

Identifying Opioid Overdose Risk in Mesa, USA Using Spatial Regression

Spatial Analysis and Geocomputation

Ayina Anyachebelu, Yik Kien Tse, Humza Hussain, Yiru Xu, and Girish Joiode

Table of Contents

Introduction	3
Context	3
Research Objectives and Hypotheses	4
Data.....	5
United States Census Bureau Data	5
City of Mesa Open Data.....	5
List of Potential Risk Factors	6
Data Exploration	7
Overdoses.....	7
Overdose Distribution	7
Global Autocorrelation	9
Local Autocorrelation	9
Examining Independent Variables	11
Spatial Distributions.....	11
Frequency Distributions	12
Correlations Between Potential Risk Factors.....	13
Correlations Between Overdose Counts and Potential Risk Factors.....	14
Modelling.....	16
OLS and Poisson Baseline Models	16
Methodology	16
Results	16
Spatial Lag Model.....	20
Methodology	20
Reasoning for Spatial Modelling.....	20
Spatial Lag Model	21

Results	22
Spatial Error Model.....	26
Methodology	26
Spatial Durbin Watson Model.....	29
Methodology	29
Results	29
Geographically Weighted Regression	32
Methodology	32
Results	32
Geographically Weighted Poisson Regression.....	35
Methodology	35
Results	36
Conclusion: Discussion of Modelling Results	41
Limitations and Further Work.....	42
References	43

Word Count: 7192

Introduction

Context

The Opioid epidemic is a global health crisis (Krausz et al., 2021). The effectiveness of this drug class in pain treatment, modern anaesthesia and palliative care (Rosenblum et al., 2008, Rosner et al., 2019) is widely regarded in the medical world. In the United States, there was widespread over-prescription of opioids such as OxyContin during the 1990s (Lyden and Binswanger, 2019). This led to increased dependency on the drug due to its highly addictive nature (Waldhoer et al. 2004), and people began to turn to illegal synthetic opioids when prescriptions began to stop. Among the many side effects of opioid usage, such as digestive disorders, constipation and depression (Adams, 1889), the most notable is respiratory depression, which is the main cause of opioid overdose related deaths (Montandon and Slutsky, 2019). Hospitalisations related to this drug increased 64% between 2005 and 2014, and in 2016 over 42,000 Americans died from opioid overdoses (Lyden and Binswanger, 2019). Governments on the local and national levels of the country are actively pursuing solutions to mitigate the risks that this crisis poses to society.

There has been a great deal of research undertaken to understand the socio-economic and environmental factors exacerbating opioid overdoses and related deaths. Socio-economic marginalisation is an important contributor to opioid overdoses amongst the people using them (Draanen et al., 2020). Several studies have found opioid abuse to be more common amongst people recently released from prison (Brinkley-Rubinstein et al., 2018), those living in poverty (Marshall et al., 2019) and the unemployed (Hollingsworth et al., 2017), whereas higher measures of educational achievement (Marshall et al. 2019) and social support in the form of healthy relationships (Burns et al., 2004) have been found to be negatively associated with overdose cases. All of these factors point to the increasing need for policy in uplifting users of opioids with social and economic insecurity (Draanen et al., 2020) and using these associated risk factors to predict potential areas for future overdose prevalence.

While many of the mentioned studies focused on linear regression models and multivariate analysis to get to these conclusions, they often failed to take into consideration the spatial relationship that opioid overdoses have with these variables. Studies that have accounted for spatial variability have been able to highlight high levels of clustering, identifying areas which require immediate policy attention. For example, a study in Cincinnati, Ohio, not only found hotspots of heroin-related incidents via spatial clustering, but what variables were common amongst them (Choi et al., 2022). These hotspots were often found in areas with lower education rates, higher poverty rates and higher male percentages, consolidating some of the key findings in existing literature. A previous study in Cincinnati implemented spatial regression methods to factor in characteristics of the built environment as well as socio-economic variables, something that may have been more difficult to accomplish without the implementation of location data and spatial analysis (Li et al. 2019). While the proportion of parks, manufacturing districts and commercial sectors was positively associated with heroin-related emergency calls, there was a negative association seen with proximity to fast food restaurants, hospitals and opioid-treatment programs. Overall,

advancements in spatial analytics have significantly progressed research in this field, with the aim to direct change in areas most needed and identify potential characteristics of areas that are most vulnerable.

Adding to the previous work in this field, we aim to spatially characterise the risk factors associated with overdoses in the City of Mesa, located in the state of Arizona. Mesa is an ideal study site as the Arizona Department of Health Services has stated that every day over 5 people die from opioid overdoses within the state (Opioid Prevention, 2022). The governor of Arizona himself, Doug Ducey, has stated clearly how Arizona, and the nation as a whole, is facing a public health emergency which requires targeted solutions to alleviate pain (Office of Governor Ducey, 2017). The City of Mesa publishes publicly available data on overdose incidents, providing an opportunity to determine the key characteristics of overdose hotspots which will inform future policymaking.

Research Objectives and Hypotheses

The objective of our research is to improve the targeting of interventions against heroin overdoses in Mesa. We aim to identify how demographic, socioeconomic, and geographic factors are associated with overdose deaths in order to inform the City about where to take proactive measures to prevent more incidences. To achieve this, we model the relationship between various potential risk factors and the number of overdoses using spatial regression.

We hypothesize that there will be higher overdose mortality in areas with greater minority and low-income populations. Our predictions are based on studies showing that heroin-related incidents were negatively correlated with median household income in Cincinnati (Li et al., 2019) and that incidents were growing most rapidly amongst Black and Latina communities across US metropolitan areas (Lippold et al., 2019). We also predict that characteristics of the built environment such as distance from greenspace, zoning assignments, or indications of plight will influence opioid cases, based on the aforementioned study (Li et al., 2019). Finally, we expect a lower number of incidents with regions of lower crime rates in accordance to results in similar work (Choi et al., 2022).

Data

United States Census Bureau Data

We obtained all our demographic and socioeconomic data from the [American Community Survey \(ACS\)](#) which is an annual data collection program run by the United States Census Bureau (United States Census Bureau 2022). ACS is the premier source for detailed demographic and housing information in the United States. We obtained the most recent 5-year estimates (2020) on the census block group level which is the smallest geographical unit for which the Census Bureau provides data. There are 345 census block groups in Mesa with a substantial population. The information we collected includes the total population, the white, Black, and Hispanic population, the median household income, the total unemployed and the total labor force, the number of homeowners, the number of renters, the number of vacant properties, and the total number of properties. For our main variables excluding income, we calculated the statistic as a proportion to account for varying population sizes across the census block groups.

City of Mesa Open Data

The majority of our data sets come from the The City of Mesa Data Portal maintained by the municipal government. Our primary dataset is the [“Fire and Medical Opioid Overdose Incidents”](#), which contains all confirmed cases of opioid overdoses and their locations rounded to the 1/3 mile increments. The overdose was either confirmed by the patient or witness, an opioid found on scene, or a positive response to Narcan treatment (City of Mesa 2023). We obtained the number of overdose incidents from 2018 to 2020 and aggregated them by census block group.

We also collected the [“Zoning Districts”](#) data from the portal. The dataset contains delineated geographic areas in the city and their land use assignments (City of Mesa 2020). We specifically extracted land parcels assigned as high-density residential, low-density residential, and commercial. Additionally, we utilized the [“City Owned Property”](#) dataset which lists the properties owned or sold by the city and their locations (City of Mesa 2022). From this dataset we obtained point data on the location of parks and cemeteries (which we labelled as greenspace), libraries, arts centers, museums, crisis centers, and senior centers (which we labelled as community development resources), fire department and police stations (which we labelled as public safety), and municipal vacant property.

To investigate the degree of plight in neighborhoods, we used the [“Transportation Graffiti”](#) dataset, which includes graffiti reports in the city (City of Mesa 2023). We obtained reports from 2018 to 2020 and filtered them by the date and location in order to collect only distinct reports.

To establish a smoother relationship to factors of our point datasets across space, we calculated average nearest neighbor features. Adapting code from Ken Steif’s Public Policy Analytics we calculated the distance from the centroid of each census block group to the 3 nearest factor points (Steif, 2021).

Finally, to investigate crime rates, we utilized the “Police Incidents” dataset, which includes incidents based on police reports and arrests and their locations rounded to the nearest hundred block (City of Mesa 2023). We excluded all non-crime-related incidents such as car collisions, medical emergencies, or animal-related issues. We also excluded crimes that were related to driving such as DUIs, missing driver’s license or speeding, because of the uncertainty about if the location of the incident is associated with the individual charged.

List of Potential Risk Factors

After feature engineering, our list of potential independent variables include population, White population (as percent), Hispanic population (as percent), Black Population (as percent), median household income, unemployed population (as percent), homeowner population (as percent), renter population (as percent), vacant homes (as percent), land zoned as high-density residential (as percent), land zoned as low-density residential (as percent), land zoned as commercial (as percent), number of crime-related police incidents, average distance to the 3 nearest: graffiti reports, community development centers, greenspace, public safety, and vacant property.

Data Exploration

Figure 1 shows a list of the variables used in the analysis and what they represent:

Variable Name	Risk Factor Measured
pop	total population
arrests_count	number of arrests
median income	median household income
perc_white	percent of population that is white
perc_black	percent of population that is black
perc_hispanic	percent of population that is hispanic
perc_unemployed	percent of population that is unemployed
perc_owners	percent of population that are homeowners
perc_renters	percent of population that are renters
perc_vacant	percent of properties that are vacant
resid_high_per	percent of land zoned as high-density residential
resid_low_per	percent of land zoned as low-density residential
commercial_per	percent of land zoned as commercial
dist_public_safety	average distance to the 3 nearest public safety facilities
dist_greenpace	average distance to the 3 nearest public green spaces
dist_vacant_prop	average distance to the 3 nearest city-owned vacant lands
dist_comm_dev	average distance to the 3 nearest community/arts centers
dist_graffitti	average distance to the 3 nearest graffiti reports

Figure 1: List of Variables and descriptions

Overdoses

Overdose Distribution

Figure 2 shows the frequency distribution of opioid overdose for Mesa census blocks. We find that the distribution of overdoses is heavily right-skewed. The mean and median of overdose cases are 5.1 and 3 respectively. However, a few census blocks have a relatively large number of overdose cases compared to others. It may represent that opioid overdose in Mesa may be more prevalent in several areas compared to most places where there are only a few cases. As a result, our modelling of the overdose counts will have to account for the prevalence of census block groups with no cases.

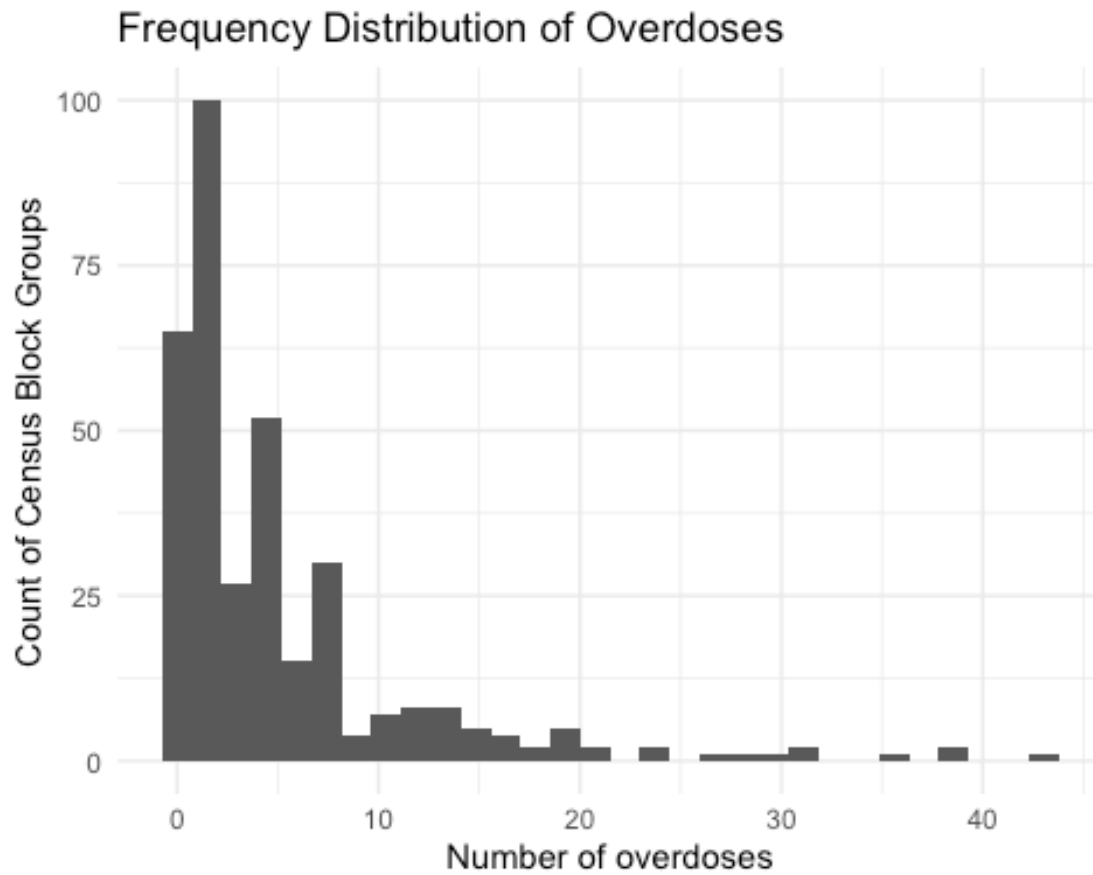


Figure 2: Frequency distribution of opioid overdoses in Mesa (2018 to 2020)

Figure 3 reveals the spatial distribution of opioid overdoses among Mesa census blocks from 2018 to 2020.

