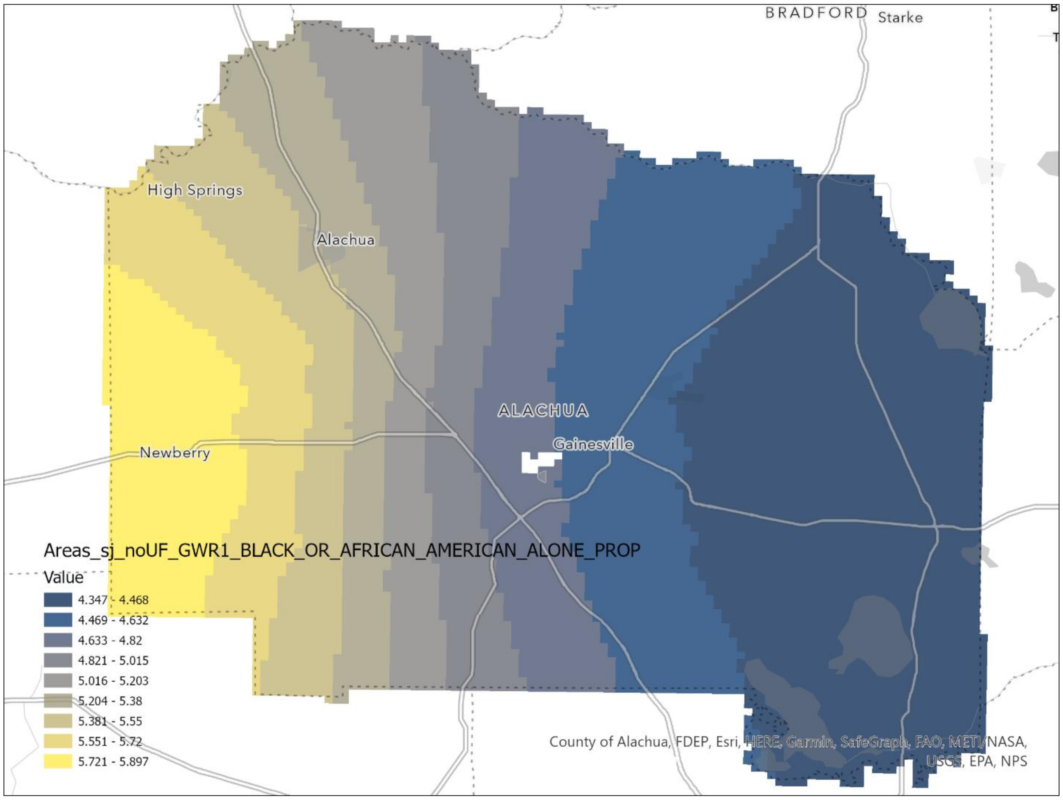
$\alpha$ HISPANIC_LATINO $+ \beta$ BLACK_AA $+ \gamma$ NON_SINGLE_FAMILY_DETACHED $+$ $\delta$ BELOW_POVERTY_LINE. The maps below illustrate the spatial variability in these coefficients.



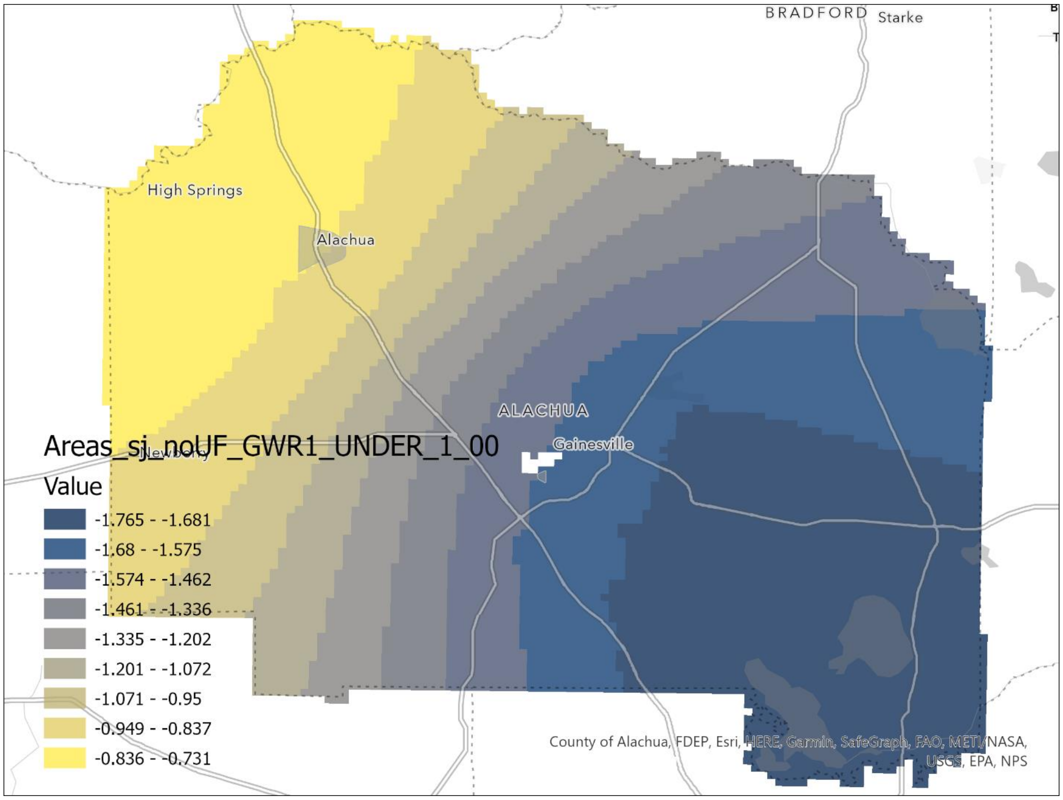

The intercept, $C$ , can be thought to represent the latent eviction pressure present at certain locations with the conditions of 0 Hispanic/Latino population, 0 Black population, all single-family housing, and 0 below the poverty line. At these conditions, the raw (log-untransformed) predicted eviction levels range from 2.86 to 3.95, a low and relatively insignificant difference on intercept. Therefore, the “latent eviction pressure” all things equal barely differs across the county.



