Predicting Homicides and Shootings
and Allocating Police Resources to Prevent Them

On Beyond Zebra:
Challenges in Predictive Modeling

Thomas Darling, Ph.D., Associate Professor
School of Public and International Affairs
and Schaefer Center for Public Policy

Debra Stanley, Ph.D., Executive Director
School of Criminal Justice

University of Baltimore

July 22, 2011

Predicting Homicides and Shootings
and Allocating Police Resources to Prevent Them

Project purpose:

- Distinguish areas within three previously identified high crime districts in the city of Baltimore that are more likely (less likely) than other areas to have a homicide or shooting within the next week.
- Use these predictions to better allocate scarce police resources to prevent the occurrence of homicides, shootings, and other crimes.

Predicting Homicides and Shootings
and Allocating Police Resources to Prevent Them

Areas (hexagons) within districts:

- Each district was "tiled" with hexagons that were one-tenth of a mile wide (flat-side to flat-side). A ring of adjacent (*"peri-meter"*) hexagons surround each district's *"core"* hexagons.
  - the Eastern District has 178 core hexagons, and ≈96 perimeter hexagons (9,256 core hex-weeks in a year);
  - the Western District has 187 core hexagons, and ≈91 perimeter hexagons (9,724 core hex-weeks in a year); and,
  - the Northwestern District has 117 core hexagons, and ≈60 perimeter hexagons (6,084 core hex-weeks in a year).
- Dependent events (homicides and shootings) and independent variables (other crimes and 911 call data) are mapped into the core and perimeter hexagons.
- A specific hexagon is referred to as the *"focal"* hexagon. The six (6) surrounding hexagons are the focal hexagon's *"neighbors."*

Predicting Homicides and Shootings
and Allocating Police Resources to Prevent Them

Dependent Variable: Homicide/Shooting Incidents

- Includes all non-domestic homicide incidents within the high crime districts, involving a firearm, knife, or other means.
- Includes all non-domestic shooting incidents. Does not include cuttings or assaults where a weapon was not fired.
- During the 52-week period between ≈July 1, 2009 and ≈June 30, 2010 there were:
  - 72 homicides and shootings in Eastern core hexagons;
  - 67 homicides and shootings in Western core hexagons; and,
  - 33 homicides and shootings in Northwestern core hexagons.
- During the 52-week period between ≈July 1, 2008 and ≈June 30, 2009 there were:
  - 77 homicides and shootings in the Eastern core hexagons;
  - 68 homicides and shootings in Western core hexagons; and,
  - 26 homicides and shootings in Northwestern core hexagons.

Predicting Homicides and Shootings

And Allocating Police Resources to Prevent Them

Seven independent "factors":

- Three of the factors relate to ***crimes***:
  - homicides and shootings (previous occurrences of the dependent variable);
  - armed assaults; and,
  - armed robberies.
- Four of the factors relate to ***calls*** reporting:
  - aggravated assaults;
  - armed persons;
  - "open air" (outside) drugs; and,
  - "on-view" drugs.

Predicting Homicides and Shootings
and Allocating Police Resources to Prevent Them

For each factor, three independent variables were calculated (21 variables total):

- The number of crimes/calls during the preceding 52 weeks in the *focal* hexagon.
- The change in crimes/calls in the *focal* hexagon between the preceding week and the average of the three weeks before the preceding week.
- The number of crimes/calls during the preceding 52 weeks in the *neighboring* hexagons.
- The *new version* of the analysis will add the change in crimes/calls in the *neighboring* hexagons between the preceding week and the average of the three weeks before the preceding week.

