Mapping the Risk Terrain for Crime using Machine Learning

09-23

Andrew Wheeler, PhD

andrew.wheeler@crimede-coder.com

CRIME
De-Coder

My background

CRIME De-Coder

Academic Background

- All degrees in Criminal Justice
- Phd @ SUNY Albany [08-15]
- Professor of Crim. at UT-Dallas [16-19]

Private Sector / Consulting

- Data Scientist @ *Gainwell Technologies* [19-current]
- Created CRIME De-Coder to continue work with police/CJ and tech

Mapping the Risk Terrain

CRIME De-Coder

Journal of Quantitative Criminology (2021) 37:445–480
https://doi.org/10.1007/s10940-020-09457-7

Check for updates

Mapping the Risk Terrain for Crime Using Machine Learning

Andrew P. Wheeler[1] · Wouter Steenbeek[2]

Published online: 24 April 2020

Problem & Motivation

CRIME De-Coder

Traditional Police Approaches to hotspots

- **Short term forecasts (nudges)**
  - **Best served via short term models (Self-exciting PredPol)**

- **Long term forecasts (problem oriented approaches)**
  - **Traditional hot spots (simple clusters or rank methods)**
  - **Risk Terrain Modelling (RTM), regression based approach**
    - **Identifies *contributing factors* to a hotspot**

Problem & Motivation

CRIME De-Coder

RTM has 3 steps

- Encodes spatial factors via *distance* or *density*
- Recodes them to binary variables
- Uses Regularization/model selection to find simple model

So start with:

	A	B	C	D	E
1	Crime	Bar Dist < 500	Bar Dist < 1000	Bar Dens < 0.5	Bar Dens < 1.0
2	2	0	0	1	1
3	4	0	1	1	1
4	10	1	1	0	1

And end up with:

$$\hat{\lambda} = \exp(\beta_0 + \beta_1 \cdot I(\text{Bar}_d < 500ft))$$

Problem & Motivation

CRIME De-Coder

Problems with RTM approach

- Encoding into binary violates distance decay
- Variable selection inconsistent with interaction effects (e.g. bars in some area of the city have a larger effect)
- Predictions are spatially invariant (gas station has the same effect across entire city)

Solution

- Non-linear random forest model
- Interpretable explanations *for each forecasted hotspot* using Shapley values

Data & Modelling

CRIME De-Coder

Application – Forecasting Robberies in Dallas

- Open Data, can provide replication code (code in R)

Data

- Robbery counts aggregated to small grid cells (200 by 200 ft), total N 217,745 cells covering Dallas
- Train set (June 2014 – May 2016), Test set (July 2016- May 2018)
- 6682 robberies in train set, 5931 in test set

Independent Variables

- 18 different crime generator/attractor variables (e.g. gas stations, apartments, large box stores, ATMs, train stops)
- Census Demographics (e.g. poverty, female headed households)
- X & Y coordinates of grid cell

Data & Modelling

CRIME De-Coder

Outcome metrics

- PAI (Predictive Accuracy Index)
  - % Crime Capture / % Area, e.g. 0.5/0.05 = 10
  - Can be translated to ROC curve
- PEI (Predictive Efficiency Index)
  - Actual PAI / Max PAI (under oracle model)
  - Crime is spread out, cannot get 100% recall given fixed target area
- RRI (Recapture Rate Index)
  - Crimes Predicted / Crimes Observed
  - Should display on log scale, calibrated model ~ 1

Data & Modelling

CRIME De-Coder

Different Models

- Random Forest
  - Default implementation in R *ranger* package
  - 500 trees, no limit on tree depth

- Kernel Density Estimate (normal kernel & 600 ft bandwidth)

- Naïve (prior crime rankings)

- RTM
  - Coded myself from public description
  - Can replicate entirely based on description minus some elastic net search parameters

