Predicting Fatal Opioid Overdoses in Rhode Island

The novel use of Emergency Medical Services, Prescription Drug Monitoring Program, land use, and American Community Survey data to predict fatal opioid overdoses.

Jiaqi Dong
Nicholas Liu-Sontag
Brandon Pachuca
Yicong Wang

2021 Urban Science Intensive Capstone
NYU Center for Urban Science and Progress
Professor Victoria Alsina

Sponsored by
Professor Magdalena Cerdá
Professor Daniel B. Neill
Bennett Allen

Abstract

The opioid epidemic is one of the largest public health crises in the United States; since 1999, over 814,000 people have died from a drug overdose in the US. Rhode Island has been hit particularly hard and regularly has some of the country's highest overdose death rates. To improve forward-looking targeting of interventions the state seeks to utilize a prediction model to target areas at higher risk of an overdose outbreak. As a subset of a larger team working on this effort, we developed four models to predict overdose risk at the census block group level utilizing the following algorithms: gaussian processes, random forest, gradient boost, and graph convolutional networks. The first three of these achieved the project's baseline target performance and we believe the graph convolutional network shows promise. Our model results will be folded into the larger project and this information will then be supplied to the Rhode Island Department of Health (RIDOH) and community organizations to deploy targeted resources to higher-risk areas. If this method proves successful, it could serve as a model for states and municipalities across the country to identify and target interventions to reduce opioid overdose risk.

Introduction

Opioid use disorder is a public health crisis in the United States. Since 1999, more than 814,000 people have died from a drug overdose, the vast majority of which involved opioids.[1] The epidemic began in the 1990s due to the over-prescription of opioids tied to aggressive marketing tactics by drug companies.[2] This epidemic has dramatically changed families and communities across the U.S and has exacted a staggering economic toll:

estimates show the epidemic's total costs exceed 500 billion dollars annually.[3] With the epidemic growing exponentially, there is a dire need for expanded intervention to treat individuals with opioid use disorder and prevent opioid overdose deaths.

Three waves characterize the epidemic: an initial wave of prescription opioid overdoses followed by a second wave of illicit opioid overdoses (namely, heroin) in 2010. The second wave was largely a result of municipalities intervening on prescription opioid use and the limitation of prescriptions by doctors, which made illicit opioids more readily available at a cheaper cost.[4] In 2013, a third wave of illicit semi-synthetic opioid overdoses primarily in the form of fentanyl, a deadly opioid often mixed into heroin, presented itself.[5]

There has been significant research on the underlying factors that contribute to opioid use disorder. Many of these are rooted in social determinants such as poverty, isolation, and social upheaval. Medical factors, including the concurrent use of sedatives/hypnotics (e.g., benzodiazepines) and mental health disorder comorbidities, also contribute to overall risk.[6,7]

Since 2016, there has been a focus on data collection for monitoring and targeting interventions starting with the Centers for Disease Control's (CDC) creation of the Enhanced State Opioid Overdose Surveillance (ESOOS). This program assisted states in setting up opioid overdose data surveillance systems and increased the accuracy and timeliness of overdose reporting. The CDC's Overdose Data to Action (OD2A) initiative, the State Unintentional Drug Overdose Reporting System (SUDORS), was also created to "support jurisdictions in collecting high quality, comprehensive, and timely data on nonfatal and fatal overdoses".[8] This SUDORS data is a valuable resource for informing prevention and response efforts.

In addition to the CDC's effort, multiple federal agencies have offered grant opportunities to local governments in response to the nationwide opioid crisis.[9] These grants allowed municipalities to advance research on the opioid crisis and deploy interventions such as naloxone kit distribution for overdose reversal, targeted response model development, and community outreach.[1*]

There is a growing body of research applying machine learning to social problems such as this one. Random forest (RF) models have been used in solving spatio-temporal urban issues such as crime prediction.[10, 11] This method has been shown to perform efficiently in predicting imbalanced spatio-temporal event prediction, which is suitable for our problem where the opioid overdoses fatality distribution varies significantly between urban and non-urban areas across RI.[10] There are also applications of gaussian processes (GP) in addressing social problems: overdose cluster detection in NYC,[12] epidemiological investigations on flu and seasonal influenza,[13] 311 reports of damaged trees and sewer issues,[14] and crime activity[15] prediction. Lastly, gradient tree boosting has been successfully used to predict cardiovascular events, delirium and sepsis events.[16, 17, 18]

In the realm of deep learning, long-short term memory neural networks have been used to effectively forecast other public health crises including the spread of coronavirus,[19] human immunodeficiency virus infection,[20] influenzas,[21] and diabetes detection[22] to offer timely and effective support for public health interventions. The more novel technique of graph convolutional networks (GCNs) has shown promising results for non-grid spatio-temporal prediction in the areas of transportation and short-term weather forecasting.[23, 24, 25] As opposed to using recurrent neural network (RNN) layers for temporal forecasting, these studies use temporal convolution via a temporal GCN layer to model temporal neighbors and capture patterns.

[1*] Naloxone is a medication that can reverse opioid overdoses, naloxone kits are often distributed to community members, opioid users, and medical personnel to administer in the event of an overdose.

In recent years, several studies have applied machine learning techniques with the intent of predicting opioid overdoses. Much of this work has focused on patient-level prediction: both Lo-Ciganic and Dong et al. used neural networks to categorize patients at varying risk levels.[26, 27] There has been far less research on temporal prediction of the epidemic, however. Sumetsky et al. used logistic growth models to predict overdose growth more accurately.[28] Spatio-temporal prediction on this topic is also sparse; however, in a novel use of social data, Young et al. accurately predicted hospital visit increases using internet search data.[29]

Problem Definition

Opioid-related deaths have significantly increased over recent years in the United States: annual deaths involving opioids increased by four times since 2000.[30] This crisis is a public health emergency; datasets to characterize the problem are only recently beginning to show value as predictors of community risk. Our literature review found health experts using various indicators to predict fatal opioid overdose events (OOEs) such as; emergency service room visits,[31] census tract-level demographic analysis,[7] patient medical histories,[32, 26] internet search terms,[29] and prescription drug monitoring programs.[33] The novel process of using high-resolution spatial and temporal municipal datasets has just recently been explored to predict opioid overdose events for targeted health and policy interventions in communities.

Problem statement

In 2020, Rhode Island recorded over 350 opioid-related deaths, and the state has regularly shown some of the highest overdose death rates per capita in the country. In 2015, the Governor created the Overdose Prevention and Intervention Task Force and set a target of reducing opioid overdose deaths by one-third within three years. The Task Force has identified a four-pronged approach to addressing the epidemic:[34]

- Treatment: Increase the number of people receiving medication-assisted treatment for opioids
- Rescue: Increase the number of naloxone kits distributed across the state
- Prevention: Decrease the number of people receiving opioid & benzodiazepine subscriptions every year
- Recovery: Increase the number of peer recovery coaches and contacts

Extracting trends and understanding areas at higher risk of overdose is critical to effectively implementing the four-pronged approach. Thus far, Rhode Island has faced challenges in effectively prioritizing areas facing higher risk of opioid overdose events before an outbreak. The state has historically analyzed trends at the local municipality level but has been unable to create a highly spatial and temporal resolution of risk across the state. Though various interventions have been deployed in the field, for example, 25,742 naloxone kits had been distributed by the end of 2020, the increasing trend of opioid overdose deaths continues.

As part of a National Institute of Health (NIH) funded project, deemed “PROVIDENT”, RIDOH, in partnership with the Task Force and researchers from Brown University and New York University (NYU), seek to target fatal opioid overdose outbreaks in advance using predictive modeling to allocate resources amongst the state’s census block groups (CBGs). These models will predict the CBGs with the highest opioid fatality risk on a six-month rolling basis. This information will be distributed to RIDOH and community organizations to deploy targeted interventions to prevent overdoses. Community organizations will have the power to choose what type of intervention, such as community outreach and educational workshops, is suitable in local neighborhoods.

As a subset of PROVIDENT, our goal is to predict fatal opioid overdose risk in Rhode Island by CBG on a six-month rolling basis. We did a comparative analysis of ensemble, gaussian processes, and deep learning models evaluating the accuracy, interpretability, and computational effectiveness in predicting the rank of each CBG's risk of overdoses.

Our research questions are:

1. Which machine learning models provide the highest accuracy for predicting fatal opioid overdose events in Rhode Island by Census Block Group?
2. What specific features are strong predictors of fatal opioid overdoses?

Data and methods

Datasets

Our target variable is the normalized rank of overdose fatalities in each CBG during the six-month target period. We use normalized rank instead of overdose counts as the target variable because the project is focused on how CBGs rank relative to one another, and not the absolute count of overdoses.[2*] This statistic is aggregated semi-annually from the overdose deaths reported in SUDORS from the first half of 2016 (2016.1) through the first half of 2020 (2020.1).

In addition to the SUDORS dataset, we use the following datasets as feature sources:

1. ACS 5-year Estimates
   Demographic and socio-economic indicators at the CBG level using ACS 5-year estimates. After reviewing RIDOH's press release and workforce update,[35, 36, 37, 38, 39, 40, 41] literature specifying the vulnerabilities of subgroup populations,[42, 43, 44] and interviews with public experts on the team, we identified a selected set of features for further modeling effort (Appendix A).
2. Emergency Medical Service (EMS) overdose-related runs
   The EMS dataset is a restricted dataset available through Brown University, provided by RIDOH. The dataset records all EMS calls in RI related to overdoses, a key indicator which captures nonfatal OOEs. The EMS calls are aggregated in six-month windows by CBG. The research team at Brown University has indicated that this dataset may have inconsistent reporting and data collection across local jurisdictions but the variations between jurisdictions will likely have a negligible impact on finding target CBGs.
3. Prescription Drug Monitoring Program (PDMP)
   The PDMP is an electronic database maintained by the CDC. It tracks controlled substance prescriptions in every state. The system provides healthcare practitioners with a comprehensive platform to check if a patient is possibly over-prescribed opioids and whether the patient is using other prescriptions that may increase the risk of an overdose.[45] The database records the prescription information, including drug type and product information, amount, daily doses, sequences, and patient information. Opioid prescription data are spatially aggregated by CBG and temporally every six months. The indicators from this dataset in Appendix A were informed by previous research.[33, 46, 47]
4. Land use

[2*] We also experimented with predicting overdose counts and overdose share (percent of overdoses occurring in the given CBG). However, normalized rank performed the best.