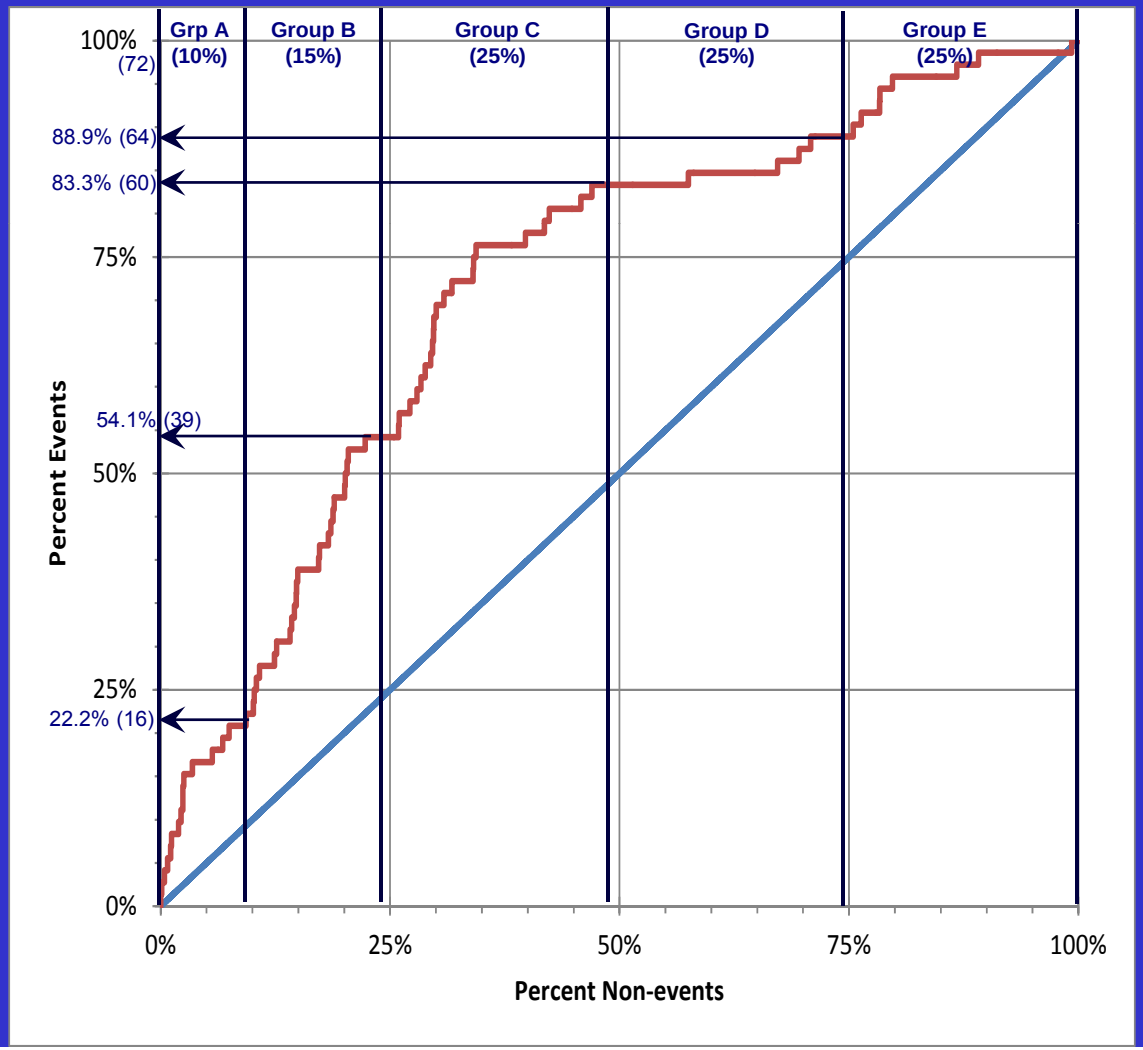
Predicting Homicides and Shootings
and Allocating Police Resources to Prevent Them

- All 21 variables were used to create a scoring rule *"trained"* on the YE2010 Eastern District data. For AY2010, applying the scoring rule leads to an A-score of 0.717 (see ROC curve below). In other words the scoring rule accounts for more than 40% of the uncertainty →
  - The 10% of hex-weeks in Category A (highest scores) include 22.3% of the 72 homicides and shootings during the year,
  - the 15% of hex-weeks in Category B account for another 31.9% of the homicides and shootings during the year, and
  - only 16.7% of the homicides and shootings occurred in the 50% of the hex-weeks in Categories D and E combined, fewer than in the 10% of hex-weeks in Category A.