Overdoses in Mesa (2018-2020) by Census Block Group

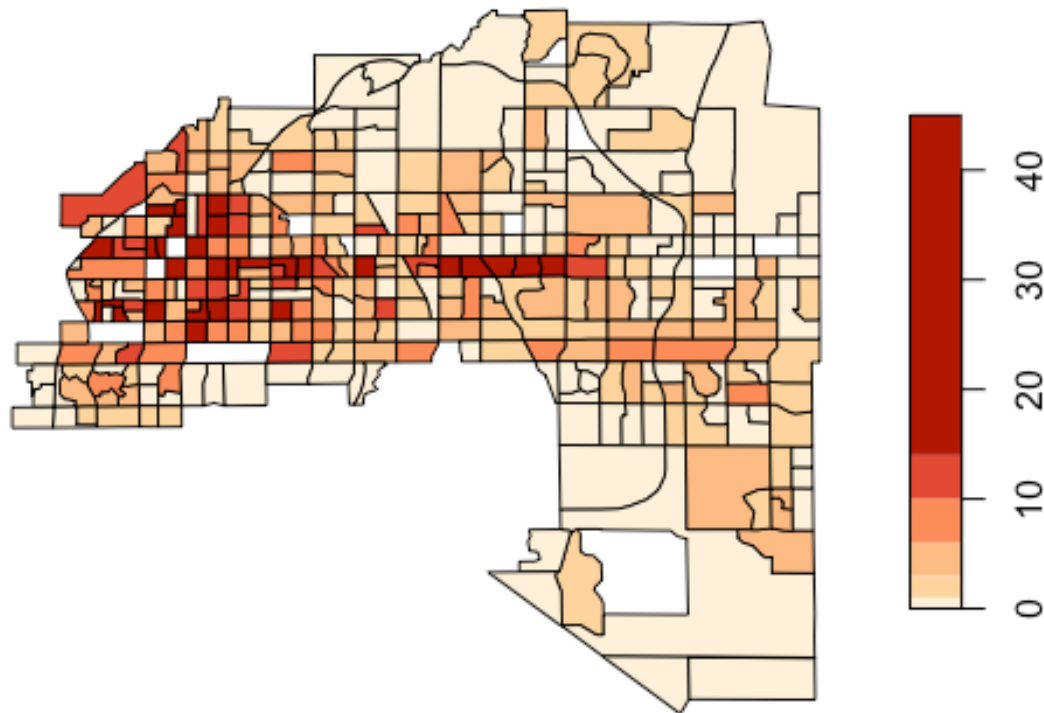


Figure 3: Spatial distribution of opioid overdoses in Mesa (2018 to 2020)

By observation, the census blocks with a more significant number of overdose cases tend to cluster. For example, particularly around the west and the centre of Mesa. On the other hand, those with a relatively lower number of cases also show a general pattern of clustering towards the Northern and Southern parts of Mesa. Therefore, it is worth studying if those with higher and lower values tend to cluster statistically. Global and Local Moran's I will be performed to test this.

Global Autocorrelation

Global Moran's I was performed to check if autocorrelation of opioid overdoses exists globally (figure 4). We find that at the census block group level, counts of overdose incidents display significant positive autocorrelation.

Figure 4: Global Moran's I of opioid overdose cases (2018 to 2020)

Local Autocorrelation

Next, we investigate local autocorrelation using Local Moran's I. The Moran scatterplot (figure 5) shows how the overdose counts of each census block are compared to the neighbour's weighted counts. The scatterplot reveals that the majority of autocorrelation is

from high overdose values neighboring high values (top right). However, there are also several instances of low values neighboring low values (bottom left).

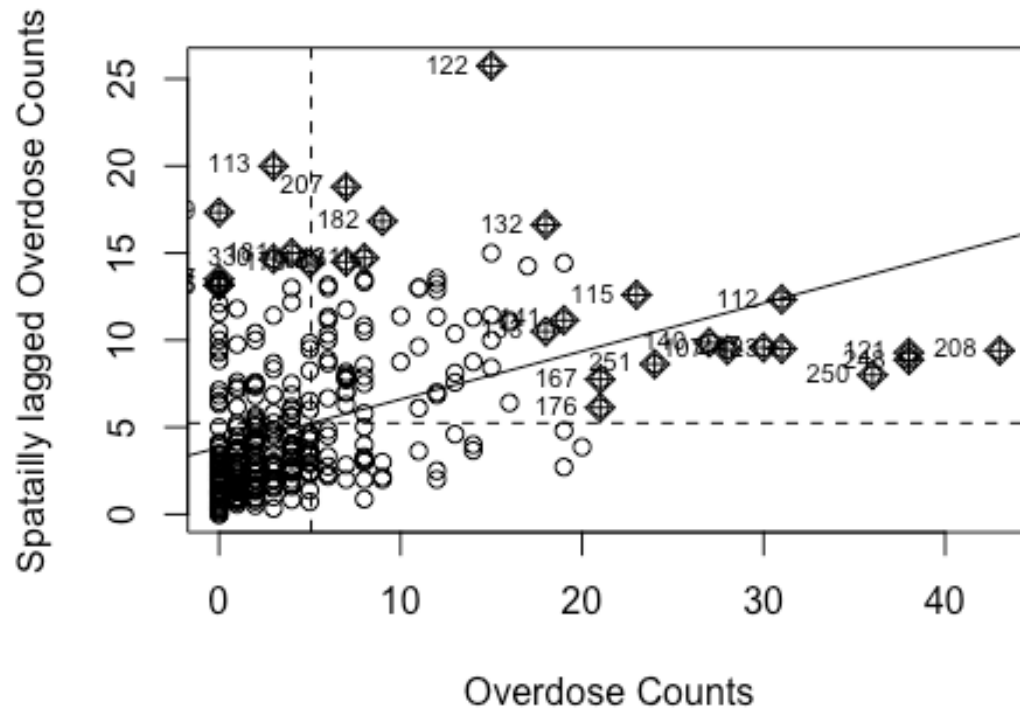


Figure 5: Moran scatterplot for opioid overdoses (2018 to 2020)

We map the various clusters to examine the spatial distribution of the autocorrelation. The result shows that in most of the city the autocorrelation is not significant. However, there is significant high-high autocorrelation in the west and center of the city. There is one case of low-low autocorrelation in the northeast of the city.

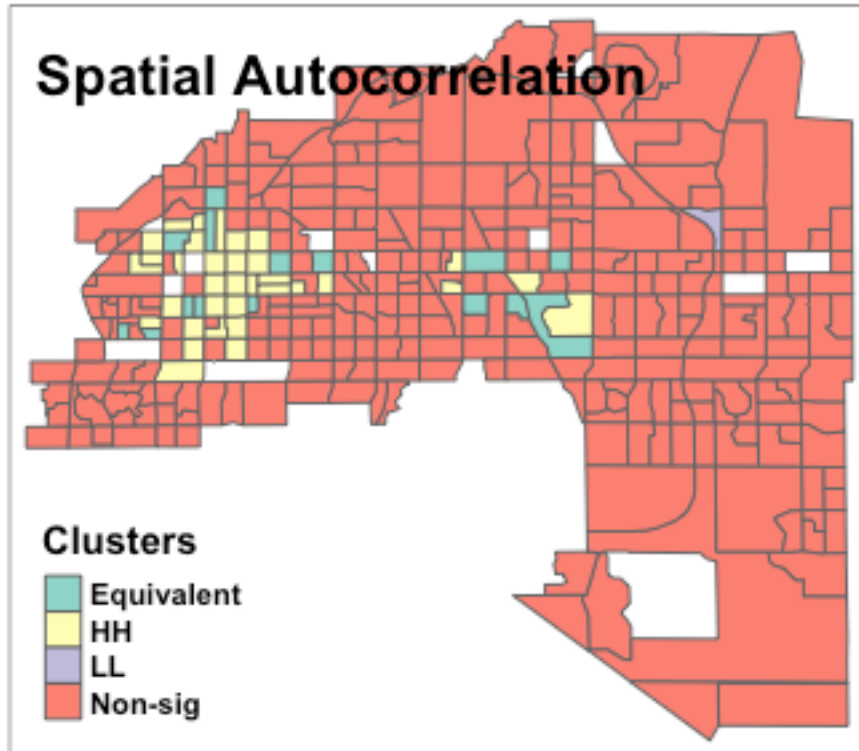


Figure 6: Local Moran's I clusters for opioid overdoses (2018 to 2020)

Examining Independent Variables

Spatial Distributions

We explored the spatial distribution of our potential independent variables by mapping them (figure 7).

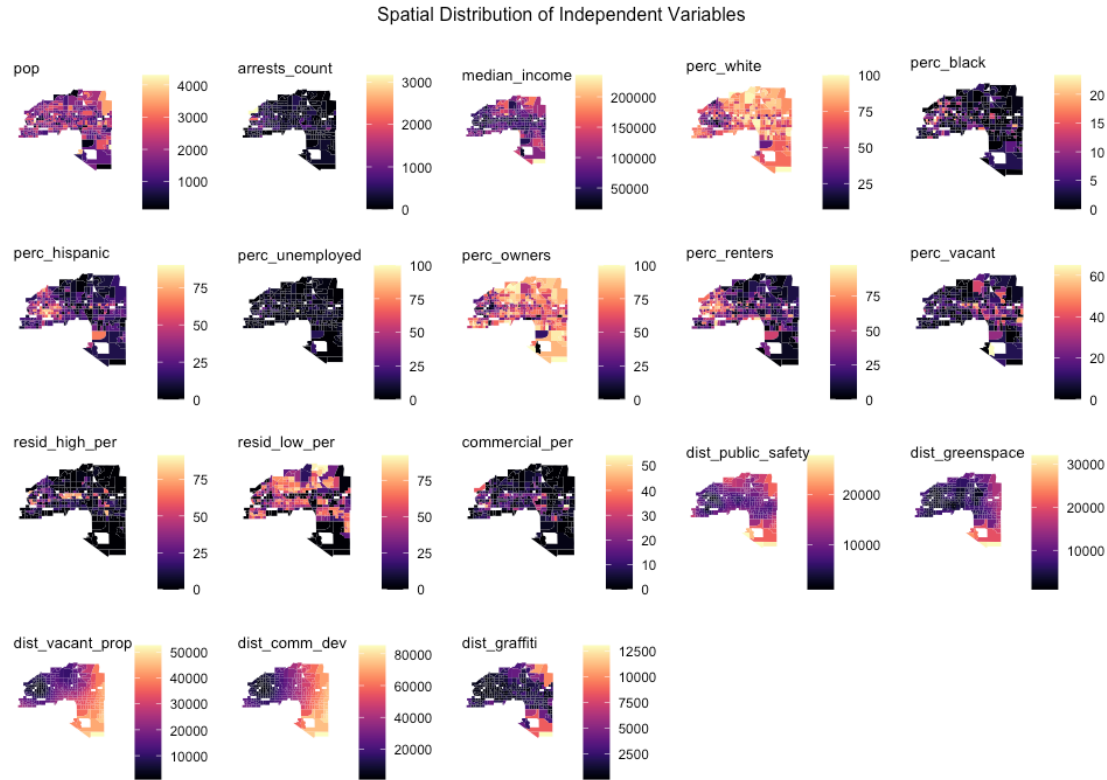


Figure 7: Spatial distribution of potential independent variables

Our results show that there are several trends associated with the western part of the city. In this part of the city, we find the most extreme number of arrests, the largest proportions of Hispanic residents and lowest proportions of White residents, as well as the lowest proportions of homeowners. In general, we find that most of the city is zoned for low-density residential housing, while most of the commercial and high-density residential zoning is in the northwest or across the central, downtown part of the city. Most of the census block groups with large amounts of areas zoned as vacant land are in the eastern part of Mesa. On the other hand, most of the city-owned vacant properties (no inhabitants) are closer to the west of the city. We find that the average distance to greenspace, public safety facilities and graffiti are generally similar across the cities, except in the northeast and southeast of the city which are also the areas of the highest incomes. Distance to community development resources increases as we move to the west of the city. This is expected given that the downtown area is near the east. Finally, we find an extreme outlier in one census block group in the center of the city with close to a 100% unemployment rate, which we will remove from our data set.

Frequency Distributions

We also explored the frequency distributions of the potential independent variables (Figure 8). Most of the distributions are heavily right-skewed or left skewed. Several of

variables have several observations in the zero bucket because many census block groups do not have areas that zoned for activities other than low-density housing.