Using multiple spatial data sources from Rhode Island Land Use 2025, Rhode Island Public Transit Authority (RIPTA), the Brown University team compiled a set of features reflecting local infrastructure availability. Public health expert interviews and literature provided a basis for selecting land use or facilities for our models.[48]

5. Public access
The Brown University team compiled the count of the following businesses, service providers and facilities in each CBG, using datasets acquired through RI Department of Business Regulation, RI Division of Taxation Department of Revenue, RI Department of Health Licensing, and National Association for the Education of Young Children. Some of these public access indicators could be a proxy for population health issues such as tobacco and alcohol consumption, mental health issues, and pain relief. It is worth noting that access is not limited to the CBG boundaries and, thus, aggregating by CBG does not perfectly reflect the actual demand of such services in each CBG.

Descriptive Statistics

The mean overdose deaths per CBG throughout our dataset remains relatively stable from 2016.2 until 2020.1, where there is a spike of overdose fatalities with higher disparities among CBGs in 2020.1 (Figure 1). Between 2016.2 through 2020.1 the mean fatal overdose counts were around 0.18-0.19, while the mean rose to 0.23 in 2020.1. In addition, the histogram below indicates that the distribution of fatal overdose counts are extremely skewed towards the left (Figure 2). Overdose deaths are concentrated in fewer than 25% of the CBGs.

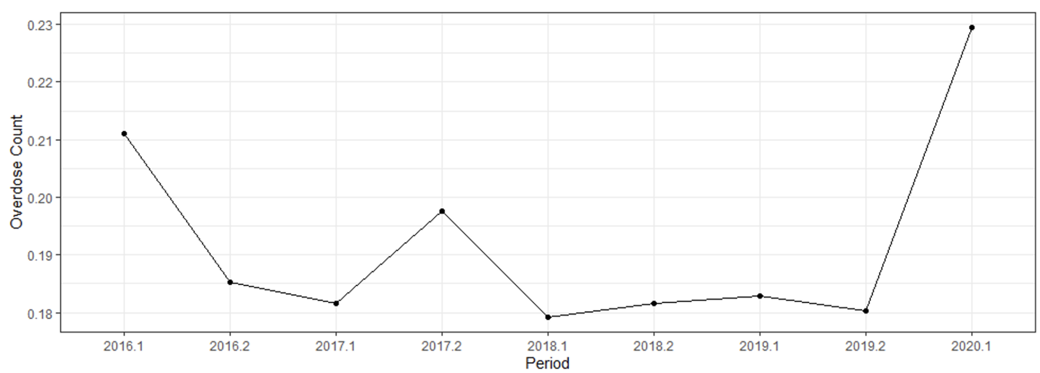

Figure 1: Average overdose count per CBG over different observation periods

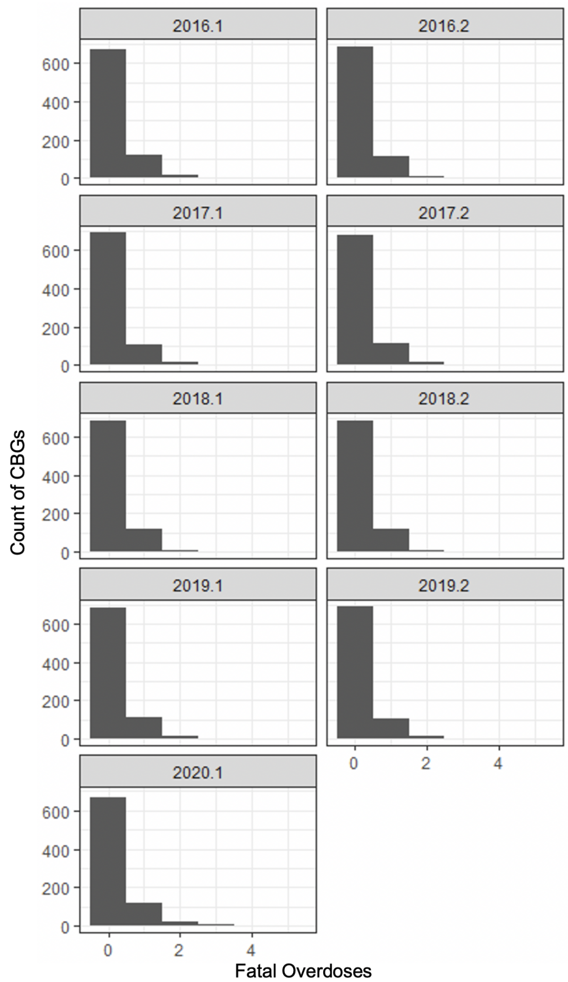

Figure 2: Distribution of overdose count over different observation periods

ACS indicators are relatively static since 5-year estimates are produced using the samples of the rolling five years. Demographic indicators show that RI is extremely homogeneous. For example, less than 25% of the CBGs have a share of white population to be lower than 55%; more than 25% of the CBGs have more than 90% of the population as non-Hispanic white.

Descriptive statistics of some PDMP indicators reflect previous efforts in combating the opioid crisis, particularly in overdose prevention medicine distribution and responsible opioid prescription. For instance, over the years, there has been an increasing number of naloxone prescribed to patients (Figure 3) and a decreasing number of large-dose (> 90 morphine milligram equivalents) opioid prescriptions that overlap with Benzodiazepine uses (Figure 3).

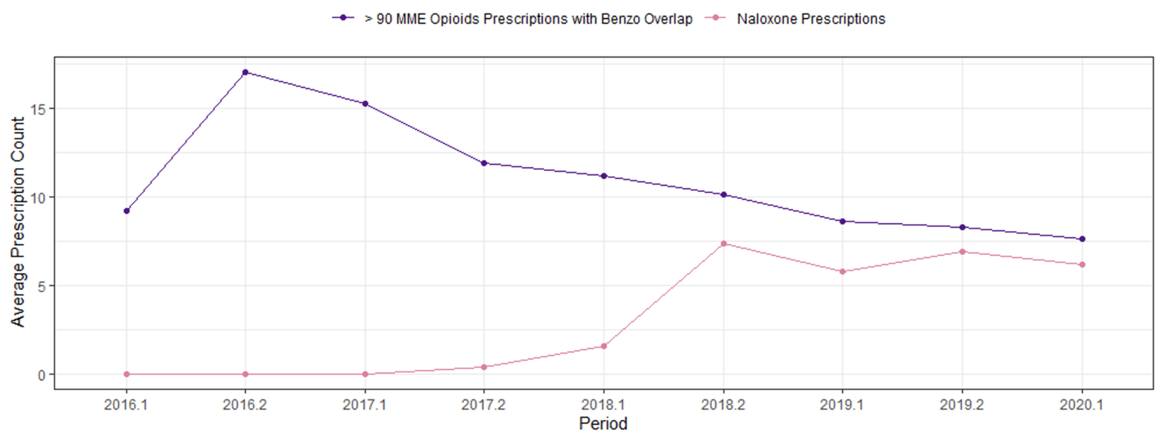

Figure 3: Average prescription count per CBG over different observation periods

Methods

Feature Selection

We first performed a qualitative review on the features from ACS, EMS, PDMP and Land Use datasets informed by literature and expert interviews to narrow down the number of features to half of the original size, this feature set is referred to as feature set A (see Appendix A, Table 1). Feature set A contains 141 features. From this, we narrowed the feature list further using a variety of methods to three additional feature sets (Appendix A, Table 2):

- Feature set B: 59 features
  - Features were removed, with a strict filter, qualitatively based on literature review and expert interviews.
- Feature set C: 109 features
  - Features were removed utilizing *sklearn*'s recursive feature extraction algorithm and a naive random forest, utilizing overdose counts as the target variable.
- Feature set D: 64 features
  - Features were removed utilizing *sklearn*'s recursive feature extraction algorithm and a naive random forest, utilizing normalized overdose rank as the target variable.

The features removed from B, C and D feature sets are mainly “overlapping” features with others, for example, total count of opioid prescriptions dispensed versus total days of supply dispensed.

Principal Component Analysis

Selecting features reduced the dimensionality of the space by about half, however, our final feature sets still contain highly correlated features. To address this, we performed both linear principal component analysis (PCA) and kernel PCA (KPCA) using an eight-degree *poly* kernel for model testing. Using linear PCA, 50 components were required to explain 80% of the variance while using KPCA the first six components explained 80% of variance.

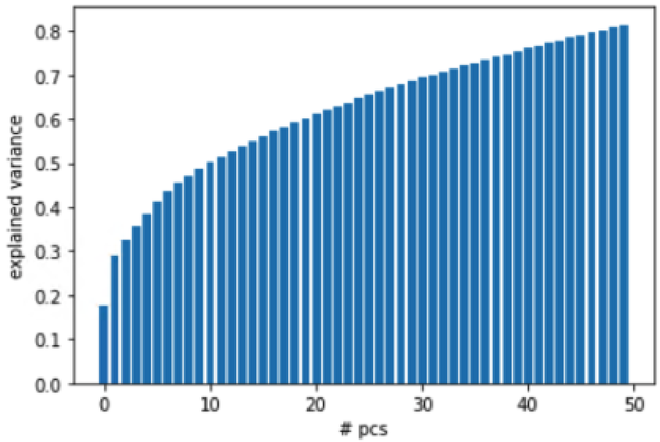

Figure 4. Linear PCA explained variance

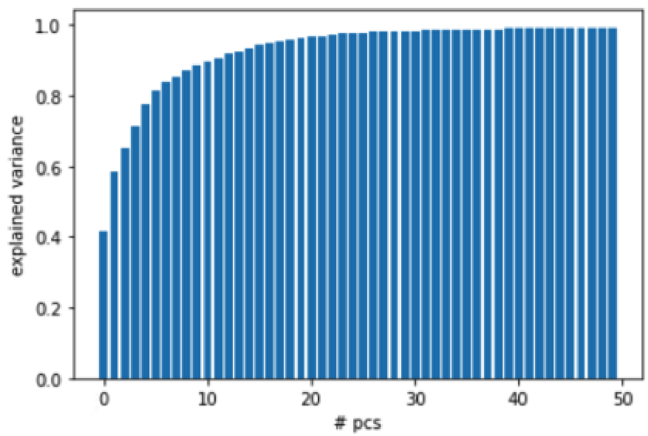

Figure 5. Poly KPCA explained variance

Spatial Aggregation

To incorporate the potential impact of CBG spatial auto-correlation due to adjacency - we performed a weighted average spatial aggregation of features for each CBG. First, we developed a normalized, inverse distance matrix based on euclidean distance between CBG centroids. Higher values in the matrix corresponded to closer CBGs, with values proportional to 1 / distance. For each CBG in each timestep, we then multiplied the features of all other CBGs by the inverse distance to the given CBG and summed the result, providing an average of neighboring features weighted by distance.