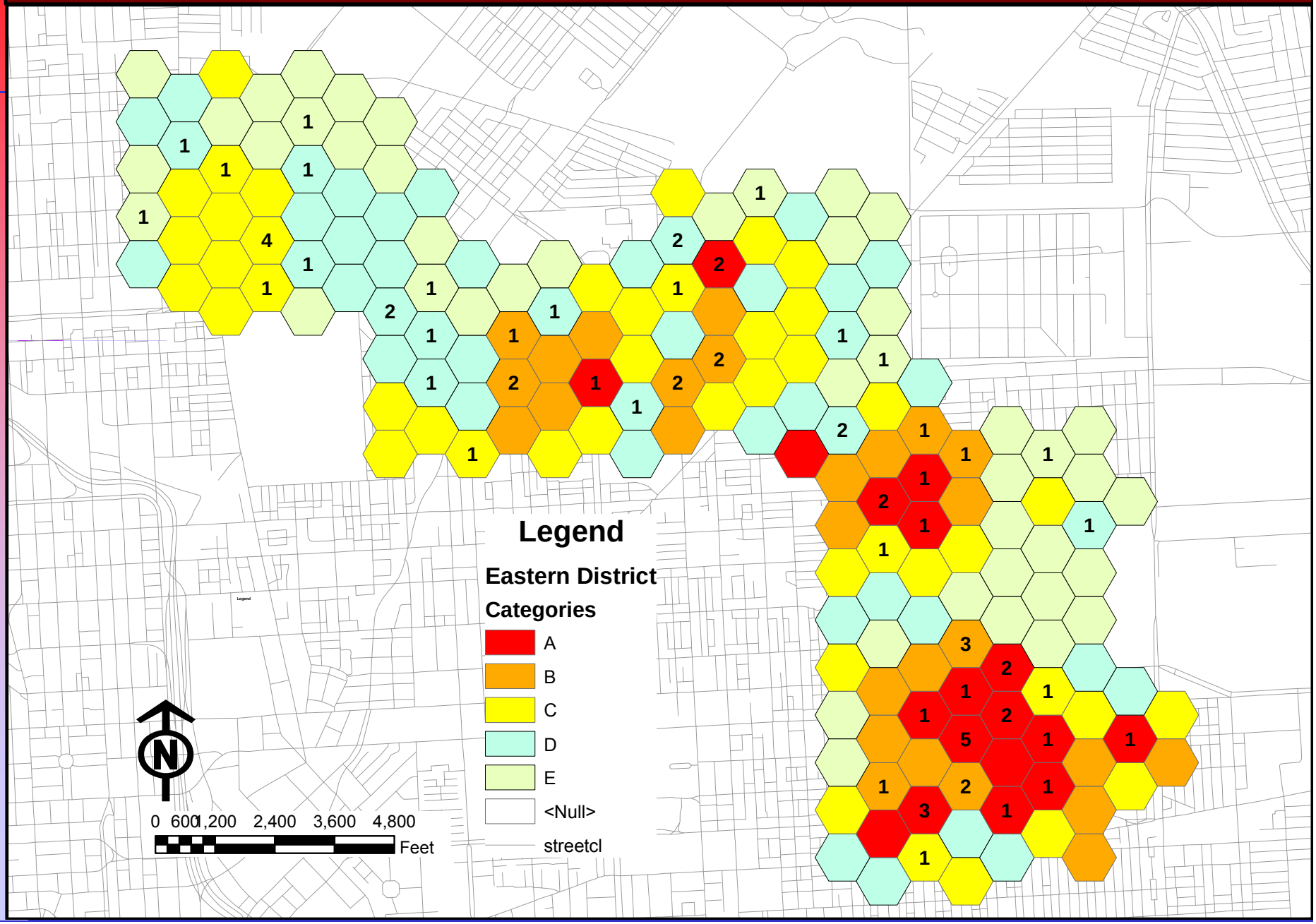
Category	*Likelihood*	*Count of Events*	*Count of Non-evts*	*Percent Events*	*Percent Non-evts*	*Percent Hex-Weeks*	*Odds*
A	0.0173	16	909	22.2%	9.9%	10.0%	2.2
B	0.0166	23	1,366	31.9%	14.9%	15.0%	2.1
C	0.0091	21	2,293	29.2%	25.0%	25.0%	1.2
D	0.0017	4	2,310	5.6%	25.2%	25.0%	0.2
E	0.0035	8	2,306	11.1%	25.1%	25.0%	0.4
Total	***0.0078***	***72***	***9,184***	***100.0%***	***100.0%***	***100.0%***	1.0

Predicting Homicides and Shootings
and Allocating Police Resources to Prevent Them