Results

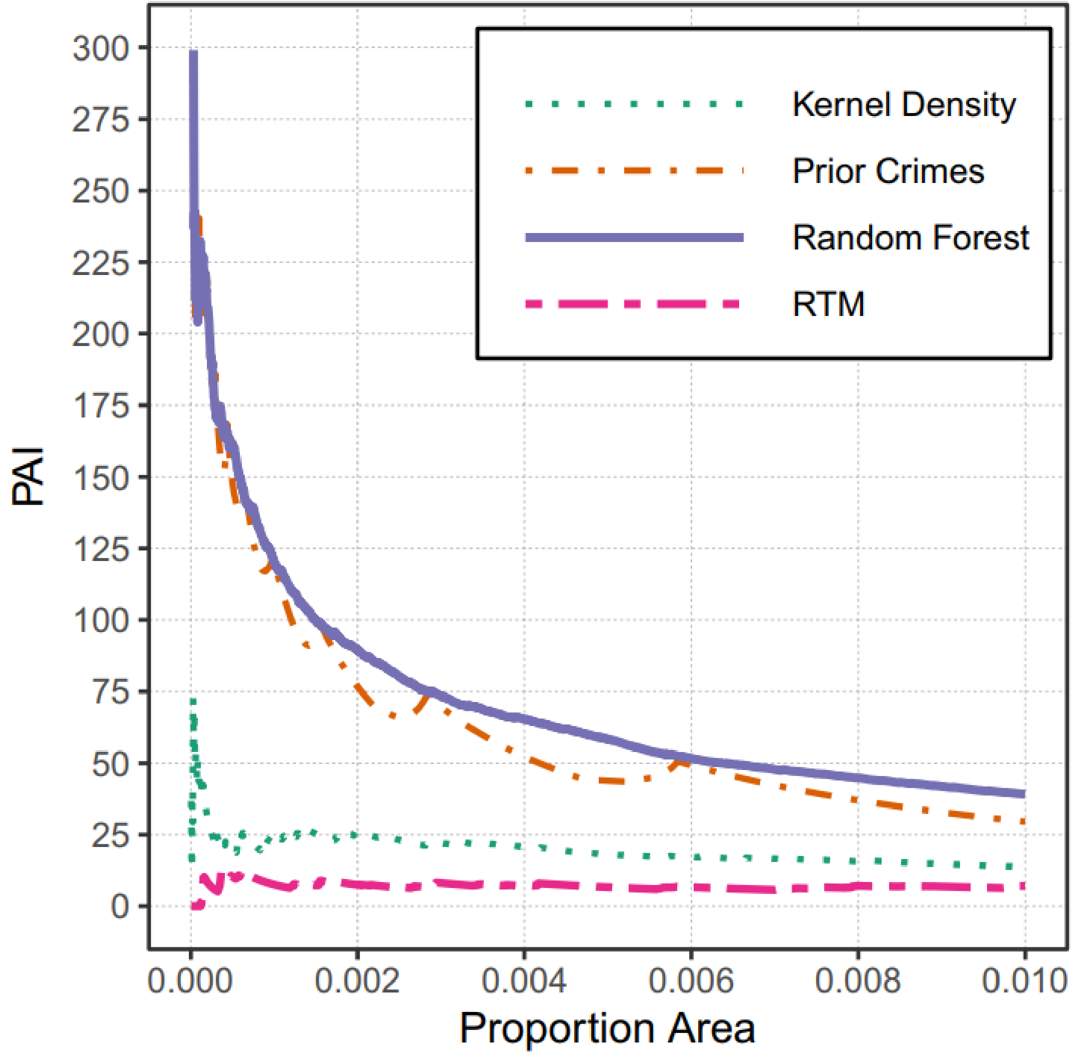
CRIME De-Coder

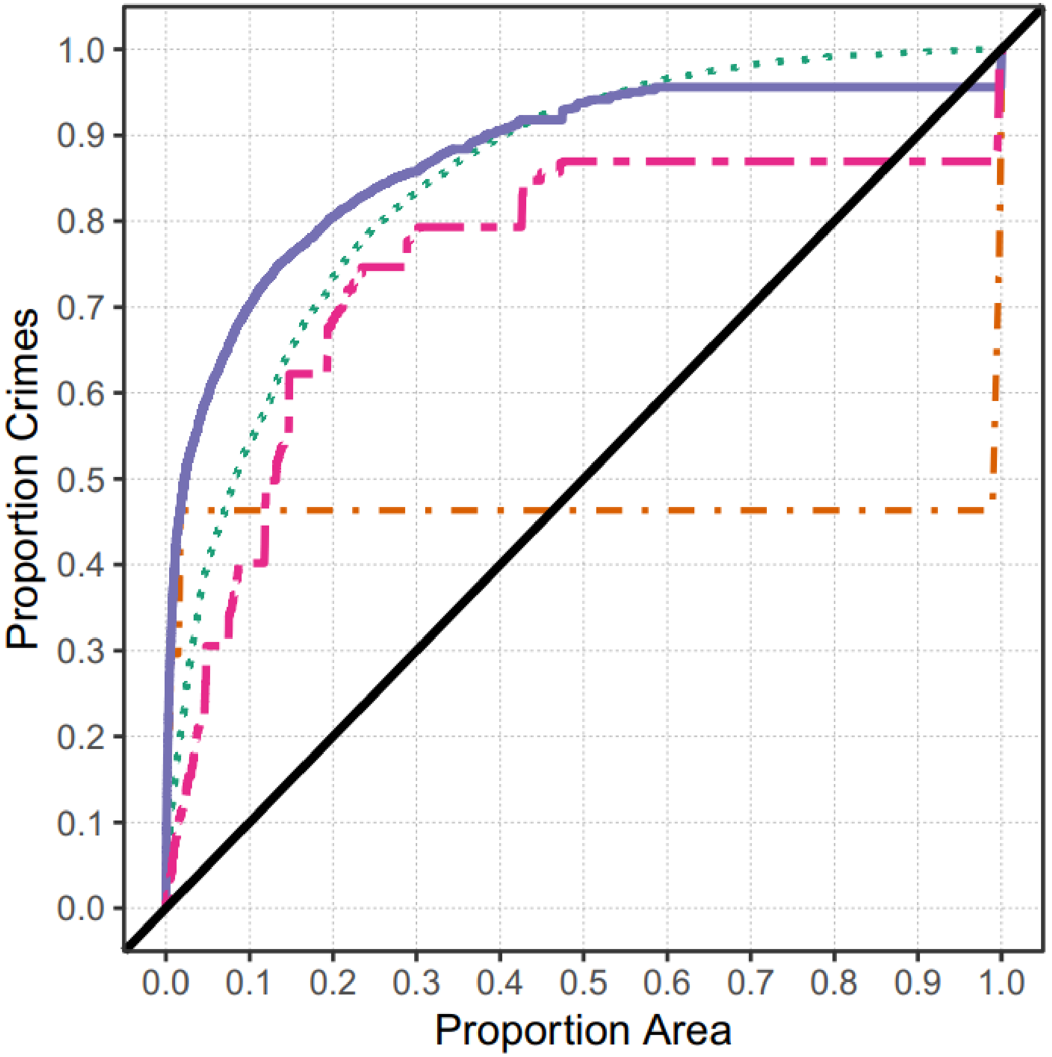
Table 2 Accuracy metrics for fixed thresholds

Number areas	RTM		Kernel density		Random forest		Prior Crimes	
	Cum. Crime	PAI	Cum. crime	PAI	Cum. crime	PAI	Cum. crime	PAI
1	4	146.9	1	36.7	9	330.5	9	330.5
10	9	33.0	16	58.8	66	242.3	60	220.3
50	23	16.9	37	27.2	270	198.3	260	190.9
100	56	20.6	59	21.7	437	160.5	436	160.1
500	262	19.2	324	23.8	1145	84.1	930	68.3
1000	391	14.4	515	18.9	1662	61.0	1239	45.5