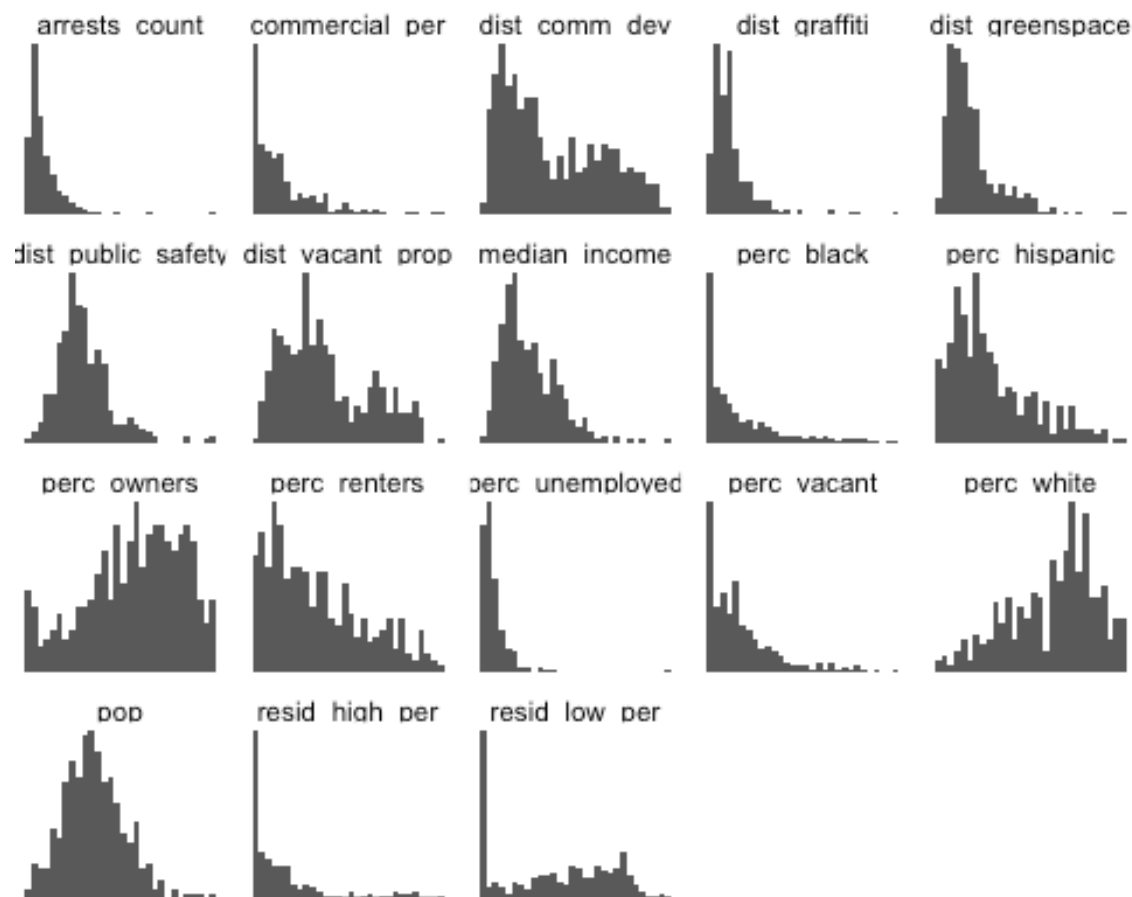


Figure 8: Frequency distributions of potential independent variables

Correlations Between Potential Risk Factors

We examine the correlations between the independent variables to check for possible multicollinearity before running our regression models (figure 9). As expected, some variables are highly correlated such as the proportion of Hispanic residents and White residents or the proportion of homeowners and renters. We find that the proportion of White residents and the proportion of renters are highly negatively correlated with each other but also correlated with several other potential risk factors, so we decide to drop them from our list of independent variables. Several of our distance-related risk factors are also highly positively correlated like the distance to public safety, greenspace and community development resources. This situation is likely because these are all municipal facilities so they are all located closer to the city center.

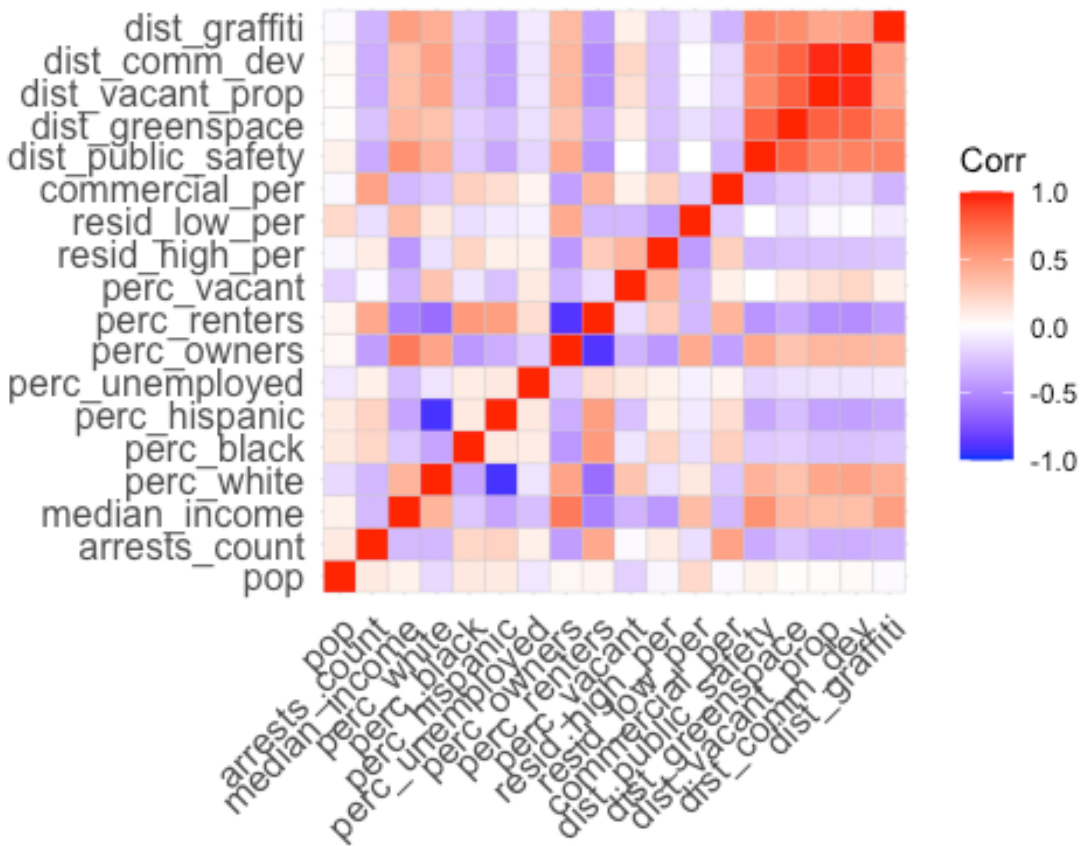


Figure 9: Correlations between independent

Correlations Between Overdose Counts and Potential Risk Factors

Finally, we explore the relationship between the potential risk factors and the number of overdoses in each census block group (figure 10). We find that there are some variables with very low linear correlations to the number of overdoses, such as the population, the percent of land zoned as low-density residential, the percent of land zoned as vacant, and the unemployment rate. On the other hand, variables such as the number of arrests and the percentage of homeowners are much more strongly correlated with overdoses. The scatterplots show that several of these relationships do not appear to follow a linear pattern, but this phenomenon may be influenced by the strong spatial relationships we previously observed.

Correlations between Number of Overdoses and Risk Variables

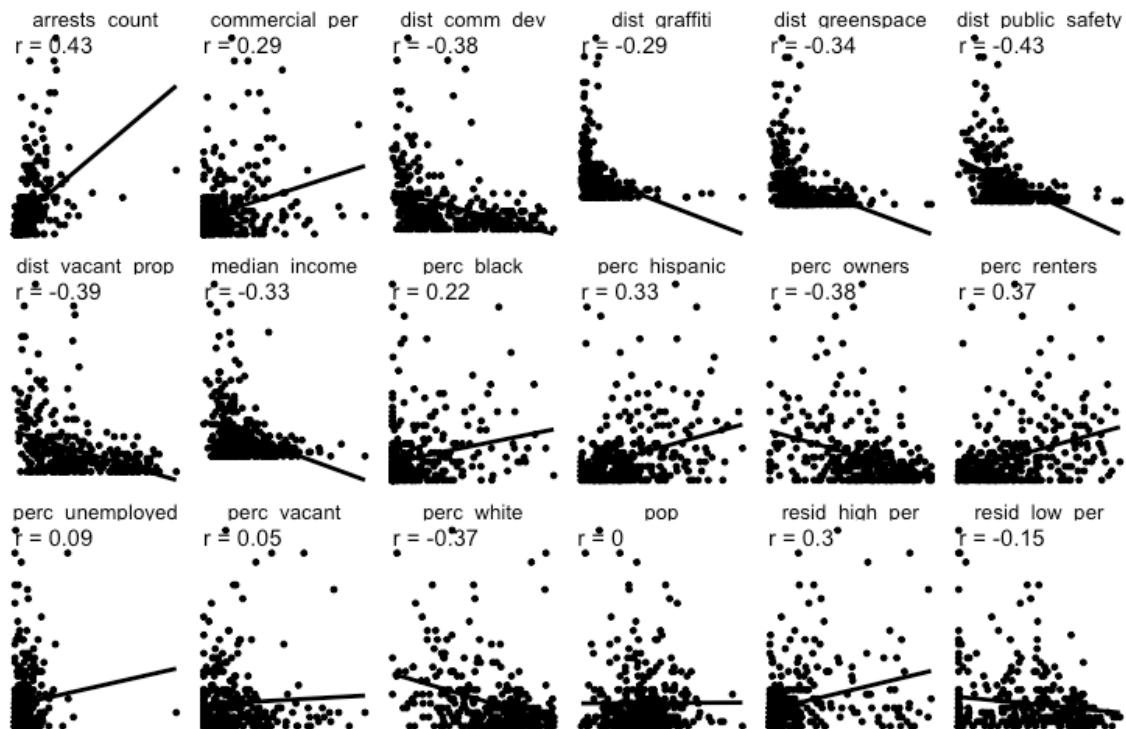


Figure 10: Relationships between opioid overdoses and potential independent variables

Modelling

After taking into account multicollinearity and the relationship with the overdose count, the variables chosen for the model were: median household income, Black population (as percent), Hispanic population (as percent), number of police incidents, homeowner population (as percent), land zoned as commercial (as percent), land zoned as high-density residential (as percent), average distance to vacant property, average distance to graffiti reports, average distance to greenspace and average distance to public safety.

OLS and Poisson Baseline Models

Methodology

We first build our baseline models using Ordinary Least Squares and Poisson Regression. The formula for the linear regression is as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_n x_{in} + \epsilon_i$$

where y_i is the i th observation of number of overdoses, x_i corresponds with each of our n potential risk factors, β_0 represents the intercept term, β_n is the n th estimated parameter, and ϵ is the error term.

Considering our dependent variable consists of non-negative overdose counts, we also use a Poisson regression with the following equation:

$$\ln(y_i) = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_n x_{in} + \epsilon_i$$

where $x_1 = 1$ so β_1 represents our constant term and for any values of $x_1 \dots x_n$ we can use the regression to predict the odds of an overdose happening.

Results

Below are the results of the OLS model (Figure 11).


```
##
## Call:
## lm(formula = overdose_count ~ arrests_count + median_income +
##      perc_black + perc_hispanic + perc_owners + resid_high_per +
##      commercial_per + dist_greenpace + dist_vacant_prop + dist_graffiti +
##      dist_public_safety, data = mesa_all)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.1507  -2.8612  -0.8112   1.5282  30.0266
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.252008260  1.853815332   2.294  0.022433 *
## arrests_count  0.006032666  0.001254947   4.807 0.00000232 ***
## median_income  0.000005764  0.000014174   0.407  0.684506
## perc_black     0.070938516  0.071251288   0.996  0.320162
## perc_hispanic  0.053501444  0.018188173   2.942  0.003494 **
## perc_owners    -0.003865112  0.018783864  -0.206  0.837098
## resid_high_per  0.058222957  0.016440101   3.542  0.000455 ***
## commercial_per  0.015697167  0.039800303   0.394  0.693540
## dist_greenpace  0.000060832  0.000141520   0.430  0.667582
## dist_vacant_prop -0.000052372  0.000044145  -1.186  0.236319
## dist_graffiti   0.000315769  0.000280584   1.125  0.261229
## dist_public_safety -0.000381148  0.000140613  -2.711  0.007064 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.585 on 333 degrees of freedom
## Multiple R-squared:  0.3365, Adjusted R-squared:  0.3146
## F-statistic: 15.35 on 11 and 333 DF, p-value: < 0.00000000000000022
```

Figure 11: Regression results from OLS Model

The OLS model has an Adjusted R-squared of 0.31, suggesting our factors only explain about 30% of the variation in number of overdoses. At the 5% level, independent variables with statistically significant coefficients were the number of arrests, the percentage of Hispanic residents, the percentage of the area zoned as high-density residential and the distance to public safety resources. All variables except the distance to public safety resources are positively correlated with number of overdoses.