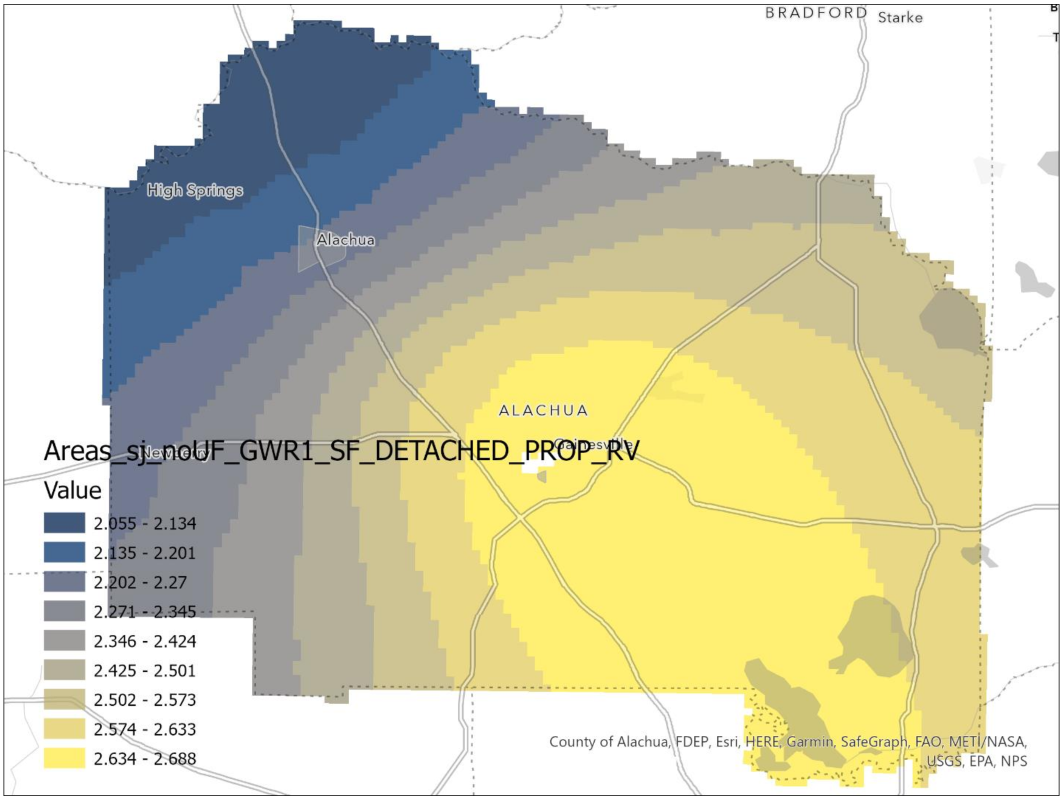

This map represents the effect (without causation) of Hispanic/Latino populations on eviction rates. This coefficient ranges from 4.53 to 6.61, so at its lowest a 10% increase in the Hispanic/Latino population corresponds with a 1.57 times increase in evictions, and at its highest the same increase associates with a 1.99 times increase in evictions, all things remaining equal. The area of highest impact seems to be rural north Alachua County, and the lowest a zone stretching from northwest Gainesville to Newberry and Archer.




This map represents the effect (without causation) of Black populations on eviction rates. This coefficient ranges from 4.34 to 5.90, so at its lowest a 10% increase in the Black population corresponds with a 1.53 times increase in evictions, and at its highest the same increase associates with a 1.80 times increase in evictions, all things remaining equal. This coefficient exhibits a strong east-west spatiality – towards the east increasing Black populations has less of an effect on evictions than the west side.



