

Predicting child maltreatment risk in Richmond, VA: An open source framework

*Ken Steif, Ph.D,
Matthew D. Harris,
Sydney Goldstein, M.C.P.*

March 6, 2019

*Urban Spatial - [Website](#) / [Other Work](#)
Predict Align Prevent - [Website](#)*

1. Executive Summary

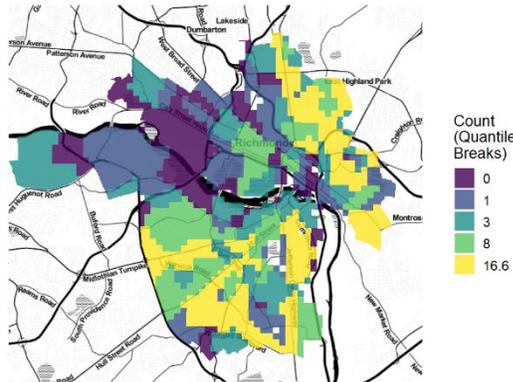
Motivation

There are approximately 40,000 children in Richmond, VA, and between July 2013 and July 2017 there were 6,500 accepted cases of child maltreatment.¹ Figure 1.1a shows the count of maltreatment events between 2013 and 2017 by neighborhood. Not all neighborhoods are affected by maltreatment in the same manner, as neighborhoods in the southwest and northeast of Richmond experience much higher counts of Child Protective Service (CPS) events. Can we do a better job ensuring that our limited child welfare resources are properly deployed to the communities where they are needed most?

To understand which communities have the greatest need, we must first determine where child maltreatment is likely to occur. There are a host of individual, family, and household level factors associated with child maltreatment, and research shows that community and social factors play an important role in understanding where maltreatment may occur.² At high densities, externalities associated with 'neighborhood effects' can influence peers in close proximity.³

These externalities appear as maltreatment event clusters, as visualized in Figure 1.1b, which maps the rate of child maltreatment events in Richmond, VA between 2013 and 2017.⁴ Recent work shows that variation in these spatial clusters can be predicted by environmental factors such as crime, blight, and nuisance land uses like bars and restaurants.⁵

Number of Child Protective Service events by neighborhood



Child Protective Service rate per 100 people

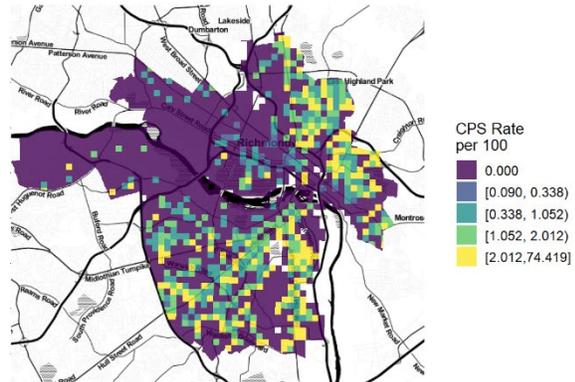


Figure 1.1

There are two goals of the work presented here. First, to create a comprehensive open source framework for developing a child maltreatment predictive model to estimate maltreatment risk across Richmond.

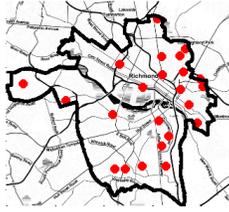
The intuition for the model is to borrow the observed maltreatment 'experience' in Richmond and test whether those experiences are generalizable to places where maltreatment may occur but is not observed directly. If so, it is reasonable to confidently forecast maltreatment risk across space.

Second, to document a strategic planning process for converting maltreatment risk predictions into actionable intelligence stakeholders can use to efficiently allocate limited child welfare resources. In this 'Align' phase of the framework, we assess the extent to which the current supply of child welfare services is properly distributed relative to the demand for these services.

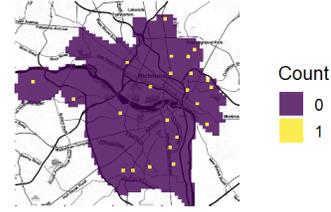
Modeling

The predictive model is based on this notion of spatial externalities, positing that maltreatment risk is a function of exposure to certain environmental factors like crime and blight. Point data describing these environmental phenomena are aggregated to a regular lattice grid and engineered into a set of spatial 'features' from which predictive relationships can be mined. Figure 1.2 provides an example of the feature engineering process, mapping community center locations and recoding these locations so that every grid cell citywide receives three complementary measures of geographic 'exposure'.

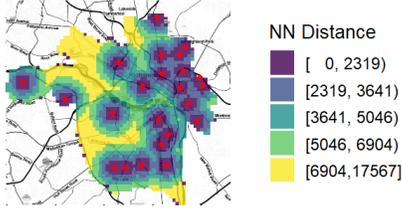
Community center locations



Community center count by fishnet



Community center euclidean distance



Community center average nearest neighbor distance

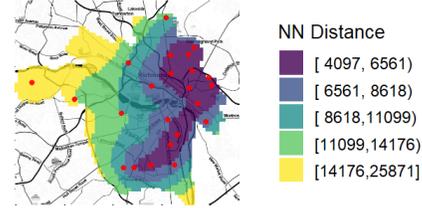


Figure 1.2

Prior to building any models, maltreatment events in Richmond are explored using various methods. We begin by analyzing maltreatment clusters to better understand the relevant spatial process. Figure 1.3 illustrates the location of statistically significant maltreatment event clusters across Richmond. These are locations where maltreatment occurs in higher densities than we might expect due to random chance alone.

Statically significant maltreatment clusters

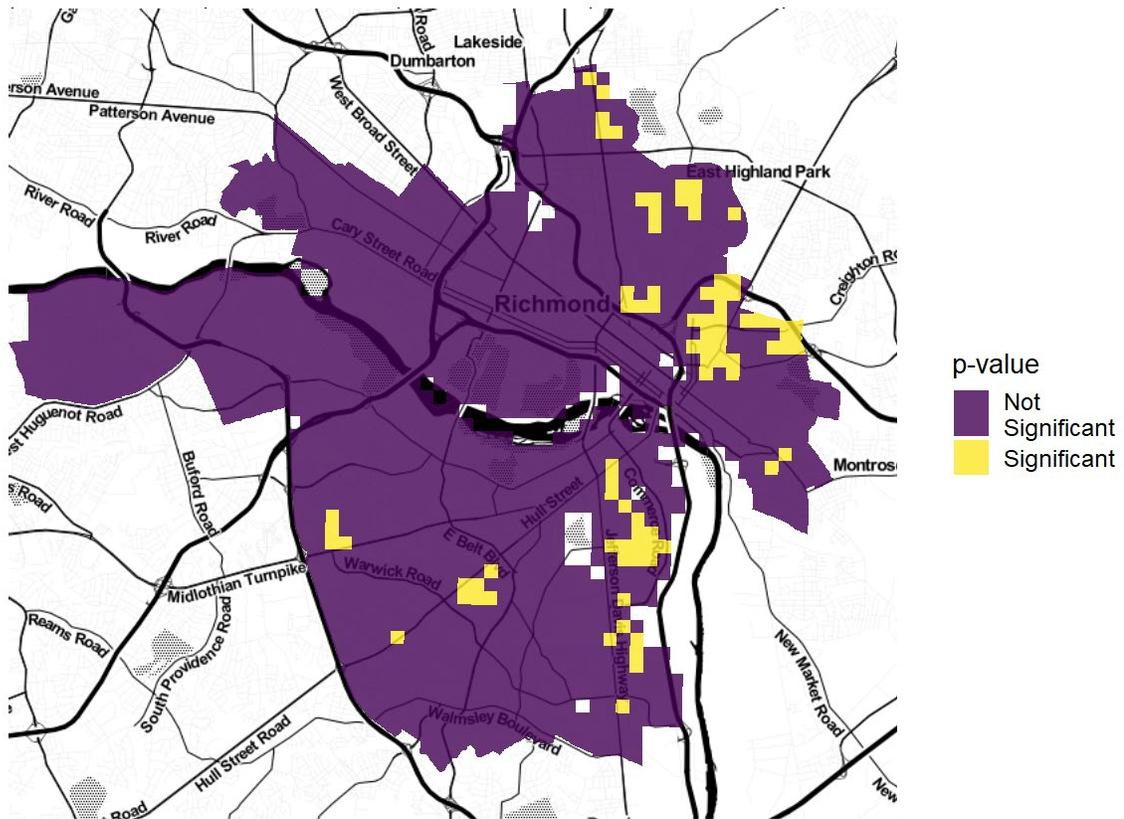


Figure 1.3

We also analyze the extent of colocation between maltreatment and measures of exposure to determine whether a given factor is 'closer' to a maltreatment event than what we might otherwise presume due to random chance alone. This is a useful, policy-centric approach for understanding how phenomena are associated with maltreatment across space.

For example, Figure 1.4 ranks crime-related factors by their average distance to maltreatment. We provide more statistical intuition for this test in Section 5, however for now, it is interesting to point out that several domestic-oriented crimes occur in close proximity to maltreatment.

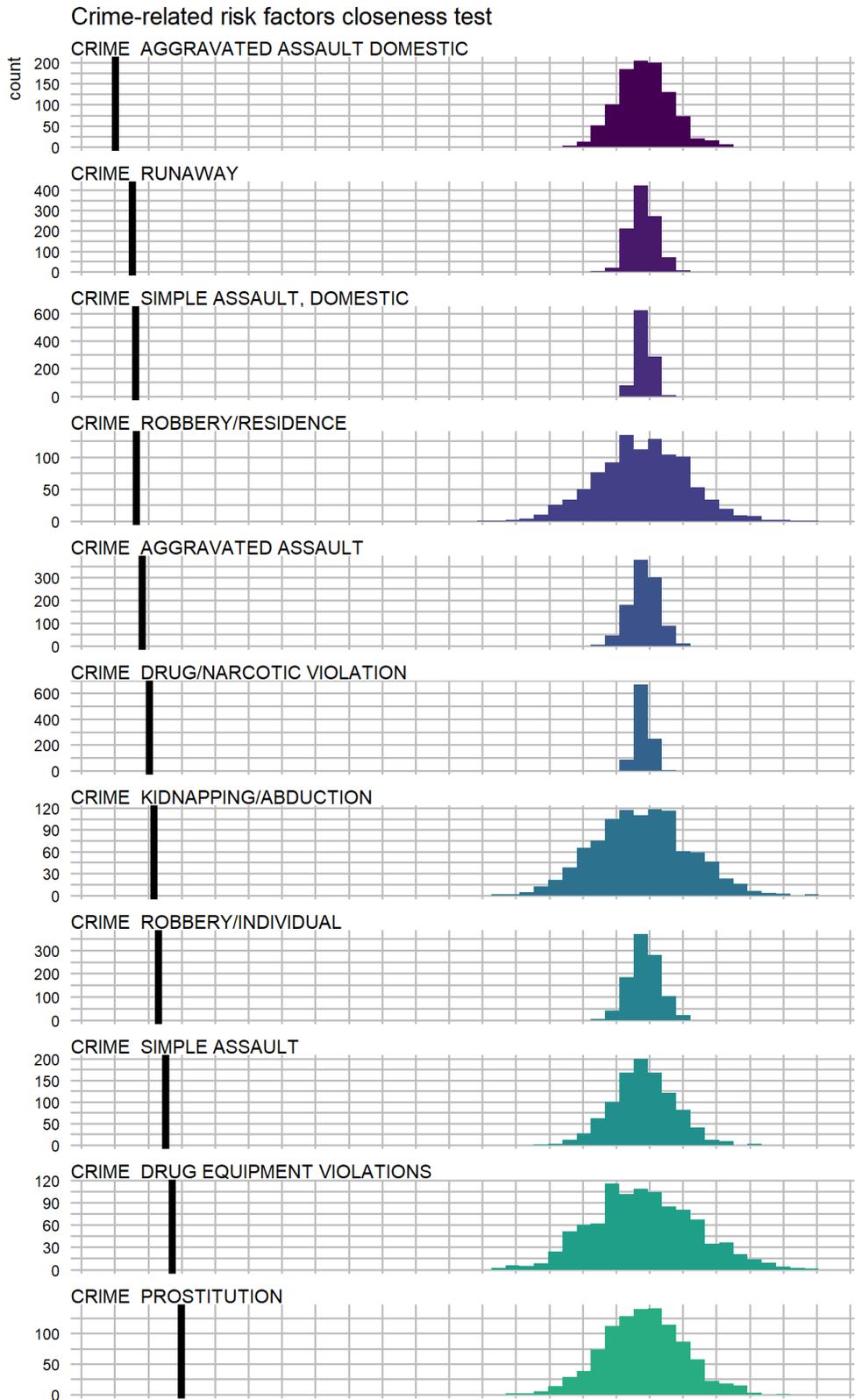


Figure 1.4 (continued below)

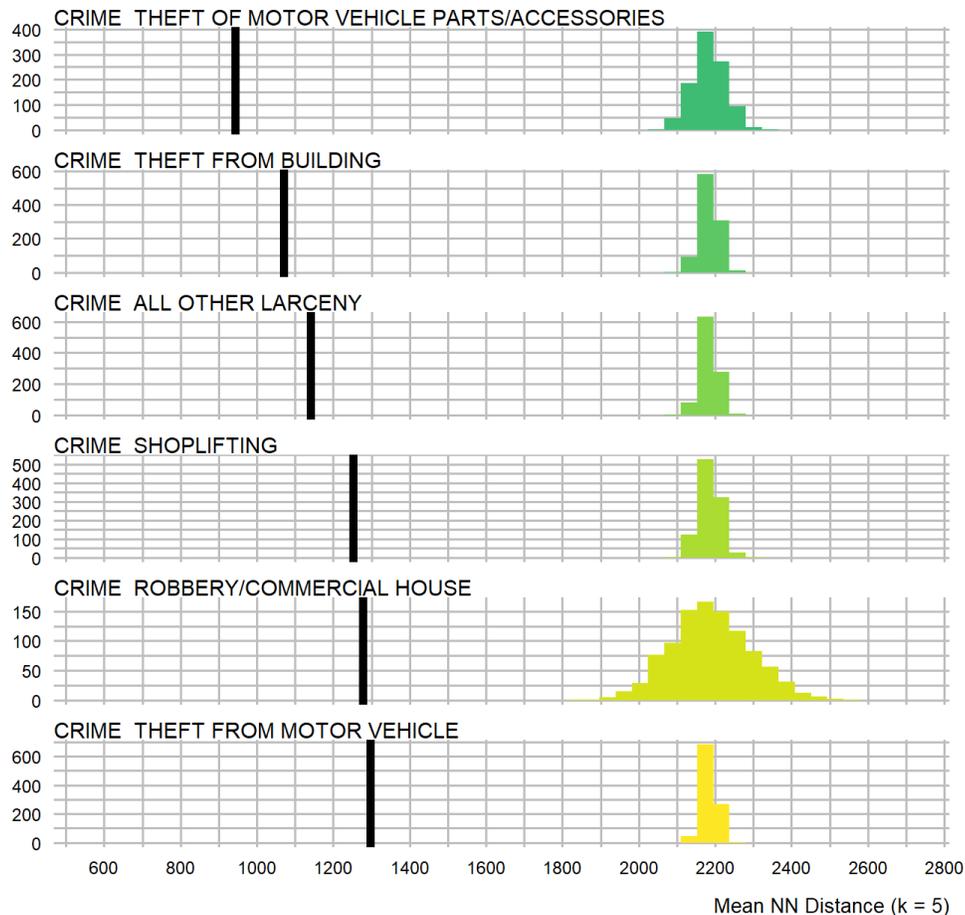


Figure 1.4 (continued)

Once we understand the patterns within the maltreatment data, several linear and non-linear machine learning algorithms are estimated. For each model, a set of 'goodness of fit' metrics are recorded that describe the tradeoff between model accuracy and generalizability across space.

At the heart of the analysis is the tradeoff between accuracy and generalizability. A model that is perfectly accurate will not reveal areas where children may be at risk despite a lack of reported maltreatment. A model that is perfectly generalizable may fail to provide targeted intelligence to identify areas at serious risk for maltreatment.

Accuracy is defined as the model's ability to minimize the difference between the observed count of maltreatment events and the predicted count. These differences are referred to as 'errors' and significant attention is paid to their nuances in Section 4 and Section 6. Generalizability refers to the model's ability to make comparable predictions across neighborhoods regardless of differences in factors like income and race. Several of these 'cross-validation' metrics are calculated, visualized, and explained in Section 6.

Results

By focusing on the tradeoff between accuracy and generalizability, the final predictive model produces highly targeted spatial risk predictions. These risk predictions are split into five risk categories - 5 being the greatest risk. Figure 1.5 maps these categories and overlays a set of maltreatment events that were withheld from the model for the purposes of validation. For privacy purposes, some cartographic information is redacted.⁶ The map suggests that our model predicts well, evidenced by the many holdout maltreatment events that fall in the highest predicted risk category.

Risk categories from meta-model

Privacy Controls: Maltreatment events in grid cells with 1 point are masked; Remaining event locations are offset at random.

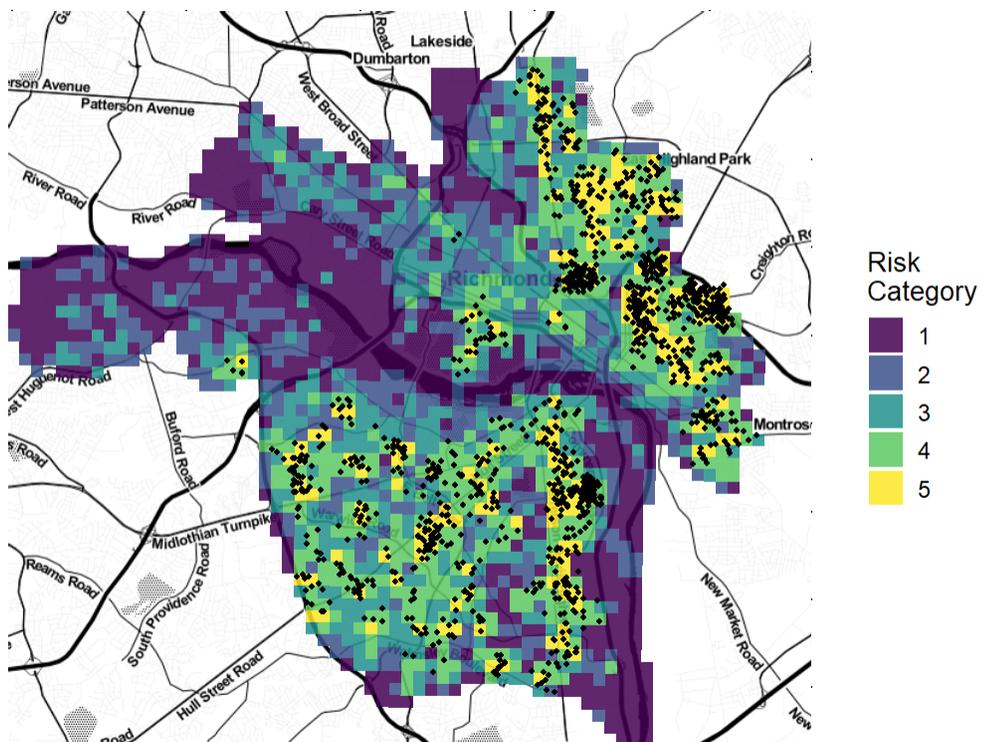


Figure 1.5

This visual relationship is further formalized by comparing the model predictions to those derived from a simple Kernel Density Estimation (KDE) - a common tool used to convert point data into hot spot maps. The ubiquity of KDE as a tool for spatial targeting makes it a useful baseline to which model predictions can be compared.

Both the KDE and the risk predictions are split into five risk categories, and the proportion of holdout maltreatment events are calculated for each category. The plot below shows that for the highest risk category, the meta-model captures approximately 70% of the observed maltreatment events, whereas the KDE captures only about 35%. For context, approximately

8,200 children live in the high-risk category which represents about 21% of total children in Richmond, but only about 10% of the total area of the city.

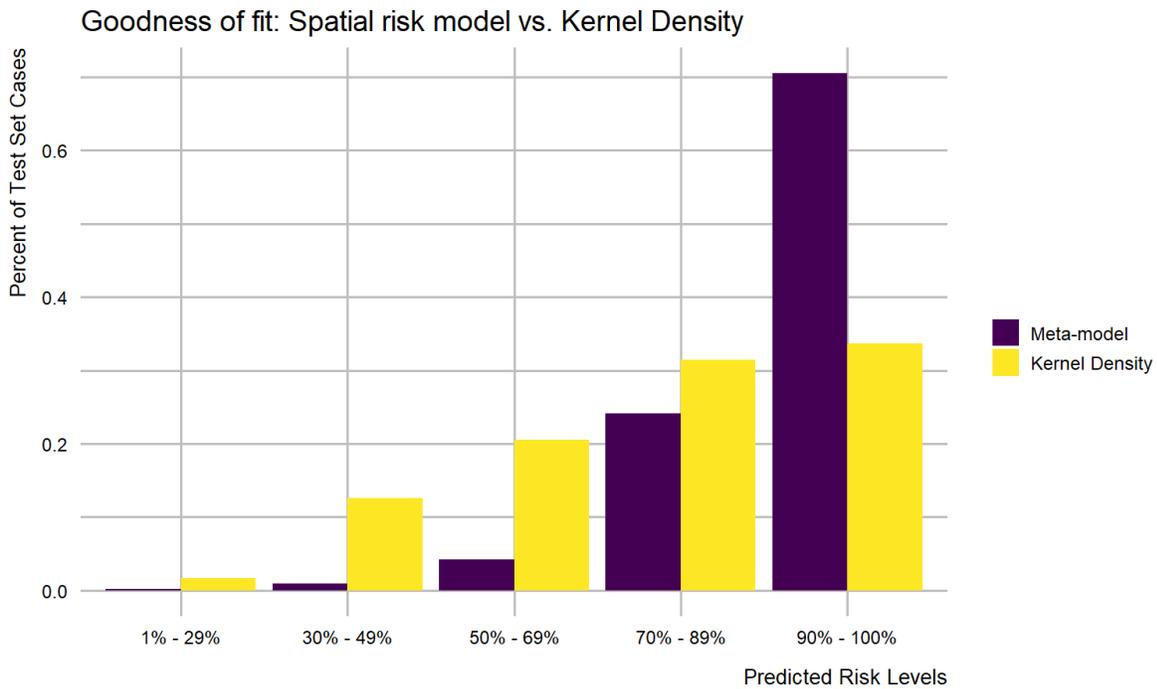


Figure 1.6

Other useful metrics are created to evaluate model performance. Our most useful predictive model has an average (absolute value) error of 0.533 of one maltreatment event. We believe this outcome is representative of a highly accurate model that is capable of producing policy-relevant risk predictions. However, it is difficult to evaluate model performance across communities with varying maltreatment rates using only this metric.