Figure 12 shows the results of the Poisson regression. Using this model, in addition to the significant variables in the OLS regression, the percentage of Hispanic residents and percentage of Black residents and the distance to public greenspace were also significant at the 5% level. Particularly interesting results are that every 1-unit increase in the percentage of land zoned as high-density residential corresponds to a ~ 0.98% increase in overdose cases. Similarly every 1-unit increase in the percentage of black residents and hispanic residents corresponded to overdose case increases of 1.17% and 0.70% respectively.

```
##
## Call:
## glm(formula = overdose_count ~ arrests_count + median_income +
##      perc_black + perc_hispanic + perc_owners + resid_high_per +
##      commercial_per + dist_greenpace + dist_vacant_prop + dist_graffiti +
##      dist_public_safety, family = "poisson", data = mesa_all)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.9643  -1.5895  -0.4465   0.7398   6.8914
##
## Coefficients:
##              Estimate      Std. Error z value      Pr(>|z|
## )
## (Intercept)      2.0595832100  0.1566417221  13.148 < 0.0000000000000000
2 ***
## arrests_count      0.0006844624  0.0000686423   9.971 < 0.0000000000000000
2 ***
## median_income    -0.0000008436  0.0000013771  -0.613      0.540
1
## perc_black        0.0116940923  0.0048565931   2.408      0.016
0 *
## perc_hispanic     0.0070092185  0.0013777334   5.087      0.00000036
3 ***
## perc_owners       0.0009652980  0.0014383300   0.671      0.502
1
## resid_high_per    0.0098350225  0.0011898858   8.266 < 0.0000000000000000
2 ***
## commercial_per    0.0045587548  0.0026782416   1.702      0.088
7 .
## dist_greenpace    -0.0000579105  0.0000143220  -4.043      0.00005266
4 ***
## dist_vacant_prop   0.0000025045  0.0000043933   0.570      0.568
6
## dist_graffiti     -0.0000081344  0.0000421016  -0.193      0.846
8
## dist_public_safety -0.0001149590  0.0000129847  -8.853 < 0.0000000000000000
2 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 2322.2  on 344  degrees of freedom
## Residual deviance: 1269.0  on 333  degrees of freedom
## AIC: 2206.1
##
## Number of Fisher Scoring iterations: 5
```

Figure 12: Regression results from Poisson regression model

Unlike our hypotheses the farther away from public greenspace and public safety resources, the lower the overdose counts. While there are no grounds for causation, the reason for this phenomenon might be the fact that many of these public facilities are nearer to the city center and farther from suburban outskirts, so more overdoses may be happening in these areas.

The Moran's correlation test was conducted to check for residual spatial dependence. Below are the results for the OLS model and the Poisson model respectively (figures 13 and 14). As seen below, the test resulted in a Moran's I of 0.014 and 0.0068 respectively with corresponding Moran I statistics that were not significant on the 5% level. As a result, we cannot retain the null hypothesis that there is no spatial clustering of the residuals.

```
##
## Global Moran I for regression residuals
##
## data:
## model: lm(formula = overdose_count ~ arrests_count + median_income +
## perc_black + perc_hispanic + perc_owners + resid_high_per +
## commercial_per + dist_greenspace + dist_vacant_prop + dist_graffiti +
## dist_public_safety, data = mesa_all)
## weights: mesa_weights
##
## Moran I statistic standard deviate = 0.98346, p-value = 0.1627
## alternative hypothesis: greater
## sample estimates:
```

## Observed Moran I	Expectation	Variance
## 0.0143625012	-0.0155922393	0.0009277194

Figure 13: Residual spatial dependence from OLS model

```
##
## Global Moran I for regression residuals
##
## data:
## model: glm(formula = overdose_count ~ arrests_count + median_income +
## perc_black + perc_hispanic + perc_owners + resid_high_per +
## commercial_per + dist_greenspace + dist_vacant_prop + dist_graffiti +
## dist_public_safety, family = "poisson", data = mesa_all)
## weights: mesa_weights
##
## Moran I statistic standard deviate = 0.41218, p-value = 0.3401
## alternative hypothesis: greater
## sample estimates:
```

## Observed Moran I	Expectation	Variance
## 0.006842371	-0.006272096	0.001012336

Figure 14: Residual spatial dependence from Poisson regression model*

Although we cannot reject the null hypothesis based on our Moran correlation test results, an examination of the residuals of the OLS model (figure 15), shows that their distribution

appear to be non-random with large positive residuals in the west and center of the city. The poisson model has residual values of smaller magnitude which are more evenly spread out spatially. However, there still seems to be some clustering of positive residuals in the west and center of the city. For this reason, we proceed with spatial regression models.

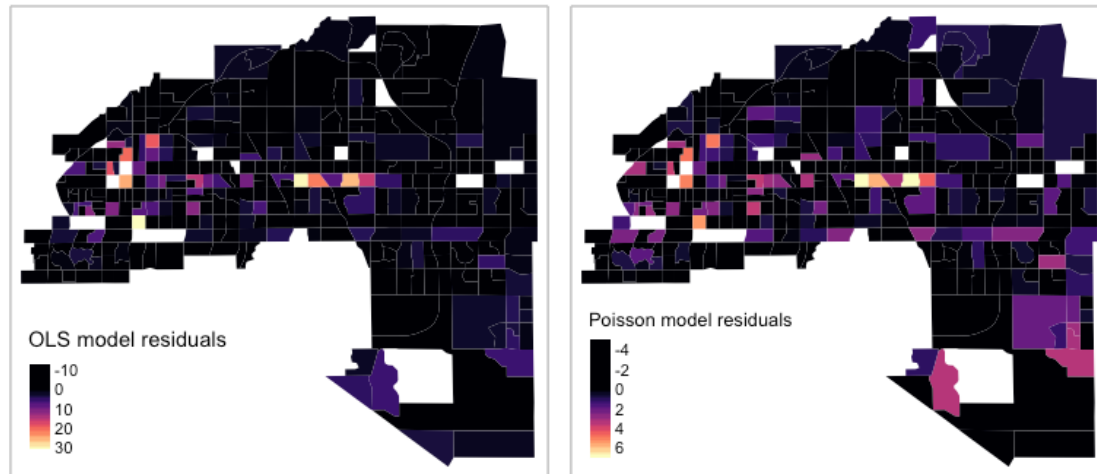


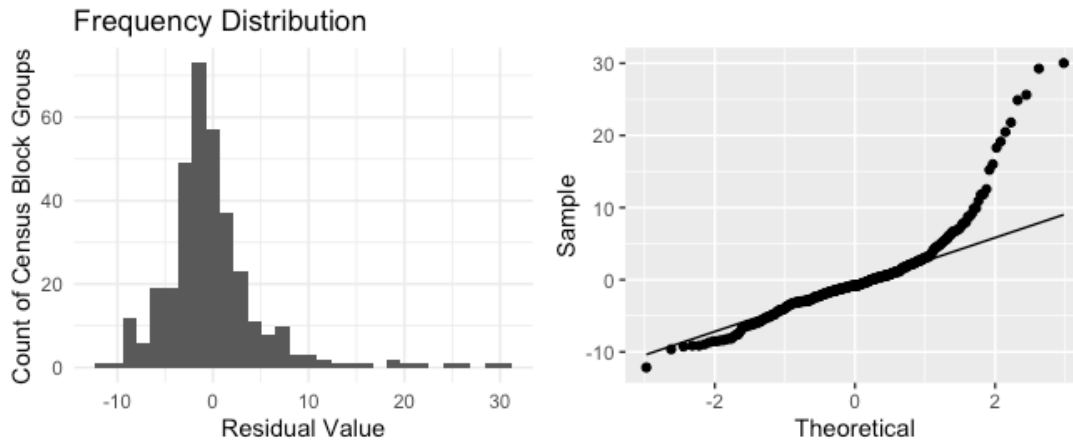
Figure 15: Spatial distribution of residuals from OLS model (left) and Poisson regression model (right)

Spatial Lag Model

Methodology

Reasoning for Spatial Modelling

The residual values from the OLS regression were observed to see whether any significant autocorrelation existed in these values. Figure 15 does show significant clusters of positive residuals to the west and centre of the city. While this histogram of residuals shows that an approximately normal distribution on the most part (figure 16), outliers to the positive side of the spectrum lead to deviation at the right tail. This is further evidenced by deviation at the right tail of the QQ-plot (figure 16).



Figure

16: Frequency distribution of residual values from OLS model, shown with histogram (left) and QQ-plot (right)

The results of the OLS model point to key assumptions of linear regression being violated, most notably the assumption that residuals are independent and not spatially autocorrelated. We attempt to integrate a spatial lag model as one of the spatial regression models to account for spatial autocorrelation in our data.

The spatial lag model assumes that autocorrelation is present in the dependent variable, that being opioid overdose cases. It assumes that the dependent variable is correlated with the same variable at nearby locations, above and beyond other covariates (Ward and Gleditsch 2008).

Spatial Lag Model

The spatial lag model is defined as:

$$Y = \rho Wy + X\beta + \epsilon$$

Where y is a vector of observations, ρ is an autocorrelation parameter, W is an $n * n$ spatial weight matrix and ϵ is an $n * 1$ vector of independent and identically distributed errors (iid). The spatial lag model follows the iid assumption, that being that the errors are normally distributed with a mean of 0 and a variance of σ^2 .

Before fitting a spatial lag model, it was assessed whether any of the independent variables required transformation prior to fitting a model. A Generalised Additive Model (GAM) was used to fit a smooth non-parametric function to the opioid overdoses variable using predictor variables (Zuur et al. 2007), and these results were tested against the distribution of independent variables to assess which transformations were appropriate. A spatial lag model was run with both predictor variables left untransformed and one with appropriate transformations, to assess which model had the best fit.

Results

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## overdose_count ~ s(new_arrests_count) + s(median_income) + s(perc_black) +
##   s(perc_hispanic) + s(perc_owners) + s(new_resid_high_per) +
##   s(commercial_per) + s(dist_greenpace) + s(dist_vacant_prop) +
##   s(dist_graffiti) + s(dist_public_safety)
##
## Parametric coefficients:
##               Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   5.0754      0.2705  18.76 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##               edf Ref.df   F        p-value
## s(new_arrests_count)  7.754  8.570 6.639 < 0.0000000000000002 ***
## s(median_income)      1.000  1.000 0.157      0.692037
## s(perc_black)         4.454  5.451 1.824      0.098363 .
## s(perc_hispanic)      1.000  1.000 2.781      0.096409 .
## s(perc_owners)        1.000  1.000 0.047      0.829217
## s(new_resid_high_per) 3.661  4.532 3.328      0.007001 **
## s(commercial_per)     6.453  7.573 1.825      0.081319 .
## s(dist_greenpace)     1.000  1.000 1.174      0.279398
## s(dist_vacant_prop)   1.000  1.000 0.493      0.483210
## s(dist_graffiti)      1.000  1.000 0.246      0.620306
## s(dist_public_safety) 3.483  4.352 5.176      0.000349 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.445   Deviance explained = 49.7%
## GCV = 27.901   Scale est. = 25.248     n = 345
```

Figure 17: Summary of GAM results

Above are the results from the GAM test on predictor variables (figure 17). Only 3 variables showed statistical significance to the 0.05% significance level, that being number of arrests, percentage of land zoned as high-density residential and average distance to public safety. In the context of a GAM, coefficients of 1 for the effective degrees of freedom coefficient (edf) tend to correspond to smooth functions that are more flexible, a good indicator that the relationship between this predictor variable and the dependent is non-linear. With all three statistically significant variables having edf coefficients above 1, their distributions were analysed further to assess if transformations were appropriate.

While distance to public safety replicated an approximately normal distribution, both police incidents and percentage of land zoned as high-density residential showed

distributions heavily skewed to the left. Therefore, a logarithmic transformation was only applied to police incidents and percentage of land zoned as high-density residential, with the results from this model summarised below (figure 18).

```
##
## Call: lagsarlm(formula = overdose_count ~ log(new_arrests_count) +
##   median_income + perc_black + perc_hispanic + perc_owners +
##   log(new_resid_high_per) + commercial_per + dist_greenspace +
##   dist_vacant_prop + dist_graffiti + dist_public_safety, data = mesa_all
##
##   listw = mesa.W)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.81490  -2.75076  -0.74777   1.44217  33.25962
##
## Type: lag
## Coefficients: (asymptotic standard errors)
##              Estimate      Std. Error z value Pr(>|z|)
## (Intercept)    7.64119338915    1.84770544783   4.1355 0.00003542
## log(new_arrests_count)  0.16575710888    0.08463837119   1.9584  0.05018
## median_income    0.00000087482    0.00001464226   0.0597  0.95236
## perc_black      0.05883983953    0.07222495625   0.8147  0.41526
## perc_hispanic    0.03817260280    0.01827009987   2.0893  0.03668
## perc_owners     -0.02672142689    0.01847452372  -1.4464  0.14807
## log(new_resid_high_per)  0.04925813137    0.06572052057   0.7495  0.45355
## commercial_per    0.07080835929    0.03849831961   1.8393  0.06588
## dist_greenspace    0.00017187779    0.00014300374   1.2019  0.22940
## dist_vacant_prop  -0.00009874233    0.00004515946  -2.1865  0.02878
## dist_graffiti     0.00035539923    0.00029159967   1.2188  0.22292
## dist_public_safety -0.00040184005    0.00014692989  -2.7349  0.00624
##
## Rho: 0.095113, LR test value: 1.2743, p-value: 0.25897
## Asymptotic standard error: 0.084784
##      z-value: 1.1218, p-value: 0.26194
## Wald statistic: 1.2585, p-value: 0.26194
##
## Log likelihood: -1088.217 for lag model
## ML residual variance (sigma squared): 32.108, (sigma: 5.6664)
## Number of observations: 345
## Number of parameters estimated: 14
## AIC: 2204.4, (AIC for lm: 2203.7)
## LM test for residual autocorrelation
## test value: 4.1952, p-value: 0.040538
```

Figure 18: Results from Spatial Lag model with transformed variables

A model was also run with all predictor variables left untransformed, with the results from this model shown below (figure 19).