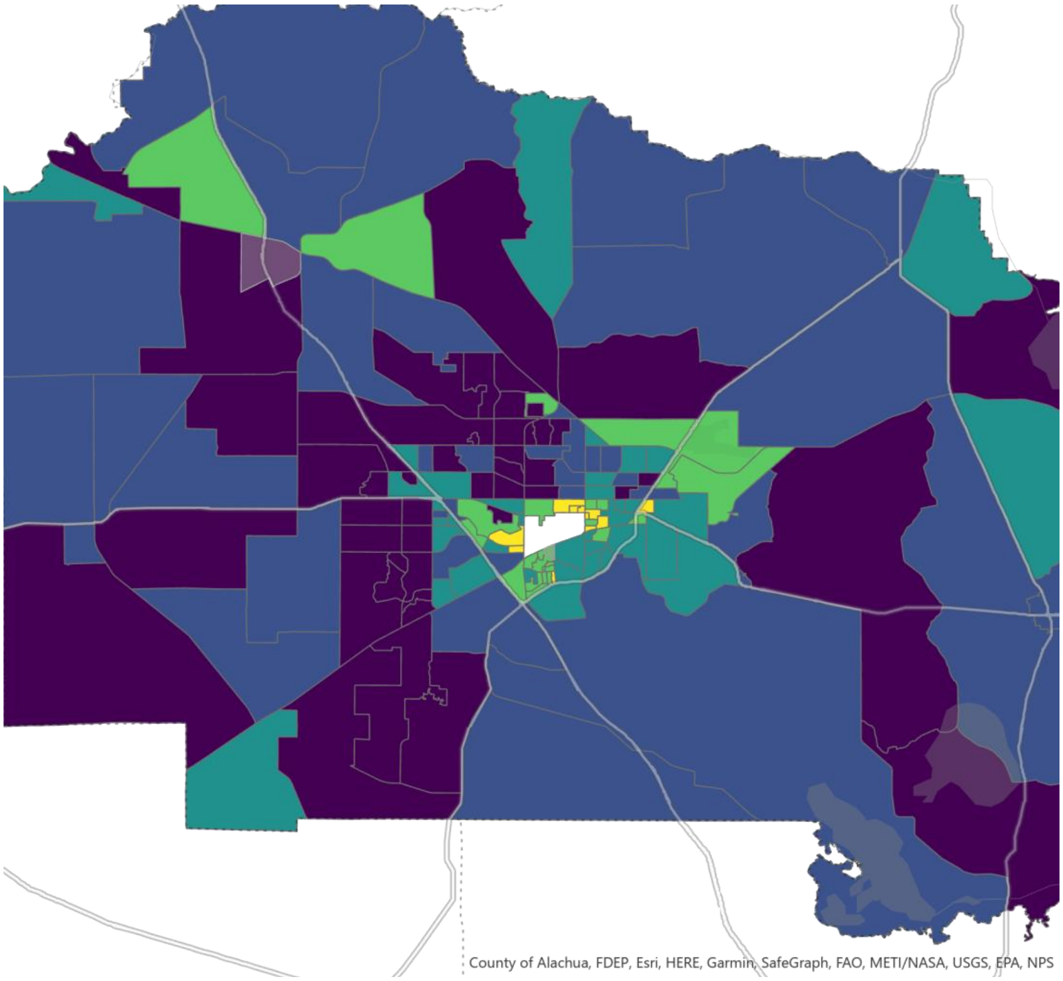

This map represents the effect (without causation) of people with incomes below the poverty line on eviction rates. This coefficient ranges from -1.77 to -0.73, so at its lowest a 10% increase in this population corresponds with a 0.84 times **decrease** in evictions, and at its highest the same increase causes a 0.92 times **decrease** in evictions, all things remaining equal. This coefficient also exhibits a strong east-west spatiality – towards the east increasing impoverished populations has stronger of an effect on evictions than the west side of the County.




Lastly, this map represents the effect (without causation) of *non-single family detached housing* on eviction rates. This coefficient ranges from 2.06 to 2.69, so at its lowest a 10% increase in non-single family detached housing corresponds with a 1.23 times increase in evictions, and at its highest the same increase associates with a 1.31 times increase in evictions, all things remaining equal. While this coefficient exhibits a strong northwest-southeast spatiality, the effect is ultimately not large enough to cause a significant difference.

While all the coefficients have spatial dimensions, the constant/intercept (representing latent eviction pressure given all other variables are the same) and the proportion of single-family detached housing, are much less dynamic spatially and exhibit a much smaller range of difference than Hispanic/Latino, Black, and poor populations.

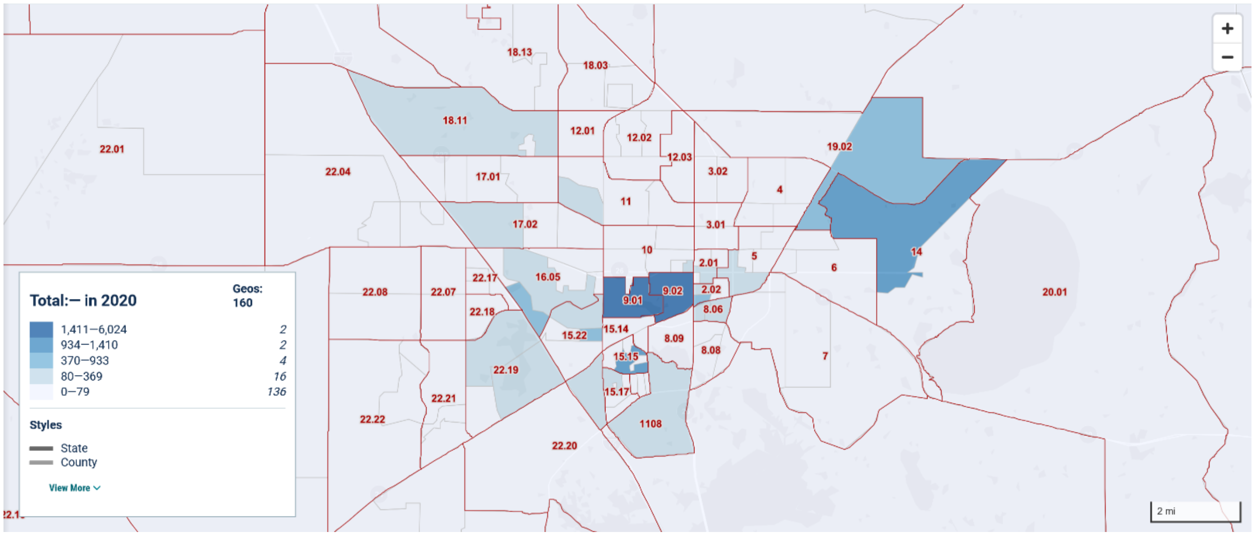
The contradictory point about this model is that increasing the proportion of people with income below the poverty line actually corresponds with a decrease in evictions. In the chance that the poverty level truly has a negative impact on evictions, one could surmise that perhaps more people at this poverty level are subsidized by public benefit, or in the case of the impact of students, subsidized by outside sources. Whereas there's the idea that the poverty rate should be representative of people not privately subsidized, there are many confounding factors in play.




