SDAL SOCIAL & DECISION ANALYTICS LABORATORY

Data Science for the Public Good

RESIDENTIAL SMOKE ALARM NEED IN ARLINGTON COUNTY

Mark Almanza (Virginia Tech), Will Sandholtz (UC Berkeley), Sam Tyner (Iowa State)
with Gizem Korkmaz, Joshua Goldstein, and Vicki Lancaster (SDAL),

Research Question

Are there characteristics that can be used to identify homes that do not have smoke alarms installed? Can these characteristics be used in a model for prediction to aid the Arlington County Fire Department (ACFD) with their Operation FireSafe program?

BACKGROUND: The ACFD runs the Operation FireSafe initiative to increase the number of residences in the county with smoke alarms. Building a model to identify homes most in need of an alarm may increase the effectiveness of the program. A similar data-driven effort was implemented in New Orleans [1].

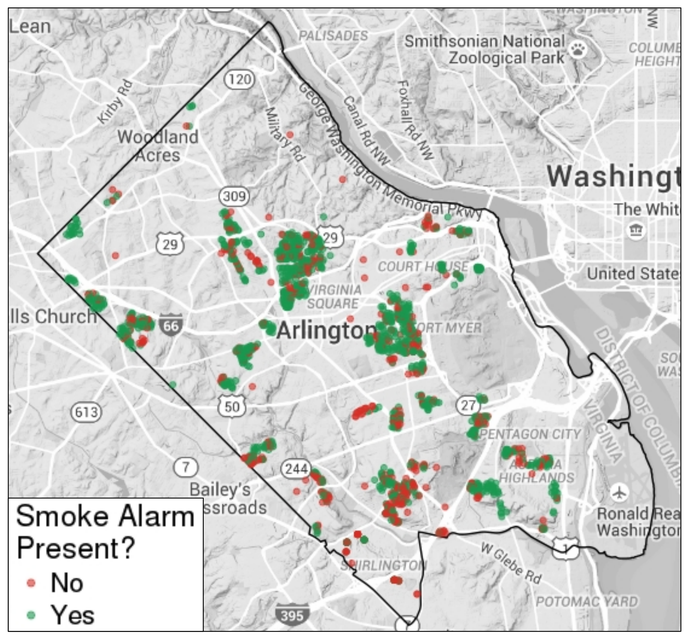
Data Sources

Operation Firesafe

- ACFD visited 5623 unique addresses in Arlington.
- 1,733 visits contained information on presence of smoke alarms in the home.
- 32 percent of these homes did not previously have smoke alarms installed.
- The dataset includes time and date of the visit, address, outcome, number of alarms installed, and number of alarms in the home before the visit.

CoreLogic

- The dataset contains information, such as house value, size, age, number of bedrooms, and ownership, for 60,343 housing units in Arlington County.



Operation FireSafe visits in Arlington County.

Modelling Smoke Alarm Need

Taking into account the locations of residential units in Arlington County and incorporating the merged Operation Firesafe and CoreLogic data sources, we build a model to predict regions of the county in need of smoke alarms.

We fit a Bayesian logistic regression model [2] with conditionally autoregressive spatial effects. Two factors effect the predicted probabilities:
1. the *covariates of interest* and
2. the location (census block group) of the home.

The response is a Bernoulli random variable at the level of the home:
$Y_i = 0$ no smoke alarm, ACFD installed one or more alarms and
$Y_i = 1$ one or more smoke alarms already present.

$$Y_i \sim \text{Bernoulli}(p_i)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \mu_i + \phi_i \qquad \mu_i = \mathbf{X}_i^T\boldsymbol{\beta} \qquad \boldsymbol{\beta} \sim N(\mathbf{0}, \mathbf{I}_{1000})$$

$$\phi_i|\boldsymbol{\phi}_{-i}, \mathbf{W}, \tau^2, \rho \sim N\left(\frac{\rho\sum_{k=1}^K w_{ik}\phi_k}{\rho\sum_{k=1}^K w_{ik} + 1 - \rho}, \frac{\tau^2}{\rho\sum_{k=1}^K w_{ik} + 1 - \rho}\right)$$