```
##
## Call:lagsarlm(formula = overdose_count ~ arrests_count + median_income +
##   perc_black + perc_hispanic + perc_owners + resid_high_per +
##   commercial_per + dist_greenpace + dist_vacant_prop + dist_graffiti +
##   dist_public_safety, data = mesa_all, listw = mesa.W)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.68858  -2.77495  -0.78857   1.59872  30.08123
##
## Type: lag
## Coefficients: (asymptotic standard errors)
##              Estimate      Std. Error z value    Pr(>|z|)
## (Intercept)    3.6479249244    1.9864004869    1.8364    0.0662911
## arrests_count    0.0059981206    0.0012320037    4.8686    0.000001124
## median_income    0.0000056458    0.0000139092    0.4059    0.6848120
## perc_black      0.0687310881    0.0699373380    0.9828    0.3257293
## perc_hispanic    0.0511849427    0.0179704183    2.8483    0.0043955
## perc_owners     -0.0026866509    0.0184493429   -0.1456    0.8842189
## resid_high_per    0.0565217873    0.0162627489    3.4755    0.0005098
## commercial_per    0.0144096230    0.0390875343    0.3687    0.7123886
## dist_greenpace    0.0000477373    0.0001398952    0.3412    0.7329257
## dist_vacant_prop -0.0000443705    0.0000448593   -0.9891    0.3226118
## dist_graffiti     0.0003057286    0.0002753647    1.1103    0.2668837
## dist_public_safety -0.0003512521    0.0001426084   -2.4631    0.0137760
##
## Rho: 0.06336, LR test value: 0.57552, p-value: 0.44807
## Asymptotic standard error: 0.082324
##      z-value: 0.76965, p-value: 0.44151
## Wald statistic: 0.59236, p-value: 0.44151
##
## Log likelihood: -1076.576 for lag model
## ML residual variance (sigma squared): 30.038, (sigma: 5.4807)
## Number of observations: 345
## Number of parameters estimated: 14
## AIC: 2181.2, (AIC for lm: 2179.7)
## LM test for residual autocorrelation
## test value: 0.51267, p-value: 0.47398
```

Figure 19: Results from Spatial Lag model with variables left in their natural form

From a first glance, we can see that the model with non-transformed predictor variables has both a lower AIC coefficient (2181.2 vs 2204.4) and a higher log likelihood (-1.076.6 vs -1.088.2). These both indicate that the model with untransformed predictor variables fits our data much better, and further analysis focused on this model.

From the individual variables, only four variables showed statistical significance at the 5% significance level, these being police incidents, Hispanic population percentage, percentage of land zoned as high-density residential and average distance to public safety. While all of their respective coefficients were substantially low, they were all positively associated with

opioid overdose cases. For example, the spatial lag model indicates that a one increase in police incidents corresponded to a 0.006% increase in opioid overdoses, and a 1 increase in percentage of Hispanic population corresponded to a 0.05% increase in overdoses. While these observations reaffirm our hypotheses of higher opioid overdoses being found in areas with increased arrests (Choi et al. 2022), higher Latina communities (Lippold et al. 2019) and increased distances from public safety (Li et al. 2019), it is important to assess the suitability of the spatial lag model as a whole before making any final conclusions.

The p-value from the rho autocorrelation parameter of 0.448 is a strong indicator that the additionality of ρ does not significantly improve the model fit. Furthermore, the p-value of the Wald test of 0.442 indicates that there is insufficient evidence to reject the null hypothesis that the parameters of the spatial lag variable are equal to zero.

The residuals from this spatial lag model further prove that the spatial lag model did a poor job in accounting for the observed spatial autocorrelation across Mesa, given their almost identical nature to residuals from the linear OLS regression model. These can be seen bellow in the map, histogram and QQ-plot of the spatial lag model residuals (figure 20).

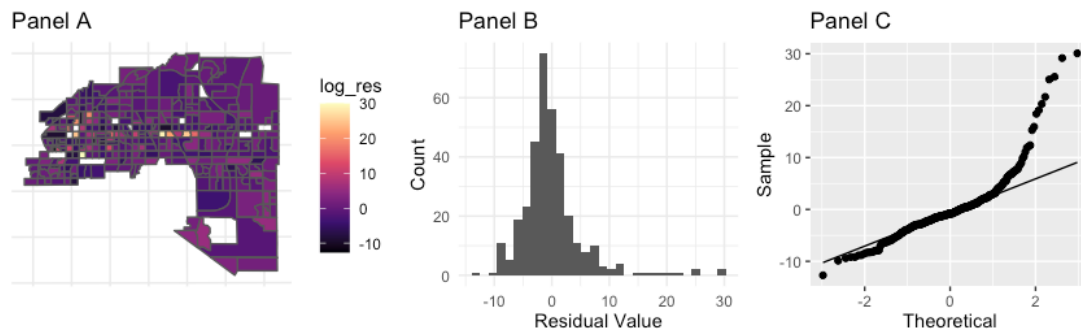


Figure 20: Spatial Lag Model Residuals: Panel A: Spatial distribution of residuals, Panel B: Frequency distribution of residuals, Panel C: QQ-plot.

All in all, these observations indicate that there is insufficient evidence to reject the null hypothesis. Therefore, we cannot say with confidence that the dependent variable (opioid overdoses) is correlated with the same variable at nearby locations. Different spatial regression models may better explain the spatial autocorrelation observed from the exploratory spatial data analysis.

Spatial Error Model

Methodology

The spatial error model assumes autocorrelation is a result of some unobserved variable(s) and is present in the residuals of the model, which means that it assumes that only the error terms in the regression are correlated. The basic form of the model is as follows:

$$y = X\beta + \epsilon$$

$$\epsilon = \lambda W\epsilon + \xi$$

Therefore, the final spatial error model can be specified as:

$$y = X\beta + \lambda W\epsilon + \xi$$

Similarly to the OLS model, y is an $n * 1$ vector of independent variable; X is an $n * m$ matrix of independent variables; β is an $m * 1$ vector of parameters to be estimated. In the OLS regression we estimate a single error term, ϵ . While in a spatial error model, we split this error into both random error and spatially structured error. Here, random error (unexplained by our model) is ξ , independent identically distributed (i.i.d.) errors. Spatially structured error is composed of the the spatial error parameter λ , and the original ϵ error term is now weighted by the weight matrix W . If $\lambda = 0$, there is no spatial correlation between the errors. If $\lambda \neq 0$, OLS is unbiased and consistent, but the standard errors will be wrong and the betas will be inefficient (Emily Burchfield, 2020).

The results of the Spatial Error Model are shown below (figure 21).

```
##
## Call:errorsarlm(formula = overdose_count ~ arrests_count + median_income +
##   perc_black + perc_hispanic + perc_owners + resid_high_per +
##   commercial_per + dist_greenpace + dist_vacant_prop + dist_graffiti +
##   dist_public_safety, data = mesa_all, listw = mesa.W)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.4582  -2.8639  -0.8219   1.6000  30.1115
##
## Type: error
## Coefficients: (asymptotic standard errors)
##              Estimate      Std. Error z value    Pr(>|z|)
## (Intercept)    4.2614776776    1.8433192983    2.3118    0.0207860
## arrests_count    0.0060331619    0.0012317820    4.8979 0.0000009686
## median_income    0.0000055683    0.0000140256    0.3970    0.6913585
## perc_black      0.0702173663    0.0701665911    1.0007    0.3169604
## perc_hispanic    0.0527355887    0.0180735822    2.9178    0.0035248
## perc_owners     -0.0031715858    0.0184911664   -0.1715    0.8638157
## resid_high_per    0.0581002948    0.0162936608    3.5658    0.0003627
## commercial_per    0.0152319922    0.0393635549    0.3870    0.6987882
## dist_greenpace    0.0000583975    0.0001427766    0.4090    0.6825302
## dist_vacant_prop -0.0000529833    0.0000446651   -1.1862    0.2355296
## dist_graffiti     0.0003068704    0.0002793663    1.0985    0.2720073
## dist_public_safety -0.0003777596    0.0001418712   -2.6627    0.0077518
##
## Lambda: 0.040382, LR test value: 0.20057, p-value: 0.65426
## Asymptotic standard error: 0.090198
##      z-value: 0.44771, p-value: 0.65436
## Wald statistic: 0.20044, p-value: 0.65436
##
## Log likelihood: -1076.764 for error model
## ML residual variance (sigma squared): 30.083, (sigma: 5.4848)
## Number of observations: 345
## Number of parameters estimated: 14
## AIC: 2181.5, (AIC for lm: 2179.7)
```

Figure 21: Results from Spatial Error model

As is shown in the result, at the 5% significance level, four independent variables show significant correlation with opioid overdose cases. The four factors are the number of arrests, the percent of Hispanic residents, the percentage of the area zoned as high-density residential and the average distance to public safety resources. Except for the distance to public safety resources, which is negatively correlated with the numbers of overdoses cases, the other three factors all show positive correlation with the dependent variable. To quantitatively illustrate the correlation, every one-unit increase in the percent of Hispanic residents corresponds to a 0.05% increase in overdose cases, and every one-unit longer distance to public safety resources corresponds to a 0.0003% decrease in overdoses case. While the influence of these independent variables are rather weak as their coefficients are

significantly low compared to the other two influential factors, the percentage of the area zoned as high-density residential and the percentage of Hispanic residents still have much higher coefficients, which means they have stronger impact on the number of overdose cases. This result is consistent with the OLS model and the spatial lag model. It also affirms most of our hypotheses that census blocks with greater minority population and higher crime rates would have more overdose cases. Discussion of the results that fail to coincide with our hypotheses was presented in the OLS model and it is also applied in the spatial error model.

Looking at the model fitting effect, the autocorrelation parameter λ didn't pass the LM test, which tests whether the addition of λ significantly improves the model fit. In this case, the p-value is not significant at the 5% level so we cannot conclude that the autocorrelation parameter significantly improves the model fit. The large negative log likelihood value also indicates a failure of better fit of the regression model. Furthermore, the 0.65 p-value of the Wald statistic shows that we have insufficient evidence to reject the null hypothesis that the parameters of the spatial error model are equal to zero. This suggests that the independent variables we are using to account for the dependent variable in this model didn't produce a good fitting effect, which is consistent with the results of aforementioned indices.

Normally, a larger log-likelihood and a larger negative AIC indicate better model fits. However, as we compare these two indices between spatial lag model and spatial error model, we find that they essentially have the same poor result, which means neither of the models have better model fits.

While we visualise the spatial distribution of residuals (figure 22) of the error model, we still see that some level of spatial autocorrelation in the west and centre of Mesa. It suggests that the spatial error model didn't fully account for the spatial dependence in our data.

Furthermore, as we compare the map, histogram and QQ-plot with OLS and spatial lag models (figure 22), it is obvious to see that they are essentially identical among these three models. It is another suggestion that the spatial error model poorly accounted for the variance that couldn't be explained in the previous OLS and spatial models, which means the spatial error model is still not a suitable method to answer our research question.

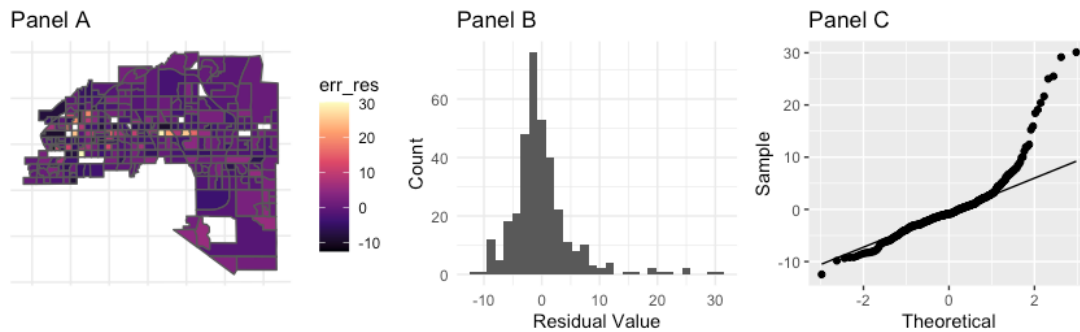


Figure 22: Spatial Error model residuals: Panel A: Spatial distribution of residuals, Panel B: Frequency distribution of residuals, Panel C: QQ-plot.

Thus, at this part, we can conclude that the spatial error model is unable to make significant improvements on the OLS model and spatial lag models. A better regression model should be applied, or more appropriate variables should be added to the model to achieve a better fitting effect.

Spatial Durbin Watson Model

Methodology

The formula of the Spatial Durbin Model is as follows:

$$Y = \rho W y + X\beta + WX\theta + \epsilon$$

The Spatial Durbin Model (SDM) is a model that extends from the Spatial Lag Model, that takes account of the spillover effects of both the independent variables and the dependent variable in one model (Atikah et al., 2021). Y is the vector of observations, ρW is the scalar autocorrelation parameter and $n * n$ is the spatial weight matrix that allows neighboring parameters for the dependent variable. $X\beta$ is the $n * m$ matrix and $n * 1$ estimated parameters of the independent variables respectively. WX creates a weighed sum of independent variables neighboring the location, which are then each multiplied by the $m * 1$ autocorrelation parameter θ . Lastly, the ϵ is the expected normally distributed errors.

Results

The results of the Spatial Durbin model are shown below (figure 23).