Model Performance Evaluation & Performance Goal

Traditional loss and error functions (e.g. MSE loss, R^2) are being used for feature selection and deep learning and gaussian process model tuning. We employ a unique method for final model performance evaluation: the percent of actual overdoses captured in CBGs targeted by the model (i.e. capture rate). Utilizing a target scenario provided by RIDOH wherein the model targets at least one CBG in each town (“lightly constrained”) and targets 20% of the highest risk CBGs. This evaluation metric is referred to as LC20.

Brown University’s grant application set an LC20 capture rate of 40% as a baseline goal to deploy model predictions in the PROVIDENT trial occuring in October 2021. The goal of our models is to achieve the LC20 target of 40% or higher.

Modeling

The project team explored the application of the following models using data from 2016.2 through 2019.2 and testing on 2020.1 and evaluated model performance using the LC20 evaluation criteria. For all models, we use *t-1* (or *t-1* and *t-2*) features to predict the overdose outcome at time *t*.

1. Random Forest
2. Gaussian Processes
3. Graph Convolutional Neural Networks
4. Gradient Boosting

All models predict normalized overdose rank, however, we experimented with predicting overdose counts and overdose share (percent of overdoses occurring in the given CBG) with normalized rank performing the best.

Random Forest

We used the top twenty-five features derived from feature set A using *t-1* and *t-2* timesteps via feature importance eliminations. We did two rounds of training, validation, and testing to get a model that is more generalizable. The first round training set uses 2017.1 through 2019.1 and hyperparameter tuning is performed using a validation set from 2019.2.[3*] The second round used 2017.1 through 2018.2, tuned parameters with 2019.1, and tested on 2019.2. The highest out-of-sample (average over two test sets) capture rate achieved with random forest was 40.2%.

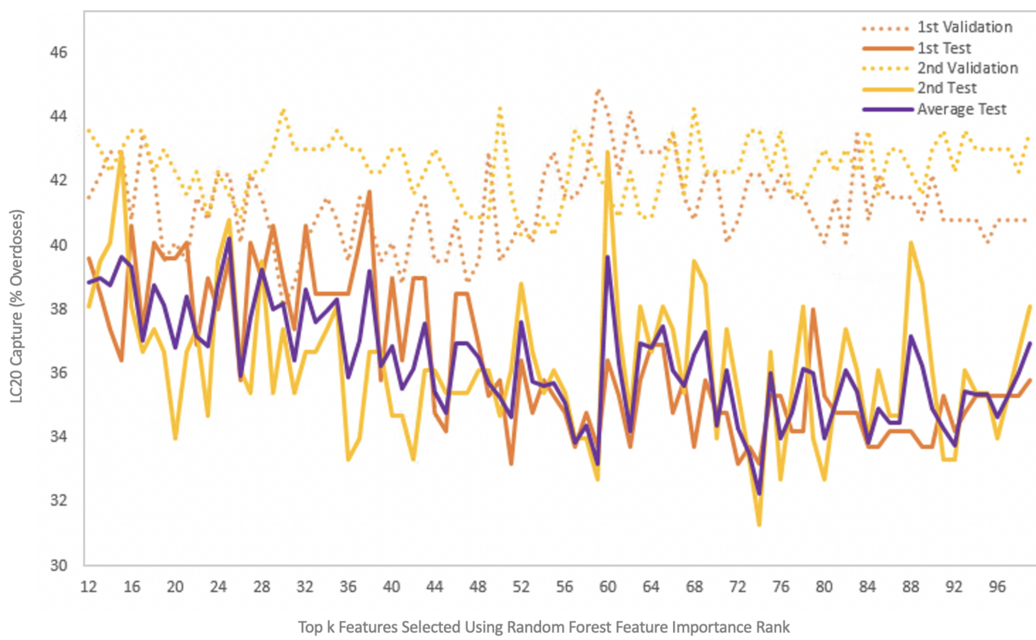

Figure 6: Validation and test sets evaluation result (LC20) with top k features

Utilizing dimensionality reduction, RF models performed below 40% LC20. Applying the top-k feature importance method to the KPCA dataset did not improve performance. RF model performance with linear PCA was also poor (< 35%). Lastly, feeding in CBG's centroid coordinates or spatial neighbors as features did not boost performance.

3* To interpret the period 2017.1 here: it means using the features from 2016.1 and 2016.2 to model the target variable from 2017.1

Gaussian Process

The GP model achieved a 40.1% LC20 using dataset D *t-1* with no spatial adjacent information. The model utilized the following kernels: radial basis function (RBF), white kernel, and rational quadratic kernel. The kernels were tuned using a restart optimizer function to reduce the log likelihood during training. PCA and KPCA resulted in a 2% - 3% decrease in capture rate compared to non-abstracted features.

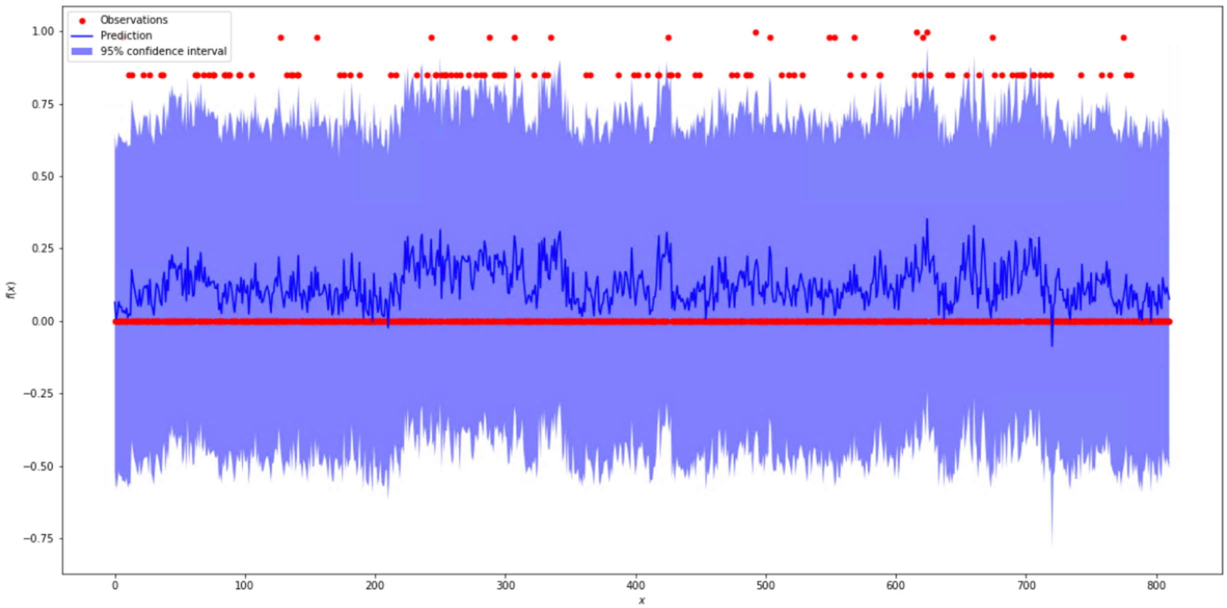

Figure 7. Gaussian Process Dataset A Overdose Count

Neural Networks

Despite the lack of interpretability of neural networks, the project sponsors have indicated they may be acceptable if they provide substantially more accurate predictions. An initial exploration of GCNs based on the framework proposed by Stanczyk & Mehrkanoon (2021) did not yield strong results. However, a from-scratch GCN model using the below architecture did show promise.

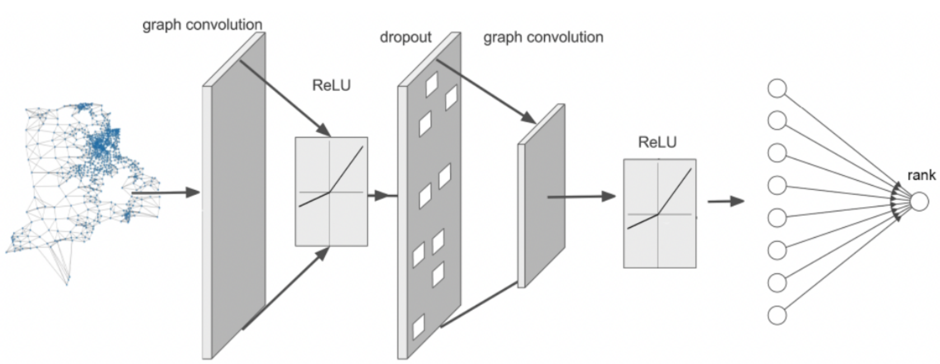

Figure 8: GCN Architecture

The model did not capture the temporal nature of our data, but rather treats all timesteps equally and focuses on finding patterns between neighbors and CBGs. Utilizing *PyTorch Geometric* we first built a graph of the data finding the *k* nearest neighbors (knn) to each CBG. We tested two methods for finding neighbors: geographic centroids and feature similarity. We fed the resulting graph to a series of graph convolution layers (*GCNConv*) followed by dropout and ReLU activation functions. Finally, the results were fed to a dense layer to predict the final rank.

The best performing GCN model used dataset A with spatial aggregation of all features. The knn graph was built by feeding in the entire feature space (and not geographic centroids) to find the five nearest neighbors. Hence, the model found the five "neighboring" CBGs within the feature space and used their features (and their spatial aggregates) as predictors for each CBG in question.

Gradient Boosting

We trained a gradient boosting (GB) model using the *XGBoost* library on normalized *t-1* features, linear PCA *t-1* features, and kernel PCA *t-1* features on all four datasets independently with and without spatial aggregation to predict normalized rank. The model utilized 2016.2 through 2018.2 as a training set and 2019.1 and 2019.2 as a validation set to tune parameters. The model showed the best performance using the first 16 KPCA components without spatial aggregation on dataset A.

Results

The table below listed the percentage of fatal overdoses captured using LC20 evaluation criteria. See Appendix B for an exhaustive table of model specifications. GP, RF and GB all show promise in predicting fatal OOEs per CBG.