Rather, Figure 1.7 displays a simple evaluative measure for the final model - visualizing predicted maltreatment counts as a function of observed counts. The black line that intersects the point cloud represents a perfect prediction, while the blue line represents actual predictions. The model is accurate for most of the data (i.e. most of the city), but loses some fidelity in areas with very high maltreatment event counts.

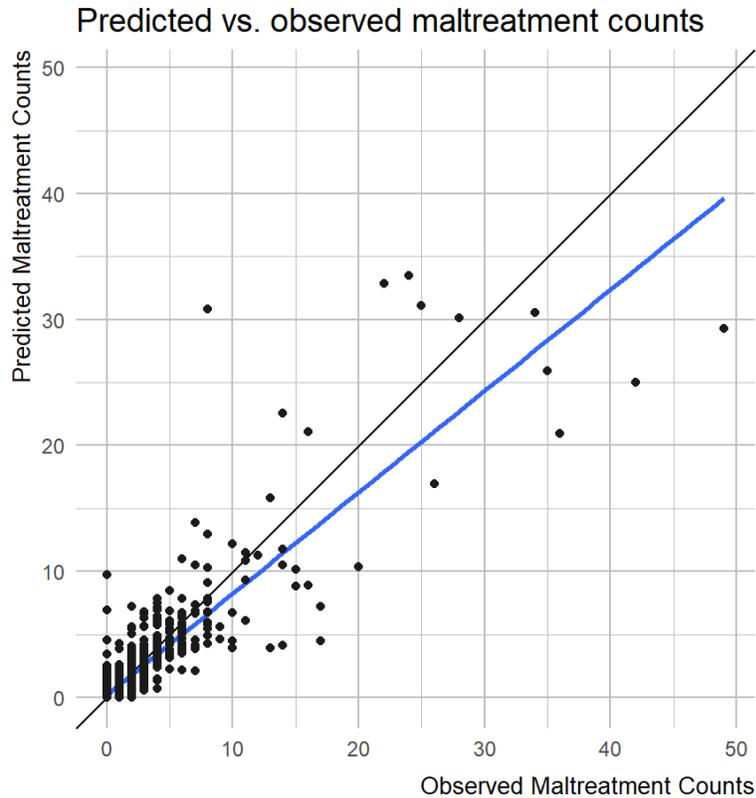


Figure 1.7

We also develop several algorithmic fairness metrics to describe generalizability across space. Despite the model's ability to rightfully predict high maltreatment risk in areas with high occurrence rates, we do see relatively higher errors in these places. The variability of (absolute value) error across Richmond neighborhoods is 0.746 of one event.

Another way of judging how well the model generalizes across space is to ask how well it performs in neighborhoods of varying typologies. We test the model looking for relative differences across high and low poverty rates and high and low minority rates, respectively. If the model generalizes well to these places, we should see similar accuracies across the varying typologies.

Results suggest that the model predicts equally well across neighborhoods of varying poverty levels. However, these metrics also indicate that the model predicts less accurately in certain minority neighborhoods, particularly those with disproportionately large maltreatment counts. The nature of this finding is discussed in detail in Section 6, and additional approaches for adding more equity into the model are proposed. An independent ethicist has reviewed the approach and found that the framework delivers 'genuine benefits while avoiding some of the familiar risks of alternative approaches to targeting child protection services.'

Align

Using the knowledge from the predictive model, we develop a strategic planning framework, embedding the risk predictions into a larger analysis describing whether the supply of child welfare services is properly aligned with the demand for these services. Demand in an area is defined by the risk predictions. We define supply as the presence of protective land uses - places stakeholders could deploy education, treatment, and prevention resources.

One way to characterize alignment is to scale these indicators of supply and demand and relate them to each other at the neighborhood level. Figure 1.8 displays the resulting relationship. Light red and light green neighborhoods are those where there is a strong alignment between the supply of and demand for child welfare services. Dark green places are those where there are more protective resources than risk, and dark red areas are places where maltreatment risk outweighs the availability of protective resources.

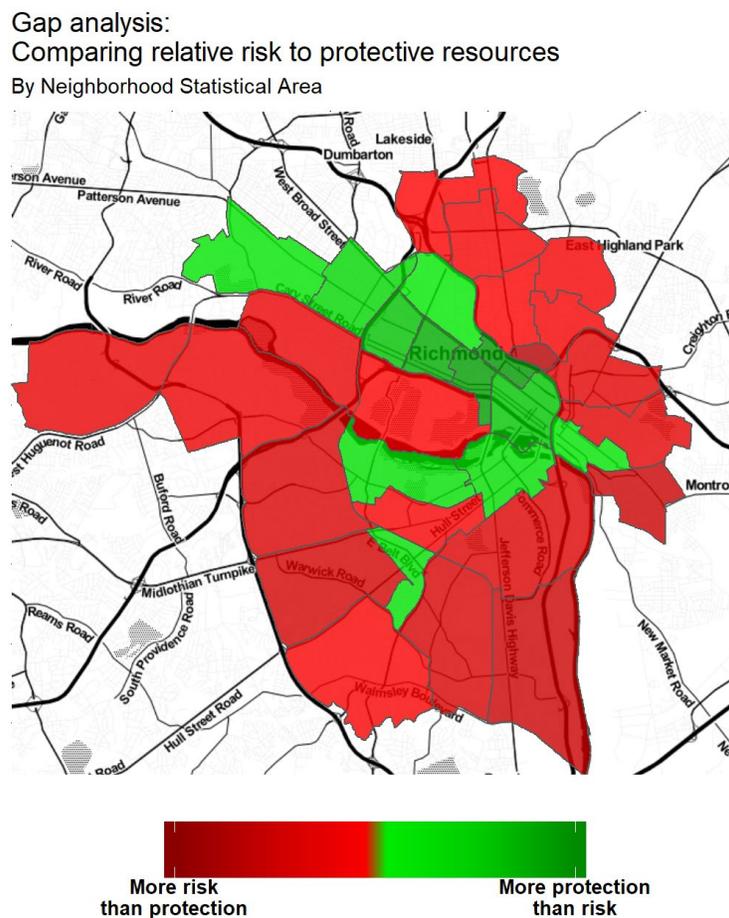


Figure 1.8

This method suggests that protective resources are not always in optimal locations. To get a clearer picture we examine Stop Child Abuse Now (SCAN) Centers (Figure 1.9). The SCAN [mission](#) is to 'prevent and treat child abuse and neglect throughout the Greater Richmond area by protecting children, promoting positive parenting, strengthening families and creating a community that values and cares for its children.' While this organization provides an invaluable

service, when mapped in the context of risk, it appears some of these centers could be better allocated to communities at greater risk for maltreatment.

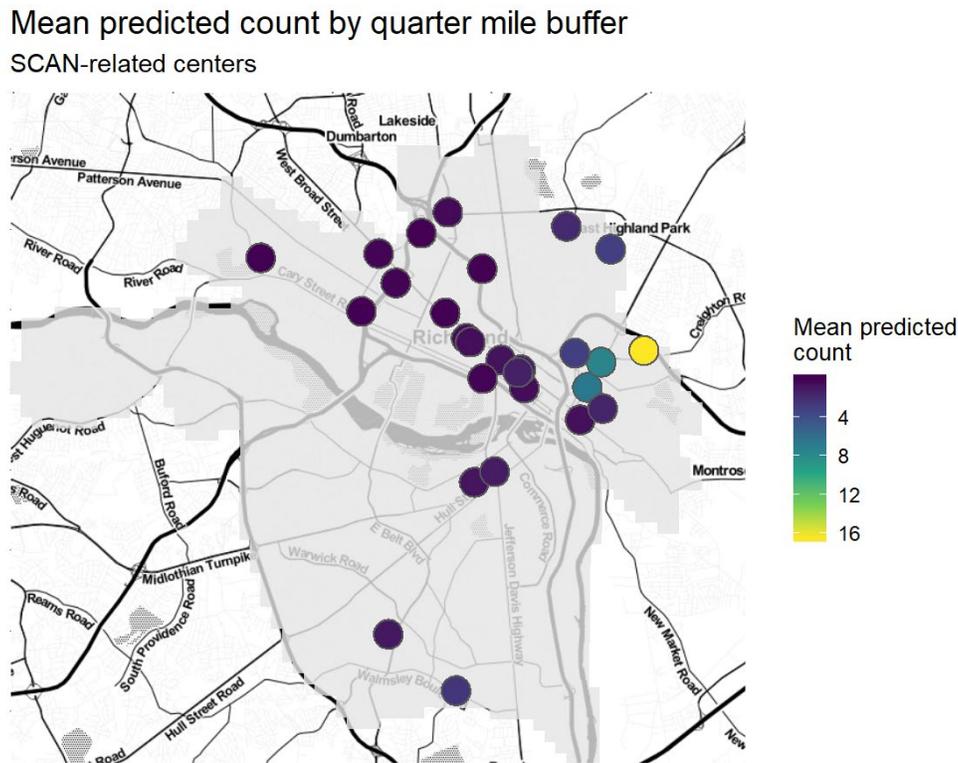


Figure 1.9

The next section provides and discusses more of these Align analytics, which stakeholders can easily develop by bringing a digital map file of risk predictions into a standard desktop GIS system. We believe this framework can have important impacts on the child welfare strategic planning process in Richmond.

In the sections to follow, we delve deeply into the spatial risk predictive model developed for this project. We provide some motivation, contrasting predictive modeling to traditional research, and present a very robust methods section situating our framework in the literature on spatial analysis and spatial prediction. Next, we discuss our exploratory analysis and finally, results are examined at length.

2. Align

This work extends beyond the creation of an algorithm for predicting maltreatment risk, toward a workflow for generating actionable intelligence. The purpose of this phase is to harness the predictions generated above in the context of strategic planning.

Specifically, the goal is to understand the extent to which the supply of child welfare services are properly aligned with the demand for child welfare services. To define supply, a host of

administrative datasets that either proxy or directly relate to service delivery are employed. To define demand, maltreatment risk predictions are used.

Aside from direct child welfare provisions like social services, other important service delivery types include community resources such as churches and recreation centers. The purpose of the Align phase in part, is to identify resources that are optimally located relative to maltreatment risk. It is at these locations where stakeholders may wish to deploy education, outreach, and treatment programs.

The reader should consider the analysis outlined below to be a framework that can be adopted and replicated by both public and non-profit stakeholders. All one needs is a digital map file (i.e. shapefile) of risk predictions and the various supply-side oriented datasets.

The subheadings below outline different approaches that we feel could be useful for the planning process. We report analytics that:

1. Look at the number of people living in areas of high maltreatment risk
2. Explore the relationship between poverty rate and predicted maltreatment risk
3. Examine correlations between risk and important public health outcomes like asthma and obesity
4. Explore the spatial relationship between child removals and predicted maltreatment risk
5. Examine the spatial relationship between child fatalities and predicted maltreatment risk
6. Assign risk scores to protective land uses (churches, rec centers etc.) to get a better sense of existing resources in the community
7. Examine whether current social worker home visits are situated in a high risk places
8. Explore maltreatment risk around Stop Child Abuse Now (SCAN) network centers
9. Perform a gap analysis to understand where and to what extent the supply of child welfare services co-locate with the demand for those services

Risk category population totals

The demand for child welfare services is related to the number of people living in high risk areas. Figure 2.1 shows that approximately 48,500 people live in the highest risk category, an area which comprises about 10% of the city. Another 30% of the population lives in the second highest risk category, indicating, in total, over 50% of the population live in areas of potentially high demand for child welfare services.

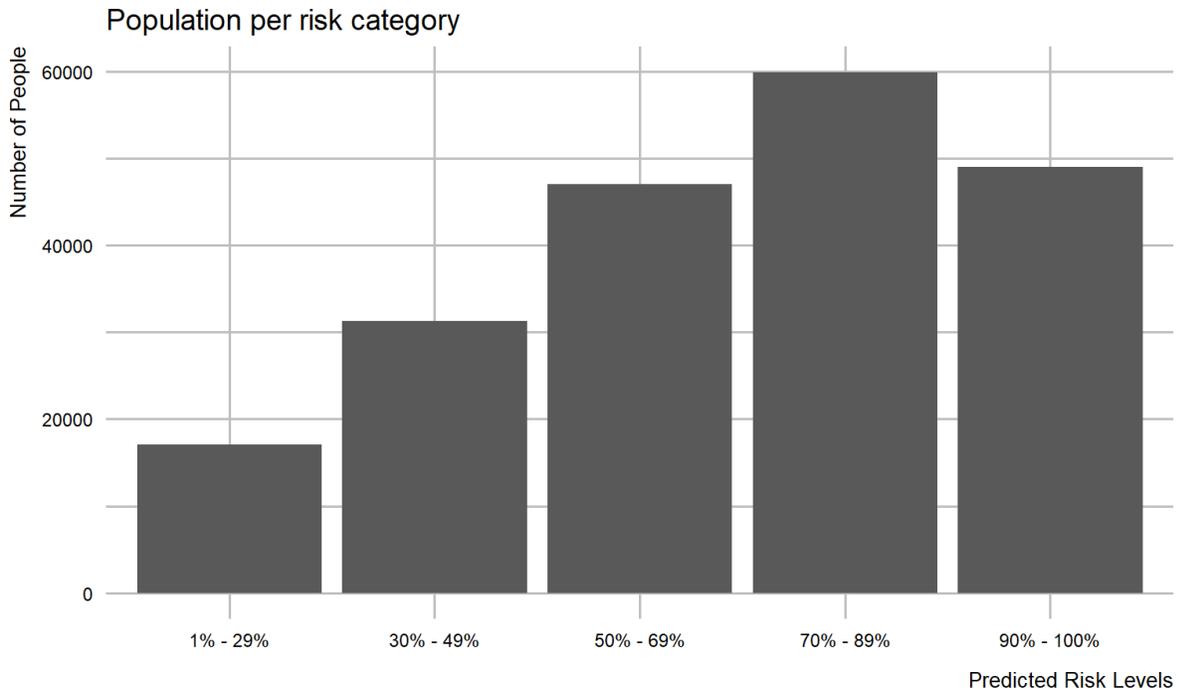


Figure 2.1

Is poverty related to predicted maltreatment events?

Figure 2.2 maps the weighted poverty rate by fishnet grid cell. How does the distribution of poverty relate to maltreatment events?

Weighted poverty rate

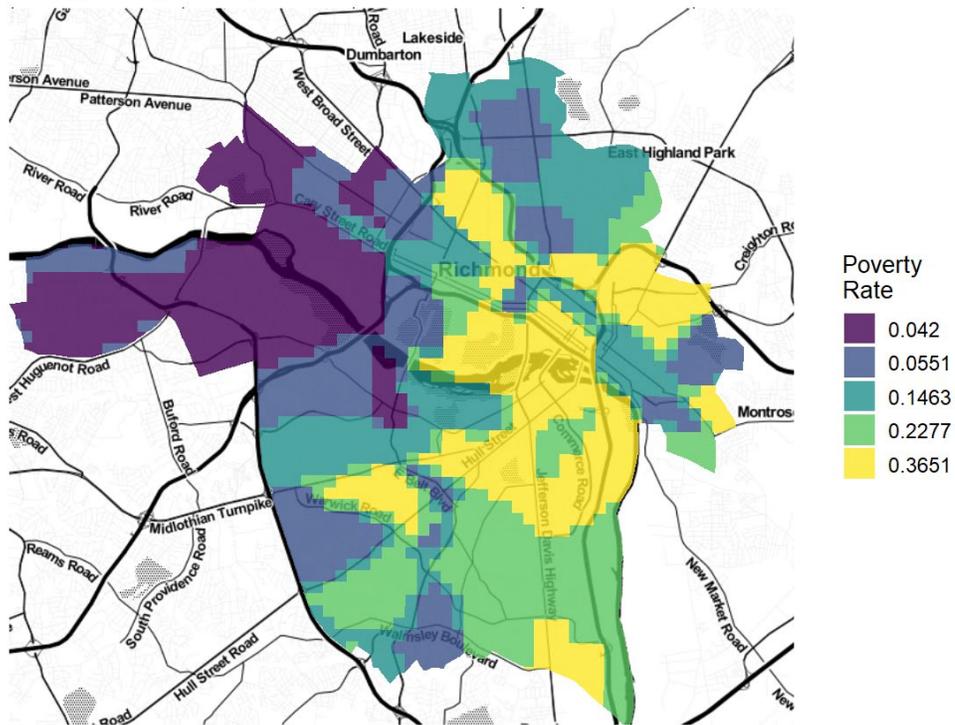


Figure 2.2

Figure 2.3 illustrates the relationship between poverty rate and predicted maltreatment count. The scatter plot shows that the correlation between poverty and predictive risk is marginal. This visual relationship is confirmed by a correlation coefficient of 0.29. The weak relationship persists even when the zero count grid cells are removed.

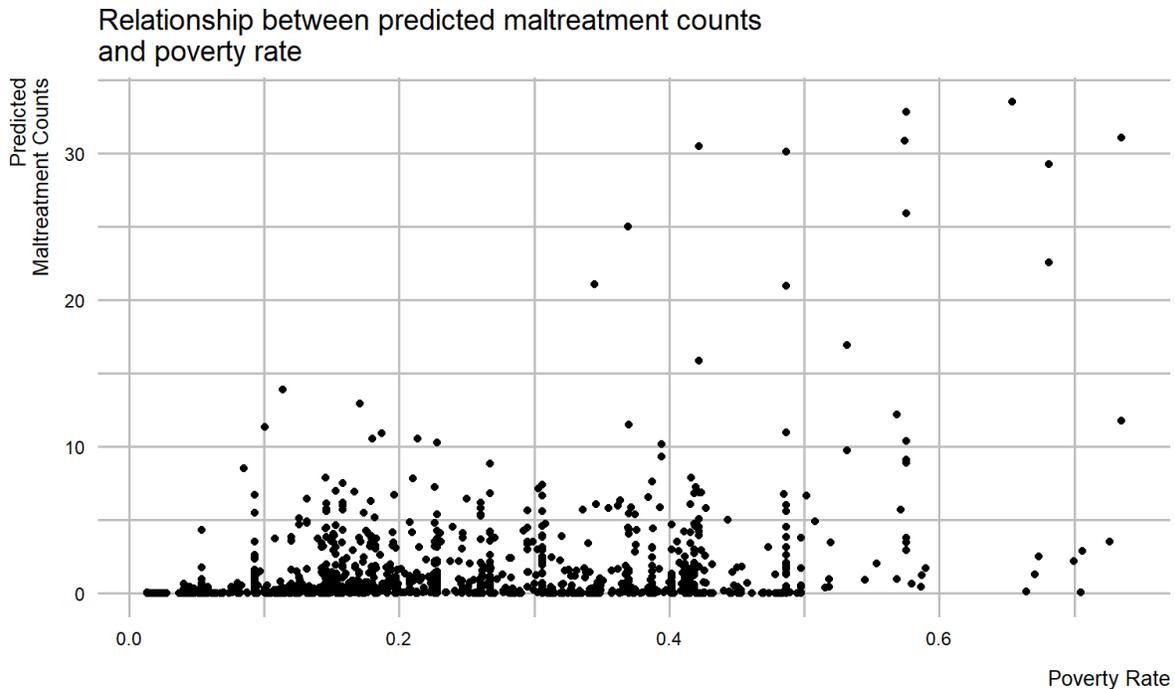


Figure 2.3

Maltreatment risk and health outcomes

Studies suggest that maltreatment is correlated with a host of other health outcomes like asthma and obesity. While we don't currently have data in hand on rates of these other health outcomes, it would be interesting to test whether maltreatment risk is correlated with these other health outcomes across space. If they are, we might be able to use maltreatment risk as a way to target interventions associated with these other outcomes.

Maltreatment risk and removals of a child from the household

About 300 removal events occur in Richmond during the study period. For privacy reasons the map of risk predictions and removal events is redacted. Figure 2.4 aggregates these data and illustrates that an overwhelming proportion of removals occur in the highest risk category.

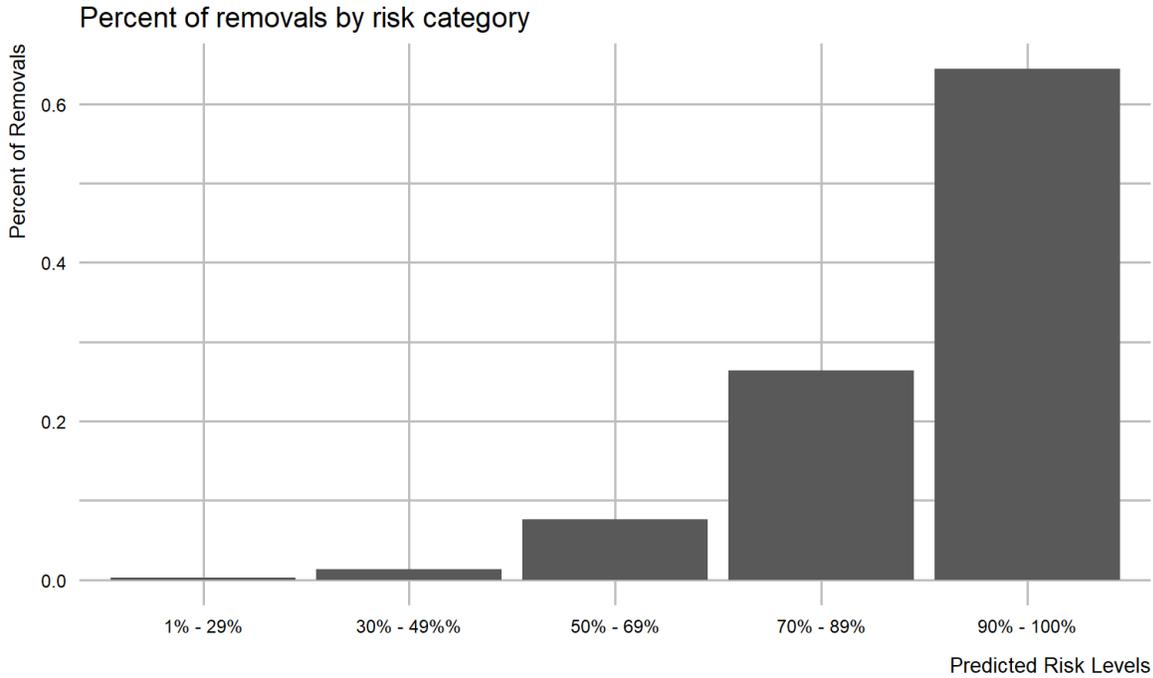


Figure 2.4

Maltreatment risk and child fatalities

Just 24 child fatality events occur in Richmond during the study period. For privacy reasons the map of risk predictions and fatalities is redacted. Figure 2.5 aggregates these data and illustrates that the majority of fatalities occur in the highest risk category.