The above map illustrates poverty rates across the county. Surrounding UF are block groups where the proportion of people experiencing poverty is the highest quantile in the county. There is also an informal cluster of high poverty proportion in the northeast of Gainesville. I have the suspicion that poverty in these areas is inflated owing to two separate factors.

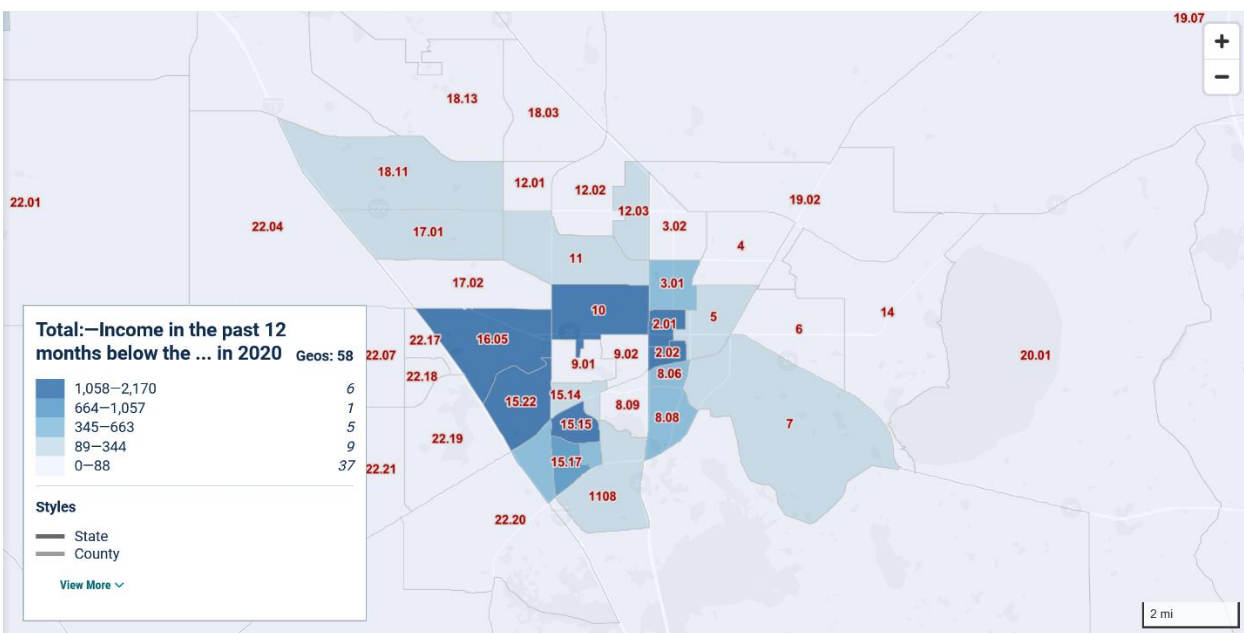
The smaller of the two relates to the institutional population. In the northeast quadrant of Gainesville, GRACE, Tacachale, and the County Jail (all group quarters and institutional housing) all contribute to raising the poverty rate due to their lack of income.

Below is a screenshot from the census also illustrating rough spatial impact of group quarters populations who are not theoretically subject to eviction at the time of data collection. Notice impacts in the northeast part of Gainesville, impacts of UF, and smaller proportions of group quarters in scattered areas.




The larger of the two deals with the nature of Gainesville as a college town. We see the highest poverty rates surrounding UF, block groups where many students live. While these students may not receive income yet (some do; many don't), they may also be otherwise subsidized by family or loans, inflating the poverty rate and confounding what the poverty rate should represent.