Model	**LC20 Capture**
Random Forest	40.2%
Gaussian Process	40.1%
Gradient Boost	40.1%
Graphic Convolutional Network	37.4%

Table 1. Best model performance

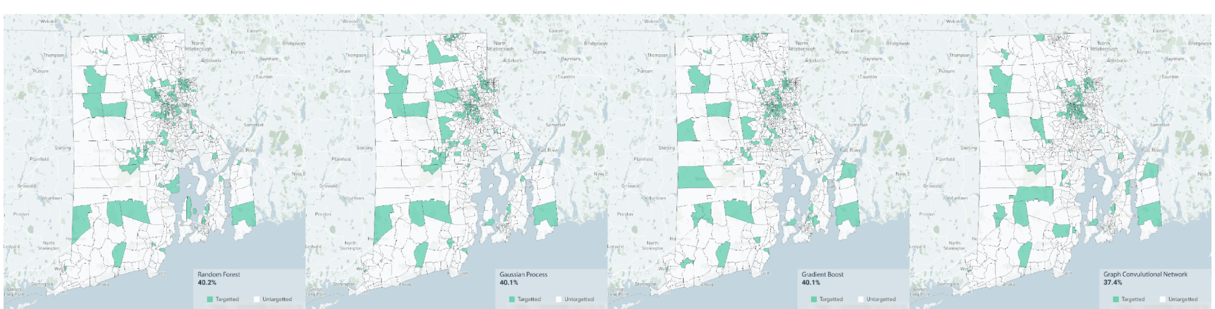

Figure 9: Model prediction results (Colored CBGs are selected by LC20., From left to right, RF, GP, GB and GCN)

Given the similar LC20 performance of GP, GB and RF, we are unable to recommend a single model for use in the trial. We believe given more observations, a single model may distinguish itself and we recommend the study continue to test these models with the release of more data. Seventy-five CBGs (46% out of the 163

CBGs within the LC20 threshold) were selected by the three top-performing models. More than 80 CBGs were selected by two out of the three models.

Random forest calculates feature importances which provides insight into the most influential features. These are measured with the average impurity decrease within each tree. In the best random forest model there are no feature importances above 0.03, indicating that no single feature was found to be highly determinant of the target variable.

Feature	**Importance**
Total persons receiving buprenorphine last 7 days (t-1)	0.021
% renter occupied units (t-2)	0.020
Total EMS calls (t-1)	0.020
% population hispanic (t-1)	0.017
Total days supply buprenorphine (t-2)	0.016

Table 2. Random forest top feature importances

Discussion

The best models captured roughly 40% of all opioid related fatalities, achieving the project's baseline performance target. However, the chosen evaluation method, although applicable to the research question, obscures the actual accuracy of the models. The best models obtained R^2 scores of approximately 0.05, a marginal improvement over randomly selecting CBGs. That the performance objective can be achieved with objectively inaccurate models is indicative of the nearly stochastic nature of the target variable. In short, although the models perform adequately well for the problem in question, they should not be taken as highly accurate in their own right.

Although objective model accuracy was low, examining the random forest most important features suggests the model did find relevant patterns in the data. Two of the top five most important features measured the distribution of buprenorphine, a drug used to treat opioid addiction; one measured the total number of overdose EMS calls to a CBG; one measured the number of renter occupied units, an indicator of overall socio-economic status; and one measured the percentage of hispanic-identifying individuals, a population with rising opioid overdose deaths.[49] These important features are valid and consistent with existing research. We believe that with more data that reinforces these patterns, model performance and accuracy will improve, a primary takeaway from this study.

There were, however, many features that were found to be unimportant despite our expectations. For example, the overdose ranks of CBGs in prior periods (*t-1* and *t-2*) were the 181st and 146th most important features and, hence, were excluded from the final random forest model of top 25 features. We had expected the overdose history of the CBG would be a primary predictor. PDMP indicators also did not perform well with random forest methods in pre-existing models made by the PROVIDENT team, however, in our models they show up as top important features.

Challenges & Limitations

The challenges outlined below created limitations on our modeling.

- Dataset size: By far the most significant modeling challenge, the project's 6,488 observations in 8 time periods is too small for even the best machine learning models to train on. We expect model performance to improve as the project's modeling continues beyond the scope of this study and obtains more data.
- Spatial library availability: The Stronghold[4*] environment did not have the capability to run spatial packages like Shapely or Geopandas, limiting our ability to perform spatial analytics.
- Non-identically distributed variables: We found patterns in the data that persisted for only some time periods as a result of state interventions (e.g. naloxone distribution increases) or broader public health trends (e.g. COVID-19 pandemic). These trends hamper the ability of our models to identify consistent patterns.

Each six month time period of observations presents opportunities to investigate emergent trends. We discovered trends which existed in 2018 and 2019 that did not continue into 2020. Our hypothesis is this could be due to the COVID-19 pandemic which affected data collection, data accuracy, and population behaviour. Additional research is suggested to account for new trends presented in future data releases.

Conclusion

In achieving the LC20 target of at least 40%, our models show promise in leveraging EMS, ACS, PDMP and land use datasets to predict fatal OOEs by CBG. Despite their relatively low objective accuracy, they are able to achieve performance sufficient for real world deployment. Furthermore, we expect performance to improve as trends continue to be studied and additional observations become available. However, the true efficacy of our models will not be revealed until the completion of the PROVIDENT study.

The methods used in this project are replicable in other state and local municipalities. The ACS, PDMP and SUDORS datasets are available nationally while the EMS dataset is becoming available in a growing number of cities across the country. With little adaptation, these models could be deployed elsewhere across the country to combat the opioid epidemic.

Acknowledgements

We would like to thank our project sponsors Dr. Daniel Neill (New York University), Dr. Magdalena Cerda (New York University Langone Health), and Bennett Allen (New York University Langone Health) for their guidance and collaboration on this project. We would also like to thank the following for their expertise and assistance: Dr. Brandon Marshall (Brown University), Dr. Will Goedel (Brown University), Claire Pratty (Brown University), Maxwell Krieger (Brown University), Konstantin Klemmer (University of Warwick), Abigail Cartus (Brown University), Rhode Island Department of Health, and NYU Center for Urban Science and Progress.

[4*] Stronghold is the secure, virtual research environment managed by Brown University on which all modeling was performed.

Bibliography

1. CDC WONDER. (n.d.). Retrieved April 04, 2021, from https://wonder.cdc.gov/
2. Mohamadi, A., Chan, J. J., Lian, J., Wright, C. L., Marin, A. M., Rodriguez, E. K., Nazarian, A. (2018). Risk factors and pooled rate of prolonged opioid use following trauma or surgery. *Journal of Bone and Joint Surgery, 100*(15), 1332-1340. doi:10.2106/jbjs.17.01239
3. The US opioid epidemic AND Prime-Age labor force participation. (2018). *Population and Development Review, 44*(2), 411-414. doi:10.1111/padr.12148
4. Jalal, H., Buchanich, J. M., Roberts, M. S., Balmert, L. C., Zhang, K., & Burke, D. S. (2018). Changing dynamics of the drug overdose epidemic in the United States from 1979 through 2016. *Science, 361*(6408). doi:10.1126/science.aau1184
5. Rudd, R. A., Aleshire, N., Zibbell, J. E., & Matthew Gladden, R. (2016). Increases in drug and opioid Overdose Deaths-United STATES, 2000-2014. *American Journal of Transplantation, 16*(4), 1323-1327. doi:10.1111/ajt.13776
6. Fulton-Kehoe, D., Sullivan, M., Turner, J., Garg, R., Bauer, A., Wickizer, T., & Franklin, G. (2015). (428) opioid poisonings in Washington state Medicaid: Trends, dosing, and guidelines. *The Journal of Pain, 16*(4). doi:10.1016/j.jpain.2015.01.347
7. Park, C., Clemenceau, J. R., Seballos, A., Crawford, S., Lopez, R., Coy, T., Hwang, T. H. (2021). A spatiotemporal analysis of opioid Poisoning mortality in Ohio from 2010 to 2016. *Scientific Reports, 11*(1). doi:10.1038/s41598-021-83544-y
8. Overdose data to action. (2021, March 02). Retrieved April 04, 2021, from https://www.cdc.gov/drugoverdose/od2a/index.html
9. https://docs.house.gov/meetings/IF/IF02/20200114/110367/HHRG-116-IF02-Wstate-Alexander-ScottMDN-20200114-SD002.pdf
10. Kadar, C., Maculan, R. & Feuerriegel, S. (2019). Public decision support for low population density areas: An imbalance-aware hyper-ensemble for spatio-temporal crime prediction. Decision Support Systems, Volume 119, Pages 107-117. Retrieved May 04, 2021, from https://doi.org/10.1016/j.dss.2019.03.001

11. Alves, L.G.A, Ribeiro, H.V. & Rodrigues, F.A. (2018). Crime prediction through urban metrics and statistical learning, *Physica A: Statistical Mechanics and its Applications*, Volume 505, Pages 435-443. Retrieved May 04, 2021, from https://doi.org/10.1016/j.physa.2018.03.084
12. Neill, D. B., & Herlands, W. (2018). Machine learning for drug overdose surveillance. *Journal of Technology in Human Services, 36*(1), 8-14. doi:10.1080/15228835.2017.1416511
13. Senanayake, R., O'Callaghan, S., & Ramos, F. (n.d.). Predicting spatio-temporal propagation of seasonal Influenza USING Variational Gaussian process regression. Retrieved April 04, 2021, from https://ojs.aaai.org/index.php/AAAI/article/view/9899
14. Herlands, W., McFowland, E., Wilson, A. G., & Neill, D. B. (2018). Gaussian process subset scanning for anomalous pattern detection in non-iid data. 425-434. Paper presented at 21st International Conference on Artificial Intelligence and Statistics, AISTATS 2018, Playa Blanca, Lanzarote, Canary Islands, Spain.
15. Morimoto, S., Kawamukai, H., & Shin, K. (2019). Prediction of Crime occurrence using information propagation model and Gaussian process. *2019 14th Asia Joint Conference on Information Security (AsiaJCIS)*. doi:10.1109/asiajcis.2019.000-2
16. Zhao J, Feng Q, Wu P, et al. Learning from Longitudinal Data in Electronic Health Record and Genetic Data to Improve Cardiovascular Event Prediction. *Sci Rep* 2019;9:717. 10.1038/s41598-018-36745-x
17. Delahanty RJ, Alvarez J, Flynn LM, et al. Development and Evaluation of a Machine Learning Model for the Early Identification of Patients at Risk for Sepsis. *Ann Emerg Med* 2019;73:334-44. 10.1016/j.annemergmed.2018.11.036
18. Wong A, Young AT, Liang AS, et al. Development and Validation of an Electronic Health Record-Based Machine Learning Model to Estimate Delirium Risk in Newly Hospitalized Patients Without Known Cognitive Impairment. *JAMA Netw Open* 2018;1:e181018. 10.1001/jamanetworkopen.2018.1018
19. Ibrahim, M. R., Haworth, J., Lipani, A., Aslam, N., Cheng, T., & Christie, N. (2021). Variational-LSTM autoencoder to forecast the spread of CORONAVIRUS across the globe. *PLOS ONE, 16*(1). doi:10.1371/journal.pone.0246120