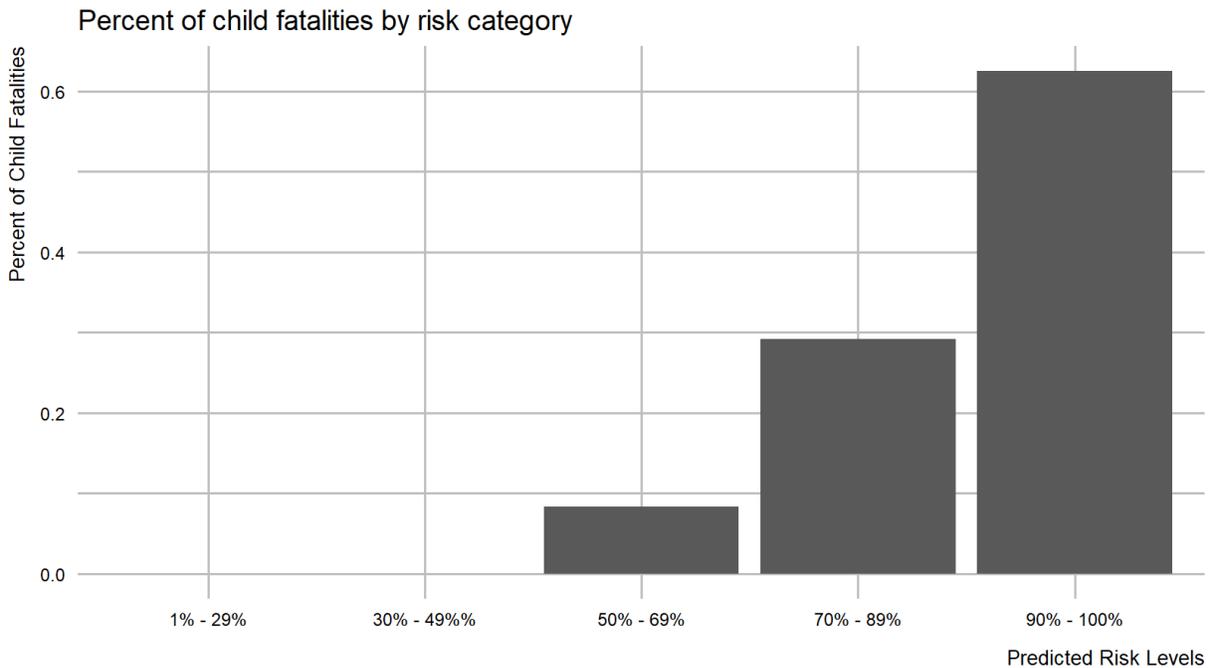


Figure 2.5

Assign risk scores to protective land uses

The maltreatment predictions suggest where education, outreach, and prevention efforts should occur. What resources are available at these locations? We answer this question using some of the original protective factors data gathered for the predictive model. Stakeholders can replicate this approach on a more finite list of sites that could host these interventions. Figure 2.6 shows the distribution of protective land uses by type.

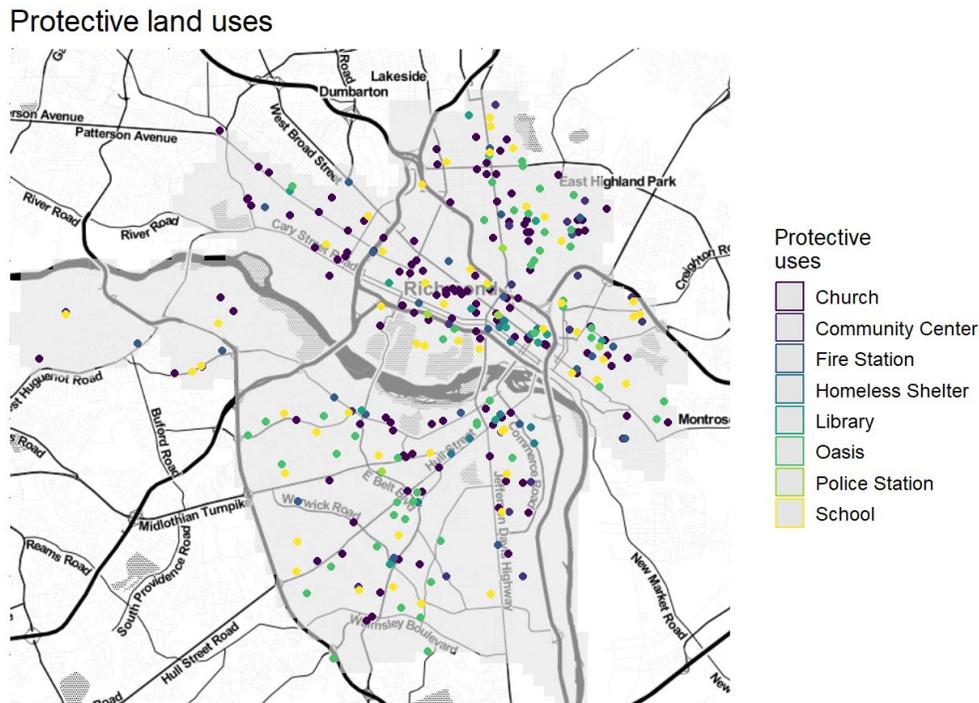


Figure 2.6

We then calculate a relative measure of risk exposure for each protective land use by drawing quarter mile buffers around each site and taking the mean count of predicted events. Figure 2.7 plots these buffers and the relative measure of risk exposure. The table that follows lists the top 20 individual protective locations sorted by type and mean predicted count.

Mean predicted count by quarter mile buffer
Protective uses



Figure 2.7

use	name	address	Mean Predicted Count
Church	Shalom Baptist Fellowship Church	None given	13
Church	Sixth Mount Zion Baptist Church	None given	13
Church	Ebenezer Baptist Church	None given	7
Community_Center	Fairfield Court Community Center/RRHA	None given	15
Community_Center	Creighton Court Community Center	None given	12
Community_Center	Calhoun Community Center and Playground	None given	12
Fire_Stations	Richmond Fire Station 5	None given	9
Fire_Stations	Richmond Fire Station 16	None given	6
Fire_Stations	Ambulance Station 40	None given	3
Homeless_Shelter	None given	180 E Belt Blvd, Richmond, VA, 23224	4
Homeless_Shelter	None given	1201 Hull St, Richmond, VA, 23224	2
Homeless_Shelter	None given	1201 Hull St, Richmond, VA, 23224	2
Library	North Avenue Branch Library	None given	5

use	name	address	Mean Predicted Count
Library	East End Branch Library	None given	4
Library	Hull Street Branch Library	None given	2
Oasis	None given	Redacted (HIPPA)	8
Oasis	None given	Redacted (HIPPA)	7
Oasis	None given	Redacted (HIPPA)	7
Police_Station	First Precinct	None given	4
Police_Station	Second Precinct	None given	4
Police_Station	Third Precinct	None given	2
School	Woodville Elementary School	None given	12
School	Fairfield Court Elementary School	None given	11
School	Preschool Development Center	None given	11

As mentioned, this process can be repeated for a specific list of sites. In Figure 2.8 we look at the relative measure of risk exposure for churches and other faith organizations. The following table lists the ten most optimally located churches based on mean predicted count of maltreatment events within a quarter mile.

Mean predicted count by quarter mile buffer
Churches and other faith organizations



Figure 2.8

name	address	Mean Predicted Count
Shalom Baptist Fellowship Church	None given	13
Sixth Mount Zion Baptist Church	None given	13
Ebenezer Baptist Church	None given	7
Swansboro Baptist Church	None given	6
New Canaan Baptist Church	None given	5
Bible Way Church	None given	5
Saint Philips Protestant Episcopal Church	None given	5
Bethlehem Baptist Church	None given	5
Mount Olivet Baptist Church	None given	4
Saint John Baptist Church	None given	4

Figure 2.9 evaluates the relative risk exposure for licensed child care providers. The table that follows lists the top provider locations sorted by type and mean predicted count.

Mean predicted count by quarter mile buffer

Licensed child care providers



Figure 2.9

license_type	address	Mean Predicted Count
LICENSED: CHILD DAY CENTER	1004 ST JOHN ST, RICHMOND, VA, 23220	19
LICENSED: CHILD DAY CENTER	1606 E 18TH ST, RICHMOND, VA, 23224	12
LICENSED: CHILD DAY CENTER	2012 SELDEN ST, RICHMOND, VA, 23223	11
LICENSED: FAMILY DAY HOME	2700 NEWBOURNE ST, RICHMOND, VA, 23223	13
LICENSED: FAMILY DAY HOME	3416 DELAWARE AVE, RICHMOND, VA, 23222	4
LICENSED: FAMILY DAY HOME	3200 UTAH PL, RICHMOND, VA, 23222	4
RECREATIONAL/INSTRUCTIONAL EXEMPT PROGRAM	900 Mosby St, RICHMOND, VA, 23223	7
RECREATIONAL/INSTRUCTIONAL EXEMPT PROGRAM	238 E 14TH ST, RICHMOND, VA, 23224	3
RECREATIONAL/INSTRUCTIONAL EXEMPT PROGRAM	4011 MOSS SIDE AVE, RICHMOND, VA, 23222	3

license_type	address	Mean Predicted Count
RELIGIOUS EXEMPT CHILD DAY CENTER	4212 CHAMBERLAYNE AVE, RICHMOND, VA, 23227	5
RELIGIOUS EXEMPT CHILD DAY CENTER	3401 CHAPEL DR, RICHMOND, VA, 23224	4
RELIGIOUS EXEMPT CHILD DAY CENTER	1515 Chamberlayne Ave N, RICHMOND, VA, 23222	4
VOLUNTARILY REGISTERED DAY HOME	1814 FAIRFAX AVE, RICHMOND, VA, 23224	4
VOLUNTARILY REGISTERED DAY HOME	3322 DELAWARE AVE, RICHMOND, VA, 23222	3
VOLUNTARILY REGISTERED DAY HOME	703 N 33RD ST, RICHMOND, VA, 23223	3

Finally, Figure 2.10 shows the relative measure of risk exposure for resource homes. Resource homes are locations certified to accept foster care children. The top resource home locations are listed in the table below.

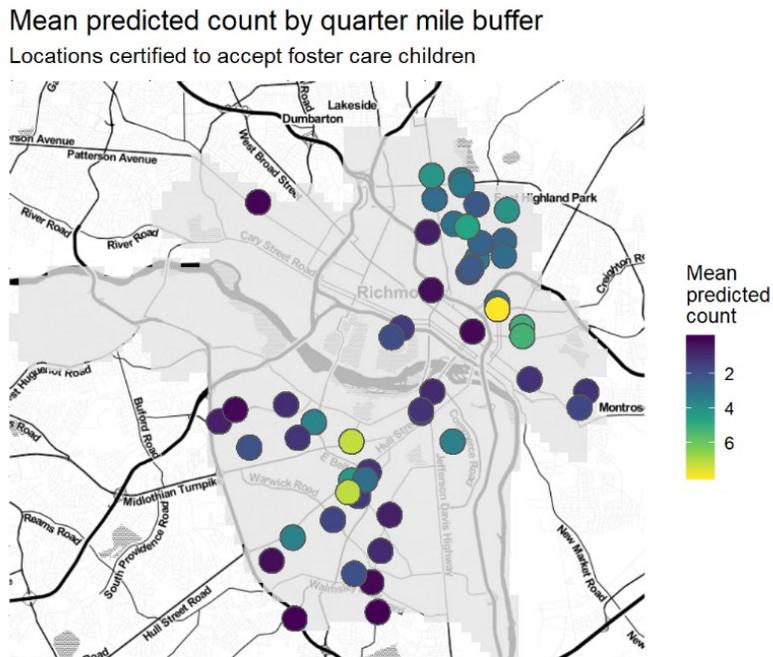


Figure 2.10

name	type	Address	Mean Predicted Count
Richmond City Jail	Detention/Correctional Facility	1701 Fairfield Way	8
Richmond Detention Home	Detention/Correctional Facility	1700 17th Street	3
Medical College of Virginia (MCV)	Hospital (Medical)	401 12th Street	1
Children's Hospital of Richmond	Hospital (Medical)	2924 Brook Road	1
Chippenham Hospital	Hospital (Medical)	7101 Jahnke Road	1
Richmond City DSS ILP	IL/Supervised Program	900 Marshall Street	0
Richmond City Independent Living Program	IL/Supervised Program	900 MARSHALL Street	0
Gates to Success ILA	IL/Supervised Program	1374 Greenmoss Drive	0
Individual	Individual	Redacted (HIPPA)	7
Individual	Individual	Redacted (HIPPA)	7
Individual	Individual	Redacted (HIPPA)	5
Elk Hill - Northside	Residential	3802 Chamberlayne Avenue	4
Kids in Focus II	Residential	3801 Lake Hills Road	0

Do home visits occur in high risk places?

Our data indicates 44 home visit events occur in Richmond during the study period. For privacy reasons the map of risk predictions and home visits is redacted. Figure 2.11 aggregates these data and illustrates that the majority of home events occur in the highest risk category.

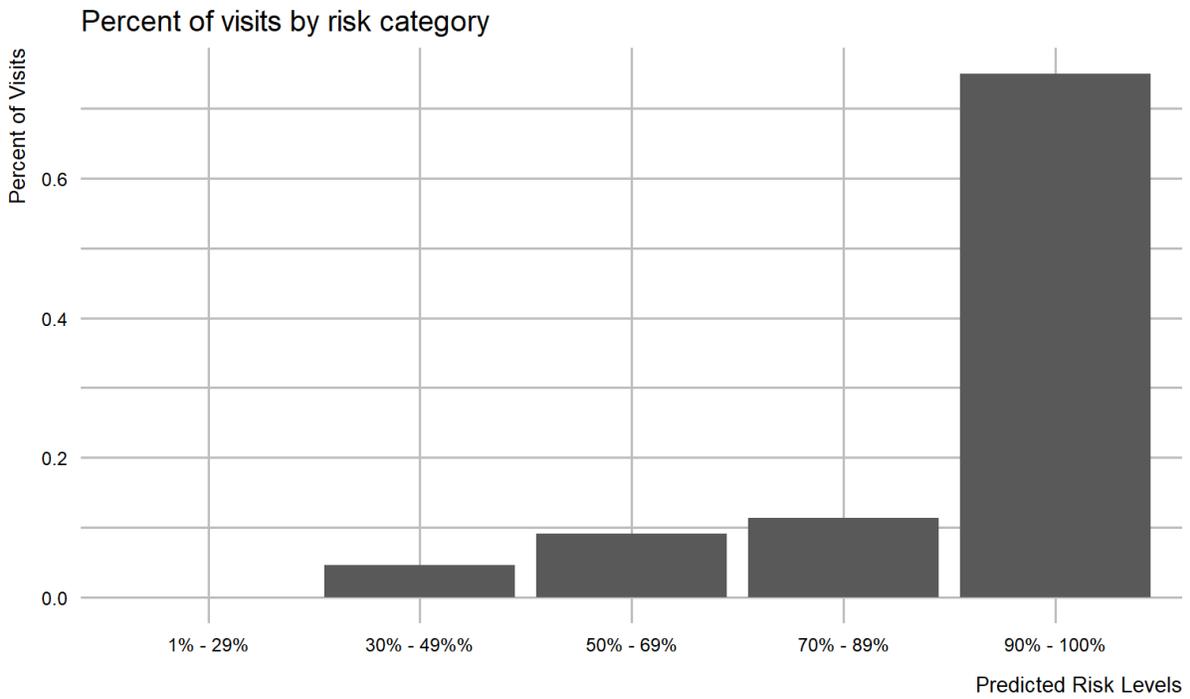


Figure 2.11

Maltreatment risk and SCAN Centers

Stop Child Abuse Now (SCAN) Centers are an important family resource in Richmond. Are they located in optimal locations? To answer this question, we overlay SCAN-related centers atop risk categories and calculate a relative exposure score for each (Figure 2.12).

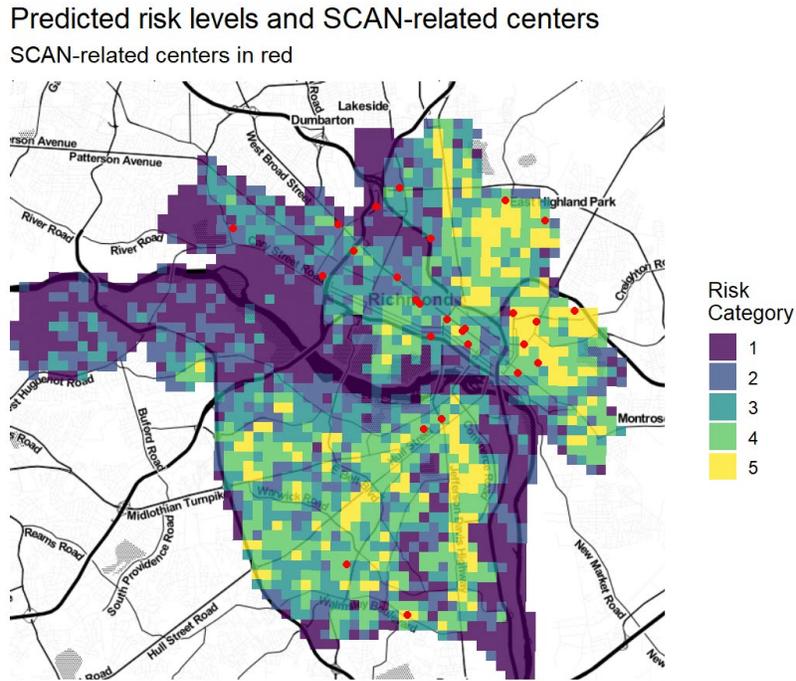


Figure 2.12

Figure 2.13 maps the mean predicted maltreatment count within a quarter mile of each SCAN-related center. The low count of maltreatment events within this radius suggests that some of these centers are not optimally located relative to maltreatment risk.

Mean predicted count by quarter mile buffer
SCAN-related centers



Figure 2.13

Finally, below is a table showing the top three optimally located SCAN-related centers.

Scan Center	Street Address	Mean Predicted Count
Boys & Girls Clubs of Metro Richmond	2506 Phaup Street Richmond, Va.	16
Richmond Peace Education Center	3500 Patterson Avenue, Richmond VA Richmond, Va.	8
Boys & Girls Clubs of Metro Richmond	1000 Mosby Street Richmond, Va.	7

Gap analysis

To reiterate, the purpose of this analysis is to ensure that the supply of child welfare services is properly aligned with the demand for these services. We approach this in a piecemeal fashion above but attempt bring it all together in one spatial analytic that can compare relative maltreatment risk to the relative supply of protective resources by neighborhood.

To define relative risk, we create a density metric defined as the sum of the highest risk category grid cells in a neighborhood divided by the total number grid cells in that

neighborhood. To define the relative supply of protective resources, we take a density of protective resources from above and divide by the total area of the neighborhood. Both metrics are scaled into percentiles to run from 1 to 100, where 100 equates to a relatively high density.

We then calculate the gap with using the following formula: $gap = RelativeProtectiveSupply - RelativeRisk$. For example, a neighborhood with a protective score of 100 (lots of protective resources) and a risk score of 10 (relatively little relative risk), would have a gap score of 90, suggesting more resources are present than needed.

Conversely, a neighborhood with a protective score of 10 (few protective resources) and a risk score of 90 (relatively high maltreatment risk), would have a gap score of -80, suggesting a tremendous misalignment of resources given the predicted need.

We calculate this and map it below (Figure 2.14). Light red and light green neighborhoods are those where there is a strong alignment between protective resource and risk. Dark green places are those where there are more protective resources than risk, whereas dark red areas are places where there is far more risk than protective resources.

This metric can be tuned further by gathering other data that better describe the protective resources available to children in Richmond.

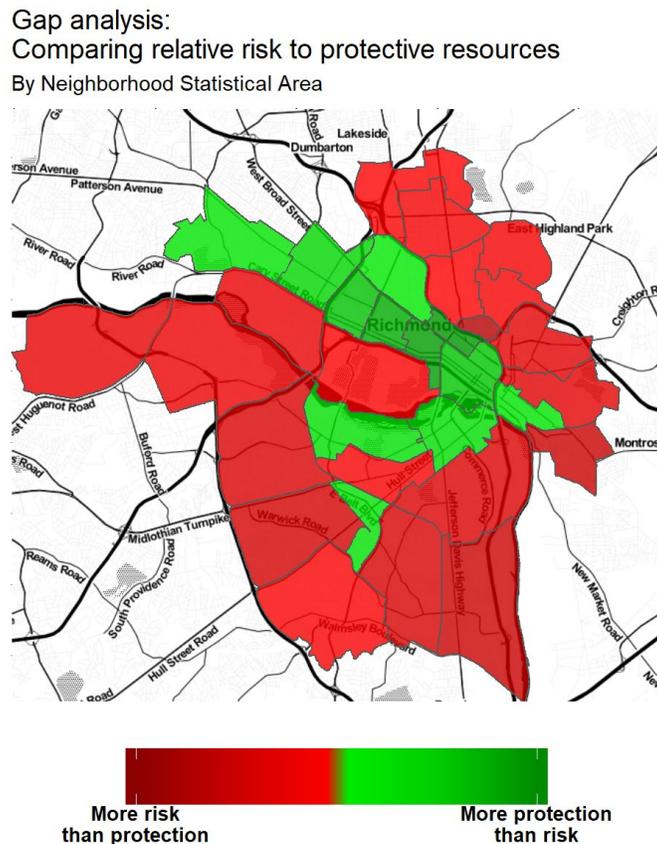


Figure 2.14

3. Motivation

Program evaluation versus predictive modeling

Much of the related child welfare research is focused on the causal relationship between maltreatment and associated exogenous factors. A well designed experiment, or quasi-experiment, can have important implications on the policy responses to child maltreatment. Conclusions from these sorts of program evaluations are often used to drive major programmatic or budgetary decisions.

In contrast, predictive modeling is best thought of as a tool for operational decision-making, helping stakeholders more effectively allocate their limited resources. A well calibrated causal model is one that avoids confounding bias when identifying how a variable of interest affects the outcome, whereas a well calibrated predictive model leads to more effective decision-making relative to the current resource allocation process.

This is in the spirit of Figure 1.6 above which compares model predictions to kernel density.