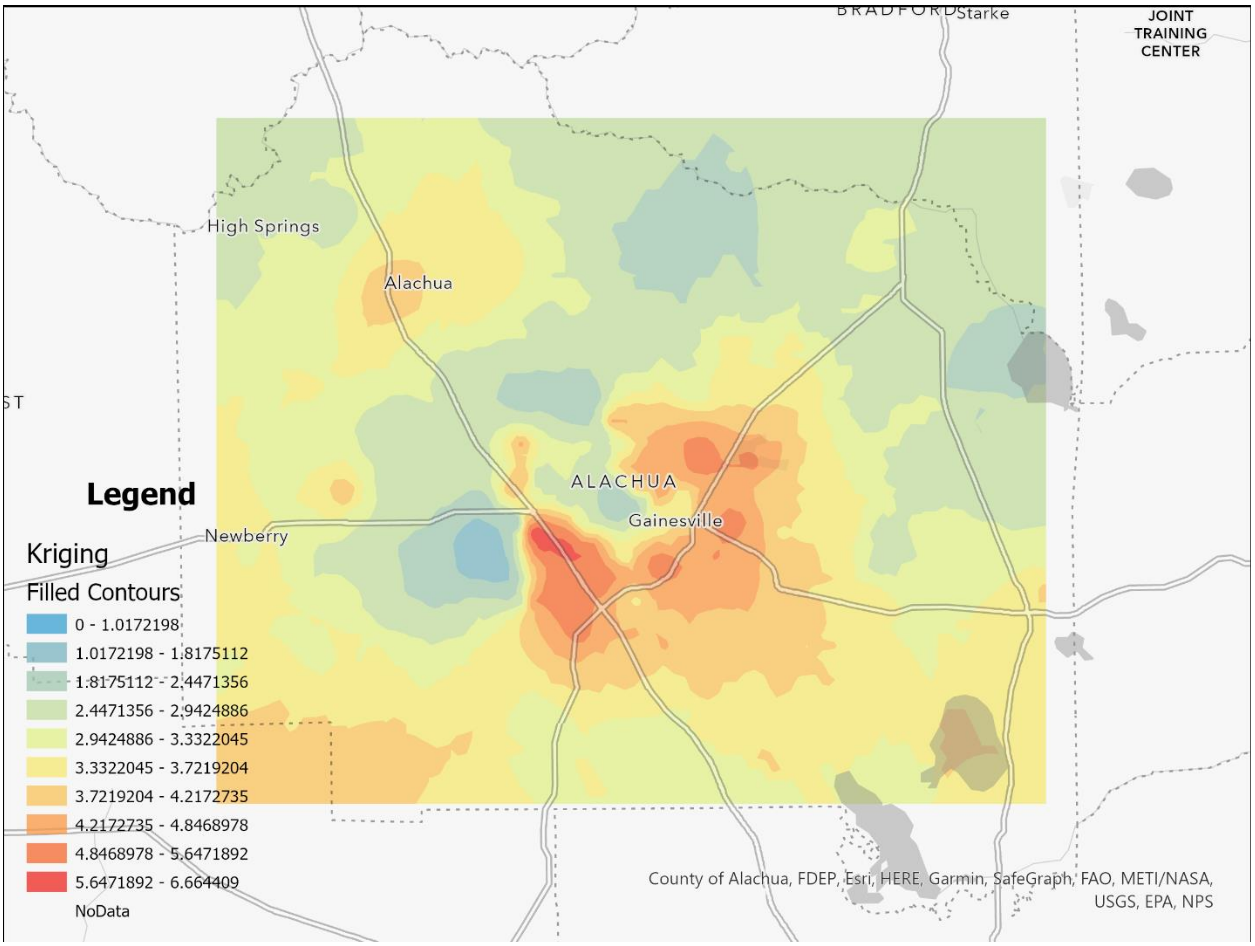
Indeed, a quick stab at Census county-level statistics for poverty level by educational attainment reveals that while the entire county-level poverty rate is 0.207, discounting students enrolled in undergraduate reduces that rate to 0.134. Looking at the tract-level data (the lowest level available) for poverty status by educational attainment, we realize that the phenomenon of college students at the poverty level is also centered around UF:



The idea of what poverty should represent on a census is not straightforward. In a non-college county such as Bradford, poverty rates typically represent household units who otherwise (may be able to) support themselves financially and/or receive public assistance directly from the State or through nonprofits. Confounding the variable is the impact of the large institutionalized (incarcerated) population of Bradford – it hosting three prisons and one jail. In a college county like Alachua, students get thrown into the mix as well. For the narrow purposes of this report I would suggest only considering determinants of evictions based on independent, non-institutional households, but we're unable to separate these various effects from either the independent variables under consideration as well as from the evictions themselves (we can't filter evictions by impact on someone enrolled in undergraduate courses, for example).

## Interpolation

Using Ordinary kriging, I completed an interpolation of the data across most of the county (the exact extent is the bounding box describing the centroids of each block group), using the log-transformed eviction count as our dependent variable. The results of the kriging are below:



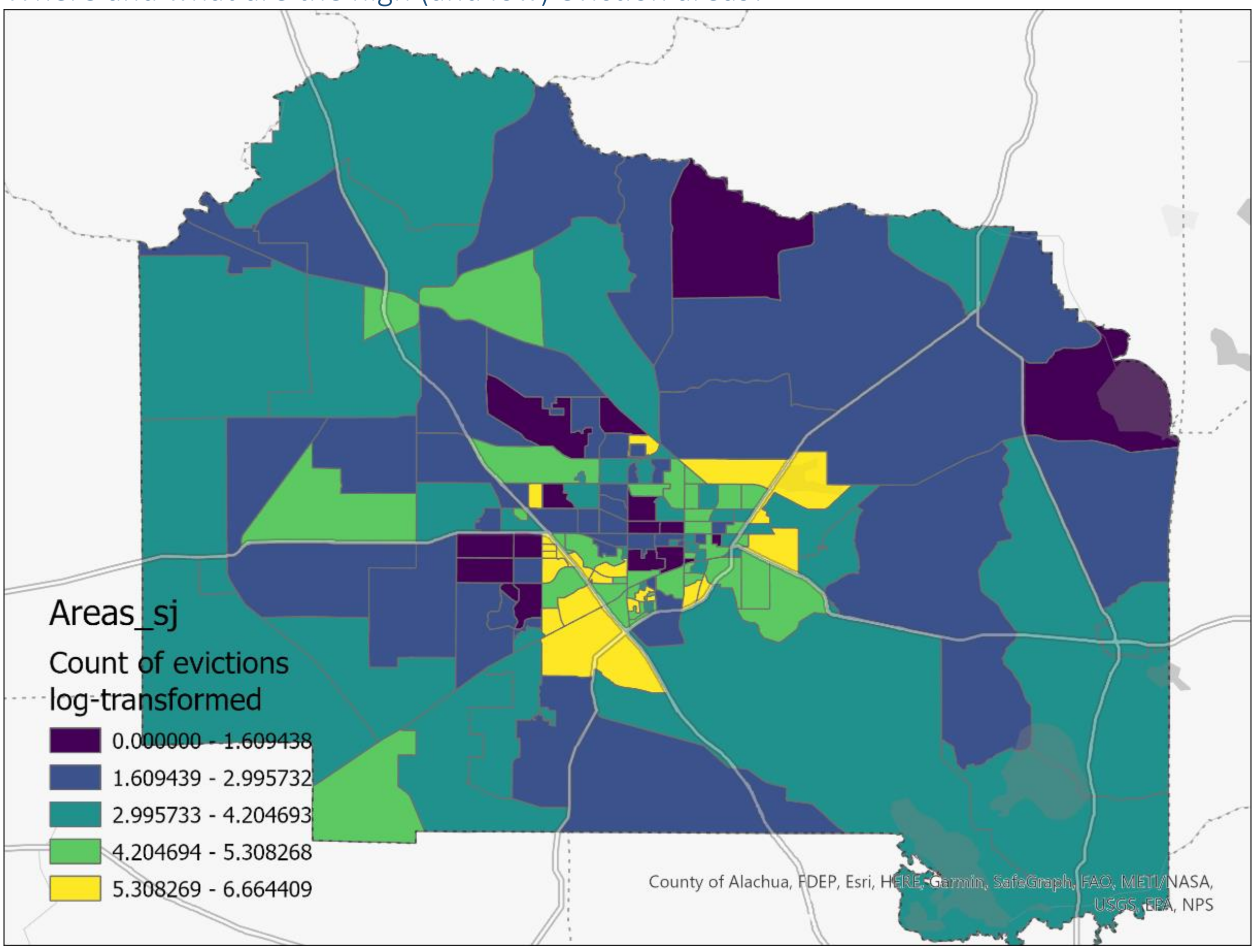

The optimized kriging model semivariogram reports a 0 nugget, a range of 9.65 km, and a sill of about 10^2.5 = 316 evictions. The semivariogram map is radially symmetric. The standard error regression trend is negative, meaning the model better predicts higher values rather than lower values.

In the sense of qualitative intensity of evictions, the kriging performs well in capturing the general pattern in the same areas as indicated by previous hotspot and outlier analyses, including the outlier region just west of the most intense area of evictions. Quantitatively, it doesn't do particularly well in predicting actual numbers of eviction, especially where no evictions occurred over the past 20 years.

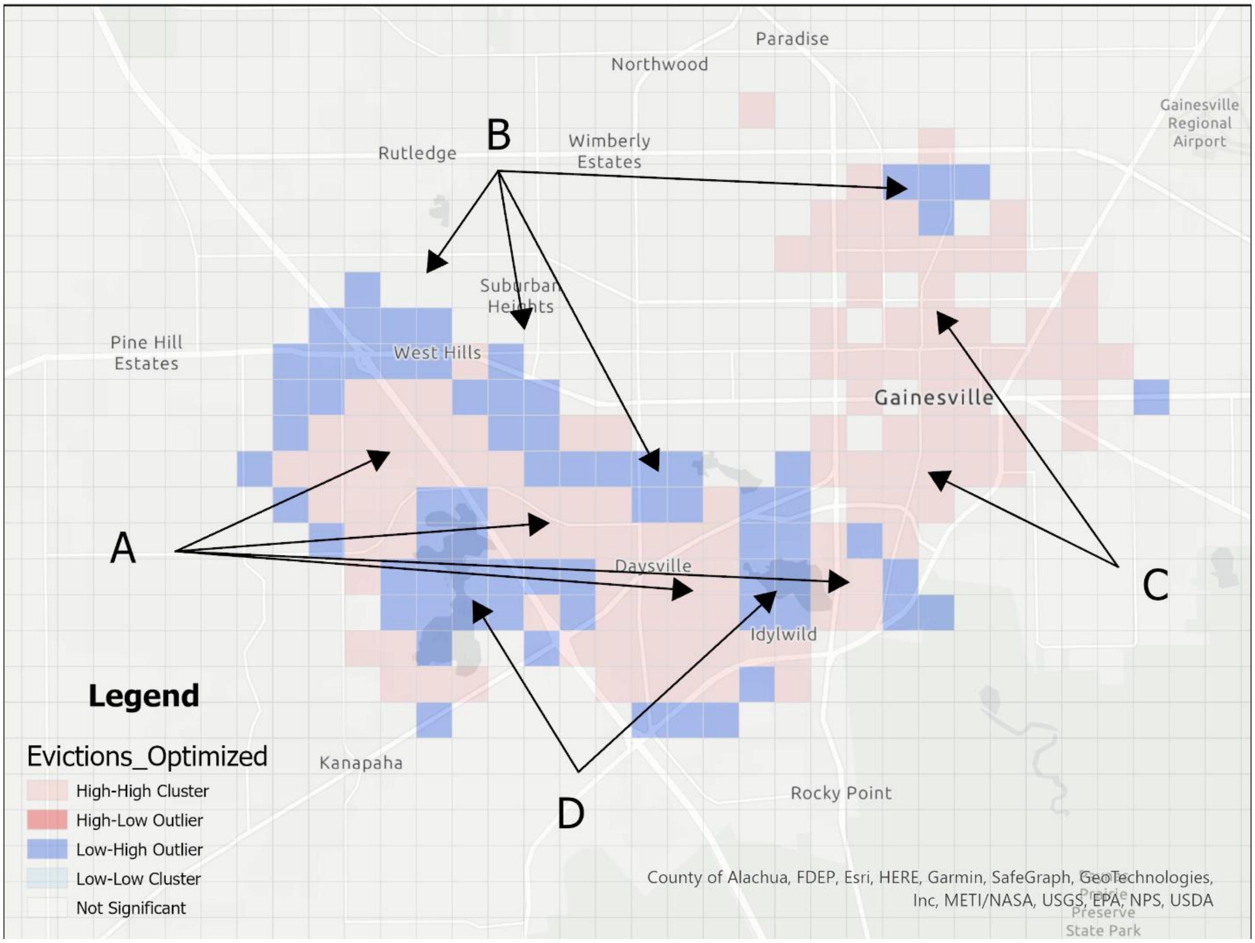
## Conclusion

I will conclude this paper by summarizing answers to the most salient questions of this report one-by-one:

### Where and what are the high (and low) eviction areas?



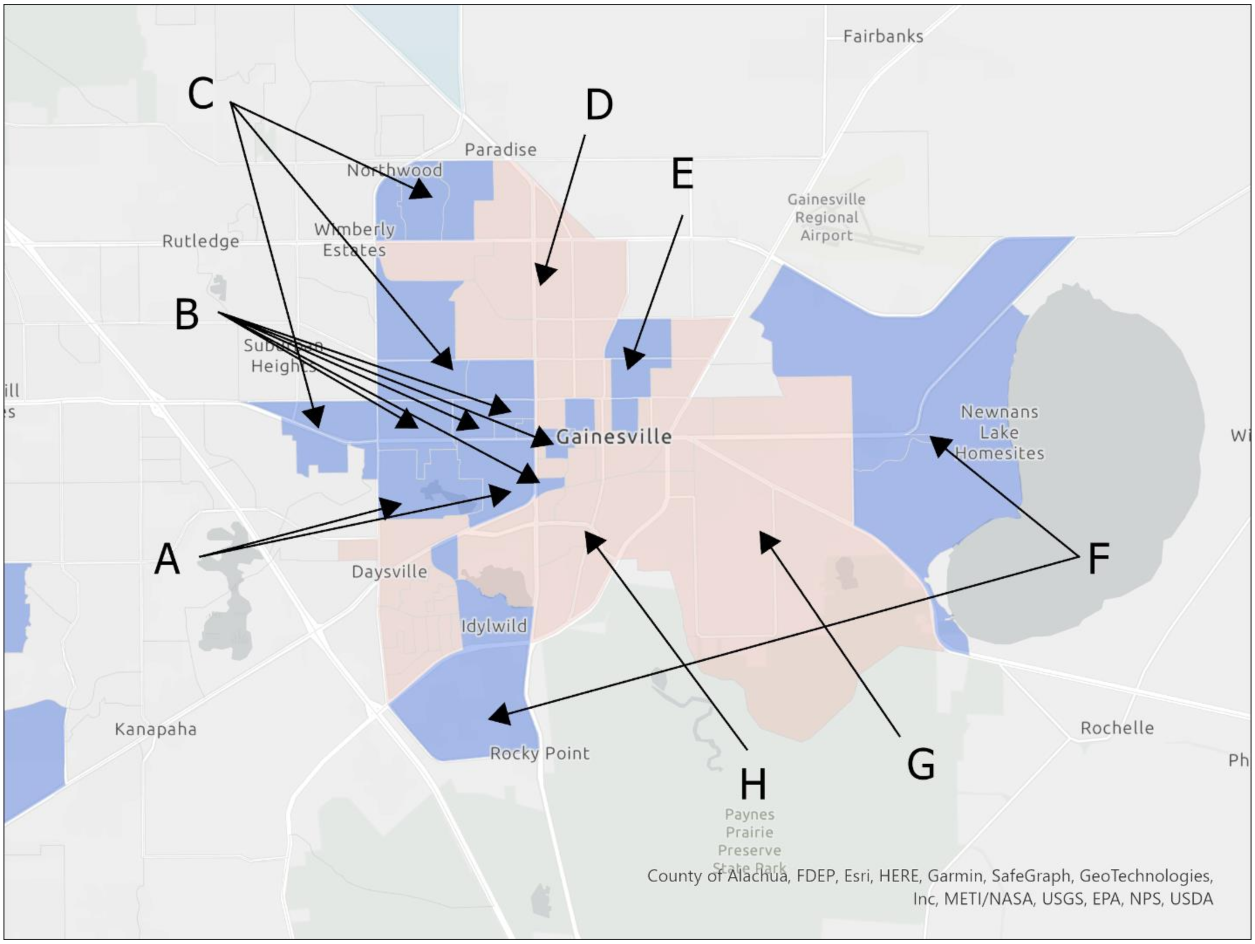

The overall structure of evictions is hard to illustrate. Broadly, we might say that, within Gainesville, evictions are lowest at a core stretching from UF and surrounding block groups into northwest Gainesville, and increasing as one exits the city. The larger county scope shows no true pattern. In any case, it is more fruitful to show local clustering patterns of evictions, as according to our Spatial Autocorrelation, the data are more clustered than random.




Above is a map of Optimized Outlier Analysis results using Evictions and fishnet aggregation, this time annotated. There seem to be four different patterns happening as to why the results are as they are:

- A: these areas are mostly zoned multi-family and developed mostly from the 1980s onward. West to East these areas are less to more student-oriented and student-populated, and more to less populated by people of color. A high proportion of the population of Gainesville is concentrated in these areas, which gives sense to the area also having high evictions.
- B: these low-high outlier zones are mostly commercial or institutional – they do not have housing to be evicted from.
- C: These areas are mostly mixed with single family, missing middle, and low-rise apartment buildings. It mostly makes up Gainesville's historical core and semi-recent suburban development. The area tends to skew mostly white-populated
- D: These areas have features such as lakes (Kanapaha and Bivens Arm) that make building untenable, or commercial areas such as Shands north of Bivens Arm.

It seems almost self-evident that Optimized Outlier Analysis produced results that say "Evictions occur where housing units exist to be evicted from", and then that areas B and D were coincidentally next to these. However, Optimized Outlier Analysis by fishnet aggregation failed to capture certain coldspots/low-low clusters of evictions that also contain housing. Aggregating by block group performs a little better:




Block group Optimized Outlier Analysis brought new areas into focus at different scales than fishnet aggregation, probably owing to the larger distance scale of aggregation and the effects of differing neighborhood/contingency structures. I focus on Gainesville because the higher density of features requires more explanation. The fact that coldspots appear as outliers is part of this higher density issue. Unfortunately, rural areas are a little too large of an area to make many broad generalizations on a single block group.