20. Wang, G., Wei, W., Jiang, J., Ning, C., Chen, H., Huang, J., . . . Ye, L. (2019). Application of a long short-term memory neural network: A burgeoning method of deep learning in forecasting HIV incidence in Guangxi, China. Epidemiology and Infection, 147, E194. doi:10.1017/S095026881900075X
21. Xi, G., Yin, L., Li, Y., & Mei, S. (2018). A deep Residual NETWORK Integrating spatial-temporal properties to predict Influenza trends at AN Intra-urban Scale. *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*. doi:10.1145/3281548.3281558
22. Temurtas, H., Yumusak, N., & Temurtas, F. (2009). A comparative study on diabetes disease diagnosis using neural networks. *Expert Systems with Applications, 36*(4), 8610-8615. doi:10.1016/j.eswa.2008.10.032
23. Kim, T. S., Lee, W. K., & Sohn, S. Y. (2019). Graph convolutional network approach applied to PREDICT Hourly BIKE-SHARING DEMANDS Considering spatial, temporal, and global effects. *PLOS ONE, 14*(9). doi:10.1371/journal.pone.0220782
24. Dai, R., Xu, S., Gu, Q., Ji, C., & Liu, K. (2020, August). Hybrid Spatio-Temporal Graph Convolutional Network: Improving Traffic Prediction with Navigation Data. Retrieved April 4, 2021, from https://export.arxiv.org/pdf/2006.12715
25. Stanczyk, T., & Mehrkanoon, S. (2021, January 26). Deep graph convolutional networks for wind Speed prediction. Retrieved April 4, 2021, from https://arxiv.org/pdf/2101.10041
26. Dong, X., Deng, J., Hou, W., Rashidian, S., Abell-Hart, K., Rosenthal, R. N., . . . Wang, F. (2020). Predicting opioid overdose risk of patients with opioid prescriptions using electronic health records based on temporal deep learning. *SSRN Electronic Journal*. doi:10.2139/ssrn.3708326
27. Lo-Ciganic, W., Huang, J. L., Zhang, H. H., Weiss, J. C., Wu, Y., Kwoh, C. K., . . . Gellad, W. F. (2019). Evaluation of machine-learning algorithms for predicting opioid overdose risk among medicare beneficiaries with opioid prescriptions. *JAMA Network Open, 2*(3), e190968. doi:10.1001/jamanetworkopen.2019.0968

28. Sumetsky, N., Mair, C., Wheeler-Martin, K., Cerda, M., Waller, L., Ponicki, W., & Gruenewald, P. (2021). Predicting the future course of opioid overdose mortality: An example from two US states. *Epidemiology (Cambridge, Mass.), 32*(1), 61-69. doi:10.1097/EDE.0000000000001264
29. Young, S. D., Zhang, Q., Zhou, J., & Pacula, R. L. (2021). Internet search and medicaid prescription drug data as predictors of opioid emergency department visits. *Npj Digital Medicine, 4*(1). doi:10.1038/s41746-021-00392-w
30. National Center for Health Statistics (NCHS). (n.d.). Overdose Death Rates. Retrieved May 04, 2021, from https://www.drugabuse.gov/drug-topics/trends-statistics/overdose-death-rates#:~:text=Opioid%2Dinvolved%20overdose%20deaths%20rose,in%202018%20with%2046%2C802%20deaths
31. Li, Y., Hyder, A., Southerland, L. T., Hammond, G., Porr, A., & Miller, H. J. (2020). 311 service requests as indicators of NEIGHBORHOOD distress and opioid use disorder. *Scientific Reports, 10*(1). doi:10.1038/s41598-020-76685-z
32. Park, T. W., Lin, L. A., Hosanagar, A., Kogowski, A., Paige, K., & Bohnert, A. S. (2016). Understanding risk factors for opioid overdose in clinical populations to inform treatment and policy. *Journal of Addiction Medicine, 10*(6), 369-381. doi:10.1097/adm.0000000000000245
33. Castillo-Carniglia, A., Ponicki, W. R., Gaidus, A., Gruenewald, P. J., Marshall, B. D., Fink, D. S., Cerdá, M. (2019). Prescription drug monitoring programs and opioid overdoses. *Epidemiology, 30*(2), 212-220. doi:10.1097/ede.0000000000000950
34. Rhode Island Gov (RI.GOV), R. (n.d.). The action plan. Retrieved April 04, 2021, from https://preventoverdoseri.org/the-action-plan/
35. Rhode Island Governor's Overdose Prevention and Intervention Task Force (2020). Prevent Overdose RI. Retrieved May 04, 2021, from https://preventoverdoseri.org/wp-content/uploads/2020/05/2019FatalOD_April2020_Final2.pdf
36. Rhode Island Governor's Overdose Prevention and Intervention Task Force (2020). Evidence Update: Plans + Progress. *Prevent Overdose RI*. Retrieved May 04, 2021, from

https://preventoverdoseri.org/wp-content/uploads/2020/10/2020-10-14_Overdose-Evidence-Update_Presentation-to-Task-Force_Final.pdf

37. Rhode Island Executive Office of Health and Human Services (OHHS) (2021). Addiction + Overdose Evidence Update: Presentation to Governor's Overdose Task Force. *Prevent Overdose RI*. Retrieved May 04, 2021, from https://preventoverdoseri.org/wp-content/uploads/2021/02/Evidence-Update-Implementation-Report-Overdose-Task-Force-2.10.2021.pdf
38. RIDOH: Rhode Island Review of Overdose Accidental Deaths (ROAD) Team (2020). February 2020 Report. *RIDOH*. Retrieved May 04, 2021, from https://health.ri.gov/publications/reports/202002-ROAD.pdf
39. RIDOH: Rhode Island Review of Overdose Accidental Deaths (ROAD) Team (2020). August 2020 Report. *RIDOH*. Retrieved May 04, 2021, from https://health.ri.gov/publications/reports/202008-ROAD.pdf
40. RIDOH: Rhode Island Review of Overdose Accidental Deaths (ROAD) Team (2020). October 2020 Report. *RIDOH*. Retrieved May 04, 2021, from https://health.ri.gov/publications/reports/202010-ROAD.pdf
41. RIDOH (n.d.). Drug Overdose Deaths. Retrieved April 04, 2021, from https://health.ri.gov/data/drugoverdoses/
42. Yedinak, J., Krieger M.S., Li Y., Nelson E., Shihipar A., Goldman J., & Marshall, B.D. (2019). The VILLAGE Prevention Plan: Improving prevention efforts and health outcomes for people who use drugs.
43. Altekruse S.F., Cosgrove C.M., Altekruse W.C., Jenkins R.A. & Blanco C. (2020). Socioeconomic risk factors for fatal opioid overdoses in the United States: Findings from the Mortality Disparities in American Communities Study (MDAC). *PLOS ONE* 15(1): e0227966. Retrieved May 04, 2021, from https://doi.org/10.1371/journal.pone.0227966
44. Park, C., Clemenceau, J.R., Seballos & A. et al. (2021). A spatiotemporal analysis of opioid poisoning mortality in Ohio from 2010 to 2016. *Sci Rep* 11, 4692 (2021). Retrieved May 04, 2021, from https://doi.org/10.1038/s41598-021-83544-y

45. CDC. (n.d.). Checking the PDMP: An Important Step to Improving Opioid Prescribing Practices. Retrieved from https://www.cdc.gov/drugoverdose/pdf/PDMP_Factsheet-a.pdf

46. Nechuta, S. J., Tyndall, B. D., Mukhopadhyay, S., & McPheeters, M. L. (2018). Sociodemographic factors, prescription history and opioid overdose deaths: a statewide analysis using linked PDMP and mortality data. *Drug and Alcohol Dependence*, 190, 62–71. Retrieved May 04, 2021, from https://doi.org/10.1016/j.drugalcdep.2018.05.004

47. Yang, Z., Wilsey, B., Bohm, M., Weyrich, M., Roy, K., Ritley, D., Jones, C., & Melnikow, J. (2015). Defining risk of prescription opioid overdose: pharmacy shopping and overlapping prescriptions among long-term opioid users in medicaid. *The Journal of Pain,* 16(5), 445–453. Retrieved May 04, 2021, from https://doi.org/10.1016/j.jpain.2015.01.475

48. Green, T. C., Clarke, J., Brinkley-Rubinstein, L., Marshall, B., Alexander-Scott, N., Boss, R., & Rich, J. D. (2018). Post Incarceration Fatal Overdoses After Implementing Medications for Addiction Treatment in a statewide Correctional System. *JAMA psychiatry*, 75(4), 405–407. Retrieved May 04, 2021, from https://doi.org/10.1001/jamapsychiatry.2017.4614

49. Prevent Overdose RI (2021). Overdose Death Rate per 100,000 person-years by Race & Ethnicity, 2016 - 2020. *Prevent Overdose RI*. Retrieved May 04, 2021, from https://preventoverdoseri.org/race-ethnicity-data/

50. Castillo-Carniglia, A., González-Santa Cruz, A., Cerdá, M., Delcher, C., Shev, A. B., Wintemute, G. J., & Henry, S. G. (2021). Changes in opioid prescribing after implementation of mandatory registration and proactive reports within California's prescription drug monitoring program. *Drug and Alcohol Dependence*, 218, 108405. Retrieved May 04, 2021, from https://doi.org/10.1016/j.drugalcdep.2020.108405

51. CDC (n.d.). CDC Guideline for Prescribing Opioids for Chronic Pain. Retrieved April 04, 2021, from https://www.cdc.gov/drugoverdose/pdf/Guidelines_At-A-Glance-508.pdf

52. Fulton-Kehoe, D., Sullivan, M., Turner, J., Garg, R., Bauer, A., Wickizer, T., & Franklin, G. (2015). Opioid Poisonings in Washington stateMedicaid: Trends, Dosing, and Guidelines. *Medical Care,* 53(8), 679-685. Retrieved May 04, 2021, from https://www.jstor.org/stable/10.2307/2641802