People versus placed-based predictive modeling

To date, researchers and jurisdictions have built both people and placed-based predictive models of child maltreatment.² The benefit of people-based models is their ability to guide interventions to specific individuals or households, potentially preempting a maltreatment event. This level of resolution comes with costs, however. People-level predictions require people-level data, often requiring a jurisdiction to overcome the significant legal, bureaucratic, and financial costs needed to integrate highly private, across-agency administrative data.

On the other hand, although placed-based models make predictions at greater geographic scales, the only required private data is geocoded locations of maltreatment events. Placed-based modeling does not entail high resolution integrated data. Most of the predictor variables describe environmental factors such as crime, blight, nuisance land uses, etc. - public data often available for download in bulk on city or state Open Data websites. Not only does this lessen the concern over the mistreatment of private data, it also means that building, testing, and deploying placed-based predictive models is more cost effective.

While people and placed-based models vary in their inputs, outputs, and applications, the modeling process is very similar. In both instances, a well fit model is one that is both accurate and generalizable across different urban contexts.

The value of open source analytics

The code created for this analysis is licensed under the GNU General Public License (GPL). The suite of tools we have developed and document on GitHub have been written in the R Statistical Language using custom functions of our own creation, as well as a number of open source packages shared freely by their authors.

The codebase written for this project is designed with the goals of standards and transparency in mind. The public sector is still developing the internal capacity for developing machine learning algorithms. Typically, governments procure sophisticated, closed-source software at a significant expense for both initial development and ongoing maintenance. We believe that by creating the foundations for an open source suite of spatial machine learning tools as part of this project, we can lay the groundwork for standards that other jurisdictions can adopt with minimal startup costs in the future.

The proliferation of standards will make it easier for the policy community to coalesce around a set of best practices for developing and deploying these models. Perhaps most critical is the ability to interpret the accuracy of these models and identify potential sources of bias. The reader will find all the source code used to develop this analysis on the associated [GitHub repository](#).

Aligning the supply of child welfare services to the predicted demand for these services

The predictive model developed below was created specifically to foster more efficient and effective decision-making in the child maltreatment realm. The goal is to ensure that the limited child welfare resources available to children and families in Richmond are allocated where they are needed most.

Toward this goal, we use the model predictions along with a host of ancillary data provided by the State of Virginia to develop a strategic planning framework that can identify places where the supply of services and the demand for these services are misaligned. This framework can be adopted and updated either in R or with standard Desktop GIS.

How this report is organized

This document is written in the R Markdown format. The purpose of R Markdown is to embed descriptive text, data visualizations, and code into one cohesive narrative. Most readers will focus only on the case study. Others, particularly those interested in replicating the analysis, should refer heavily to the associated [GitHub repository](#).

4. Methods

The child maltreatment spatial risk prediction framework presented here was developed in the spirit of criminological platforms like [RTMDx](#) and [Hunchlab](#), with some critical differences. First, it is free and open source and released with the hope that other jurisdictions will replicate and improve our work. Second, it is more than just a predictive algorithm. The exploratory data analysis functionality, bias detection tools, and suite of comprehensive planning tools (Align) should make it easier for analysts to communicate data-driven insight to non-technical policy makers.

Spatial prediction has been a tool in the spatial analysis toolbox for decades. Predicting unknown quantities across space from observed sample data is typically known as 'interpolation.' In their exhaustive review, Li and Heap situate this brand of spatial prediction in the realm of 'spatial interpolation' suggesting, 'nearly all spatial interpolation methods can be represented as weighted averages of sampled data.'⁸

Methodological literature at the intersection of interpolative predictive modeling and geography include simple interpolation, geostatistical methods such as Kriging, Gaussian Process Regression, Geographic Weighted Regression, Generalized Linear Models, and non-linear models like Random Forests.⁹

These models have been applied across a variety of disciplines including forestry, soil science, biology, archaeology, criminology, and public health.¹⁰ There is also a compelling thread of literature on the validation of spatial prediction methods including Congalton, Vicente-Serrano et al., Li and Heap, and Brenning.¹¹ There is decades of literature on machine learning techniques, which vary from geostatistics but are mechanically similar, including Kuhn and Johnson and Hastie, Tibshirani, and Friedman.¹²

The first application of these techniques to predict child maltreatment employed The Risk Terrain Model (RTM).¹³ The original RTM model used a weighted raster overlay methodology rooted in 'Map Algebra', methods first presented by Tomlin.¹⁴ While this is a useful descriptive approach for generating spatial hotspots, it does not provide measures of statistical confidence or goodness of fit.

The RTM approach matured from geographical overlay into a more prescriptive machine learning workflow as part of a partnership with Jeremy Heffner and software development firm Azavea. The Risk Terrain Modeling Diagnostics Utility tool (RTMDx) used regularization to reduce potential collinearity when predicting crime counts.¹⁵ Heffner developed a novel feature selection approach combining Penalized regression and Stepwise regression to estimate a final Poisson regression model that predicts crime counts.

The RTMDx is similar to other 'predictive policing' applications including HunchLab and Predpol. RTMDx [ranges from](#) \$100/year for a student license to \$4,995/year for 'Agency Users.' Hunchlab, the predictive policing application developed by Azavea, [can cost](#) \$15,000/yr for every 50,000 people in a jurisdiction. According to the same source, Predpol, another predictive policing application can also range in the six figures.

Below, we review the many tools developed as part of the framework. A host of Extract, Transform, and Load tools have been developed for the purposes of data ingestion, cleaning, and standardization. Several Exploratory Data Analysis routines have been created as well that describe key relationships in the data, many of which can be visualized and explained to non-technical policy makers. Finally, a host of comprehensive, machine-learning specific tools have been generated for the purposes of both prediction and prediction assessment. We look at each in turn below.

ETL and Feature Engineering

Feature engineering is the creation of standardized observations from which predictive relationships can be mined. The process of feature engineering includes the steps of data ingestion, aggregation, and transformation. As Table 4.1 shows, we employ several datasets for this analysis, converting each into a series of individual features.

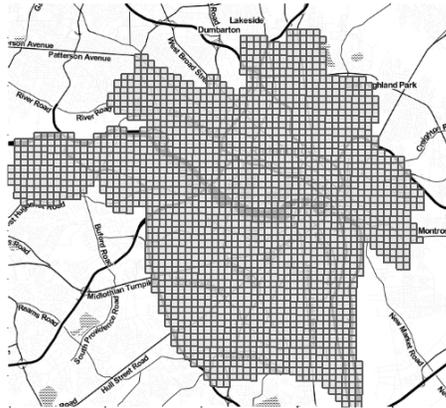
Dataset	Description
Crime	Locations of reported violent and nonviolent incidents that represent risk in the city
Points of Interest	Places in Richmond that characterize environmental factors and are categorized as either risk (i.e. Pawnbrokers, Payday loan locations, ABC stores) or protective (i.e. Libraries, Community Centers, Grocery stores) features
Businesses	Businesses that either characterize nuisance land uses or protective land uses
Code Violations	Indicates the safety and overall quality of structures
CPS Accepted Events	Locations of child maltreatment incidents

Routines have been developed that load data from various files and formats into the R programming environment for standardization and reshaping. This step begins with moving all data files to a common folder and then using R to list the files, check for the presence of ID and spatial coordinate fields, the number of rows, and the number of columns. A custom R function is written to handle these steps automatically regardless of file type or file names. Once loaded, summary statistics on each dataset are generated.

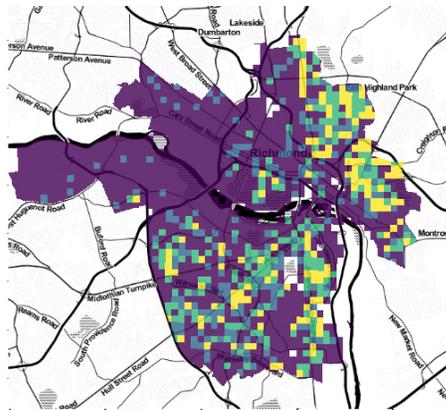
Additional data cleaning steps include the removal of events that are spatial and temporal duplicates. Additionally, maltreatment events that are labeled as 'unverified' are removed from the analysis.

In creating the dependent variable, a lattice grid, the 'fishnet', is overlain across Richmond and maltreatment events are summed for each grid cell. The fishnet consists of 1,910 individual cells with an area of 1000 ft², which allows for greater level of predictive resolution relative to say, census tracts. We also calculate for each grid cell, the population normalized rate of maltreatment events per 100 residents. This is achieved by overlaying the fishnet grid onto 2010 census tracts and attributing to each cell a weighted population proportional to the area of the tract that falls into the grid cell.

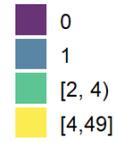
Fishnet - Richmond, VA



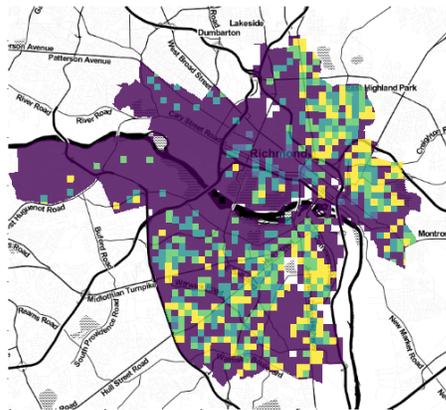
CPS count per fishnet cell



CPS Count



CPS rate per 100 people



CPS Rate per 100

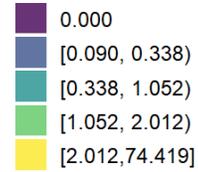


Figure 4.2

We also relate measures of exposure to the fishnet, aggregating risk and protective factors using three different strategies:¹⁶

1. Net or sum of exposure events per grid cell
2. The Euclidean distance from the center of each grid cell to the nearest exposure event
3. The average Euclidean distance from the center of each grid cell to the five nearest event neighbors

More than 200 features are created in this fashion, scaled and combined into a dataset which is then used in the modeling process.¹⁷ We also generate a series of spatial lag features, using the three feature engineering approaches above to account for the spatial externalities associated with maltreatment. Figure 4.3 visualizes examples of each feature engineering approach.

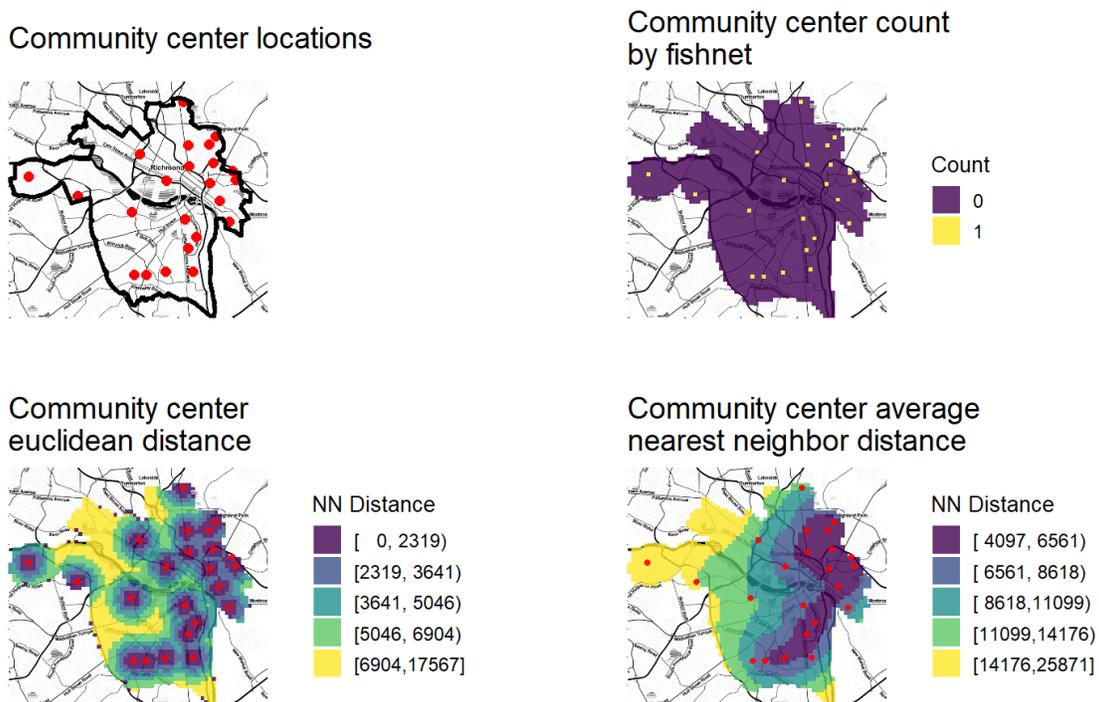


Figure 4.3

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a principled exploration of the data that allows natural patterns, structures, and relationships in the data to come to the surface.¹⁸ The tables and visuals created for this stage highlight outliers and anomalies, test assumptions, and uncover underlying structures. A desired outcome of EDA is to present data in such a way that stakeholders ask new and interesting questions that were not previously realized. This approach benefits in clearer goal setting and an atmosphere of collaboration between policy makers, technical specialists, and analysts.

EDA is employed here to gain a deeper understanding at the intersection of the contributing features, maltreatment outcomes, and subject matter expertise. This process is even more

prescient due to the complicating dimensions of space and time. The EDA methods developed here consist of:

1. Calculating summary statistics for each of the features
2. Plotting histograms and counts of categorical variables
3. Mapping the distribution and density of each feature
4. Visualizing the correlation between each variable
5. Computing Global and Local Moran's I statistics
6. Simulating random Nearest Neighbor distance
7. Estimating the distribution of the outcome variable

Summary statistics (e.g. minimum, maximum, deciles, counts, etc.) provide quantitative descriptions of each feature. Correlations inform the pairwise relationship between features and the outcome of interest.

Moran's I is a measure of spatial autocorrelation and used here to estimate the pattern of spatial association for maltreatment counts. The values of Moran's I generally range between -1 and +1, but more extreme values are possible. A value of zero typically indicates spatial randomness, a value of -1 indicates negative spatial autocorrelation (i.e. a regular pattern), and a value of +1 indicates positive spatial autocorrelation (i.e. clustering). For this study, the Moran's I is calculated on both the global scale and the local scale. Global Moran's I results in a single value describing the universal autocorrelation. This study uses the `moran.mc` function of the `spdep` R package to derive the global I value through Monte Carlo simulation.¹⁹

The local Moran's I describes the spatial autocorrelation within a local neighborhood of eight fishnet grid cells surrounding a given fishnet grid cell. This neighborhood configuration is called the 'Queen case' because it is the same eight directions as the Queen moves in chess. The local measure of Moran's I also computes the p-value indicating the significance of the local autocorrelation. Both the Moran's I value and the p-value are plotted to visualize significant positive and negative local clusters of maltreatment event counts.

A second simulation method is used to better understand whether risk and protective factors are more closely or distantly associated with maltreatment event locations.²⁰ This is achieved by simulating a set of random points equal in quantity to the number of events for a specific feature and averaging the nearest neighbor distance to n maltreatment events. This is repeated 1,000 times resulting in a distribution of randomly permuted nearest neighbor distances. Finally, this distribution is compared to the observed global average nearest neighbor distance of each contributing feature. If the observed distance of a given feature is closer than 95% of the randomly permuted distances (a p-value of 0.05), we can conclude that the feature is 'close' to maltreatment citywide.

The final step in the EDA is to gain a better understanding of the statistical distribution of maltreatment event counts. Inherently, the number of maltreatment events per fishnet cell is a count data type, which is typically modeled by either the Poisson or Negative Binomial distributions.

The intuition for the Poisson distribution is that maltreatment is a relatively rare event and that the probability of this occurrence is independent after accounting for contributing factors.

Further, it is assumed that the variance in maltreatment event counts is approximately equal to the mean of counts. This assumption limits the potential for drastic outliers in maltreatment event counts relative to the mean of all counts. By contrast, the Negative Binomial distribution can be conceptualized in a similar manner with the exception that the variance is not the same as the mean.

The `fitdist` function in the `fitdistrplus` package is employed to simulate samples from the empirical maltreatment event count distribution and compare the goodness of fit to both the Poisson and Negative Binomial distributions.²¹ The results of this test are visualized and used to choose between the two distributions.

Model fitting and stacking

Model 'fitting' describes the process by which a statistical algorithm learns about maltreatment risk by relating the interaction of risk/protective factors to maltreatment events across space. Once a model is fit and validated, the learned pattern is applied back to the contributing features in each fishnet cell to predict the count of maltreatment events across space. This prediction highlights areas where maltreatment risk may be present but not reported.

The first step in the model building process is to select the 15 most statistically important risk and protective feature sets. We select across the different feature types (Euclidean distance, average nearest neighbor distance, and aggregate counts) based upon statistical correlation. These features make up the final feature sets, which are then subjected to our models.

Three different algorithms are fit modelling different aspects of the spatial process and then combined into a fourth 'meta-model.' The three individual models are a Poisson Generalized Linear Model (Poisson GLM), a Random Forest model, and a Spatial Durbin Model (SDM). The final prediction of maltreatment events is produced from the meta-model which is created by applying the Random Forest algorithm to the predictions made by the sub-models. The use of three different model algorithms is an effort to understand different aspects of the highly complex system that contributes to the observation of a maltreatment event.

At each stage in this process, models are fit using a 'leave one group out cross validation' routine (LOGOCV). LOGOCV splits the data into spatially explicit groups, in this case neighborhoods, fits the models to all but one of the groups, and predicts maltreatment event counts for the left out group. This process, explained in greater detail below, tests how well the models generalize across neighborhoods. Below we explain the three sub-models and their inclusion in the final meta-model as well.

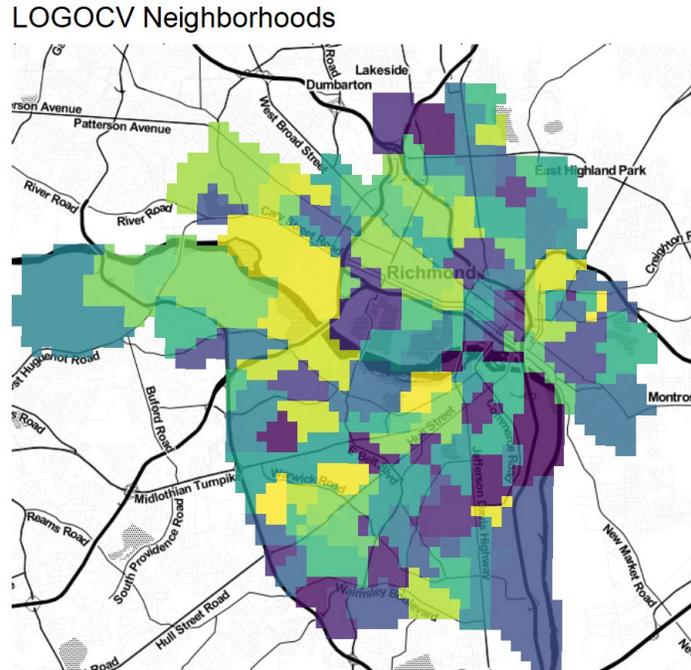


Figure 4.4

Poisson GLM

The Poisson GLM model, fit with the base R `glm` function, is an adaptation of linear regression that accounts for the characteristics of count data. The adaptations include modeling the residuals as Poisson distributed and transforming the linear model with a natural log function. As a result, the predictions from a Poisson model are positive and represent mean expected counts conditional on the contributing risk or protective features for each fishnet grid cell. In the meta-model, Poisson GLM predictions represent a linear model of a Poisson distributed count process.

Random Forest

In this study, the Random Forest algorithm is fit using the `ranger` library.²² The Random Forest algorithm builds a series of decision tree models to learn the relationship between maltreatment and exposure variables. The stochastic approach to sampling from observed data ensures that each individual tree is different from the next. The Random Forest provides a 'wisdom of many' approach, contributing non-linear interactions between maltreatment and the corresponding features to the final meta-model.

Spatial Durbin Model

To model spatial interrelationships - also referred to as 'spatial autocorrelation' - a Spatial Durbin Model (SDM) is fit using the `errorsarm` function of the `spdep` package in R.²³ In the setting of this study, the interpretation of this model is that the rate of maltreatment events is affected by both the exogenous exposure factors as well as neighboring rates of maltreatment. Further, this model assumes that there may be latent features that impact the model errors but

are not accounted by the exposure features. The key model input of spatial autocorrelation is a spatial weights matrix relating maltreatment in a given grid cell to its neighbors. Modeling the underlying spatial maltreatment process provides a powerful predictive story when input into the final meta-model. It is important to note that the SDM is not fit with the LOGOCV method due to the complications of subsetting a spatial weights matrix in a cross-validation setting.

Meta-Model

The final maltreatment count predictions are generated from a meta-model which combines predictions from the three sub-models. The process to combine the three models is straightforward; as the predicted counts from each sub-model are used as input features of a new model fit with the Random Forest algorithm. Often referred to as model 'stacking', this technique seeks to average out the variance in the three separate models. To reduce the risk of over-fitting, the stacked meta-model is fit and predicted using the same LOGOCV routine as the sub-models.

Model validation

Assessing the accuracy and spatial generalizability of model predictions is crucial when considering how to embed this model in the provision of child welfare services. A variety of approaches are used for model validation including LOGOCV and assorted goodness of fit metrics. Some of these metrics are statistical in nature, while others measure goodness of fit across space.

Leave One Group Out Cross Validation

LOGOCV is a technique for assuring that model predictions are generalizable across neighborhoods. The first step in LOGOCV is to assign each fishnet cell to the neighborhood that encompasses it. From a policy perspective, LOGOCV evaluates whether the maltreatment experience as we've modelled it is relevant to varying neighborhood contexts in Richmond. From a modeling perspective, this approach helps ensure our models are not overfit. Each of the 148 neighborhoods takes a turn as the hold out, totaling in 297 individual sub-models and 297 separate estimates of goodness of fit.²⁴

Goodness of fit metrics

Model error is defined simply as the difference between the observed count of maltreatment events and the predicted count for each grid cell. Complicating matters is that 297 models yields more than 567,000 grid cell level predictions. We derive several statistics to summarize and aggregate these errors in order to judge models and compare across them. We describe each below.

The Mean Absolute Error or MAE, measures the average absolute difference between the observed and predicted values.²⁵ An example interpretation of MAE is that, 'on average, the model is off by plus or minus 1.67 events.' MAE is simple to interpret in a policy context, however, it comes with some drawbacks, namely that the direction of the error is unknown and that every error is assumed to have the same severity. The MAE assumes that an error between

a predicted count of 5 and an observed count of 7 events should be considered the same as a prediction of 23 and an observed value of 25 events. This metric is used here due to its obvious interpretation and common usage in the assessment of predictive models.