Results

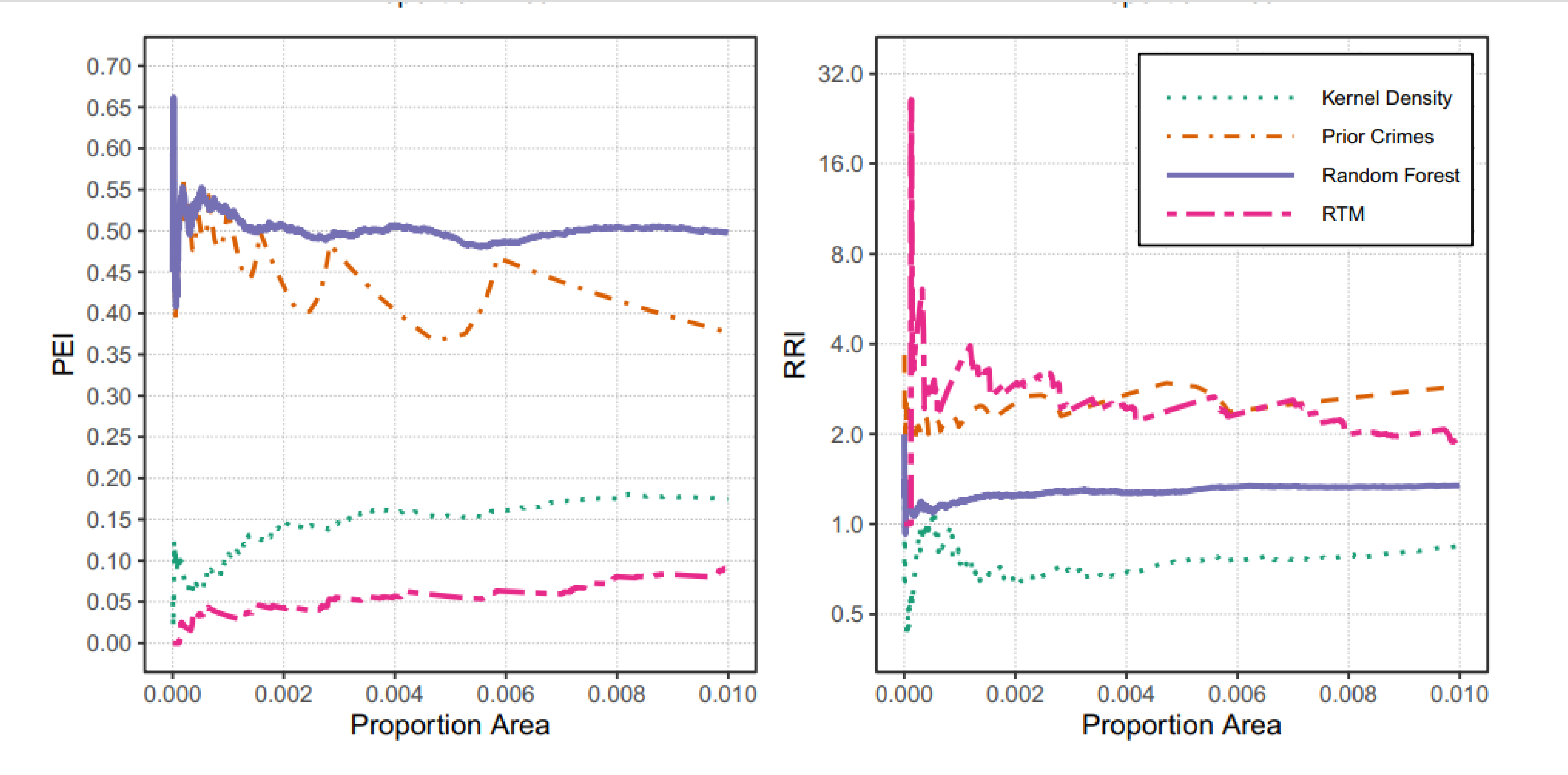
CRIME De-Coder





Results

CRIME De-Coder



Interpreting Random Forests

CRIME De-Coder

Interpreting Random Forest Models

- **Average Local Effects**
  - **Conduct a simulation, slightly change inputs, see how *average* prediction changes**
- **Shapley Value Decomposition**
  - **If a location is predicted to have 4 robberies, 0.5 due to nearby apts, 0.1 due to nearby DART station, etc.**
- **I do not like "variable importance scores" (volatile, easy to misinterpret)**