```
##
## Call:lagsarlm(formula = overdose_count ~ arrests_count + median_income +
```

```
##      perc_black + perc_hispanic + perc_owners + resid_high_per +
##      commercial_per + dist_greenpace + dist_vacant_prop + dist_graffiti +
##      dist_public_safety, data = mesa_data, listw = mesa.W, type = "mixed")
##
## Residuals:
##      Min          1Q      Median          3Q      Max
## -13.19256  -2.67268  -0.61879   1.50243   29.89404
##
## Type: mixed
## Coefficients: (asymptotic standard errors)
##              Estimate      Std. Error z value      Pr(>|z|)
## (Intercept)      5.6618323065    3.5783539464    1.5822    0.113594
## arrests_count      0.0059790260    0.0012305927    4.8587 0.000001182
## median_income      0.0000011883    0.0000154048    0.0771    0.938516
## perc_black         0.0573027454    0.0703387274    0.8147    0.415262
## perc_hispanic      0.0359409301    0.0207238656    1.7343    0.082869
## perc_owners        0.0022064861    0.0187978765    0.1174    0.906559
## resid_high_per     0.0549702158    0.0185873580    2.9574    0.003102
## commercial_per     0.0048768524    0.0423432843    0.1152    0.908307
## dist_greenpace     -0.0006275158    0.0003773956   -1.6628    0.096362
## dist_vacant_prop   -0.0003820141    0.0003825711   -0.9985    0.318016
## dist_graffiti       0.0001414306    0.0003493138    0.4049    0.685565
## dist_public_safety  0.0001997832    0.0003972708    0.5029    0.615042
## lag.arrests_count  -0.0021900418    0.0023721805   -0.9232    0.355893
## lag.median_income   0.0000181667    0.0000303154    0.5993    0.549002
## lag.perc_black      -0.0477986379    0.1564148028   -0.3056    0.759918
## lag.perc_hispanic   0.0208325298    0.0358815645    0.5806    0.561516
## lag.perc_owners     -0.0432393542    0.0431452344   -1.0022    0.316256
## lag.resid_high_per   0.0052586828    0.0361600681    0.1454    0.884373
## lag.commercial_per   0.0836544198    0.0810073721    1.0327    0.301755
## lag.dist_greenpace   0.0009034674    0.0004791819    1.8854    0.059371
## lag.dist_vacant_prop 0.0002958820    0.0003965177    0.7462    0.455546
## lag.dist_graffiti    0.0003465680    0.0005534204    0.6262    0.531165
## lag.dist_public_safety -0.0007010343    0.0004978249   -1.4082    0.159073
##
## Rho: 0.019272, LR test value: 0.046365, p-value: 0.82951
## Asymptotic standard error: 0.090316
##      z-value: 0.21338, p-value: 0.83103
## Wald statistic: 0.045532, p-value: 0.83103
##
## Log likelihood: -1071.651 for mixed model
## ML residual variance (sigma squared): 29.211, (sigma: 5.4047)
## Number of observations: 345
## Number of parameters estimated: 25
## AIC: 2193.3, (AIC for lm: 2191.3)
## LM test for residual autocorrelation
## test value: 0.097449, p-value: 0.75491
```

Figure 23: Results from Spatial Durbin model

As seen in the results, ρ , the spatial autoregressive term, is 0.019, with an insignificant p -value of 0.830. These are similar to the results of the spatial lag model, and indicate that it does not necessarily improve fitting the model with the additional p . Compared to the spatial lag model, the log-likelihood is higher (-1.071.7 vs -1.076.6), but the AIC coefficient is higher as well (2193.3 vs 2181.2), so whether adding the spatially lagged independent variable effect improves fitting the model is contradictory since the log-likelihood improved but the AIC coefficient was not as a lower value is preferred (Lee and Ghosh, 2009). Furthermore, the Wald test, another spatial dependence test, has a value of 0.046 with a p -value of 0.831, showing that it is also statistically insignificant. Lastly, for the LM test, which tests residual autocorrelation, the chi-square value is 0.097 with a p -value of 0.755. It indicates that this model does not perform better in accounting for the autocorrelation in the data, and the residuals are not fully accounted for in the SDM.

Looking at the coefficients of the independent variables as well as the lagged independent variables, only one variable showed statistical significance at the 5% confidence level, that being the percentage of land zoned as high-density residential. It has a positive relationship with opioid overdose cases, that a 1% increase in the percentage of land zoned as high-density residential corresponded to a 0.055 increase in opioid overdose cases in the area. Compared to the spatial lag model, other 3 significant variables including police incidents, Hispanic population percentage and average distance to public safety are insignificant in the spatial durbin model. SDM generally does not show positive feedback to our hypothesis other than affirming the positive relationship of opioid overdose cases and proximity to high-density residential areas.

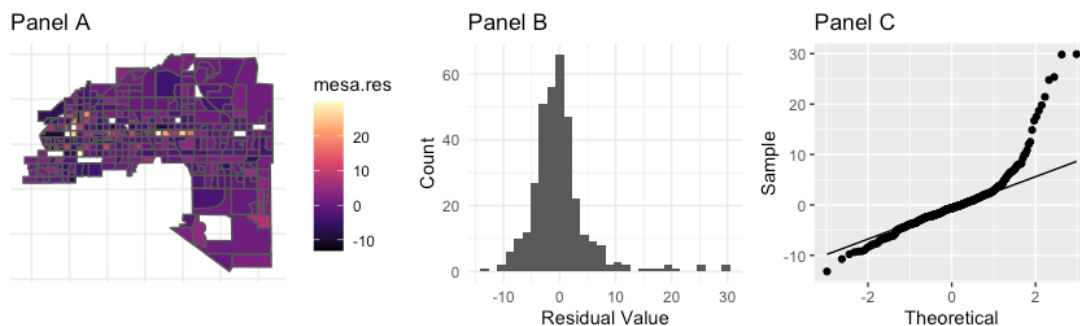


Figure 24: Spatial Durbin model residuals: Panel A: Spatial distribution of residuals, Panel B: Frequency distribution of residuals, Panel C: QQ-plot.

Similarly to the spatial lag model and spatial error model, the plotting of the residuals of the spatial durbin model shows they need to be accounted for in the model since the residuals are almost identical to those of the linear OLS regression and the spatial lag model (figure 24). Therefore, the spatially lagged dependent variable with the spatially lagged independent variables in the spatial durbin model cannot fully explain the opioid

overdose data with the independent variables in Mesa. Other spatial regression models are preferred in order to show a better fit in this case.

Geographically Weighted Regression

Methodology

Geographically weighted regression (GWR) is a method for analyzing spatially varying relationships (C. Brunsdon, Fotheringham, and Charlton 1998). It is an extension of linear regression that accounts for variation across space (Brunsdson et al., 1998). The general form of the model is:

$$y_i = \beta_{i0} + \sum_{p=1}^m \beta_{ip} x_{ip} + \epsilon_i$$

where β_{i0} is the intercept term at location i , $i = 1, 2, \dots, N$; N is the number of spatial locations; β_{ip} is the value of the p th parameter at location i , $p = 1, 2, \dots, m$; m is the number of independent variables; ϵ_i is a random error.

We first obtain the centroid points and their coordinates for all district polygons. Choice of bandwidth affects the results of GWR so it is important to select a reasonable value. In the `spgwr` package, the 'optimal' bandwidth can be estimated using the `gwr.sel` function. This function uses a method called leave-one-out cross-validation (LOO-CV) to find the bandwidth that results in the smallest error. With the acquired bandwidth of `bwG` we can fit the GWR model.

Results

The results of the Geographically Weighted Regression are shown below (figure 25).

```
## Call:
## gwr(formula = overdose_count ~ arrests_count + median_income +
##      perc_black + perc_hispanic + perc_owners + resid_high_per +
##      commercial_per + dist_greenpace + dist_vacant_prop + dist_graffiti +
##      dist_public_safety, data = mesa_a, coords = cbind(mesa_a$long,
##      mesa_a$lat), bandwidth = bwG, gweight = gwr.Gauss, hatmatrix = TRUE,
##      longlat = FALSE)
## Kernel function: gwr.Gauss
## Fixed bandwidth: 39346.39
## Summary of GWR coefficient estimates at data points:
##               Min.           1st Qu.           Median           3rd
Qu.
## X.Intercept.      2.01661628821    3.07968641233    4.63097540161    5.9727334
4535
## arrests_count      0.00557120061    0.00592299325    0.00644574060    0.0071328
8933
## median_income     -0.00000054689    0.00000248991    0.00000313323    0.0000036
```



```

0525
## perc_black          0.03733497580  0.04634311065  0.05468442433  0.0632259
4756
## perc_hispanic       0.04512094753  0.05068834961  0.05205751341  0.0530173
6527
## perc_owners         -0.01330684532 -0.00803094778 -0.00109869285  0.0078428
6618
## resid_high_per      0.04725627867  0.05403475004  0.06142864801  0.0682670
4811
## commercial_per      -0.00918068593  0.01022204671  0.02118481395  0.0327075
4974
## dist_greenspace     -0.00020857309 -0.00008650396  0.00001321014  0.0000874
8317
## dist_vacant_prop    -0.00008646740 -0.00006370662 -0.00004483716 -0.0000318
9330
## dist_graffiti       0.00006145540  0.00028778262  0.00039531766  0.0004820
8337
## dist_public_safety -0.00051129148 -0.00047462771 -0.00043191627 -0.0003459
6630
##                               Max.  Global
## X.Intercept.          7.22034767470  4.2520
## arrests_count         0.00733801434  0.0060
## median_income         0.00000554901  0.0000
## perc_black            0.07266534315  0.0709
## perc_hispanic         0.05391461656  0.0535
## perc_owners           0.01439470982 -0.0039
## resid_high_per        0.06992657835  0.0582
## commercial_per        0.04369542691  0.0157
## dist_greenspace       0.00012320822  0.0001
## dist_vacant_prop      -0.00002571506 -0.0001
## dist_graffiti         0.00057536061  0.0003
## dist_public_safety    -0.00016078807 -0.0004
## Number of data points: 345
## Effective number of parameters (residual: 2traceS - traceS'S): 21.81124
## Effective degrees of freedom (residual: 2traceS - traceS'S): 323.1888
## Sigma (residual: 2traceS - traceS'S): 5.557324
## Effective number of parameters (model: traceS): 17.87676
## Effective degrees of freedom (model: traceS): 327.1232
## Sigma (model: traceS): 5.523803
## Sigma (ML): 5.378786
## AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 2180.029
## AIC (GWR p. 96, eq. 4.22): 2157.844
## Residual sum of squares: 9981.313
## Quasi-global R2: 0.3624625

```

Figure 25: Results from Geographically Weighted Regression

The output of the GWR model shows that the Quasi-global R-squared value of 0.36, meaning it is only able to explain around 36% of the variation in the dependent variable.