The second goodness of fit metric used in this study is called the Logarithmic Score. This metric is not as straightforward as MAE, but it has qualities that make it well-suited to count-based predictions. The intuition of the Logarithmic Score is as follows: 'What is the likelihood of the observed count given the predicted count.' More descriptively stated: if the model predicts 10 events and the observed count is 7 events, then what is the probability of observing those 7 events if the prediction of 10 is indeed the correct number. In this way, the Logarithmic Score measures the deviance between the predicted and observed counts. Specifically, this is measured by calculating the probability density of the observed value from a Poisson distribution centered on the predicted value. The goodness of fit measures below report the negative log of the probability density so that the value should always be minimized. In the results portion of this report, the Logarithmic Score is converted back to a probability and aggregated. The result is a score that is interpreted as the 'average likelihood that the observed counts are true given the predicted maltreatment counts.' Closer to one means a higher relative likelihood, 0.5 equates to maximum uncertainty, and a value near zero signifies very little likelihood.²⁶

Accuracy and generalization tradeoff

The purpose of the LOGOCV and the goodness of fit metrics is to assess model errors on average and across space. A model that perfectly predicts the observed event counts for each fishnet grid cell would be very accurate, but would not generalize well to other cells because exposure changes across the city. Conversely, a model that predicts the same count of maltreatment events for every cell would generalize well, but not be relevant to the conditions of any one cell. LOGOCV and associated metrics help establish a balance between model accuracy and model generalization. Given the purpose of this study, it is important to create a model that is accurate enough to give confidence, but general enough to be applicable in areas where few cases are documented.

5. Exploratory Analysis

In this section, we derive a number of descriptive insights from our data. To begin, maltreatment events across space and time are visualized and we discuss some of the possible selection bias in the data. Next, hypotheses related to the clustering of maltreatment events across space are presented. Finally, we calculate a series of pairwise correlations between maltreatment and our features.

Analysis of maltreatment events over time and space

Figure 5.1 plots the counts of CPS events throughout the study period which began midway from 2013 and ended midway through 2017. From discussions with our partners in Virginia, the differences in counts across years may be due to changes in how maltreatment reports were

entered into the system. As we do not observe these reporting changes, time variation cannot be used for model validation.

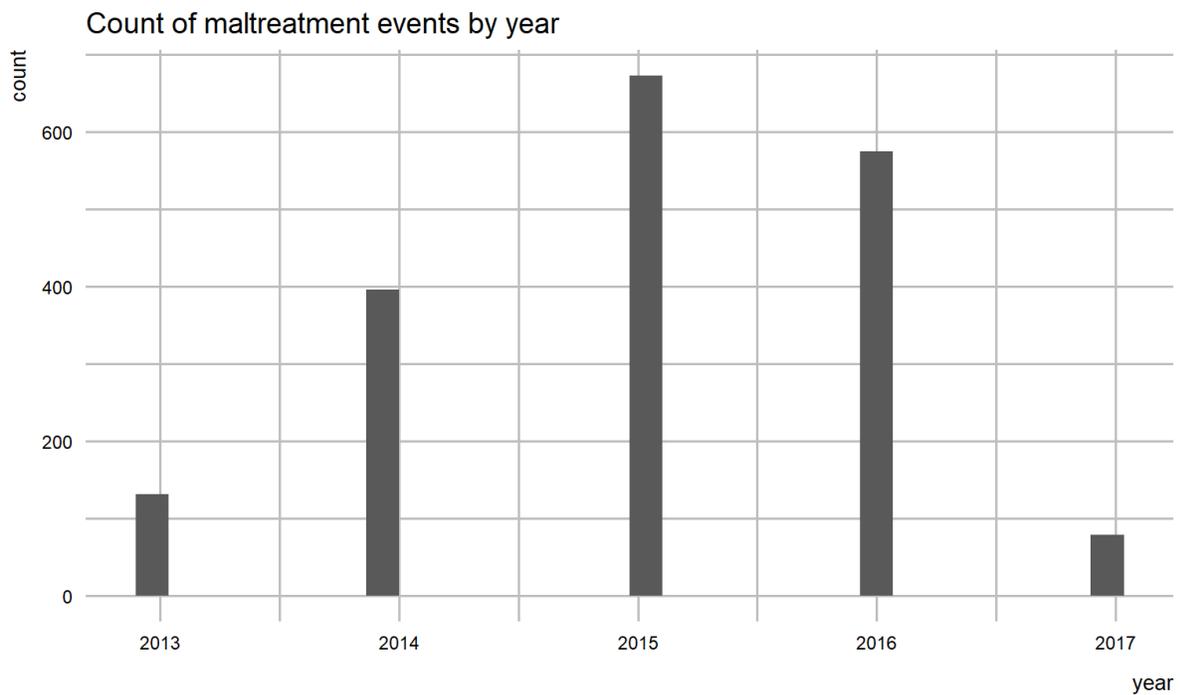


Figure 5.1

Although the counts may vary over time, they appear to be relatively consistent across space. Figure 5.2 shows the density of maltreatment events across space between 2013 and 2017. This is a strong indicator that the aforementioned reporting bias has less of an effect across space.

CPS accepted in Richmond, VA by year

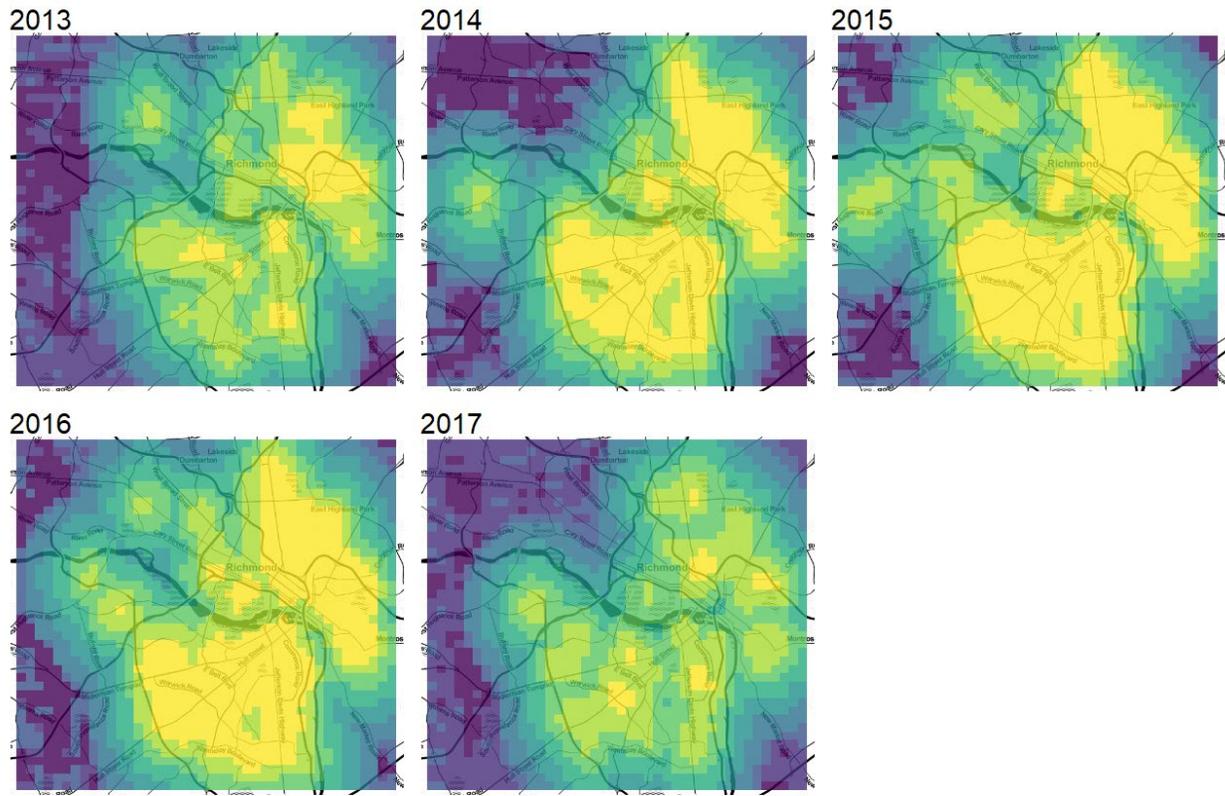


Figure 5.2

Visualizing risk and protective features

Figures 5.3 and 5.4 plot the spatial density of selected protective and risk factors, respectively.

Spatial density of protective factors

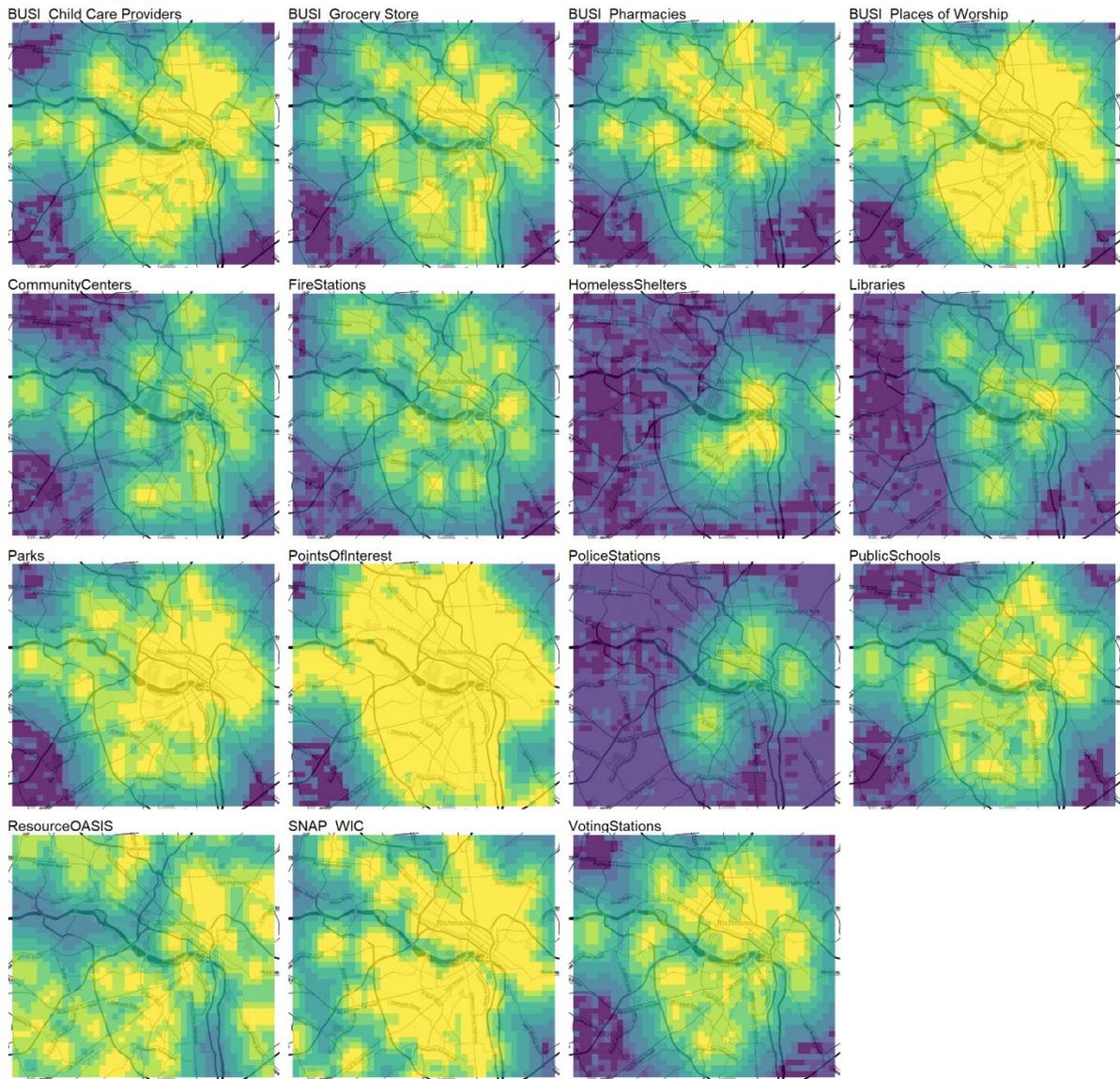


Figure 5.3

Spatial density of risk factors

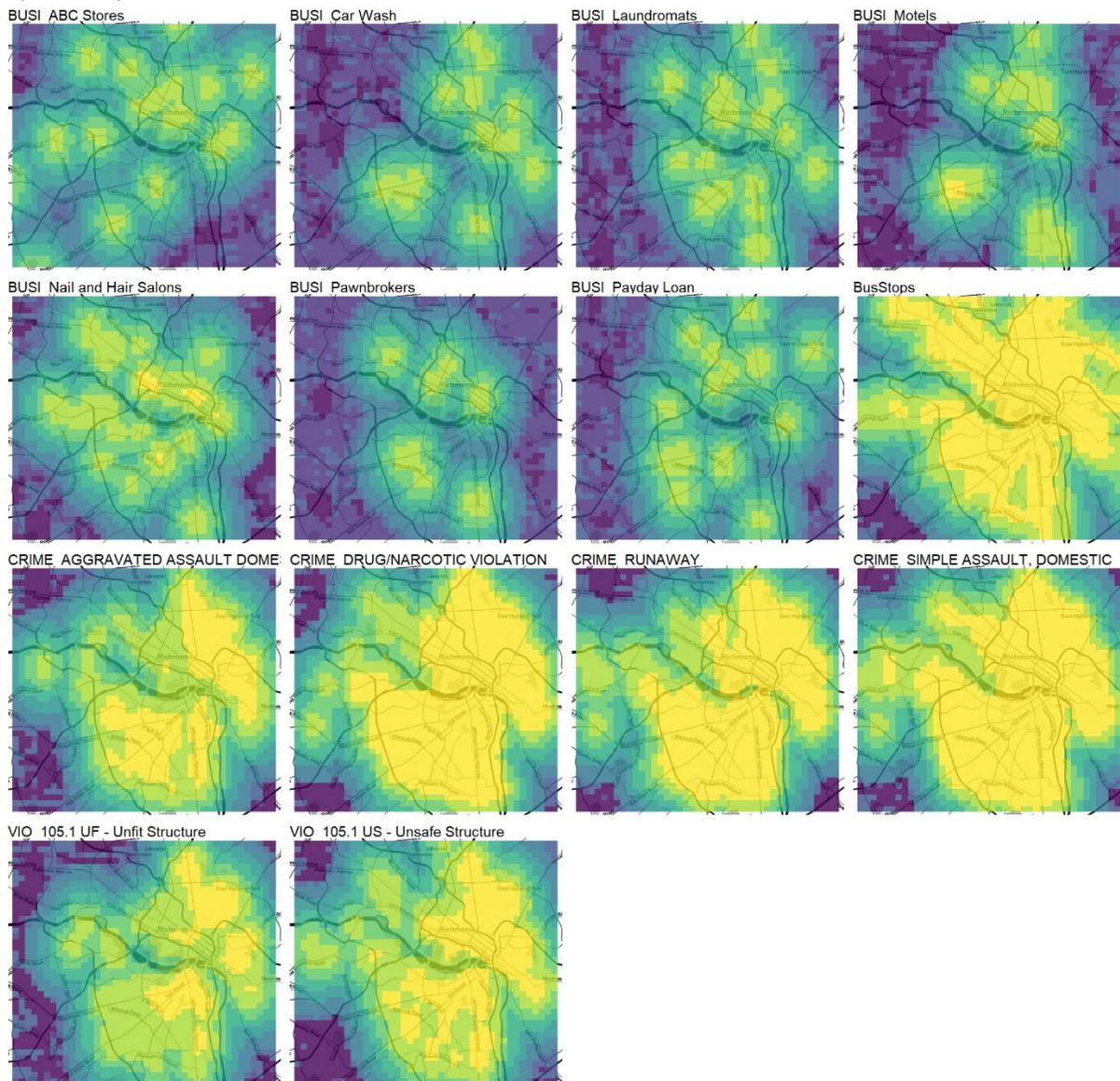


Figure 5.4

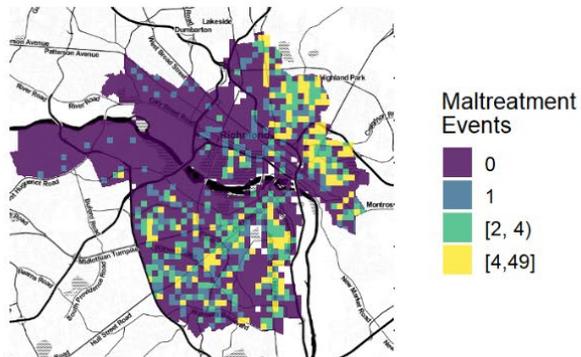
Testing maltreatment events for clusters

One overarching assumption behind our model is that maltreatment events are clustered in space. To test this hypothesis, the Local Moran's I statistic is employed.²⁷

The Local Moran's I statistic asks if the local distribution in the rate of maltreatment events, defined by a spatial weights matrix, is more clustered than we would expect due to random chance alone. Figure 5.5 plots the Local Moran's I results. The first panel shows the count of maltreatment events; Panel 2 shows the Local Moran's I value; and Panel 3 shows areas that exhibit statistically significant clustering. Panel 3 shows areas that resemble discrete clusters of maltreatment in space.

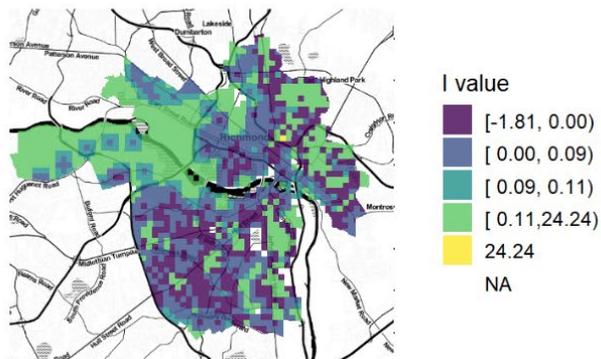
Panel 1

CPS count by fishnet



Panel 2

Local Moran's I value



Panel 3

Statically significant maltreatment clusters

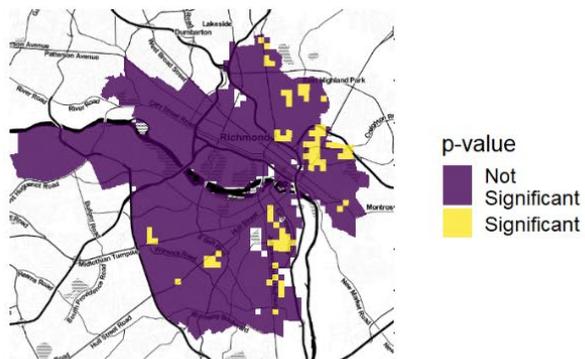


Figure 5.5

Which risk/protective features are statistically 'close' to maltreatment events?

Our predictive models below produce measures of 'variable importance', but because these estimates are produced outside of an experimental setting, they should not be interpreted as causal in nature. In this section, we generate some descriptive statistics that test the extent to which risk/protective factors are 'close' to maltreatment events.

The x-axes of these plots represent the average nearest neighbor distance from the factor to maltreatment events. The bar plots visualize the distribution of randomly generated distances, while the black lines represent the observed nearest neighbor distance. Figure 5.6 suggests that domestic instances of aggravated assault are not only far closer to maltreatment events than what would be expected due to random chance alone, but are closer than all other crime types.