Interpreting Random Forests

CRIME De-Coder

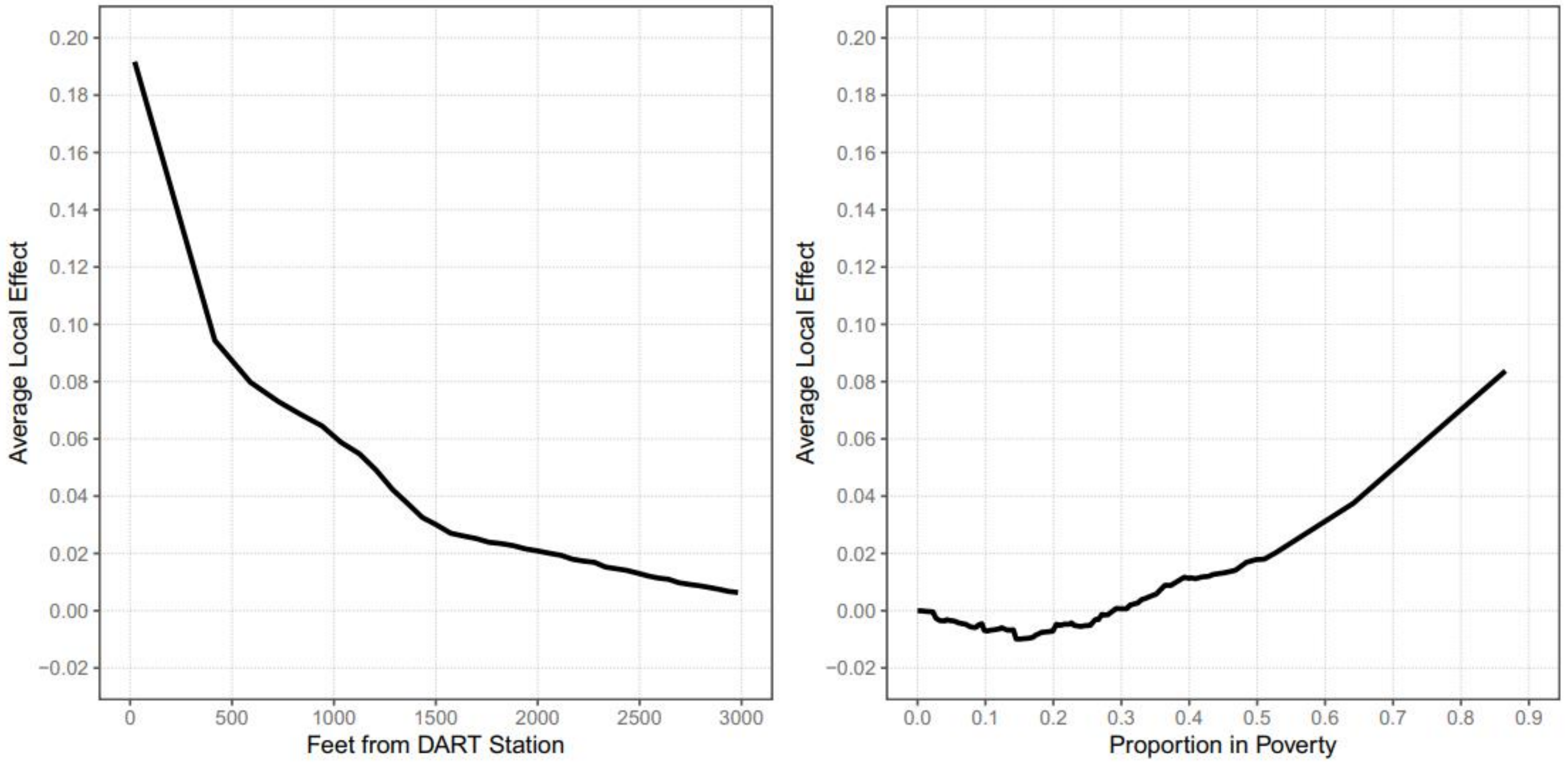


Fig. 2 The average local effect of the distance to the nearest train station (left panel), and the proportion in poverty (right panel)