```
##
## Leung et al. 2000 three moment approximation for Moran's I
##
## data: GWR residuals
## statistic = 827.38, df = 805.81, p-value = 0.2914
## sample estimates:
## I
## -0.009883989
```

Figure 26: Moran's I for GWR residuals

To check for autocorrelation in the residuals, we use the Moran's I (Figure 26). The p-value of 0.2914 is not significant at the 5% level, revealing that we cannot reject the null hypothesis that there is no spatial autocorrelation. This means that the GWR is able to account for the spatial autocorrelation in the opioid overdose incidents.

By mapping each of the varying coefficient estimates, we can examine how these statistical values change spatially (figure 27).

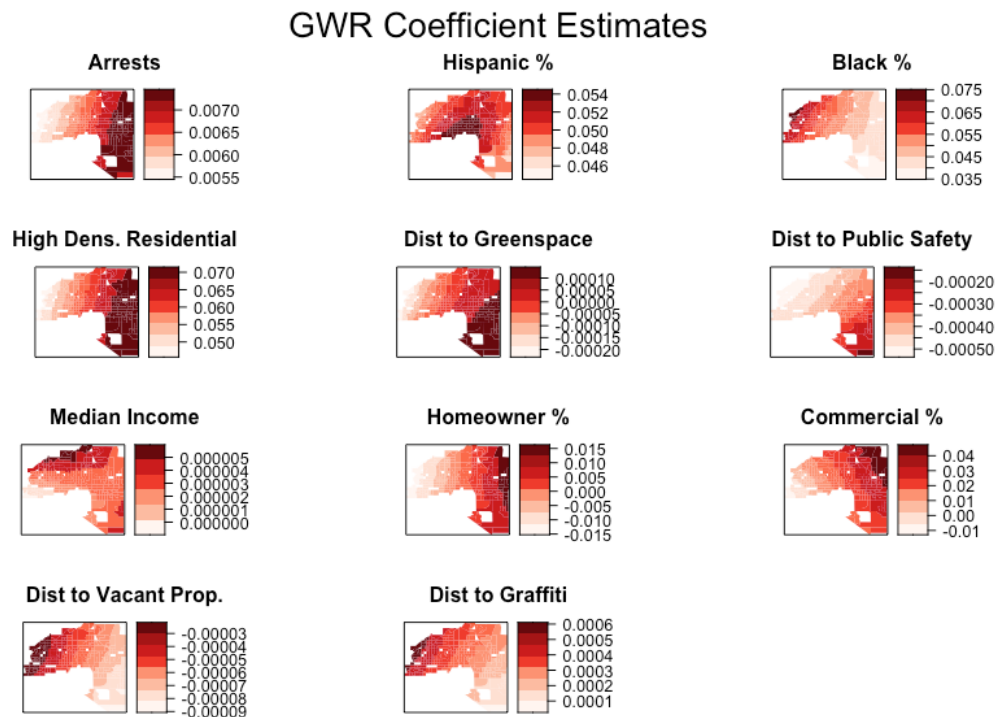


Figure 27: Spatial distribution of residuals from GWR

Geographically Weighted Poisson Regression

Methodology

Given that a Poisson regression better represents the relationship between the risk factors and the overdose count, we attempt to integrate this model and the spatial variation in overdose cases using a geographically weighted poisson regression which captures local relationships.

The geographically weighted poisson regression follows the form:

$$\ln(y_i) = \beta_0 u_i + \beta_1(u_i) x_{i1} + \beta_2(u_i) x_{i2} + \dots + \beta_n(u_i) x_{in} + \epsilon$$

where where y_i is the i th observation of number of overdoses and the β coefficients are functions of the location u_i designating the coordinates of the i th census block group, allowing the estimated β parameter to differ between census block groups (Poliart et. al, 2021).

For each census block group, the parameters were estimated by

$$\hat{\beta}_i = (X^T W(u_{xi}, u_{yi}) X)^{-1} X^T W(u_{xi}, u_{yi}) Y$$

where $W(u_{xi}, u_{yi})$ denotes a spatial weight matrix containing weights calculated based on the distance between the centroid of the i th census block group with coordinates $(u_{\{xi\}}, u_{\{yi\}})$ and its spatial neighbors j (Fortheringham et al, 2003).

The spatial weight w_{ij} allows for census block groups closer to the i th census block group to have a larger influence on the parameter estimation for β_i . The weight is based on a kernel function. We experiment with two kernel functions, the Gaussian and Exponential, which are both continuous functions of the distance between observation points.

The corresponding weight matrices are calculated as follows

$$\text{Gaussian: } w_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right) \text{ and Exponential: } w_{ij} = \exp\left(-\frac{|d_{ij}|}{b}\right)$$

where h is the kernel bandwidth and d_{ij} is the distance between points i and j (Gollini et al., 2013).

For each kernel function, we also experiment with both a fixed kernel (which use a fixed bandwidth) and adaptive kernel (which uses a varying bandwidth based on the number of nearest neighbors from a given regression point.) Given we have census block groups of different sizes, the adaptive kernels could be useful because these kernels change in size based on the density of data. Using the fixed kernels could result in calibration on few observation points for local regression in larger units and vice versa, resulting in larger standard errors (Gollini et al., 2013).

We select the optimal bandwidth for each kernel by minimizing the Akaike Information Criterion to account for both prediction accuracy as well as lower complexity of our model (Gollini et al., 2015). For the fixed Gaussian and Exponential kernels, the optimal bandwidths were 7614.3182847 and 6843.6505642 meters respectively. For the adaptive Gaussian and Exponential kernels, the optimal bandwidths was 18 nearest neighbors.

Results

We fit the 4 geographically weighted Poisson models using their corresponding optimal bandwidths. We find that the model using the exponential adaptive kernel has the lowest AIC and residual sum of squares and the highest Pseudo R-squared values. Below are the maps of the standard deviation of the local residual values for the models (figure 28). As mentioned, exponential adaptive kernel appears to have the best fit for our data, so we choose these parameters for our final model.

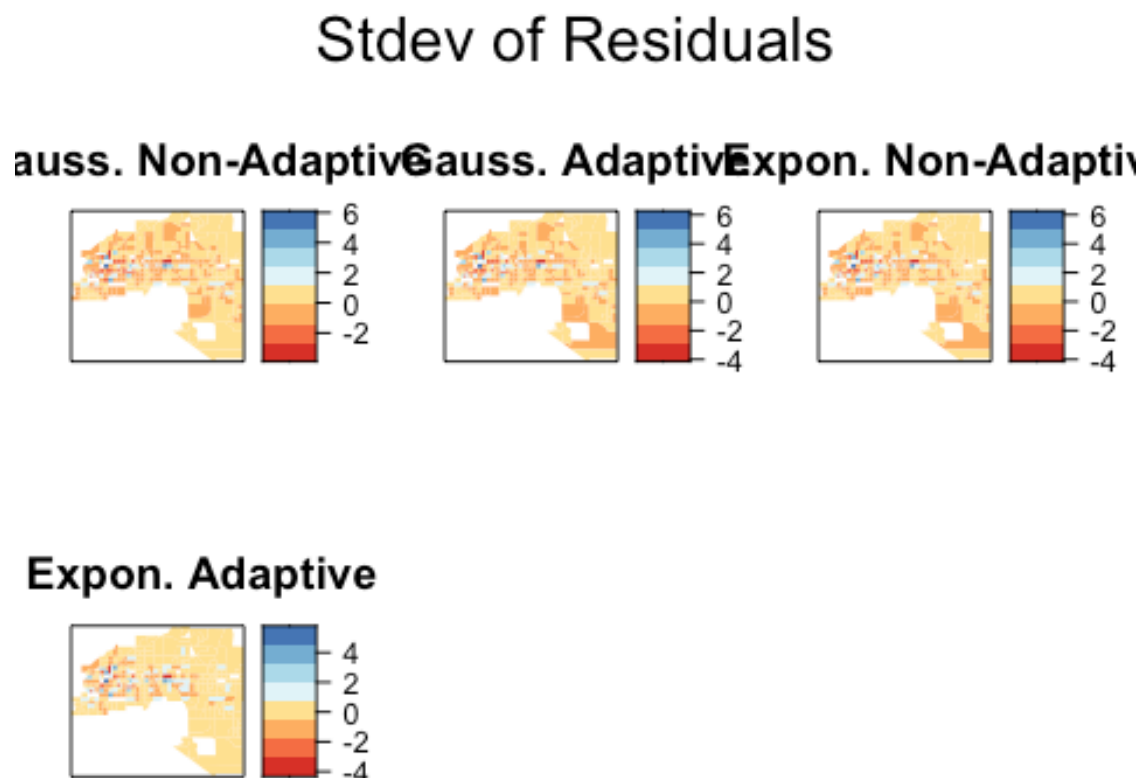


Figure 28: Standard deviation of local residual values from models

Below are the results of the geographically weighted poisson regression using an adaptive exponential kernel (figure 29). The results show that the model has a pseudo R-squared value of 0.69, meaning that it is able to explain almost 70% of the variation in overdose cases. This is a substantial improvement from the 45% R-squared value of the baseline poisson model.

The results of the GWPR provide us with better insight into how the risk factors vary across the different census block groups. For example, the baseline poisson regression indicated that 1 additional arrest corresponded with a 0.068% increase in overdoses. However, with the GWPR we see that the increase can be over 3 times higher, corresponding to a 0.23% increase in overdoses. The broadness of the ranges in local estimates are even more pronounced for the Hispanic population as well as land zoned as high-density residential. The initial estimate from the baseline model indicated that one percentage point increase in the proportion of Hispanic residents was associated with a 0.7% increase in number of overdoses, but the GWPR reveals that it can be associated with an overdose increase of as high as 2.8% in some city areas. Meanwhile, estimates for increases in incidents corresponding to high-density residential zoning range from as low as 0.1% to as high as ~2.5%. The results also reveal that a majority of our distance-related risk factors, especially the distance to public greenspace and distance to public safety, are almost always negatively correlated with overdose cases, regardless of the part of the city. This disaffirms our original hypothesis that in blocks closer to these facilities there will be less incidents.

```
## *****
## *                               Package    GWmodel                               *
## *****
## Program starts at: 2023-01-19 17:51:06
## Call:
## ggwr.basic(formula = overdose_count ~ arrests_count + median_income +
##   perc_black + perc_hispanic + perc_owners + resid_high_per +
##   commercial_per + dist_greenspace + dist_vacant_prop + dist_graffiti +
##   dist_public_safety, data = mesa_all, bw = bw.gwrEA, family = "poisson"
## ,
##   kernel = "exponential", adaptive = TRUE, dMat = DM)
##
## Dependent (y) variable:  overdose_count
## Independent variables:  arrests_count median_income perc_black perc_hispanic
##   perc_owners resid_high_per commercial_per dist_greenspace dist_vacant_p
##   rop dist_graffiti dist_public_safety
## Number of data points: 345
## Used family: poisson
## *****
## *               Results of Generalized linear Regression               *
## *****
##
## Call:
## NULL
##
```

```

## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -1.0000   -0.6807   -0.2278    0.4445   15.8330
##
## Coefficients:
##              Estimate      Std. Error z value      Pr(>|z|)
## Intercept          2.0595832100    0.1566417221   13.148    < 2e-16 ***
## arrests_count       0.0006844624    0.0000686423    9.971    < 2e-16 ***
## median_income      -0.0000008436    0.0000013771   -0.613    0.5401
## perc_black          0.0116940923    0.0048565931    2.408    0.0160 *
## perc_hispanic       0.0070092185    0.0013777334    5.087 0.000000363 ***
## perc_owners         0.0009652980    0.0014383300    0.671    0.5021
## resid_high_per      0.0098350225    0.0011898858    8.266    < 2e-16 ***
## commercial_per      0.0045587548    0.0026782416    1.702    0.0887 .
## dist_greenspace     -0.0000579105    0.0000143220   -4.043 0.000052664 ***
## dist_vacant_prop     0.0000025045    0.0000043933    0.570    0.5686
## dist_graffiti        -0.0000081344    0.0000421016   -0.193    0.8468
## dist_public_safety  -0.0001149590    0.0000129847   -8.853    < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 2322.2  on 344  degrees of freedom
## Residual deviance: 1269.0  on 333  degrees of freedom
## AIC: 1293
##
## Number of Fisher Scoring iterations: 5
##
##
## AICc: 1293.976
## Pseudo R-square value: 0.4535294
## *****
## *              Results of Geographically Weighted Regression              *
## *****
## *****Model calibration information*****
## Kernel function: exponential
## Adaptive bandwidth: 18 (number of nearest neighbours)
## Regression points: the same locations as observations are used.
## Distance metric: A distance matrix is specified for this model calibration.
##
## *****Summary of Generalized GWR coefficient estimates:*****
##              Min.          1st Qu.          Median
## Intercept      -0.22607694166    1.20067276202    1.73960798969
## arrests_count   0.00042833565    0.00074185490    0.00107242046
## median_income  -0.00001113456   -0.00000422743    0.00000013829
## perc_black     -0.02218714651   -0.00127668722    0.00569746260
## perc_hispanic  -0.00317040486    0.00372447667    0.00640199380

```