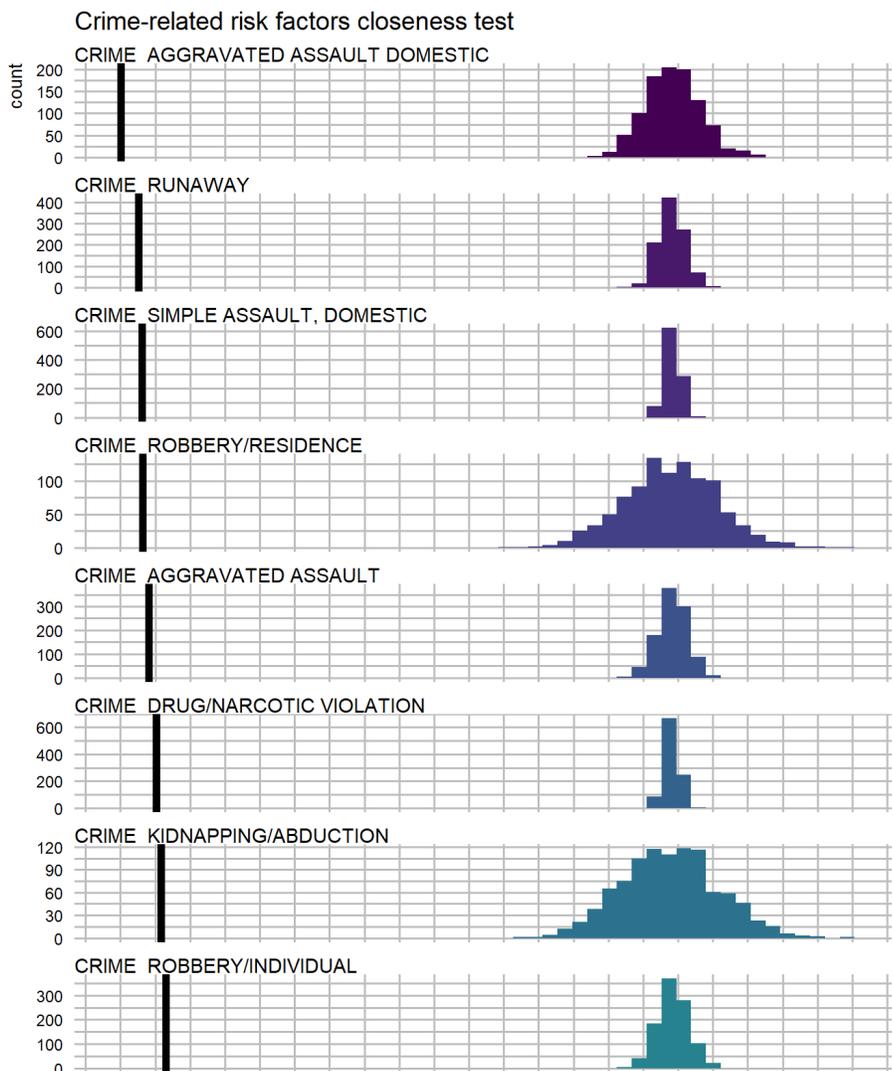


Figure 5.6 (continued below)

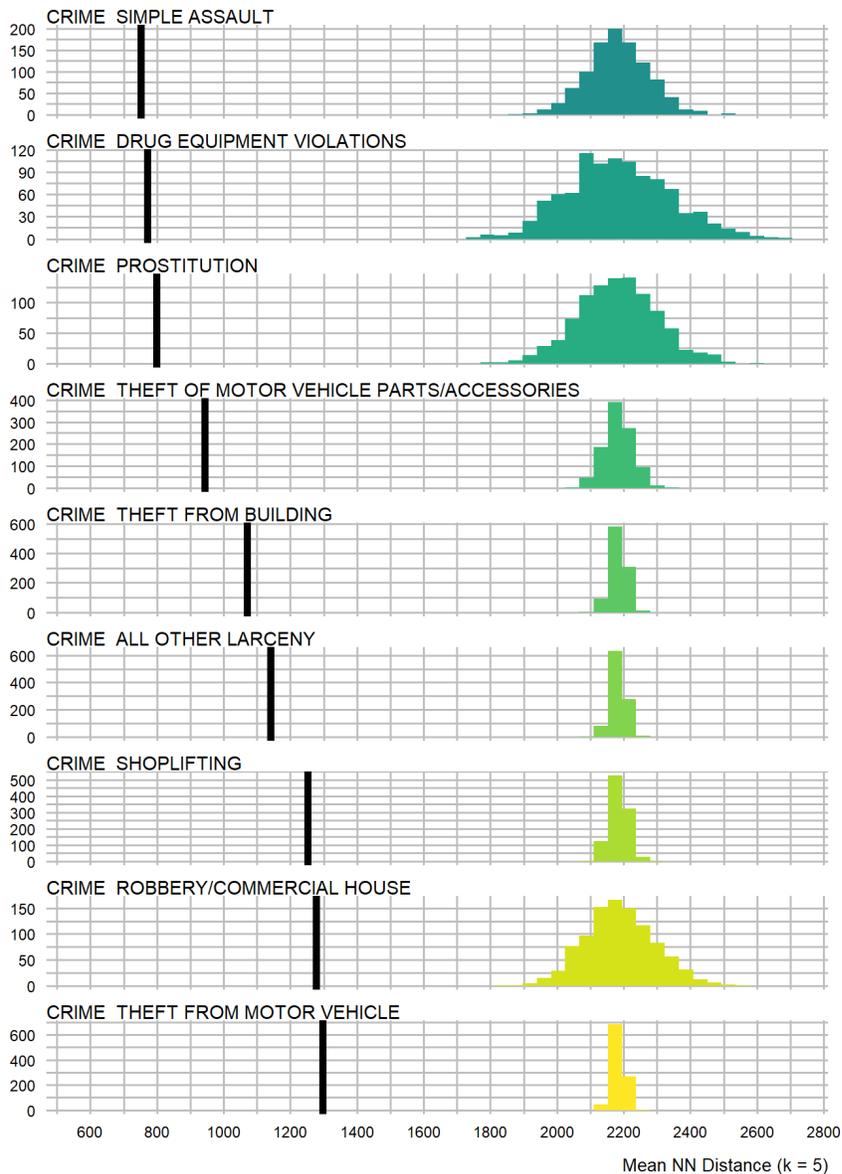


Figure 5.6 (continued)

In fact, all crimes are closer to maltreatment events than otherwise expected, suggesting that nearby crime is a significant risk factor for maltreatment.

Figure 5.7, which visualizes the results for violation-related risk factors suggests that blight-related factors, like unfit structures and refuse, are strongly associated with maltreatment in space.

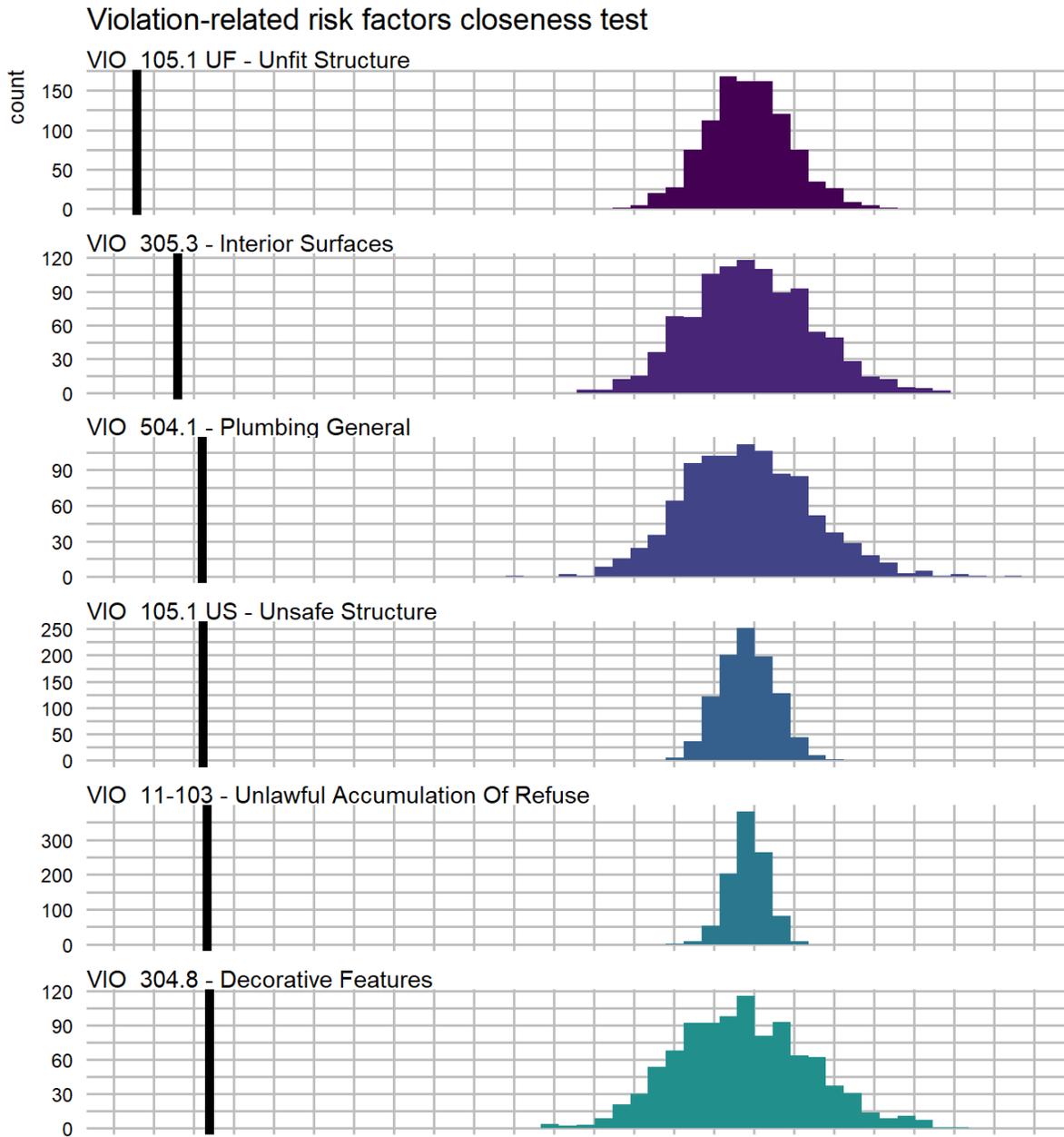


Figure 5.7 (continued below)

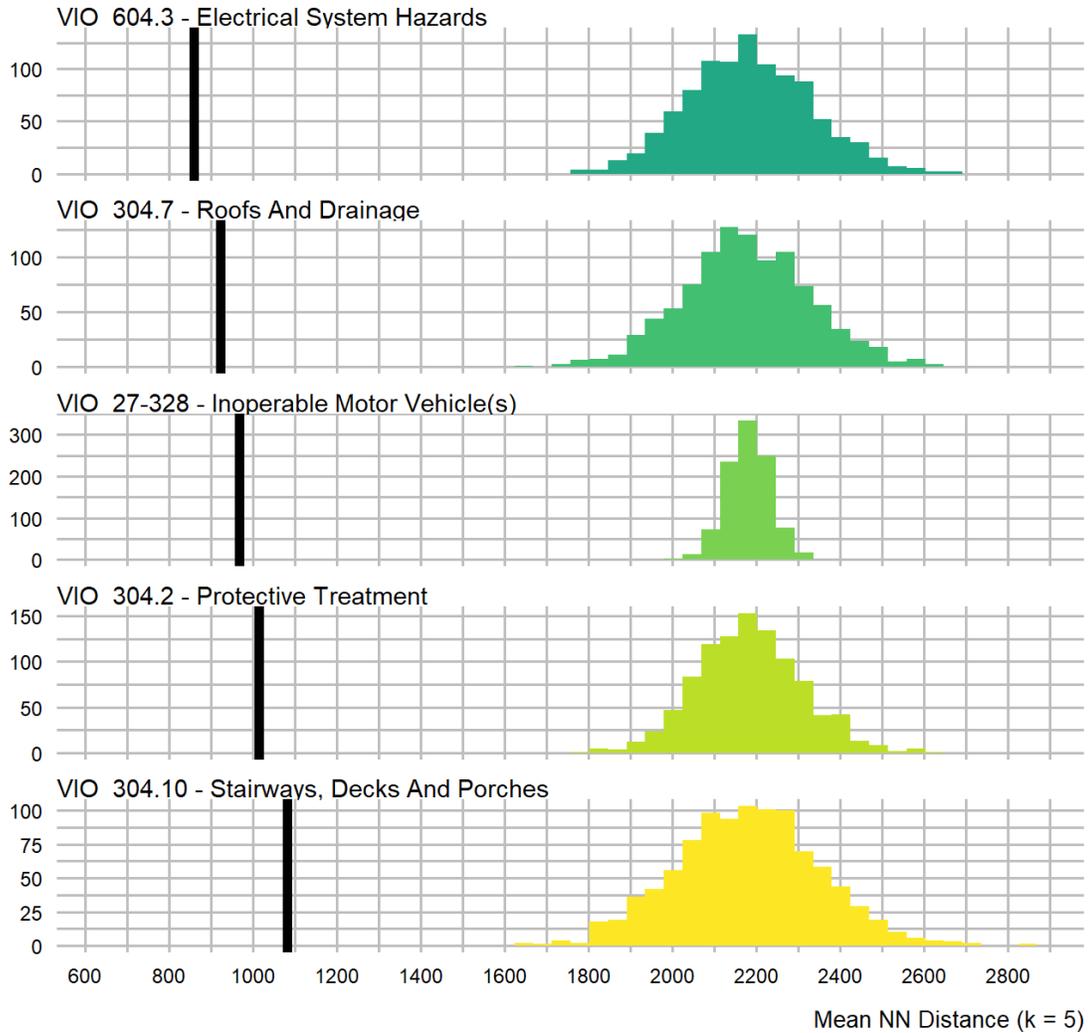


Figure 5.7 (continued)

Finally, Figure 5.8 illustrates how this test can also play a role in understanding select business features. The plot tells us that child care providers, taken as a whole, are uniquely situated in close proximity for maltreatment. Thus, when looking to align resources to fight maltreatment, stakeholders should consider deploying these resources from child care providers.

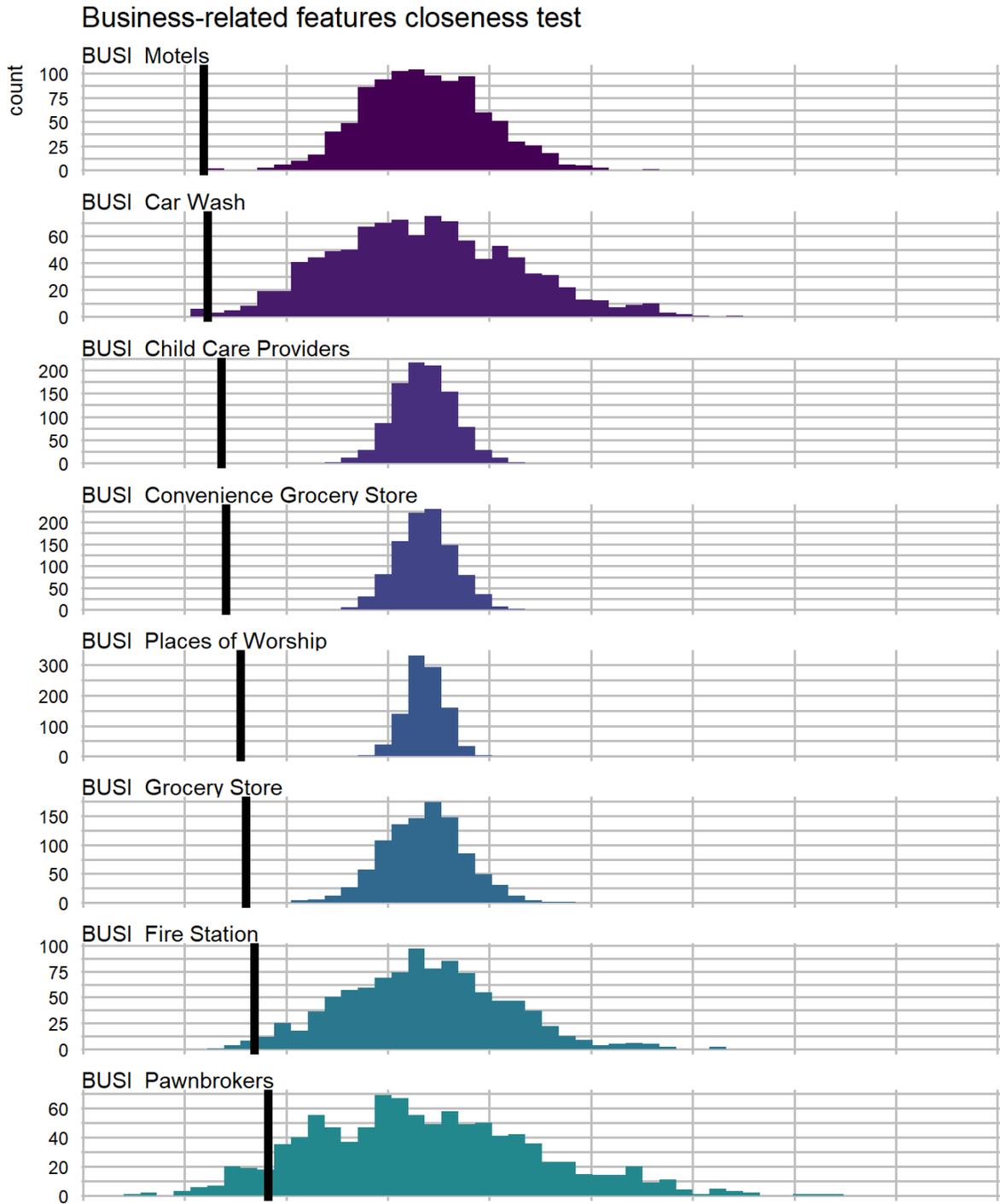


Figure 5.8 (continued below)

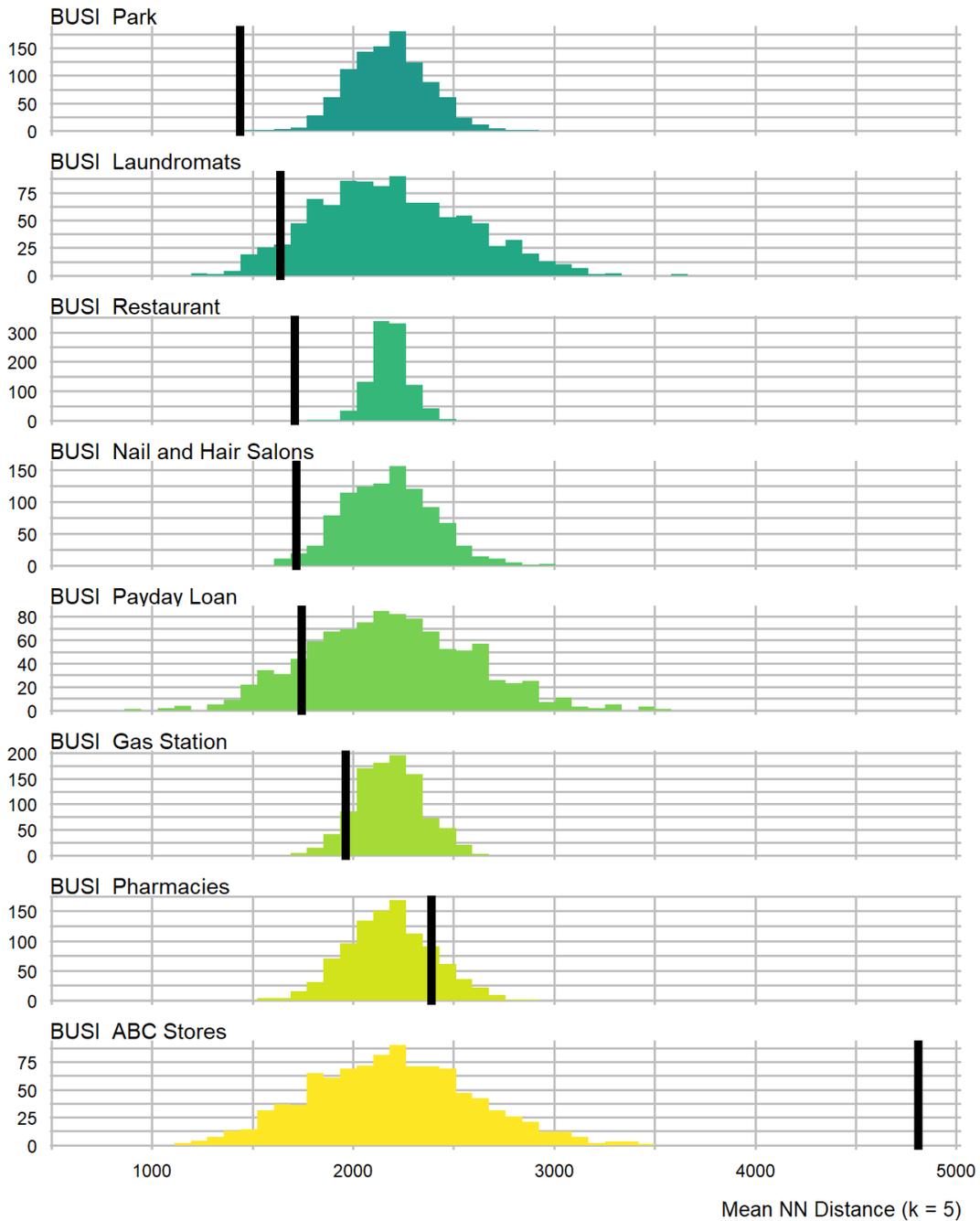
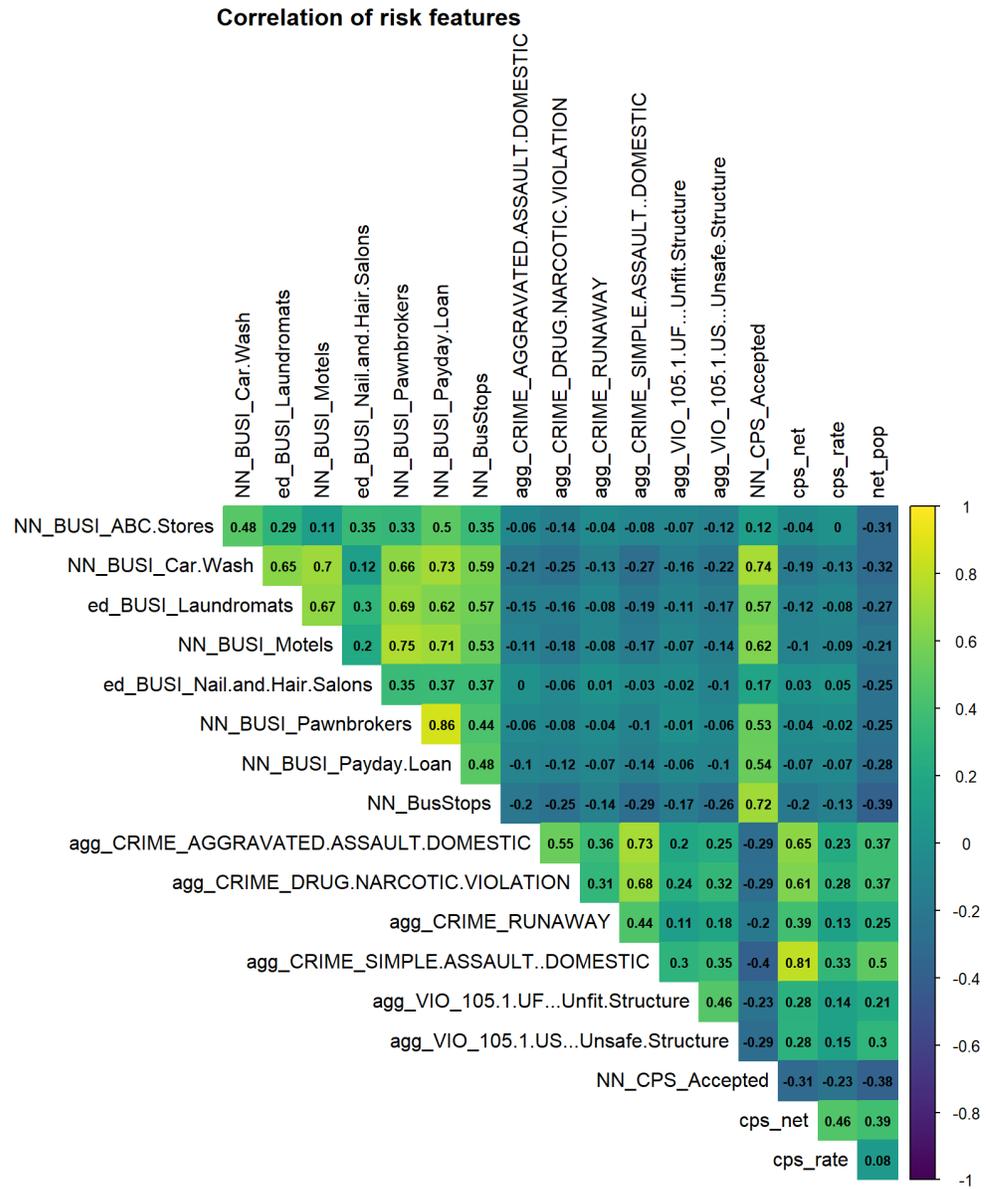


Figure 5.8 (continued)