53. Klemmer, K., Neill, D. B., & Jarvis, S. A. (2018). Modeling Rape Reporting Delays Using Spatial, Temporal and Social Features. *Computers and Society.*

54. Rowe, C., Santos, G. M., Behar, E., & Coffin, P. O. (2016). Correlates of overdose risk perception among illicit opioid users. *Drug and Alcohol Dependence*, 159, 234–239. Retrieved May 04, 2021, from https://doi.org/10.1016/j.drugalcdep.2015.12.018

55. U.S. Food and Drug Administration (2018). Opioid Analgesic Risk Evaluation and Mitigation Strategy (REMS). Retrieved May 04, 2021, from https://www.fda.gov/drugs/information-drug-class/opioid-analgesic-risk-evaluation-and-mitigation-strategy-rems

56. Volkow, N.D., Koroshetz, W.J (2019). The role of neurologists in tackling the opioid epidemic. *Nat Rev Neurol* 15, 301–305 (2019). https://doi.org/10.1038/s41582-019-0146-8

57. Wightman, R.S., Perrone, J., Scagos, R. et al. (2021). Opioid Overdose Deaths with Buprenorphine Detected in Postmortem Toxicology: a Retrospective Analysis. *J. Med. Toxicol*. 17, 10–15 (2021). Retrieved May 04, 2021, from https://doi.org/10.1007/s13181-020-00795-3

Appendices

Appendix A: Feature Selection

In summary, our use the these following groups of features:

1. ACS 5-year Estimates
    - Total Population
    - % Population by Sex
    - % Population by Age Groups
    - % Population by Race and Ethnicity
    - % Household by Family Type
    - % Population by Marital Status
    - % Population by Educational Attainment
    - % Population by High-school Enrollment Status
    - Housing Tenure
    - Family Poverty-level by Family compositions
    - Population Poverty-level
    - Unemployment

- Population by Occupation
- Veteran Status for the Civilian Population
- Health Insurance Coverage
- Public Assistance Income or Food Stamps/Snap
- Household Language by Household Limited English Speaking Status
- Accessibility:
    - Vehicle Availability
    - Telephone Service Availability
    - Presence of Internet Subscriptions in Household

2. Emergency Medical Service (EMS) overdoses-related calls
    - EMS runs total count
    - EMS runs by Sex
    - EMS runs by Age
    - EMS runs by Location
3. Prescription Drug Monitoring Program (PDMP) data
    - General prescriptions and patient counts
    - Prescriptions with large daily doses (> 90-milligram equivalents dispensed)
    - Prescriptions covering longer than 30 days
    - Medicine type and sequences
        - Prescriptions with Benzo overlap
        - Prescriptions with buprenorphine
        - Prescriptions with long-acting opioids
        - Prescriptions with Naloxone distribution
    - Patients with multiple prescribers or dispensers
4. Land Use
    - Correctional institutions
    - EMS Stations
    - Fire Stations
    - Law Enforcement
    - Ports and Harbors
    - RIPTA Park and Ride Stops
5. Public access
    - Acupuncture licensees
    - Accredited early child care center/preschool programs
    - Adult day care programs
    - Ambulatory surgery centers
    - Aquatic venues
    - Blood test screening licensees
    - Chiropractic licensees
    - Cigarette dealer licensees
    - Drug Addiction Treatment Act waived physicians
    - Drug and alcohol treatment facilities
    - Mental health facilities
    - Department of Behavioral Healthcare, Developmental Disabilities, and Hospitals (BHDDH) Licensed Mental Health and Behavioral Health Treatment Facilities Programs

- BHDDH Licensed Substance Use Behavioral Health Treatment Facility Programs
- Recreational facilities
- Physician licensees
- Other community housing services
- Religious organizations
- Skilled nursing facility businesses
- Drinking places (alcoholic beverages)
- Nursing facility licensees
- Temporary shelters

Please find the full list below:

Table 1 - Qualitative Review for Feature Reduction

Indicators	Unit of the base indicator	Category	Notes
SUDORS			
Total count of overdoses	person	Target variable	/
ACS			
Total Population	person	Population	/
Population Density (Per Sq. Mile)	person		
% Total Population: Male	person	Population by Age	RIDOH (n.d.) suggests 75% of overdose deaths are men
% Total Population: 15 to 17 Years	person		Rhode Island Governor's Overdose Prevention and Intervention Task Force (2020) suggests most overdoses occur among adults
% Total Population: 18 to 24 Years	person		
% Total Population: 25 to 34 Years	person		
% Total Population: 35 to 44 Years	person		
% Total Population: 45 to 54 Years	person		
% Total Population: 55 to 64 Years	person		
% Total Population: Not Hispanic or Latino: White Alone	person	Population by Race and Ethnicity	Rhode Island Governor's Overdose Prevention and Intervention Task Force (2020) found the majority of the overdose deaths cases are White population, and Black and Hispanic populations have the highest overdose death rates. We kept all subgroups for now just in case other minority groups might be concentrated in
% Total Population: Not Hispanic or Latino: Black or African American Alone	person		

% Total Population: Not Hispanic or Latino: Asian Alone	person		granular geographic levels so that their previous analysis at a broader level could fail to capture the signal
% Total Population: Not Hispanic or Latino: Native Hawaiian or Other Pacific Islander Alone	person		
% Total Population: Not Hispanic or Latino: Some Other Races Alone	person		
% Total Population: Not Hispanic or Latino: Two or More Races	person		
% Total Population: Hispanic or Latino	person		
% Total Population: Hispanic or Latino: Black or African American Alone	person		
% Total Population: Hispanic or Latino: Asian Alone	person		
% Total Population: Hispanic or Latino: Native Hawaiian or Other Pacific Islander Alone	person		
% Total Population: Hispanic or Latino: Some Other Races Alone	person		
% Total Population: Hispanic or Latino: Two or More Races	person		
% Households: Family Households	household	Households by Family Type and Composition	Yedinak et al. (2019) identified "Households consisting of a single member living alone" as a risk characteristics in outbreak of new HIV or hepatitis C infections related to drug overdose
% Households: Family Households: Married-Couple Family	household		
% Households: Family Households: Other Family	household		
% Households: Family Households: Other Family: Male Householder, No Wife Present	household		
% Households: Family Households: Other Family: Female Householder, No Husband Present	household		

% Households: Nonfamily Households	household		
% Households: Nonfamily Households: Male Householder	household		
% Households: Nonfamily Households: Female Householder	household		
% Population 15 Years and Over: Never Married	person	Population by Marital Status	Altekruse et al. (2021) found widowed population at risk; we suspect marital status could relate to mental health (i.e. trauma), which could be a cause of opioid overdose so we expanded the subgroup selections
% Population 15 Years and Over: Separated	person		
% Population 15 Years and Over: Widowed	person		
% Population 15 Years and Over: Divorced	person		
% Population 25 Years and Over: Less than High School	person	Population by Educational Attainment	Yedinak et al. (2019) identified these indicators as a risk characteristics in outbreak of new HIV or hepatitis C infections related to drug overdose
% Civilian Population 16 to 19 Years: Not High School Graduate (Drop out)	person	High School Dropout	
Median Household Income (In 2019 Inflation Adjusted Dollars)	household	Household Income	
% Occupied Housing Units: Renter Occupied	housing unit	Housing Tenure	Park et al. (2021) found housing characteristics, such as vacancy rate and dilapidated housing, is highly correlated with opioid overdose risk in Ohio
% Housing Units: Vacant	housing unit		
% Families: Income Below Poverty Level	family	Family by Poverty Level and Composition	Altekruse et al. (2021) found population under poverty level at risk; combined with family composition and marital status for further investigation
% Families: Income Below Poverty Level: Male Householder, No Wife Present, With Children	family		
% Families: Income Below Poverty Level: Male Householder, No Wife Present, Without Children	family		
% Families: Income Below Poverty Level: Female Householder, No Husband Present, With Children	family		
% Families: Income Below Poverty Level: Female	family		

Householder, No Husband Present, Without Children			
% Population for Whom Poverty Status Is Determined: Under 1.00 (Doing Poorly)	person	Population by Poverty Level	
% Population for Whom Poverty Status Is Determined: 1.00 to 1.99 (Struggling)	person		
% Civilian Population in Labor Force 16 Years and Over: Unemployed	person	Population by Employment Status	Altekruse et al. (2021) found unemployed population at risk
% Civilian Noninstitutionalized Population: No Health Insurance Coverage	person	Population by Health Insurance Coverage	Altekruse et al. (2021) found population without health insurance at risk
% Households: With Cash Public Assistance or Food Stamps/Snap	household	Households by Public Assistance Status	Another proxy for poverty status (but also indicates access to public esource)
% Households: Spanish Limited English Speaking Household	household	Households by Language Spoken	The language availability of Prevention Program publications are available in English, Spanish, and Portuguese; most of the intervention and local partners use those as major communication languages
% Households: Indo European Languages Limited English Speaking Household	household		
% Households: Asian Pacific Islands Languages Limited English Speaking Household	household		
% Households: Limited English Speaking Household	household		
% Civilian Population 18 Years and Over: Veteran	person	Population by Veteran Status	Jalal et al. (2020)
% Civilian Employed Population 16 years and over: Natural Resources, Construction, And Maintenance Occupations	person	Population by Occupations	Rhode Island Review of Overdose Accidental Deaths Team (2020) found Natural Resources, Construction, And Maintenance workers have a significantly higher opioid overdose death rate
% Occupied Housing Unit: No Telephone Service Available	housing unit	Accessibility	RIDOH has established two hotlines through phone call (similar to telephone service access, we believe internet availability could be another bottleneck for people to acquire information for intervention resources):
% Occupied Housing Unit: No Internet Subscription	housing unit	Accessibility	

			* People with moderate to severe opioid use disorder can access buprenorphine treatment 24/7 through telehealth hotline 401-606-5456. * Addiction counseling line: 401-414-LINK(5465)
% Occupied Housing Unit: No Vehicle Available	housing unit	Accessibility	Yedinak et al. (2019) identified this indicator as a risk characteristics in outbreak of new HIV or hepatitis C infections related to drug overdose
		EMS	
Total Runs	trip	Emergency Medical Service	/
Total Runs for male patients	trip	Emergency Medical Service by Sex	Echos ACS Sex data suggested by RIDOH (2020)
Total Runs for patients aged 15-24	trip	Emergency Medical Service by Age	Echos ACS Age Group data suggested by RIDOH (2020)
Total Runs for patients aged 25-34	trip		
Total Runs for patients aged 35-44	trip		
Total Runs for patients aged 45-54	trip		
Total Runs for patients aged 55-64	trip		
Total Runs at public locations	trip	Emergency Medical Service by Location	Interview suggests that private locations (i.e. homes) might be main locations for opioid overdose but could explore public versus private especially among urban versus rural area
Total Runs at private locations	trip		
		PDMP	
Total count of prescriptions dispensed		General statistics - prescription	/
Total count of opioid prescriptions dispensed	prescription		/
Total days of supply dispensed	prescription	Dose - prescription level	/
Average days of supply per prescription			/
Total milligram equivalents dispensed	prescription		/
Average daily milligram equivalents dispensed per prescription	prescription		/