```

##      perc_owners      -0.00384157393  0.00040092182  0.00289802946
##      resid_high_per    0.00105202619  0.00577706051  0.01019328060
##      commercial_per    -0.01377220074  0.00243908237  0.01007332235
##      dist_greenpace    -0.00023266937 -0.00014462074 -0.00004561559
##      dist_vacant_prop  -0.00004565609 -0.00002088898 -0.00000954670
##      dist_graffiti     -0.00013024107  0.00001260078  0.00004873067
##      dist_public_safety -0.00026929248 -0.00014928048 -0.00010660187
##                                     3rd Qu.    Max.
##      Intercept          2.47403573774  3.6311
##      arrests_count      0.00152285380  0.0023
##      median_income      0.00000312500  0.0000
##      perc_black         0.01224007904  0.0332
##      perc_hispanic      0.01012181509  0.0283
##      perc_owners        0.00740199548  0.0182
##      resid_high_per     0.01461808174  0.0242
##      commercial_per     0.01523986914  0.0230
##      dist_greenpace     -0.00001225402  0.0001
##      dist_vacant_prop   0.00000815925  0.0001
##      dist_graffiti      0.00009227430  0.0003
##      dist_public_safety -0.00006211557  0.0000
##      *****Diagnostic information*****
##      Number of data points: 345
##      GW Deviance: 719.0314
##      AIC : 889.3877
##      AICc : 946.1102
##      Pseudo R-square value:  0.6903717
##
##      *****
##      Program stops at: 2023-01-19 17:51:07

```

Figure 29: Results from Geographically Weighted Poisson Regression using an adaptive exponential kernel

Given that the magnitude of the coefficients is very small, it is useful to examine the test statistics to investigate how the statistical significance of these values change spatially. Below are maps of the pseudo t-values for each of the risk factors (figure 30).

From the results, we observe that the number of arrests is almost always statistically significant in all regions, but it is especially significant in the northeast, southeast and western-central areas of the city. Additionally, the hispanic population is increasingly significant closer to the center of the city, while the black population is significant in the far west of the city. Distance to greenspace has a significant negative relationship with overdose incidents in the west of the city while distance to public safety has a significant negative relationship with overdose cases in the east of the city. However, it is important to note when interpreting these results that these are municipal-run facilities, so for example, our model is only taking into consideration city-owned parks are not accounting for private greenspace which might be available in other areas. Unsurprisingly, land being zoned

commercial is only significant in the east of the city, which is reasonable as that is farther from the western downtown area where most of the neighborhoods would be zoned as commercial. On the other hand, distance to vacant property is positively statistically significant in the center of the city, but negatively statistically significant in the in the southeast. Similarly, median income is positively significant near the downtown area and central-east, but there are enclaves of blocks where the relationship is negative. Finally, homeownership is mainly statistically significant (positively) in the northwest of the city.

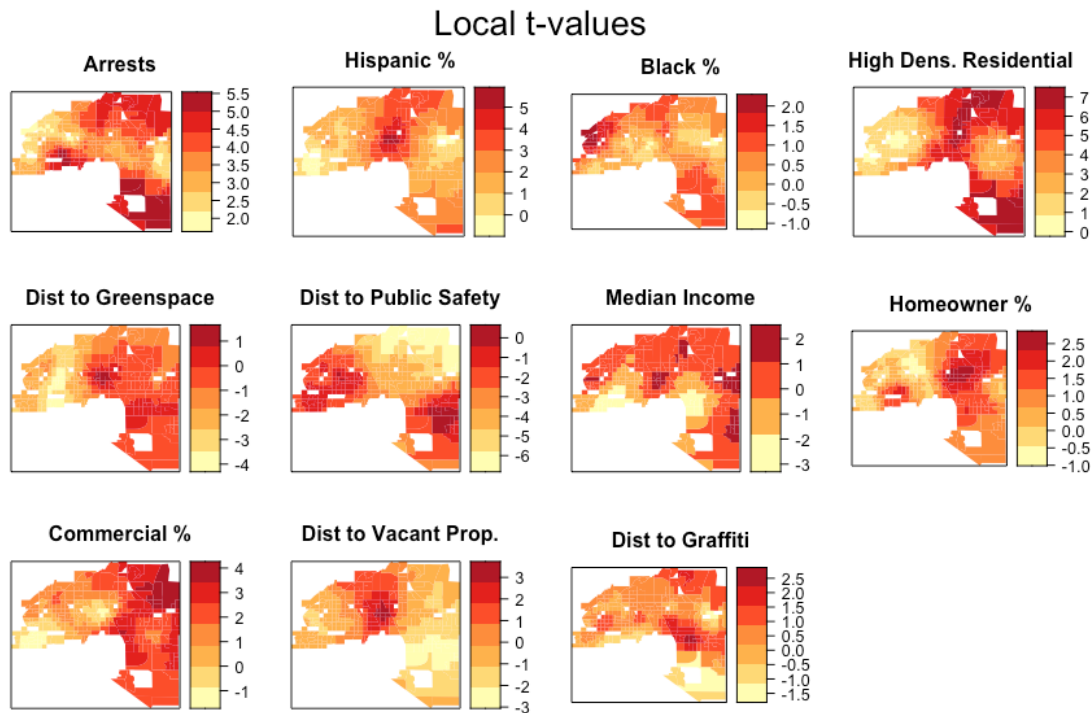


Figure 30: Pseudo t-values for each of the risk factors

Conclusion: Discussion of Modelling Results

In this study, we explored risk factors associated with overdose counts in Mesa, Arizona. We tried an OLS model, Poisson model, spatial lag model, spatial error model, spatial durbin model, geographically weighted regression, and geographically weighted poisson regression. While our Moran's tests suggested no significant spatial autocorrelation for our models, the mapping of the residuals of our initial models indicated higher residuals in the central and eastern part of the city, leading us to make increasingly more complex spatial models and finally ending with geographically weighted regression models.

After evaluating the results of the various models, we find that the geographically weighted poisson regression model was the best fit for our study. This model was the most applicable given that poisson regressions are a better fit for count data such as the number of opioid overdoses, and the fact that the overdose cases differed greatly in the eastern and central areas compared to other parts of Mesa. For this reason, a geographically weighted regression that created different models for each part of the city was the most valuable in understanding how different risk factors affect different neighborhoods of the city in distinct ways. Our final model had a pseudo-r-squared result of 0.69, meaning we are able to explain almost 70% of the variation in the data, although the amount of variation explained varies greatly based on the census block group.

Based on the results of this model, we developed suggestions for the City of Mesa as it seeks to understand factors associated with overdose incidents. Firstly, we encourage the government to focus on supporting the neighborhoods with high populations of minority groups, namely black and Hispanic residents. Specifically, the city should look into these populations in the western and central parts of the city respectively, considering these places had the largest associations between race/ethnicity and overdose cases. Similarly, in the eastern and west-central part of the city where median income is negatively correlated with overdose counts, the city should take preventive measures by aiding lower income families. In the northwest of the city, the government should be mindful of residents who might be burdened by high costs of living as homeownership is highly positively correlated with overdoses in this area. While we are not sure about the cause of this relationship, it is possible that the higher homeownership is linked to the fact that various areas of the northwest have higher incomes and are zoned for larger single-family homes, which could be creating a cost-of-living burden for some residents.

Furthermore, the government should be proactive about interventions in neighborhoods with high crime, given the number of arrests is highly correlated with overdose counts in all census block groups. This is not a call for additional policing. On the contrary, the fact that the distance to public safety is mainly negatively associated with overdoses points to the limitations of the public safety and policing system. Instead, we suggest preventive interventions in areas where crime is high or increasing. Finally, we noted that we believe that the reason distance to public greenspace has a significant negative relationship with overdose incidents in the west of the city is because most of the greenspace owned by the city is near the city center. Consequently, we suggest that the government should prioritize towards making the city center safer, especially establishing more resources to develop a more secure and inclusive downtown.

Limitations and Further Work

One major limitation in our findings is the use of opioid overdose cases as a way of measuring opioid overdoses. Given that these overdose cases only account for those reported or identified by the municipality, there is always potential for overdose counts to be an underestimation of the true level of influence of this on the population of Mesa. However, given how difficult it would be to measure the true level of opioid overdoses city wide, using this dataset was a viable alternative in the context of our study (assuming the level of unidentified overdoses was not spatially autocorrelated).

Secondly, our data set consisted of overdose cases from 2018 to 2020 since we aimed to use the most up-to-date data matching with the latest US Census reports (2020). Because the COVID-19 pandemic which corresponded to national spikes in opioid usage falls within our selected time frame, overdose cases could have been affected by various additional factors unaccounted for. Depending on the extent to which opioid usage returns to its previous rates, our findings might be overstating areas that had higher number of cases during the pandemic, but were not normally as vulnerable to opioid overdoses.

While our study aimed at informing the municipality about general trends in risk factors associated with overdose incidents, further work could extend our research to provide a temporal context. For example, a spatio-temporal analysis of incidents could investigate whether overdose spikes are seasonal, and if they occur at different periods for different neighborhoods. While the overall number of cases in Mesa has been increasing, it would also be helpful to examine if the cases are rising in only specific neighborhoods while they are falling in others. Finally, given the big effect of the pandemic, it would serve to explore how the correlations between the risk factors and overdose counts changed before and after the shutdowns as these findings can prepare the city for targeted interventions to take during future crisis situations.

References

- Adams J. F. A. (1889), "Substitutes for opium in chronic diseases", Boston Medical and Surgical Journal, 121, 15, 351-356.
- Arizona Department of Health Services. Opioid Prevention [Internet]. 2022 [cited 2022 Dec 1] Available from: <https://www.azdhs.gov/opioid/>
- Atikah N., Widodo B., Rahardjo S., Mardlijah, Kholifia N. and Afifah D. L. (2021) "The efficiency of Spatial Durbin Model (SDM) parameters estimation on advertisement tax revenue in Malang City", Journal of Physics: Conference Series, 1821, 012012.
- Brinkley-Rubinstein L., A. Macmadu, B. D. L. Marshall, A. Heise, S. I. Ranapurwala, J. D. Rich, T. C. Green (2018) "Risk of fentanyl-involved overdose among those with past year incarceration: Findings from a recent outbreak in 2014 and 2015", Drug and Alcohol Dependence, 185 (2018), 189-191.
- Burns J. M., R. F. Martyres, D. Clode and J. M. Boldero (2004) "Overdose in young people using heroin: Associations with mental health, prescription drug use and personal circumstances", Medical Journal of Australia, 128, 7, 25-28.
- City of Mesa (2020) Zoning Districts, Available from: <https://data.mesaaz.gov/Zoning-Property/Zoning-Districts/qscf-6ebm>
- City of Mesa (2022) City Owned Property, Available from: <https://data.mesaaz.gov/Engineering/City-Owned-Property/xms2-ya86>
- City of Mesa (2023) Fire and Medical Opioid Overdose Incidents, Available from: <https://data.mesaaz.gov/Fire-and-Medical/Fire-and-Medical-Opioid-Overdose-Incidents/qufy-tzv6>
- City of Mesa (2023) Police Incidents, Available from: <https://data.mesaaz.gov/Police/Police-Incidents/39rt-2rfj>
- City of Mesa (2023) Transportation Graffiti, Available from: <https://data.mesaaz.gov/Transportation/Transportation-Graffiti/9spb-749m>
- Choi J. I., Lee J, Yeh A. B., Lan Q and Kang H, (2022) "Spatial clustering of heroin-related overdose incidents: a case study in Cincinnati, Ohio", BMC Public Health, 22, 1253, Available from: <https://doi.org/10.1186/s12889-022-13557-3>
- Draanen J. V., C. Tsang, S. Mitra, M. Karamouzian, and L. Richardson (2020), "Socioeconomic marginalisation and opioid-related overdose: A systematic review", Drug and Alcohol Dependence, 214 (2020), 108-127.
- Burchfield E. (2020) Spatial Regression, Available from: https://www.emilyburchfield.org/courses/gsa/spatial_regression_lab
- Fotheringham A. S., Brunsdon C. and Charlton M., (2003), Geographically weighted regression: the analysis of spatially varying relationships, John Wiley & Sons.

Gollini I., Lu B., Charlton M., Brunsdon C., and Harris P. (2013), "GWmodel: an R Package for Exploring Spatial Heterogeneity using Geographically Weighted Models", arXiv, Available from: <https://arxiv.org/abs/1306.0413>

Gollini I., Charlton M., Brunsdon C., and Harris P. (2015), "GWmodel: an R Package for Exploring Spatial Heterogeneity using Geographically Weighted Models", *Journal of Statistical Software*, 63(17).

Guy G. P., K. Zhang, M. K. Bohm, J. Losby, B. Lewis, R. Young, L. B. Murphy and D. Dowell (2017) "Vital signs: Changes in opioid prescribing in the United States, 2006-2015", *Morbidity and Mortality Weekly Report*, 66, 26, 697-704.

Hollingsworth A., C. J. Ruhm and K. Simon (2017) "Macroeconomic conditions and opioid abuse", *Journal of Health Economics*, 56 (2017), 223-233.

Kosten T. R. and T. P. George (2002) "The neurobiology of opioid dependence: Implications for treatment", *Science and Practice Perspectives*, 1, 1, 13-20.

Krausz R. M., J. N. Westenberg and K. Ziafat (2021) "The opioid overdose crisis as a global health challenge", *Current Opinion in Psychiatry*, 34, 4, 405-412.

Lee H., S. K. Ghosh (2009) "Performance of information criteria for spatial models", *Journal of Statistical Computation and Simulation*, 79, 1, 93-106.

Lippold K. M., Jones C. M., Olsen E. O. and Giroir B. P. (2019) "Racial/Ethnic and Age Group Differences in Opioid and Synthetic Opioid-Involved Overdose Deaths Among Adults Aged ≥ 18 Years in Metropolitan Areas - United States, 2015-2017", *MMWR Morb Mortal Wkly Rep*, 68, 4, 967-73. doi: 10.15585/mmwr.mm6843a3.

Li Z. R., E. Xie, F. W. Crawford, J. L. Warren, K. McConnell, J. T. Copple, T. Johnson and G. S. Gonsalves (2019) "Suspected heroin-related overdose incidents in Cincinnati, Ohio: A spatiotemporal analysis", *PLoS Medicine*, 16, 11, 1-15.

Lyden J. and I. A. Binswanger (2019) "The United States opioid epidemic", *Seminars in Perinatology*, 43 (2019), 123-131.

Marshall J. R. S. F. Gassner, C. L. Anderson, R. J. Cooper, S. Lotfipour and B. Chakravarthy (2019) "Socioeconomic and geographical disparities in prescription and illicit opioid-related overdose deaths in Orange County, California, from 2010-2014", *Substance Abuse*, 40, 1, 80-86.

Montandon G. and A. S. Slutsky (2019), "Solving the opioid crisis; Respiratory depression by opioids as critical end point", *Chest*, 156, 4, 653-658.

Office of Governor Ducey. Governor Ducey Declares Statewide Health Emergency In Opioid Epidemic [Internet]. 2017 [cited 2022 Dec 1] Available from: <https://azgovernor.gov/governor/news/2017/06/governor-ducey-declares-statewide-health-emergency-opioid-epidemic>

Poliart, A., Kirakoya-Samadoulougou, F., Ouédraogo, M. (et al. (2021), "Using geographically weighted Poisson regression to examine the association between socioeconomic factors and hysterectomy incidence in Wallonia, Belgium." BMC Women's Health 21, 373.

Rosenblum A., L. A. Marsch, H. Joseph and R. K. Portenoy (2008), Opioids and the treatment of chronic pain: Controversies, current status and future directions", Experimental and Clinical Psychopharmacology

Rosner B., J. Neicun, J. C. Yang and A. Roman-Urrestarazu (2019) "Opioid prescription patterns in Germany and the global opioid epidemic: Systematic review of available evidence", PLoS One, 14, 8, 1-20.

Steif, Ken. Public Policy Analytics: Code & Context for Data Science in Government [Internet]. 2021.[cited 2022 Dec 1] Available from:
<https://urbanspatial.github.io/PublicPolicyAnalytics/index.html>

United States Census Bureau (2022) American Community Survey (ACS), Available from:
<https://www.census.gov/programs-surveys/acs>

Ward M. and K. Gleditsch (2008), Spatial Regression Models, SAGE Publications.

Waldhoer M., S. E. Bartlett and J. L. Whistler (2004) "Opioid receptors", Annual Review of Biochemistry, 73 (2004), 953-990.

Zuur A. F., E. N. Leno and G. M. Smith (2007) Analysing Ecological Data: Statistics for Biology and Health, Springer: New York.