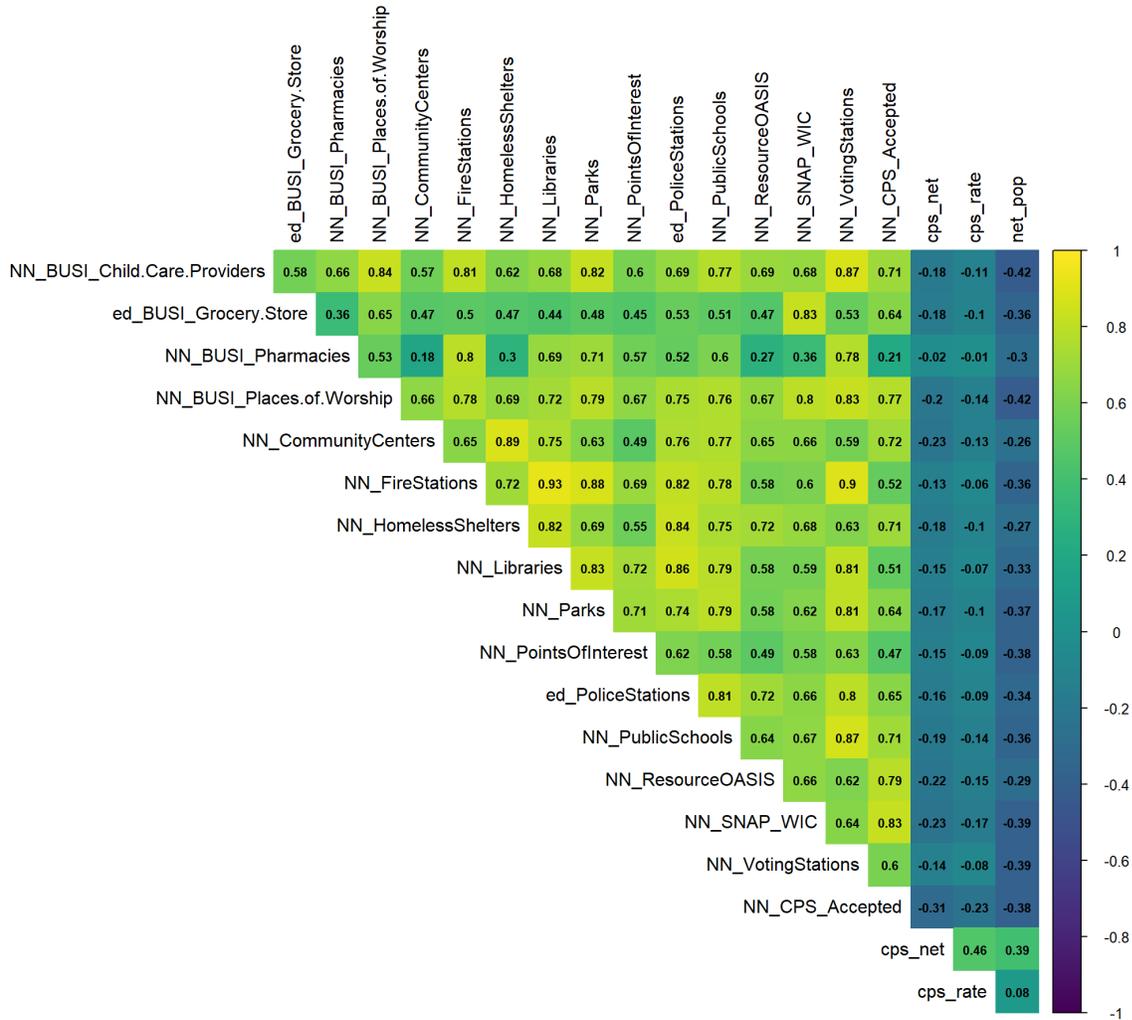
Pairwise correlations

In the Feature Engineering phase, we create more than 200 features. Figures 5.9 and 5.10 visualize pairwise correlations for the top 15 most correlative risk and protective factors respectively. Note the correlation coefficients associated with maltreatment count (`cps_net`) and maltreatment rate (`cps_rate`). The colors of the plot vary with the strength of the correlations, either positive or negative.

There are three different prefixes associated with each type of feature. NN refers to features calculated by taking the average distance between a fishnet grid cell and its n nearest risk/protective factor neighbor. ed refers to the Euclidean distance between a fishnet grid cell and its 1 nearest risk/protective factor neighbor. agg refers to the count of risk/protective factor events in a given fishnet grid cell.



Correlation of protective features



6. Modeling and Validation

The results of the modeling and validation stage of this report indicate that the three sub-models (Poisson GLM, Random Forest, and SDM) each obtain similar levels of accuracy and generalizability and the meta-model captures the nuances of each. The results indicate that the models are accurate, but there is a significant amount of variation in accuracy between neighborhoods. As we describe below, the meta-model generalizes well across neighborhoods of varying poverty rates, but does not generalize well across neighborhoods of varying race.

Below we provide both spatial and aspatial goodness of fit metrics.

Average goodness of fit results

The table below displays the goodness of fit metrics for each of the sub-models and the meta-model. Mean and standard deviation of different metrics are calculated. Means are taken to describe relative goodness of fit across each held out neighborhood. Standard deviations are taken to describe the variation in goodness of fit across each held out neighborhood. If the model generalizes well across neighborhoods, then the standard deviation should be relatively low.

Model Name	R2 mean	R2 sd	MAE mean	MAE sd	RMSE mean	RMSE sd	logdev mean	logdev sd
GLm - Poisson	0.522	0.401	0.56	0.767	0.927	1	0.685	0.24
Meta-Model	0.513	0.393	0.533	0.746	0.9	1.308	0.697	0.227
Random Forest	0.496	0.398	0.547	0.739	0.888	1.347	0.666	0.226
Spatial Durbin - sqrt	0.835	NaN	0.486	NaN	1.278	NaN	0.706	NaN

R2 or R Squared is a traditional measure of goodness of fit. Although typically not used to evaluate count outcomes, we include it here because it will be familiar to many readers.

MAE or Mean Absolute Error is the absolute difference between the observed maltreatment counts and predicted counts. The meta-model MAE equates to roughly one half of one maltreatment event on average. This suggests the model is accurate. The relatively high standard deviation of MAE suggests that greater errors can be found certain places, namely those with very high maltreatment counts.

RMSE or Root Mean Squared Error is the standard deviation of the prediction error. Like MAE, RMSE is reported on the scale of the dependent variable, but it varies in that the metric is weighted heavily by errors of high magnitude.

For the Logarithmic Score (logdev) the mean is 0.697 with a standard deviation of 0.227. This equates to a 95% confidence interval between 0.66 and 0.733 for the population average. The intuition of this result is that on average, the probability that the model estimates are correct given the documented maltreatment counts is between 0.66 and 0.733. While the population average of errors from independent LOGOCV estimates is helpful for assessing how the model generalized, it is equally important to know how these errors are distributed both statistically and across space.

Of note in the above table is the reduction in not only MAE and logdev but perhaps more importantly is a reduction in the standard deviation in those metrics across all of the 149 neighborhoods. The meta-model results in an average MAE of 0.533 with a standard deviation of 0.746. The 95% confidence interval for the meta-model MAE across the entire population of

neighborhoods is between 0.412 and 0.653. Since MAE is on the scale of absolute count of maltreatment events, this means that the population average MAE is less than one incident.

Figure 6.1 visualizes predicted versus observed maltreatment event counts for the meta-model. The black line represents a perfect fit while the blue line represents the predicted fit. This plot provides visual evidence of an accurate model. Nevertheless, the plot also shows that the model errors are much higher for the highest observed counts. In other words, the model fits most of the data well but breaks down in grid cells with far greater counts. As we discuss below, this has some ramifications with respect to generalizability.

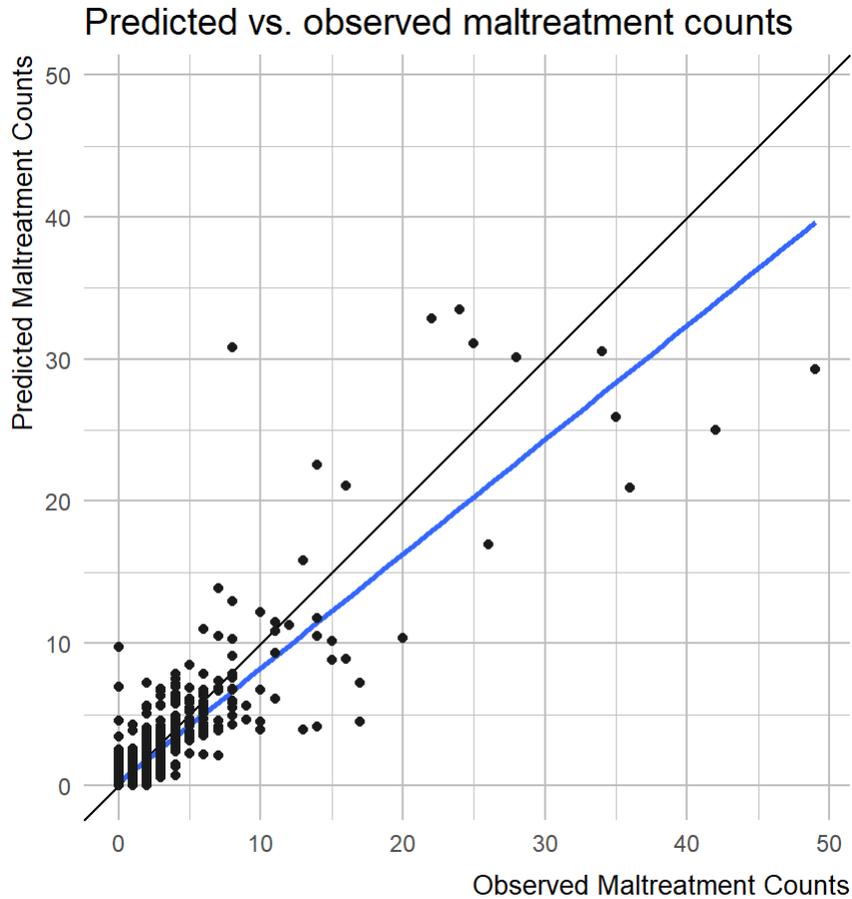


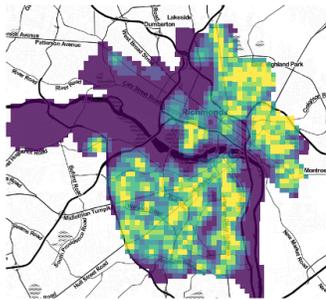
Figure 6.1

Mapping goodness of fit

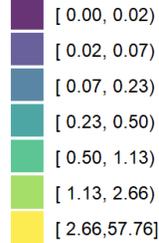
From the spatial perspective, Figure 6.2 illustrates the MAE and predicted maltreatment count for each of the three sub-models and the meta-model. The plots show that each model's approach to prediction is unlike its peers. The Poisson linear model is focused on the main areas of recorded maltreatment with lower variance outside of the City of Richmond population centers. The Random Forest model includes a higher degree of variance across areas where few incidents are documented, but where underlying risk and protective features are present. The SDM model incorporates knowledge of the underlying spatial organization of both maltreatment events and risk/protective factors. Given that the SDM model is fit and estimated on the full Richmond dataset, that information is translated into the 'patchiness' of the estimates. Finally, the meta-model exhibits a pattern of prediction that moderates the three sub-models and identifies the known areas of higher maltreatment rates. Note that the meta-model predicts areas as having high risk where the risk/protective factors are present, but recorded events are few.

Poisson Regression

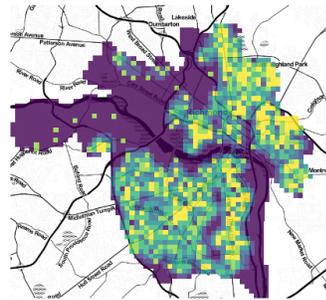
Predicted Maltreatment Count



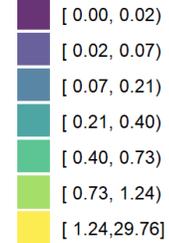
Prediction



MAE

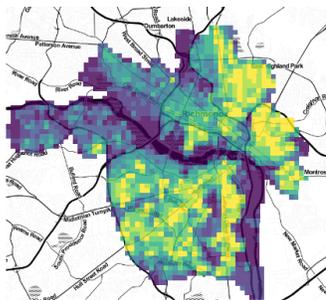


MAE

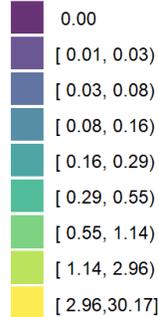


Random Forest

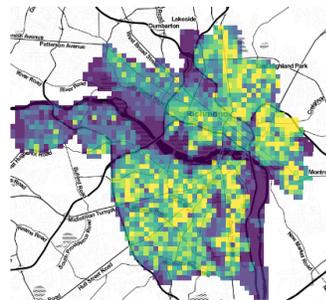
Predicted Maltreatment Count



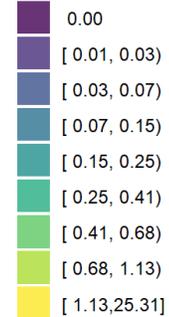
Prediction



MAE

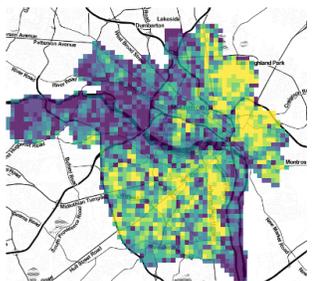


MAE

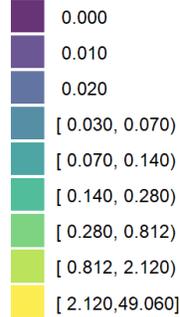


Spatial Durbin Model

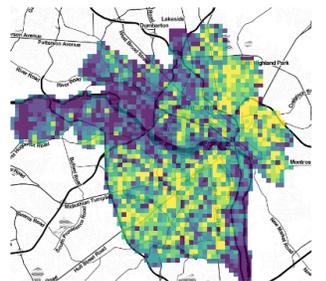
Predicted Maltreatment Count



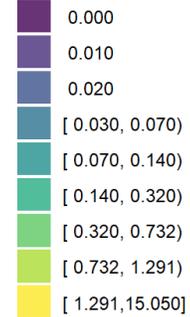
Prediction



MAE

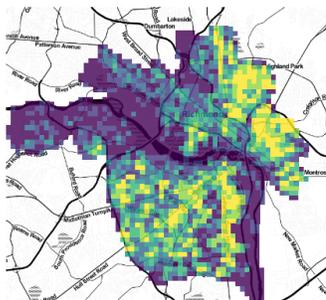


MAE

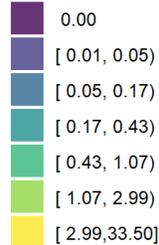


Meta-Model

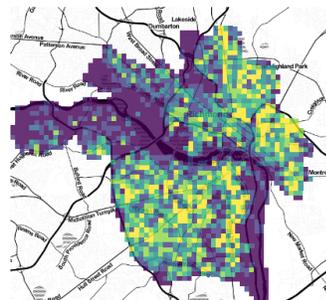
Predicted Maltreatment Count



Prediction



MAE



MAE

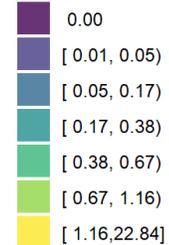


Figure 6.2

Further complementing these findings, Figure 6.3 shows the goodness of fit metrics broadened to the neighborhood level for the meta-model estimates. These goodness of fit indicators were created by way of LOGOCV. Generally speaking, the MAE and Logarithmic Score metrics follow a similar pattern with higher errors in neighborhoods with higher rates of maltreatment events.

If the model was perfectly generalizable, model errors by neighborhood would be randomly distributed. The map in Figure 6.3 shows that MAE clusters slightly, particularly in neighborhoods in the northeast of the city. A Global Moran’s I test confirms the clustering of errors, with a p-value of 0.001.

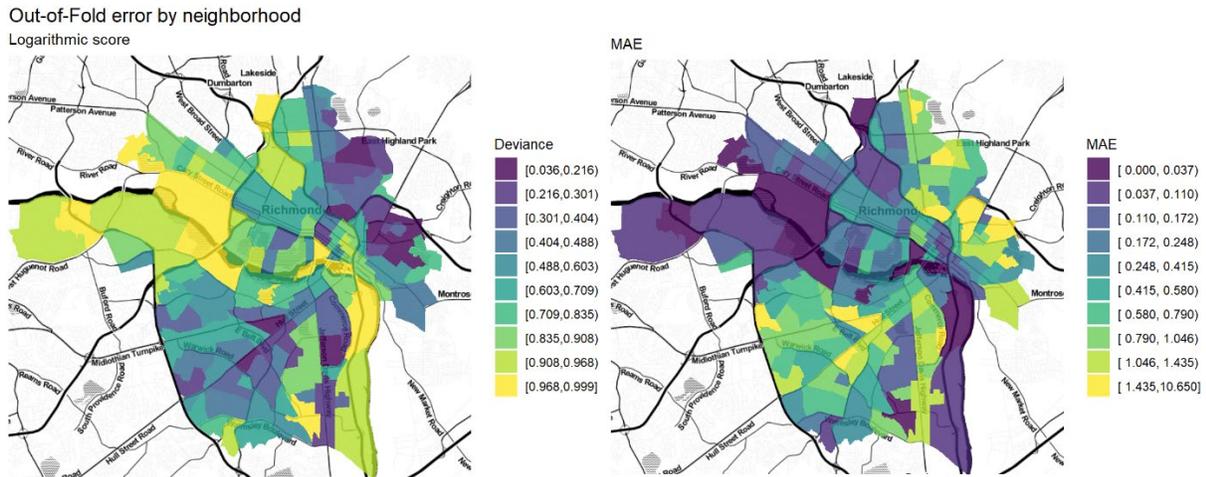


Figure 6.3

Feature importance

Next, we try to get under the hood of the model. We do this first, by visualizing feature importance. The below plot visualizes ‘feature importance’ for the Random Forest sub-model, showing which features make the greatest contribution in predicting maltreatment. We caution the reader to consider these relationships the result of correlation, not causation. The most important predictor, NN_CPS_Accepted, is the spatial lag variable used to account for the spatial externalities associated with maltreatment. Four of the top ten features relate to domestic occurrence of crime including assault and juvenile runaway. Three of the top ten are control variables describing demographic and housing unit characteristics. We also see other significant exposure factors including vacant housing, community centers, and motels.

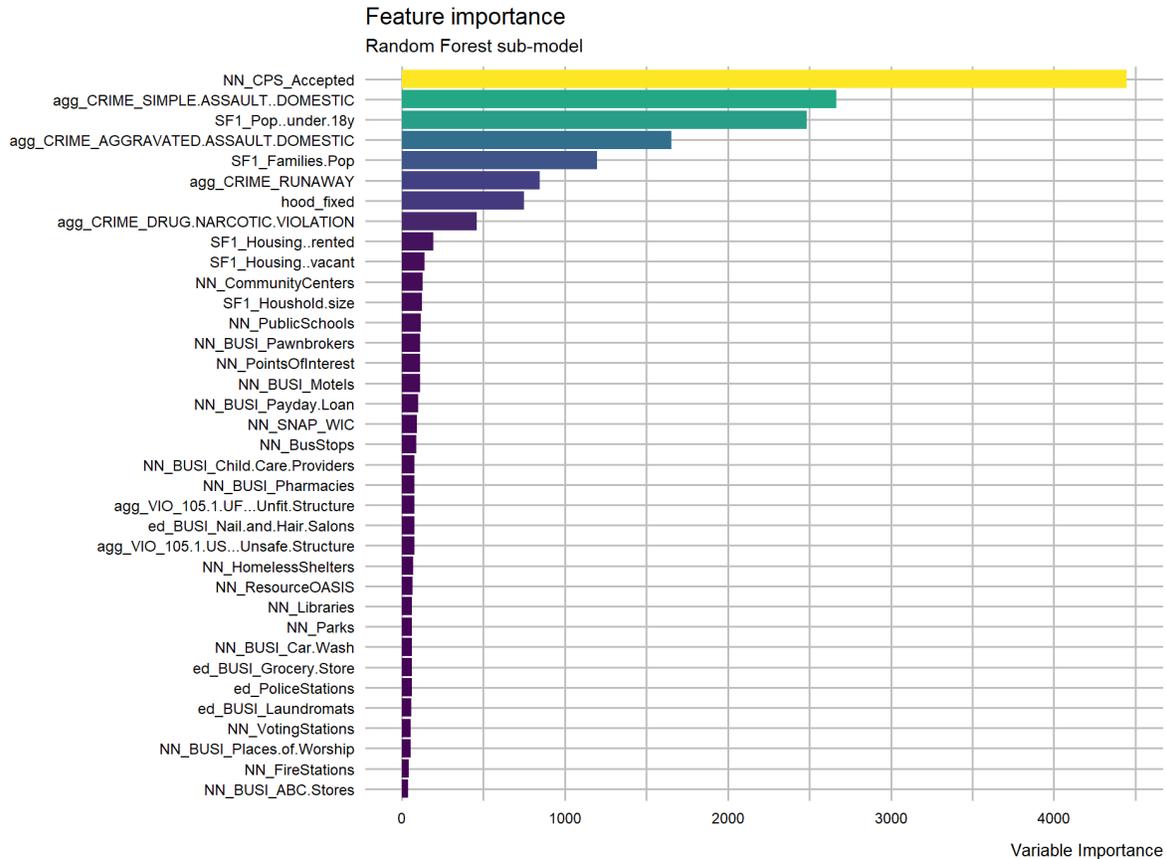


Figure 6.4

Neighborhood typology comparison

Income and race are inextricably linked to many of the census and exposure features used in the model, but no variables directly measuring race or income are included in the models. While feature importance provides some glimpses into how the model predicts, the best way to understand the inner-workings of a model is to look for patterns in how it predicts. Our approach for doing so, tests how well the model generalizes across both rich and poor neighborhoods as well as majority White and minority neighborhoods.

Two census attributes are selected for these purposes including percent living below poverty and percent non-White. These census data are joined to [Neighborhood Statistical Areas \(NSA\)](#) and high and low neighborhood typologies are created.

Next, median meta-model predictions are calculated for each NSA and goodness of fit is compared between high and low areas. The table below lists the median Logarithmic Score for both the high and low classes for each of the census variables. If the model generalizes well to both neighborhood typologies, the Logarithmic Score should be comparable across high and low categories. We find this to be true for poverty-related differences across the city but less so for race-related differences.