Total count of prescriptions > 90 milligram equivalents dispensed	prescription	Large dose - prescription level	/
Total count of prescriptions > 120 milligram equivalents dispensed	prescription		
Total new opioid prescription dispensed	prescription	New prescriptions	Behavioral perspective. If a person runs out of opioid, one might go for a new prescription when the old prescription runs out.
Total days of supply new opioid prescription dispensed	prescription		
Average days of supply new opioid prescription dispensed	prescription		
Total milligram equivalents new prescription dispensed			
Total count of new prescriptions > 90 milligram equivalents dispensed	prescription		
Total count of new prescriptions > 120 milligram equivalents dispensed	prescription		
Total count of prescriptions with benzo overlap	prescription	Benzoylecgonine overlap with opioids - prescription level	Fulton-Kehoe et al. (2015), Wightman et al. (2021) and Rowe et al. (2016)
Total days of supply of prescriptions with benzo overlap	prescription		
Average days of supply of prescriptions with benzo overlap	prescription		
Total milligram equivalents dispensed with benzo overlap	prescription		
Average daily milligram equivalents dispensed with benzo overlap	prescription		
Total count of prescriptions with benzo overlap	prescription		
Total count of prescriptions > 90 milligram equivalents dispensed with benzo overlap	prescription		
Total count of prescriptions > 120 milligram equivalents dispensed with benzo overlap	prescription		

Total count of prescriptions with short-acting opioids with benzo overlap	prescription		
Total count of buprenorphine prescriptions	prescription	Opioid use disorder medication - prescription level	Buprenorphine is a medication for treating opioid use disorder
Total days of supply buprenorphine	prescription		
Average days of supply per prescription of buprenorphine	prescription		
Total count of long acting opioids prescriptions	prescription	Long acting opioids - prescription level	FDA (2018)
Total days of supply of long acting opioids prescriptions	prescription		
Average days of supply of long acting opioids prescriptions	prescription		
Total milligram equivalents of long acting opioids	prescription		
Average daily milligram equivalents of long acting opioids	prescription		
Total count of new prescriptions with long acting opioids	prescription		
Total count of prescriptions with long acting opioids and benzo overlap	prescription		
Total count of prescriptions > 90 milligram equivalents dispensed with long acting opioids	prescription		
Total count of prescriptions > 120 milligram equivalents dispensed with long acting opioids	prescription		
Total count of naloxone prescriptions	prescription	Overdose medication	Naloxone is a medication for reversing opioid overdose
Total count of patients	person	General Statistics - patients	/
Average opioid prescriptions per person	person		/
Total persons with new opioid prescriptions covering 30 days or more	person	Long-coverage - patients	More likely for chronic pains - chances of addiction could increase

Average new opioid prescriptions covering 30 days or more per person	person		
Total count of persons with prescriptions > 90 milligram equivalent	person	Large dose - patient level	CDC (n.d): "Avoid or carefully justify increasing dosage to ≥90 MME/day. These dosage thresholds are based on overdose risk when opioids are prescribed for pain and should not guide dosing of medication-assisted treatment for opioid use disorder."
Average prescriptions > 90 milligram equivalent per person	person		
Total count of persons with benzo overlap	person	Benzoylecgonine overlap with opioids - patient level	Fulton-Kehoe et al, (2015), Wightman et al. (2021) and Rowe (2016)
Average prescriptions with benzo overlap per person	person		
Total count of persons with multiple prescribers or dispensers	person	Multiple sources	Yang et al. (2015)
Total count of persons ever receiving buprenorphine	person	Opioid use disorder medication - patient level	Buprenorphine is a medication for treating opioid use disorder
Total count of persons initiating buprenorphine	person		
Total count of persons receiving buprenorphine for at least 7 days	person	Opioid use disorder medication - patient level	
Total count of persons receiving buprenorphine for at least 180 days	person		
Land Use			
Count of Correctional Institutions	facility	Land use	Green et al. (2018)
Count of EMS Stations	facility		Echos the EMS data source
Count of Fire Stations	facility		Safe Stations program was implemented in some localities to help those struggling with addiction - residents who are victim of addiction or overdose can seek help from local fire stations
Count of Law Enforcement	facility		Interview suggests that law enforcement such as police activity could discourage people from seeking help and reinforce community stigma
Count of Ports and Harbors	facility		Interview with sponsor suggested so
Count of RIPTA Bus Stops	facility		Another proxy for transportation accessibility
Public Access			

Count of Chiropractic Licencees	service provider	Healthcare provider	Co-occurrence between opioid use disorder and anxiety, psychiatric comorbidity, and other drug use (Park et al. 2021 and Volkow et al. 2019) (potentially related to chronic pain)
Count of Physician Licencees	service provider		
Count of Mental Health Facilities	facility		
Count of Recreational Facilities	facility	Public infrastructure	Population health positive and negative influences
Count of Drinking places	business	Population health - negative influence	
Count of Cigarette Dealers Licensees	business		
Count of BHDDH Licensed Mental Health and Behavioral Health Treatment Facilities	facility	Population health facilities	Suggested by Brown University team
Count of BHDDH Licensed Substance Use Behavioral Health Treatment Facilities	facility		
Count of Religious Organizations	organization	Religious organization	Some of the interventions are deployed through religious organizations
Count of Temporary Shelters	facility	Public infrastructure	Homelessness and drug use

Table 2 - Feature Set Details

Indicator	Set A	Set B	Set C	Set D
year	TRUE	TRUE	TRUE	TRUE
period	TRUE	TRUE	TRUE	TRUE
Census Block Group FIPS ID	TRUE	TRUE	TRUE	TRUE
Drug Overdoses Normalized Rank	TRUE	TRUE	TRUE	TRUE
Total count of opioid prescriptions dispensed	TRUE	FALSE	FALSE	TRUE
Total days of supply dispensed	TRUE	FALSE	TRUE	TRUE
Average days of supply per prescription	TRUE	FALSE	TRUE	TRUE
Total milligram equivalents dispensed	TRUE	TRUE	TRUE	TRUE
Average daily milligram equivalents dispensed per prescription	TRUE	FALSE	TRUE	TRUE
Total count of prescriptions > 90 milligram equivalents dispensed	TRUE	FALSE	TRUE	FALSE
Total count of prescriptions > 120 milligram equivalents dispensed	TRUE	FALSE	TRUE	FALSE
Total new opioid prescriptions dispensed	TRUE	FALSE	TRUE	TRUE

Total days of supply dispensed among new prescriptions	TRUE	FALSE	TRUE	FALSE
Average days of supply per prescription among new prescriptions	TRUE	FALSE	TRUE	TRUE
Total milligram equivalents dispensed among new prescriptions	TRUE	FALSE	TRUE	TRUE
Average daily milligram equivalents dispensed per prescription among new prescriptions	TRUE	FALSE	TRUE	FALSE
Total count of prescriptions > 90 milligram equivalents dispensed among new prescriptions	TRUE	FALSE	FALSE	FALSE
Total count of prescriptions > 120 milligram equivalents dispensed among new prescriptions	TRUE	FALSE	FALSE	FALSE
Total count of prescriptions with benzo overlap	TRUE	FALSE	FALSE	FALSE
Total days of supply of prescriptions with benzo overlap	TRUE	FALSE	TRUE	TRUE
Average days of supply of prescriptions with benzo overlap	TRUE	FALSE	TRUE	TRUE
Total milligram equivalents dispensed with benzo overlap	TRUE	TRUE	TRUE	TRUE
Average daily equivalents dispensed with benzo overlap	TRUE	FALSE	TRUE	TRUE
Total count of prescriptions with benzo overlap	TRUE	FALSE	TRUE	FALSE
Total count of prescriptions > 90 milligram equivalents dispensed with benzo overlap	TRUE	FALSE	FALSE	FALSE
Total count of prescriptions > 120 milligram equivalents dispensed with benzo overlap	TRUE	FALSE	FALSE	FALSE
Total count of buprenorphine prescriptions	TRUE	TRUE	FALSE	TRUE
Total days of supply buprenorphine	TRUE	FALSE	TRUE	TRUE
Average days of supply per prescription of buprenorphine	TRUE	TRUE	TRUE	TRUE
Total count of long acting opioids prescriptions	TRUE	FALSE	TRUE	FALSE
Total days of supply of long acting opioids prescriptions	TRUE	FALSE	FALSE	TRUE
Average days of supply of long acting opioids prescriptions	TRUE	FALSE	TRUE	TRUE
Total milligram equivalents of long acting opioids	TRUE	FALSE	TRUE	TRUE
Average daily milligram equivalents of long acting opioids	TRUE	FALSE	TRUE	TRUE
Total count of new prescriptions with long acting opioids	TRUE	FALSE	TRUE	FALSE
Total count of prescriptions with long acting opioids and benzo overlap	TRUE	FALSE	FALSE	FALSE
Total count of prescriptions > 90 milligram equivalents dispensed with long acting opioids	TRUE	FALSE	FALSE	FALSE
Total count of prescriptions > 120 milligram equivalents dispensed with long acting opioids	TRUE	FALSE	FALSE	FALSE
Total counts of prescriptions with short acting opioids with benzo overlap	TRUE	FALSE	FALSE	FALSE