- A: UF Campus. Part of the coldspot as no evictions are able to occur on campus.
- B: Student housing areas, where much of the housing is subsidized by family or loans. An exception is the easternmost of the block groups which is student apartments in otherwise black neighborhoods.
- C: Single-family detached residential areas. The areas are mostly part of the 1960s suburban construction boom and mostly white.
- D: Mixed single- and multi-family housing, also of a mixed demographic. Somewhat lower-middle-class and middle class in nature.
- E: Mostly single-family housing, owner-occupied. A gradient from south to north of upper-middle-class white households to middle and lower-middle class mixed black and white neighborhoods.
- F: Mostly rural single-family area, some areas are mostly white and some are mostly black.
- G: Mostly black neighborhoods with single-family housing, much of which is owner-occupied.
- H: Mixed student and residential multi-family housing, along with the historic Southeast district.

## What are social and architectural determinants of evictions?

Among 14 models, I found a model that summarized racial, class, and housing determinants of evictions. This model, a geographically weighted regression is reiterated below:

$$y' = C + \alpha \text{ HISPANIC\_LATINO } + \beta \text{ BLACK\_AA } + \gamma \text{ NON\_SINGLE\_FAMILY\_DETACHED} \\ + \delta \text{ BELOW\_POVERTY\_LINE}$$

Of these, proportion of Hispanic/Latinos and Black people had the most impact on evictions, with a raw effect of increasing evictions 150-200% and 153-180% respectively per 10% change. These were also the most spatially variable, with a strong high-low north-south gradient for the former and west-east gradient for the latter. The poverty rate as noted in the Regression section had the opposite effect as expected, decreasing evictions 84-92% per 10% change. This may owe to confounding effects such as students or incarcerated/sheltered populations. Indeed, judging by exploratory regression results this variable 1/3 of the time had a negative effect on evictions, and 2/3 of the time had a positive effect. Proportion of non-single family detached housing had the effect of increasing evictions 123-131% per 10% change.

Across all models tested, other characteristics that didn't make it into to the family model but were significant across at least 90% of the models (along with their effects of eviction) were:

- Proportion of 3-4-unit housing (+)
- White proportion (-) – collinear with Black proportion
- Proportion of 5-9-unit housing (+)
- Renter occupied proportion (+)
- Single-family attached housing (e.g. townhouses, duplexes) (+)
- Single female householder (+)

I tried to balance a model that was more generalizable as opposed to a model with numerous variables that would not be as generalizable. In theory, having more variables might increase the R^2 value of the model. In practice, with these variables in addition to the original model's variables, the R^2 value only increases 0.02 points.

## Limitations

A crucial limitation to note is the conceptualization of these research questions: we are trying to understand the characteristics of the areas in which evictions do (and don't) occur and its determinants. Because court systems don't collect demographic data on either the parties who are filed against or the filers, we can't say anything directly about the defendants themselves. For example, although we have evidence to support the claim that evictions are more common in areas with higher proportions of Black and Latino residents, we cannot say that Black and Latino residents are subject to evictions at a higher rate than white people. This research speaks to the *environmental* influence on evictions as opposed to the demographics of eviction victimization.

Another smaller limitation is the previously noted confounding of the poverty variable in our models. To quickly reiterate, the desired implied subject of poverty level data is an otherwise free, financially independent (or government/NGO-supported) individual or household. However, the poverty rate in the census is confounded by the presence of incarcerated/institutionalized populations and by financially dependent students with no income otherwise. This resulted in non-consensus effects of the poverty level across explored regression models. Unfortunately the census doesn't have poverty information disaggregated by educational enrollment available at the block group level.

Lastly, due to the geographical scope (Alachua County) and aggregation method (block groups), these analyses fail to capture useful information about evictions at the rural scale. At the global scale any spikes in eviction at local levels in rural areas fail to be captured compared to higher eviction rates in Gainesville. At local scales, clustering may fail to be captured because the block group covers such a large area and the population density is low. As well, it's difficult to say much about the broad area captured in a rural block group, especially areas with small population. For example, there are many evictions that occur at One 51 Place in Alachua and little else within the block group, but that spike is diluted across the census block group.

A way to deal with this might involve splitting analyses into two different analyses: a city scale and a rural scale that excludes the county, and then using finer geometry on the rural scale. Global patterns and variations may become more important when analyzing only rural areas.

Ultimately, these results present correlations and predictions but do not address causation. It can't be said, for example, that a high percentage of Black people causes evictions. It is more correct to say broadly that eviction being correlated with higher numbers of populations of color and not white people, with non-single-family housing, and variably with poverty, is a product of planning, legislation, and economics that deprivileges black, brown, and poor people. The causes of these factors, and why they occur where they do in Gainesville and Alachua County, is left for another paper.

## Trajectories

It would be interesting to know more about the people that are directly impacted by eviction, degeneralizing from the environmental onto the demographic aspect of eviction. Currently the county does not collect this information, but with this information it would prove helpful to directing resources and understanding any disparities onto individuals, as opposed to just onto communities. It may be helpful to analyze the evictors as well – who are they and what type of properties do they generally own, are they large or small.

With all this information, government or private actors may be able to better target eviction defense resources to communities and people at most risk of eviction. Understanding areas of evictions alone may allow us to target outreach, but without understanding the nature of evictors and evictees, we do not gain a full picture of eviction and ways to create housing stability early on. Indeed, knowing that for example there is one bad actor, may allow the County to step in and prevent them from doing more business. Knowing that black single mothers are the most targeted for eviction may allow us to create social nets early on to prevent lapsing on bills, direct mothers to services, and more.