Poverty	Log_score	Count_events
Low	0.658	668
High	0.635	1145

Non_White	Log_score	Count_events
Low	0.757	524
High	0.471	1289

This suggests the model generalizes well to places with varying incomes but generalizes less well to neighborhoods of different racial composition. A portion of this finding has to do with the very high maltreatment counts that the model cannot account for. These are the areas where more than six maltreatment events have been recorded for a single apartment building. This may suggest that some spatial externalities exist at scales smaller than that of our current model. Perhaps future iterations of this model can account for this building scale variation, which may help correct for some of the bias described in Table 6.2.

Comparing meta-model predictions to Kernel Density

Perhaps the strongest method for assessing the usefulness of a predictive model is to compare its predictive power to that of the current resource allocation strategy. Although no equivalent decision making tool exists in Richmond, we can compare our model to another common spatial targeting algorithm - Kernel Density Estimation (KDE).

KDE is a simple spatial interpolation technique that calculates 'predicted risk' by taking a weighted local density of maltreatment events. No risk/protective factors are used and no measures of statistical significance can be calculated. To compare between the meta-model predictions and KDE, predictions from both are divided into five risk categories for the purposes of comparison. We then overlay held out maltreatment events that were not used to fit the original model, and calculate the percent of observed maltreatment events that fall into each predicted risk category.

Figure 6.5 maps the comparison. For privacy purposes, some cartographic information is redacted. The KDE clearly picks up the main areas of recorded events, but also interpolates high predictions for maltreatment in the areas between and beyond. The meta-model is far more targeted.

Risk categories from KDE

Privacy Controls: Maltreatment events in grid cells with 1 point are masked; Remaining event locations are offset at random.

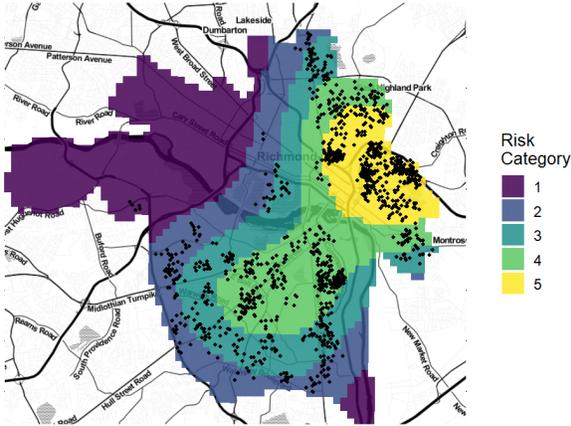


Figure 6.5

Risk categories from meta-model

Privacy Controls: Maltreatment events in grid cells with 1 point are masked; Remaining event locations are offset at random.

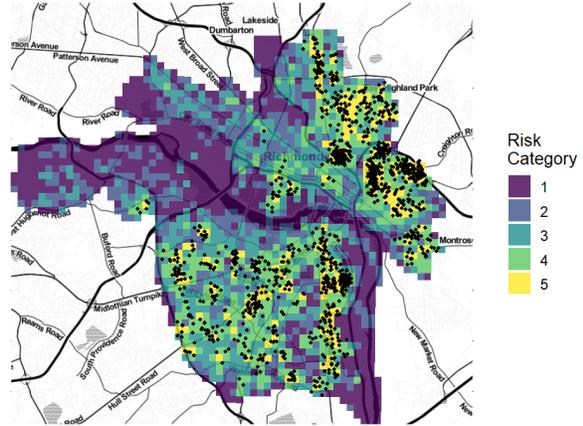


Figure 6.6 formalizes the comparison in chart form. While maltreatment events have been offset in Figure 6.5, true locations are used as inputs to Figure 6.6. The highest risk category for the meta-model captures approximately 70% of the recorded maltreatment events, whereas the KDE captures only about 35%. This suggests that the spatial risk model vastly outperforms KDE.

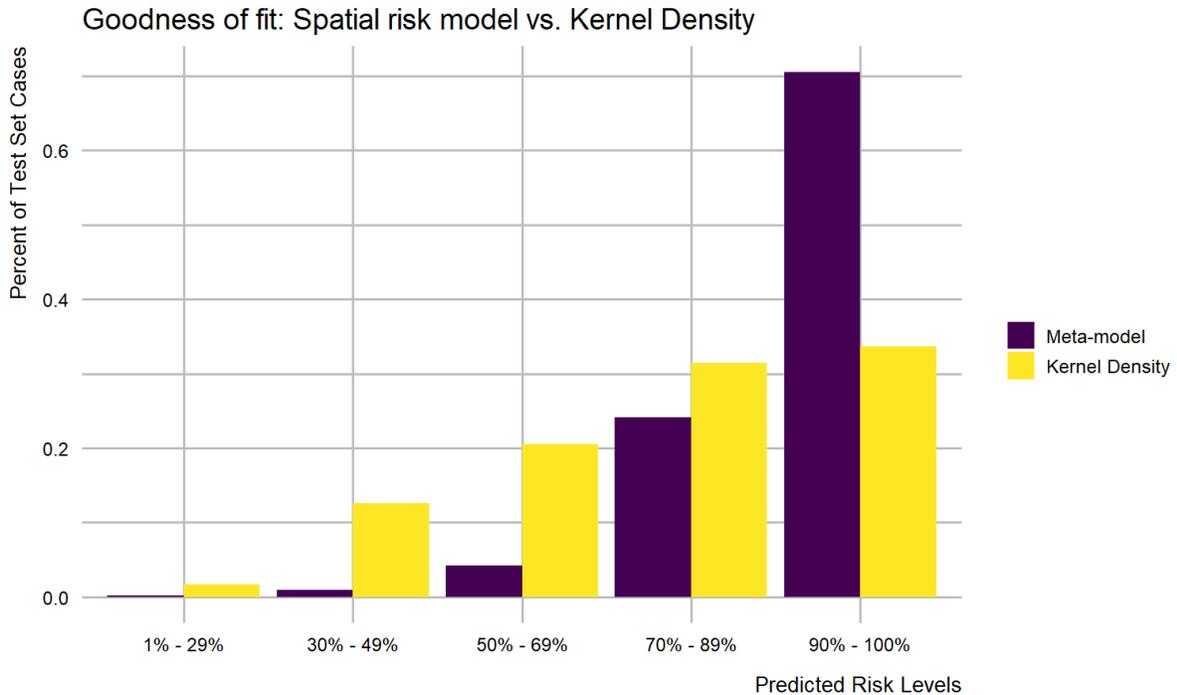


Figure 6.6

7. Conclusion

We set out to create a tool for embedding predictive modeling into the decision-making process of child welfare stakeholders. We provide several descriptive spatial and statistical analytics for understanding maltreatment across Richmond. Our model predicts child maltreatment accurately relative to a more traditional Kernel Density approach and exhibits little bias with respect to income.

The model correctly predicts high maltreatment risk in several African American communities in the northeast of Richmond. In these places, observed maltreatment events occur at much higher rates than others areas of the city. Although the model correctly highlights these areas as high risk, these outlying counts lead to higher errors which suggest, at least mechanically, that the model may exhibit some bias toward African American communities. As discussed in Section 6, some of these errors may be due to unaccounted variation in the built environment - namely that many of these events happen in the same apartment complex. Future models that control for these parcel level features may help adjust for these anomalies.

Our predictions show where maltreatment risk is at its greatest. We hope that stakeholders will use this metric to help guide education and prevention resources. In Section 2, model predictions are then used as inputs into spatial analysis that ask to what extent the supply of child welfare services align to the demand for these services. We show areas where more protective factors are needed. The only required input for this analysis from the model are the predictive risk categories. The hope is that stakeholders will repeat this analysis with other datasets towards the completion of a strategic plan for combating child maltreatment in Richmond.

Finally, the associated [GitHub repository](#) includes all the tools required for replicating our analysis. While we will continue to develop this codebase into a more easy to use package for replicating the analysis, we hope curious civic technologists will adopt the code for their own purposes.

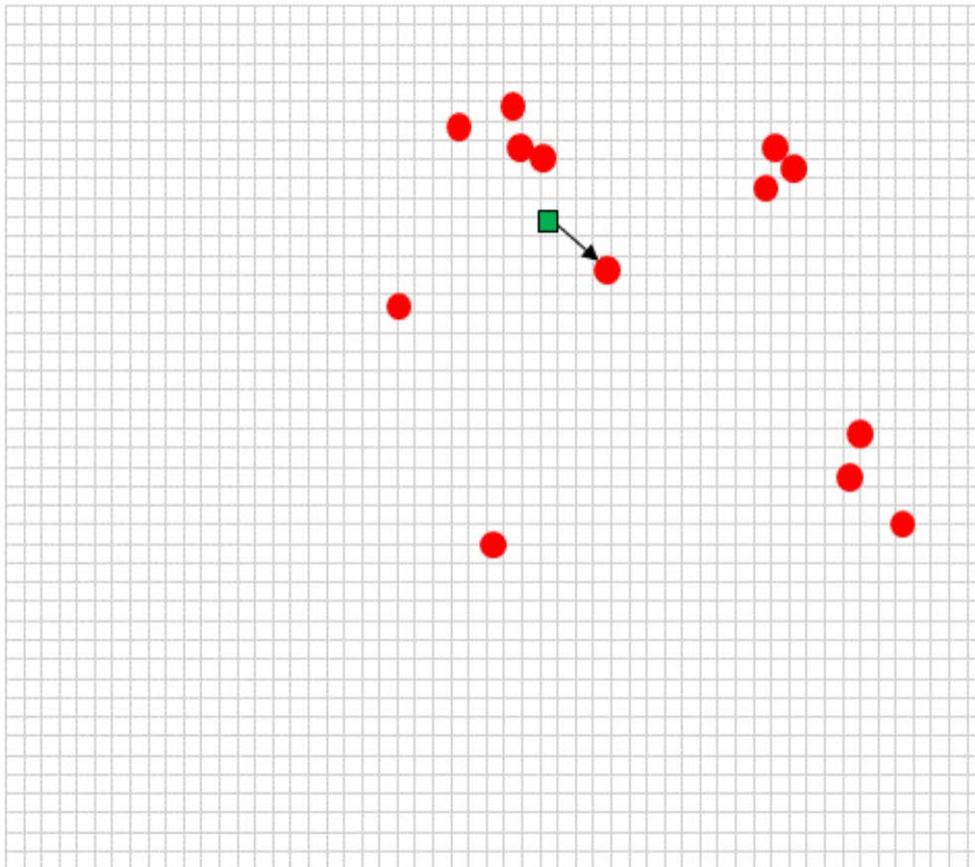
-
1. Virginia Department of Social Services. "[Abuse and Neglect by Locality Reports](#)", 2013-2017.
 2. Daley, Dyann, Michael Bachmann, Brittany A. Bachmann, Christian Pedigo, Minh-Thuy Bui, and Jamye Coffman. "Risk terrain modeling predicts child maltreatment." *Child abuse & neglect* 62 (2016): 29-38.
 3. Durlauf, Steven N. "Neighborhood effects." In *Handbook of regional and urban economics*, vol. 4, pp. 2173-2242. Elsevier, 2004.
 4. We do not map the raw point location data for privacy reasons.
 5. Daley, Dyann, Michael Bachmann, Brittany A. Bachmann, Christian Pedigo, Minh-Thuy Bui, and Jamye Coffman. "Risk terrain modeling predicts child maltreatment." *Child abuse & neglect* 62 (2016): 29-38.
 6. Privacy controls are developed to simultaneously prevent the re-identification of maltreatment events, while presenting a useful cartographic narrative. This is done for the public version of this work, but these controls are unnecessary for the client's internal use. Two general rules are followed. First, any point that is alone in a grid cell is removed from the cartographic output. Second, points that remain are randomly offset one quarter to one half mile (i.e. "jittered"),

meaning that some random noise is introduced to the XY coordinates of each point. This approach had to reconcile the fact that jittering offsets points with duplicate geometries, such as apartment buildings where multiple maltreatment events may be recorded for a single address. In such cases, duplicate geometries are removed before jittering. Without this step, the resulting maps would appear to visualize far more discrete point events than actually exist in the original data. In our maps including Figure 1.5 for instance, any point event that appears alone in a grid cell was originally in a cell with greater than 1 event and is only now unto itself as the result of jittering.

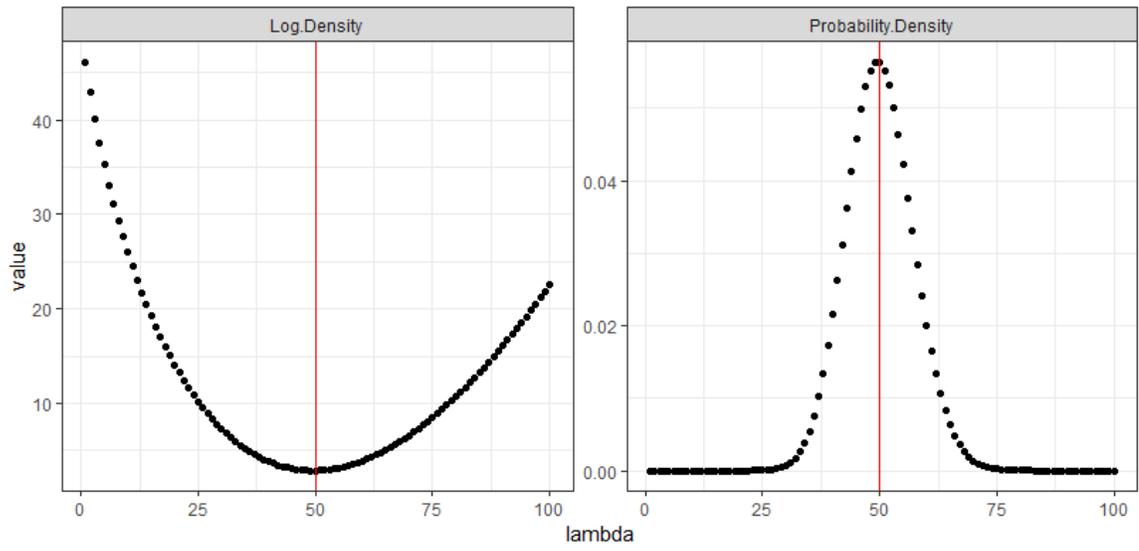
7. Hurley, Dan. "Can an Algorithm Tell When Kids Are in Danger?"; *The New York Times Magazine*, Jan. 2, 2018.
8. Li, Jin and Andrew D. Heap. "A Review of Spatial Interpolation Methods for Environmental Scientists." *Geoscience Australia*, 2008.
9. Silverman, B.W. "Density Estimation for Statistics and Data Analysis." In *Monographs on Statistics and Applied Probability*, London: Chapman and Hall, 1986; Cressie, Noel A.C. "Statistics for spatial data." 1993; Dearmon, Jacob and Tony E. Smith. "Gaussian Process Regression and Bayesian Model Averaging: An Alternative Approach to Modeling Spatial Phenomena." *Geographical Analysis* (2016): vol. 48, p.82-111; Brunson, Chris, Stewart Fotheringham, and Martin Charlton. "Geographically Weighted Regression - Modelling Spatial Non-Stationarity." *Journal of the Royal Statistical Society. Series D (The Statistician)* (1998): vol. 47, no. 3, pp.431-443; Haining, Robert, Jane Law, and Daniel Griffith. "Modelling small area count in the presence of overdispersion and spatial autocorrelation." *Computational Statistics & Data Analysis* (2009): vol. 53, no. 8, p.2923-2937; Oliveira, Sandra, Friderike Oehler, Jesus San-Miguel_Ayanz, Andrea Camia, and Jose M.C. Pereira. "Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest." *Forest Ecology and Management* (2012): vol. 275, p.117-129.
10. Oliveira, Sandra, Friderike Oehler, Jesus San-Miguel_Ayanz, Andrea Camia, and Jose M.C. Pereira. "Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest." *Forest Ecology and Management* (2012): vol. 275, p.117-129; Leib, Mareike, Bruno Glaser, and Bernd Huwe. "Uncertainty in the spatial prediction of soil texture: Comparison of regression tree and Random Forest models." *Geoderma* (2012): vol. 170, p. 70-79; Knudby, Anders, Alexander Brenning, and Ellsworth LeDrew. "New approaches to modelling fish - habitat relationships." (2010): vol. 221, no. 3, p.503-511; Harris, Matthew D., Robert G. Kingsley, and Andrew R. Sewell. "Archeological Predictive Model Set." *Pennsylvania Department of Transportation*, March 2015; Caplan, Joel M., Leslie W. Kennedy, and Joel Miller. "Risk Terrain Modeling: Brokering Criminological Theory and GIS Methods for Crime Forecasting". *Justice Quarterly* (2011): vol. 28, no. 2; Steif, Kenneth, Jordan Butz, and Annie Streetment. "Predicting Spatial Risk of Opioid Overdoses in Providence, RI." May 2018.
11. Congalton, Russell G. "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data." *Remote Sensing of Environment* (1991): vol. 37, p.35-46; Vicente-Serrano, Sergio M. M. Angel Saz-Sanchez, and Jose M. Cuadrat. "Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature." *Climate Research* (2003): vol. 24, p.161-180.; Li, Jin and Andrew D. Heap. "A Review of Spatial Interpolation Methods for Environmental Scientists." *Geoscience Australia*, 2008; Brenning, Alexander. "Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sperrorest." In *2012 IEEE International Geoscience and Remote Sensing Symposium*, July 2012.
12. Kuhn, Max and Kjell Johnson. "Applied Predictive Modeling." 2013; Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. "The Elements of Statistical Learning: Data Mining, Inference, and Prediction." 2008.
13. Daley, Dyann, Michael Bachmann, Brittany A. Bachmann, Christian Pedigo, Minh-Thuy Bui, and Jamye Coffman. "Risk terrain modeling predicts child maltreatment." *Child abuse & neglect* 62 (2016): 29-38; Caplan, Joel M., Leslie W. Kennedy, and Joel Miller. "Risk Terrain Modeling:

Brokering Criminological Theory and GIS Methods for Crime Forecasting". Justice Quarterly (2011): vol. 28, no. 2.

14. Tomlin, Dana C. "Geographic Information Systems and Cartographic Modeling." 1990.
15. Caplan, Joel M., Leslie W. Kennedy, and Eric L. Piza. "Risk Terrain Modeling Diagnostics Utility User Manual." Rutgers Center on Public Security, 2013.
16. The first approach illustrated the right image in this figure, measures the distance from each grid cell to its one nearest neighbor risk/protective factor. This relative measure of exposure is more useful than a simple aggregate count of risk/protective factors, however, it may be biased if the one nearest neighbor is an outlier. Thus, as illustrated by the left image in this figure, our third approach is to measure the average distance from each grid cell to its n nearest neighbor risk/protective factor, in this case $n = 3$. How many nearest neighbors is the 'correct' number to use? Consider the following three notions: First, for a given risk/protective factor, the 'correct' number of nearest neighbors leads to the best prediction. An optimal (but computationally intensive) predictive algorithm is one that would iteratively test each combination of risk/protective factors by each combination of n , selecting those the combination which leads to the most optimal model. Second, assumptions about n can be made according to how rare a given risk/protective event occurs in space. For instance, a feature describing exposure to crime would likely have a higher n than a feature describing distance to the nearest school. This is because crime occurs more frequently. Finally, if the scale of phenomena varies across the study area, there is likely no one correct nearest neighbor parameter. For example, imagine a city composed of a dense downtown and a less dense neighborhood along the periphery. Despite being in the same jurisdiction, the scale of the social relationships occurring in each place are fundamentally different because the scale of the built environment is fundamentally different. The figure provides an example of all three feature engineering. approaches.



17. The final step in the feature engineering phase is to normalize the contributing risk and protective features by putting all of the measurements onto a relative scale. Data are normalized in two steps: 1) centering the range of measurements for each feature on zero by subtracting the mean, and 2) by scaling the measurements of each feature by dividing by the measurements standard deviation. The end result is that each feature has a mean of zero and each value is a z-score. A primary benefit of putting each of the measurements on the same scale is to help the machine learning algorithms compute more efficiently. Additionally, it leads to scaled model coefficients that are relative to the average measurement of the feature.
18. Tukey, John W. "Exploratory Data Analysis." 1977.
19. Bivand, Roger and Gianfranco Piras. "Comparing Implementations of Estimation Methods for Spatial Econometrics. Journal of Statistical Software"(2015): vol. 63, no. 18, p.1-36.
20. Smith, Tony E. "Notebook on Spatial Data Analysis." 2016, <https://www.seas.upenn.edu/~ese502/#notebook>.
21. Delignette-Muller, Marie Laure and Christophe Dutang. "fitdistrplus: An R Package for Fitting Distributions. Journal of Statistical Software" (2015):vol. 64 no. 4, p.1-34
22. Wright, Marvin N. and Andreas Ziegler. "ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R." Journal of Statistical Software (2017): vol. 77 no. 1, p.1-17.
23. Bivand, Roger and Gianfranco Piras. "Comparing Implementations of Estimation Methods for Spatial Econometrics. Journal of Statistical Software"(2015): vol. 63, no. 18, p.1-36.
24. There 148 neighborhoods in Richmond. With each taking a turn as a hold out, there are 296 models calculated for the GLM and Random Forest models. As explained above, the Spatial Durbin model is estimated once, as the spatial weights matrix prevents LOGOCV approach. This yields a total of 297 models.
25. For example, if the predicted number of events is 5 and the observed number of recorded events is 7, then the MAE is 2. The absolute part describes that the error is the same whether it is positive or negative. The same prediction of 5 events and an observed value of 3 also has an MAE of 2. The MAE returns a single value no matter how many estimates are made. If the predictions for three different fishnet cells is 5, 8, and 10 with corresponding observed values of 3, 9, and 12, then the absolute errors are 2, 1, and 2 leading to an MAE of 1.67.
26. The graph below gives an example of how the log density and probability density measurements of the Logarithmic Score relate. In this plot, the two measurements are calculated for the same scenario in which the predicted value of maltreatment events is 50 (at the red vertical line) and the observed number of events is represented by each of the black dots from zero to 100. At the left hand side of each is the case where the recorded number of events is zero or close to zero. At this location the log density is very high relative to the baseline of zero and the probability density is near zero. In the middle of each chart where the estimated and expected values are close to the same, the log density is near zero and the probability is at its maximum. As the observed values move to the right away from the estimate, the log density again goes up and the probability goes down. The important takeaway from this plot is that when measured as a log density, the best estimate is always the one closest to zero which makes it very easy to compare models. However, the scale of the log density is not very intuitive. On the other hand, measuring on the probability side can lead to difficulties in comparing models if they are fit to different sets of data. However, it is on an intuitive scale of probability between zero and one. Given that all of the models here are fit to the same dataset, the probability density measurement is selected to represent the Logarithmic Score.



27. Anselin, Luc. "Local Indicators of Spatial Association - LISA." *Geographical Analysis* (1995): p.93-115