Total count of naloxone prescriptions	TRUE	TRUE	TRUE	FALSE
Total count of patients	TRUE	TRUE	FALSE	TRUE
Total count of prescriptions	TRUE	TRUE	TRUE	FALSE
Average prescriptions per person	TRUE	FALSE	TRUE	TRUE
Total count of persons with new prescription covering more than 30 days	TRUE	FALSE	TRUE	FALSE
Average count of new prescriptions per person covering more than 30 days	TRUE	FALSE	FALSE	TRUE
Total count of persons with prescriptions > 90 milligram equivalent	TRUE	FALSE	TRUE	FALSE
Average prescriptions > 90 milligram equivalent per person	TRUE	FALSE	FALSE	TRUE
Total count of persons with Benzo overlap	TRUE	FALSE	TRUE	FALSE
Average prescriptions with Benzo overlap per person	TRUE	FALSE	FALSE	TRUE
Total count of persons with multiple prescribers or dispensers	TRUE	FALSE	TRUE	FALSE
Total count of persons initiating buprenorphine	TRUE	FALSE	FALSE	TRUE
Total count of persons receiving buprenorphine for at least 7 days	TRUE	FALSE	FALSE	FALSE
Total count of persons receiving buprenorphine for at least 180 days	TRUE	TRUE	TRUE	FALSE
EMS runs for opioid overdose	TRUE	TRUE	TRUE	FALSE
EMS runs among men	TRUE	TRUE	TRUE	FALSE
EMS runs occurring in a public location	TRUE	FALSE	TRUE	FALSE
EMS runs occurring in a private location	TRUE	TRUE	FALSE	FALSE
EMS runs with patient aged 15 to 24	TRUE	FALSE	TRUE	FALSE
EMS runs with patient aged 25 to 34	TRUE	TRUE	FALSE	FALSE
EMS runs with patient aged 35 to 44	TRUE	FALSE	FALSE	FALSE
EMS runs with patient aged 45 to 54	TRUE	FALSE	FALSE	FALSE
EMS runs with patient aged 55 to 64	TRUE	TRUE	TRUE	FALSE
Total Population	TRUE	TRUE	FALSE	TRUE
Population Density (Per Sq. Mile)	TRUE	TRUE	TRUE	FALSE
% Total Population: Male	TRUE	TRUE	TRUE	FALSE
% Total Population: 15 to 17 Years	TRUE	FALSE	TRUE	TRUE
% Total Population: 18 to 24 Years	TRUE	FALSE	TRUE	FALSE
% Total Population: 25 to 34 Years	TRUE	TRUE	TRUE	FALSE
% Total Population: 35 to 44 Years	TRUE	TRUE	TRUE	FALSE

% Total Population: 45 to 54 Years	TRUE	TRUE	TRUE	FALSE
% Total Population: 55 to 64 Years	TRUE	TRUE	TRUE	FALSE
% Total Population: Not Hispanic or Latino: White Alone	TRUE	FALSE	TRUE	TRUE
% Total Population: Not Hispanic or Latino: Black or African American Alone	TRUE	FALSE	TRUE	TRUE
% Total Population: Not Hispanic or Latino: American Indian and Alaska Native Alone	TRUE	FALSE	TRUE	TRUE
% Total Population: Not Hispanic or Latino: Asian Alone	TRUE	FALSE	FALSE	TRUE
% Total Population: Not Hispanic or Latino: Native Hawaiian and Other Pacific Islander Alone	TRUE	FALSE	TRUE	TRUE
% Total Population: Not Hispanic or Latino: Some Other Race Alone	TRUE	FALSE	FALSE	TRUE
% Total Population: Not Hispanic or Latino: Two or More Races	TRUE	FALSE	FALSE	TRUE
% Total Population: Hispanic or Latino	TRUE	TRUE	TRUE	TRUE
% Total Population: Hispanic or Latino: White Alone	TRUE	TRUE	TRUE	TRUE
% Total Population: Hispanic or Latino: Black or African American Alone	TRUE	TRUE	TRUE	TRUE
% Total Population: Hispanic or Latino: American Indian and Alaska Native Alone	TRUE	TRUE	FALSE	TRUE
% Total Population: Hispanic or Latino: Asian Alone	TRUE	FALSE	FALSE	FALSE
% Total Population: Hispanic or Latino: Native Hawaiian and Other Pacific Islander Alone	TRUE	FALSE	FALSE	FALSE
% Total Population: Hispanic or Latino: Some Other Race Alone	TRUE	TRUE	FALSE	FALSE
% Total Population: Hispanic or Latino: Two or More Races	TRUE	FALSE	TRUE	FALSE
% Households: Family Households	TRUE	TRUE	FALSE	TRUE
% Households: Family Households: Other Family	TRUE	FALSE	TRUE	TRUE
% Households: Family Households: Other Family: Male Householder, No Wife Present	TRUE	FALSE	TRUE	TRUE
% Households: Family Households: Other Family: Female Householder, No Husband Present	TRUE	FALSE	TRUE	FALSE
% Households: Nonfamily Households: Male Householder	TRUE	TRUE	TRUE	FALSE
% Households: Nonfamily Households: Female Householder	TRUE	FALSE	TRUE	FALSE
% Population 15 Years and Over: Never Married	TRUE	TRUE	TRUE	FALSE
% Population 15 Years and Over: Separated	TRUE	TRUE	TRUE	TRUE
% Population 15 Years and Over: Widowed	TRUE	FALSE	TRUE	FALSE
% Population 15 Years and Over: Divorced	TRUE	TRUE	TRUE	TRUE

% Population 25 Years and Over: Less than High School	TRUE	TRUE	TRUE	TRUE
% Civilian Population 16 to 19 Years: Not High School Graduate, Not Enrolled (Dropped Out)	TRUE	TRUE	TRUE	TRUE
Median Household Income (In 2019 Inflation Adjusted Dollars)	TRUE	TRUE	FALSE	FALSE
% Occupied Housing Units: Renter Occupied	TRUE	TRUE	TRUE	TRUE
% Housing Units: Occupied	TRUE	TRUE	TRUE	TRUE
% Housing Units: Vacant	TRUE	FALSE	TRUE	TRUE
% Occupied Housing Units: Fuel Oil, Kerosene, Etc.	TRUE	FALSE	TRUE	TRUE
% Occupied Housing Units: Coal, Coke or Wood	TRUE	FALSE	TRUE	TRUE
% Occupied Housing Units: No Fuel Used	TRUE	FALSE	FALSE	TRUE
% Families: Income Below Poverty Level	TRUE	TRUE	FALSE	TRUE
% Families: Income Below Poverty Level: Male Householder, No Wife Present: with Related Children Under 18 Years	TRUE	TRUE	TRUE	FALSE
% Families: Income Below Poverty Level: Male Householder, No Wife Present: No Related Children Under 18 Years	TRUE	TRUE	FALSE	TRUE
% Families: Income Below Poverty Level: Female Householder, No Husband Present: with Related Children Under 18 Years	TRUE	FALSE	FALSE	TRUE
% Families: Income Below Poverty Level: Female Householder, No Husband Present: No Related Children Under 18 Years	TRUE	FALSE	TRUE	TRUE
% Population for Whom Poverty Status Is Determined: Under 1.00 (Doing Poorly)	TRUE	FALSE	FALSE	FALSE
% Population for Whom Poverty Status Is Determined: 1.00 to 1.99 (Struggling)	TRUE	FALSE	TRUE	TRUE
% Population for Whom Poverty Status Is Determined: Under 2.00 (Poor or Struggling)	TRUE	TRUE	TRUE	FALSE
% Population for Whom Poverty Status Is Determined: 2.00 and Over (Doing Ok)	TRUE	FALSE	TRUE	FALSE
Correctional Institutions	TRUE	FALSE	TRUE	TRUE
EMS Stations	TRUE	FALSE	FALSE	FALSE
Fire Stations	TRUE	FALSE	FALSE	FALSE
Law Enforcement	TRUE	FALSE	FALSE	FALSE
Ports and Harbors	TRUE	FALSE	FALSE	TRUE
RIPTA Bus Stops	TRUE	FALSE	FALSE	FALSE
RIPTA Park and Ride Stops	TRUE	FALSE	TRUE	FALSE
Chiropractic licensees	TRUE	FALSE	TRUE	FALSE

Cigarette dealers licensees	TRUE	FALSE	TRUE	FALSE
Mental health facilities	TRUE	TRUE	TRUE	FALSE
BHDDH Licensed mental health and behavioral health treatment facilities programs	TRUE	TRUE	TRUE	FALSE
BHDDH Licensed Substance Use Behavioral Health Treatment Facilities Programs	TRUE	TRUE	TRUE	FALSE
Recreational facilities	TRUE	FALSE	TRUE	FALSE
Physician licensees	TRUE	TRUE	TRUE	FALSE
Religious organizations	TRUE	FALSE	TRUE	FALSE
Drinking places (Alcoholic beverages)	TRUE	TRUE	TRUE	FALSE
Temp shelters	TRUE	TRUE	TRUE	FALSE
% Population 16 Years and Over: In Labor Force: Civilian Labor Force: Unemployed	TRUE	TRUE	TRUE	TRUE
% Employed Civilian Population 16 Years and Over: Construction, Extraction and Maintenance Occupations	TRUE	TRUE	TRUE	FALSE
% Civilian Population 18 Years and Over: Veteran	TRUE	TRUE	TRUE	FALSE
% Occupied Housing Units: No Vehicle Available	TRUE	TRUE	TRUE	FALSE
% Occupied Housing Units: No	TRUE	TRUE	TRUE	FALSE
% Households: Spanish Limited English Speaking	TRUE	TRUE	TRUE	FALSE
% Households: Indo European Languages Limited English Speaking	TRUE	TRUE	TRUE	TRUE
% Households: Asian Pacific Island Languages Limited English Speaking	TRUE	TRUE	TRUE	FALSE
% Households: Other Languages Limited English Speaking	TRUE	TRUE	TRUE	FALSE

Appendix B: Modeling Results and Details

Model	Feature set	Training, Validation and Test	LC20 Capture
Gaussian Process	Features of the previous period selected by Recursive Feature Elimination, plus distance-weighted spatial aggregates of those features, and each Census Block Group's	Train: 2016.2 - 2019.2 Test: 2020.1	40.1%

	centroid coordinates (in total 140 features).		
Graphic Convolutional Network	Features of the previous period based on literature review and export opinions, plus distance-weighted spatial aggregates of those features (in total 291 features).	Train: 2016.2 - 2019.2 Test: 2020.1	37.4%
Gradient Boosting	16 principal components extracted from an original set of 143 features of the previous period, using an 8-degree poly kernel.	Train: 2016.2 - 2018.2 Validation: 2019.1 - 2019.2 Test: 2020.1	40.1%
Random Forest	25 top important features from the previous two periods.	1st round: Train: 2017.1 - 2019.1 Validation: 2019.2 Test: 2020.1 2nd round: Train: 2017.1 - 2018.2 Validation: 2019.1 Test: 2019.2	40.2% (average of two rounds' testing result)