Interpreting Random Forests

CRIME De-Coder

 Journal of Quantitative Criminology (2021) 37:445–480

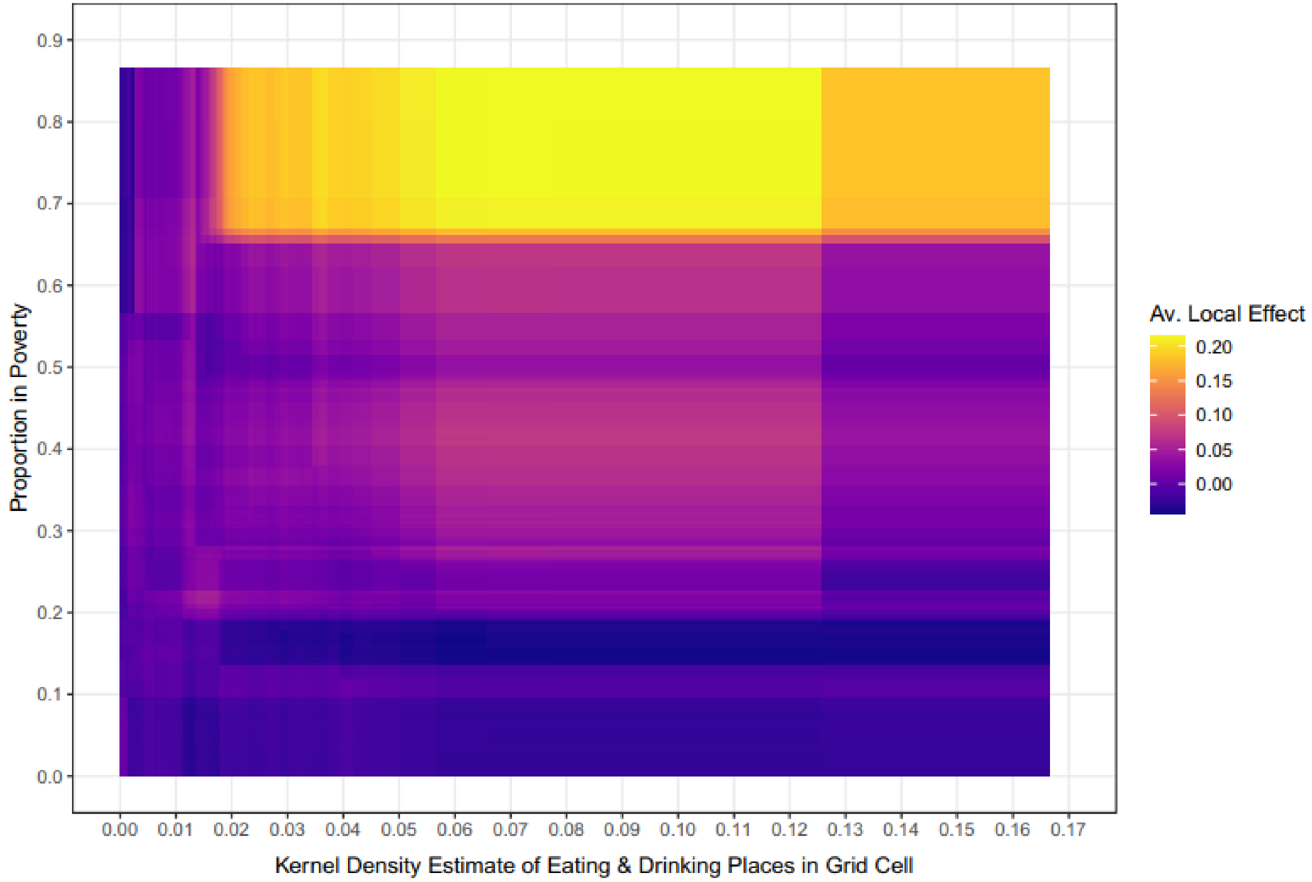


Fig. 3 The average local effect when varying two variables, the density of eating and drinking places and the proportion of poverty within a census tract