$$w_{ik} = \begin{cases} 1 & \text{if houses } i \text{ and } k \text{ are in the same census block group} \\ 0.5 & \text{if houses } i \text{ and } k \text{ are in neighboring census block groups} \\ 0 & \text{otherwise} \end{cases}$$

$$\tau^2 \sim \text{Inverse-Gamma}(5, 40)$$

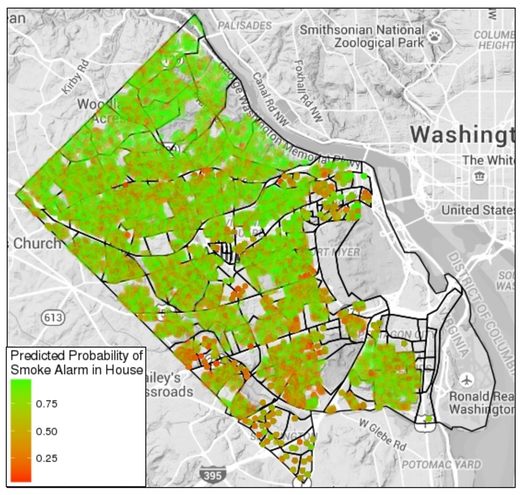
$$\rho \sim \text{Uniform}(0, 1)$$

The *covariates of interest* that we use in modelling are:
- the natural log of the value of the home (a proxy for income)
- the age of the home
- an indicator of whether or not the residence is a condo
- the number of bedrooms (a proxy for household size)
- an indicator of whether or not the residence is owner occupied

The parameter estimates and credible intervals are given below:

Parameter Value	Posterior Median	95% Credible Interval
Intercept*	-30.12	(-59.60, -20.39)*
log(HomeValue)*	2.45	(1.67, 4.83)*
Age*	-0.0109	(-0.0257, -0.0022)*
Condo Indicator	-0.48	(-1.58, 0.22)
Number of Bedrooms	-0.12	(-0.39, 0.09)
Owner Occupied Indicator	0.02	(-0.61, 0.65)

* Effect is significant at 5% level



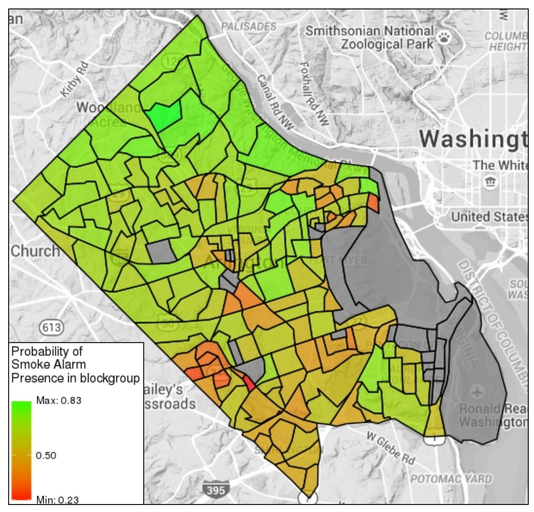
The model estimates the probability that a housing unit has a smoke alarm. Because of the spatial dependence included in the model, the predictions are made for each residential unit in the CoreLogic data source that are:
- in census block groups that appeared in the original data or in a census block group neighboring those in the original data; and
- with no missing covariate values and/or census block group information.

Consequently, predictions are made for 49,178 residential units.

The figure on the left shows the probability of each home having a smoke alarm. The lower the predicted probability (red areas) the more likely the home lacks a smoke alarm. The figure on the right shows the average probability for homes within a census block group.

By focusing on the homes with a low probability of an installed smoke alarm, the ACFD can better target future Operation FireSafe visits.

Our next step is to incorporate other sources of data (e.g., American Community Survey (ACS)) into the model to capture the effects of sociodemographic characteristics.



References

[1]"City Launches New Smoke Alarm Initiative Based on Data Analytics", Mayoral Press Release, City of New Orleans. http://www.nola.gov/mayor/press-releases/2015/20150331-smoke-alarm-data-campaign/
[2] Duncan Lee (2016). CARBayes: Spatial Generalised Linear Mixed Models for Areal Unit Data. R package version 4.6. https://CRAN.R-project.org/package=CARBayes