Interpre

CRIME De-Coder

Journal of Quantitative Criminology (2021) 37:445–480

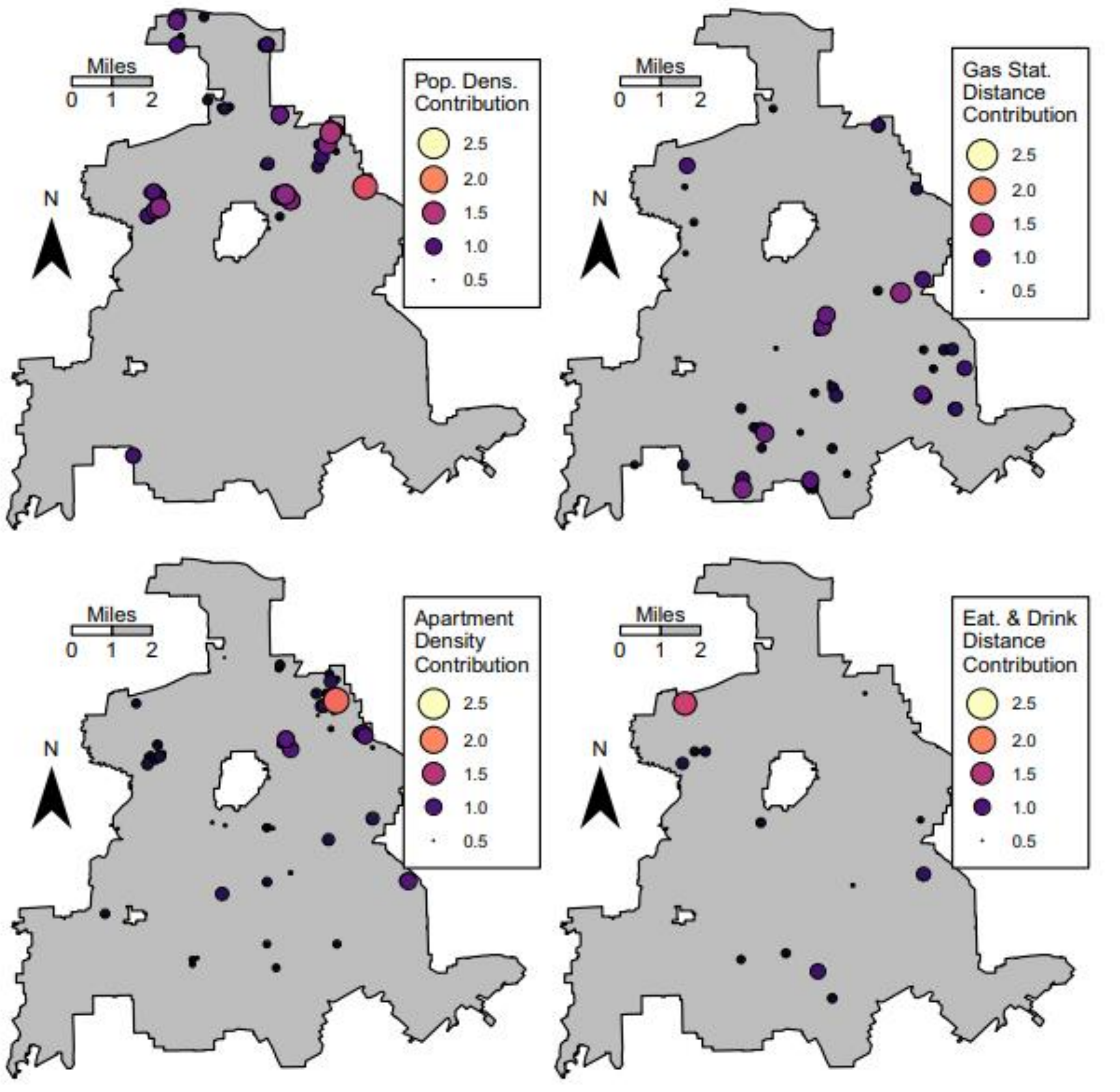
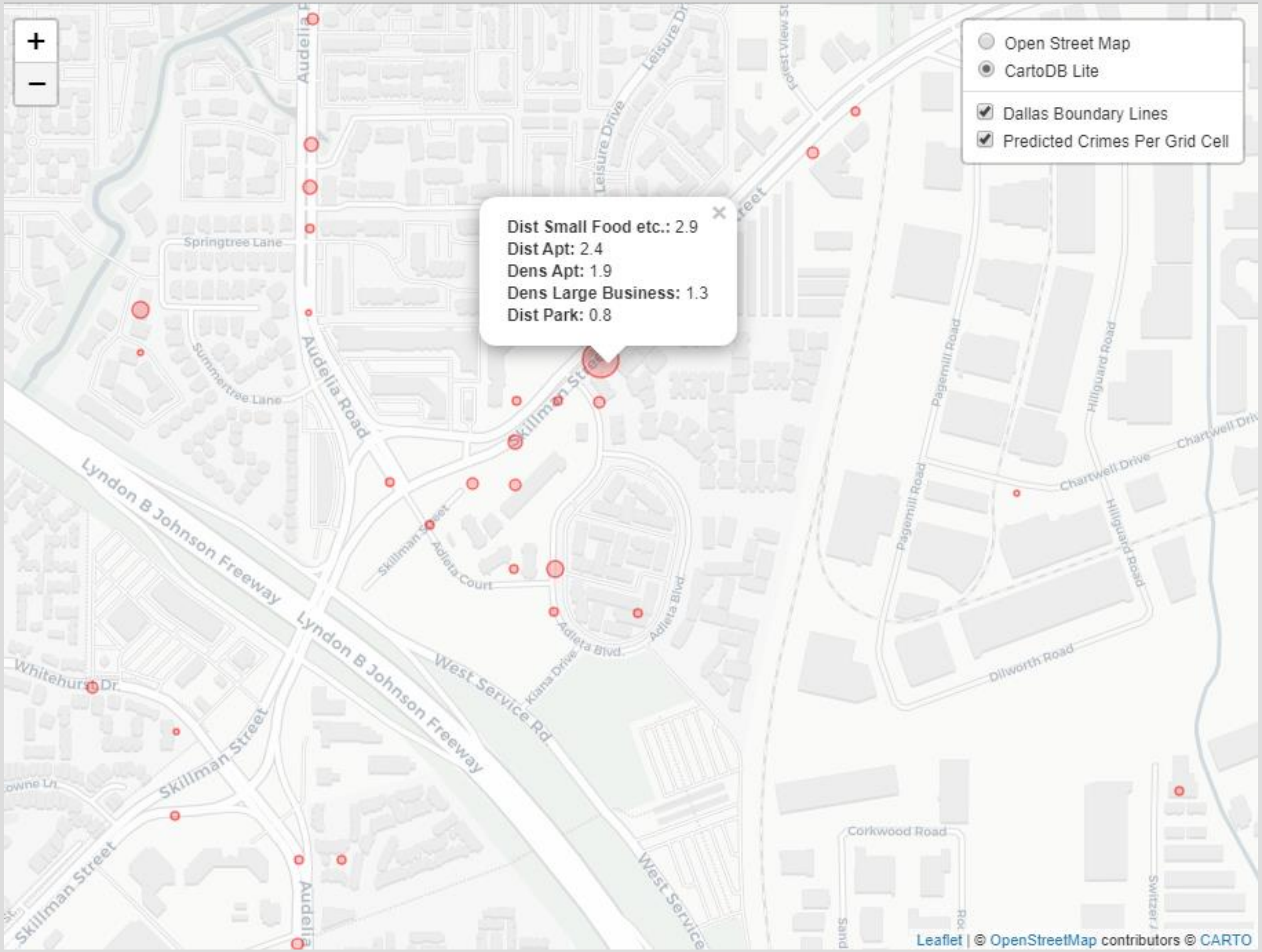


Fig. 4 Contribution of different risk factors to predicted crime counts over space. Factors are calculated using Shapley value regression, and locations with a risk factor of over 0.5 are shown

Interpreting Random Forests

CRIME De-Coder



Takeaways

CRIME De-Coder

Lessons from Analysis

- **Should always show "simple" baseline**
  - **RTM performs *much worse* than simple prior ranking**
  - **Random Forest only slightly beats simple ranking**
  - **Need to do train/test**

- **Random Forest still has some benefits**
  - **Slightly better forecasts, but more accurate *cumulative* than naïve**
  - **Much better job discriminating between prior 0 crime locations**
  - **Complicated, but can do reduced form summaries**

- **But reduced form summaries of models can be misleading (Rudin's work)**

General Advice / Future Work

CRIME De-Coder

Random Forest Tips

- Binary predictions often need to limit depth of trees (and/or sample size splits) to prevent over-fitting

- Ditto for boosted model variants

- Can use out-of-bag estimates to produce forecast intervals

- Tend to only beat traditional regression models post 20k observations in my experience

Other Work of Interest

CRIME De-Coder

- **Fairness in predictive policing allocation**
  - **Wheeler, A.P. (2020). Allocating police resources while limiting racial inequality. *Justice Quarterly*, 37(5), 842-868.**
- **Cost-benefit analysis when to allocate patrols to hotspot**
  - **Wheeler, A.P., & Reuter, S. (2021). Redrawing Hot Spots of Crime in Dallas, Texas. *Police Quarterly*, 24(2), 159-184.**
- **Optimal Spatial Districting with workload equality**
  - **Wheeler, A.P. (2018). Creating optimal patrol areas using the p-median model. *Policing: An International Journal*, 42(3), 318-333.**
- **Preventing future near-repeat crimes via arrest**
  - **Wheeler, A.P., Riddell, J.R., & Haberman, C.P. (2021). Breaking the chain: How arrests reduce the probability of near repeat crimes. *Criminal Justice Review*, 46(2), 236-258.**

Mapping the Risk Terrain for Crime using Machine Learning

09-23

Contact: andrew.wheeler@crimede-coder.com
Website: crimede-coder.com

CRIME De-Coder