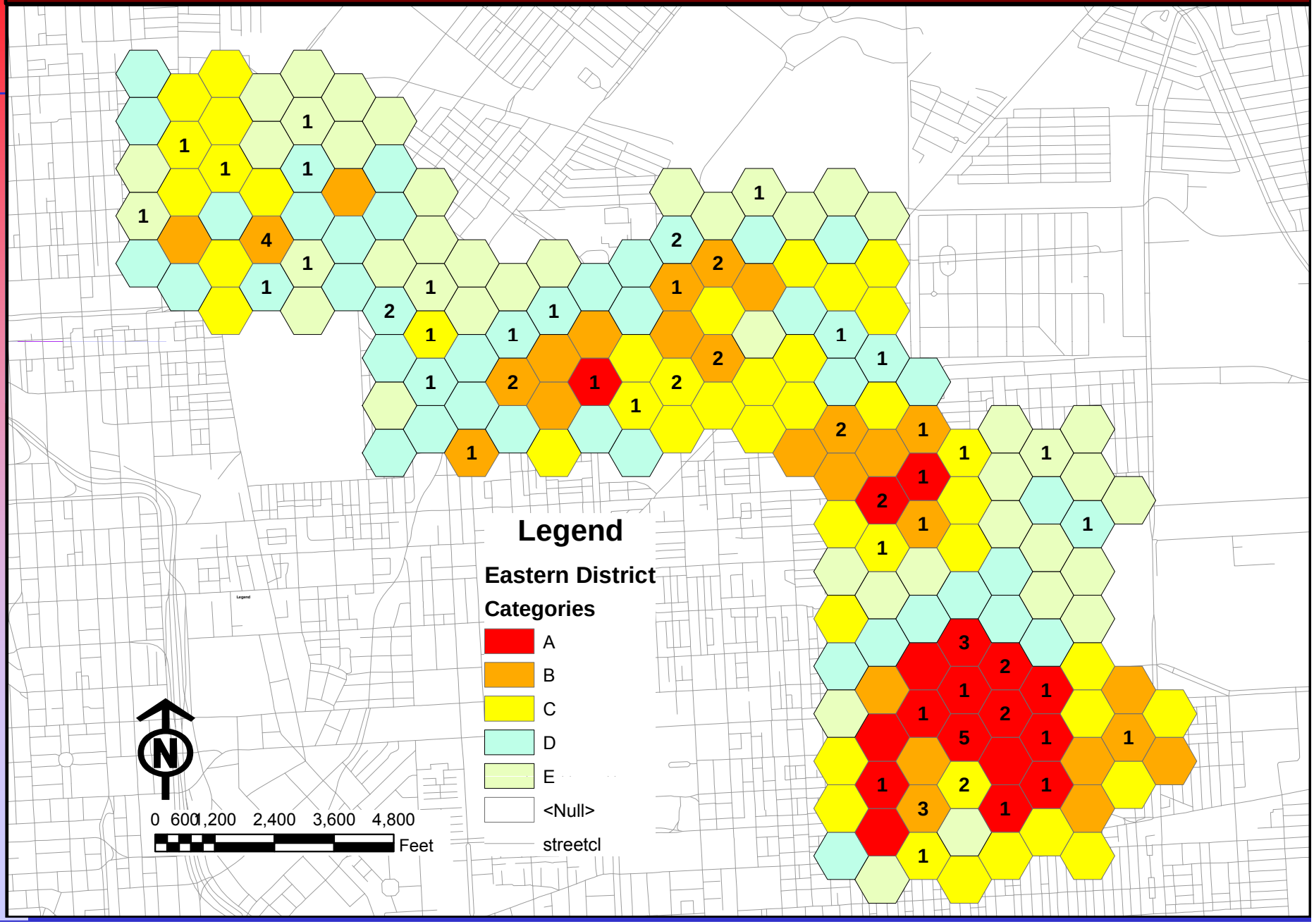
- The ROC curve below shows the results from the scoring rule using all 21 variables on the Eastern YE2010 data (A score = 0.717).



21 Variable Scoring Rule, Eastern District Effort Allocation Categories Week of Sept, 24, 2009 with Homicide and Shooting Count (YE 6/30/10)



21 Variable Scoring Rule, Eastern District Effort Allocation Categories Week of March 25, 2010 with Homicide and Shooting Count (YE 6/30/10)



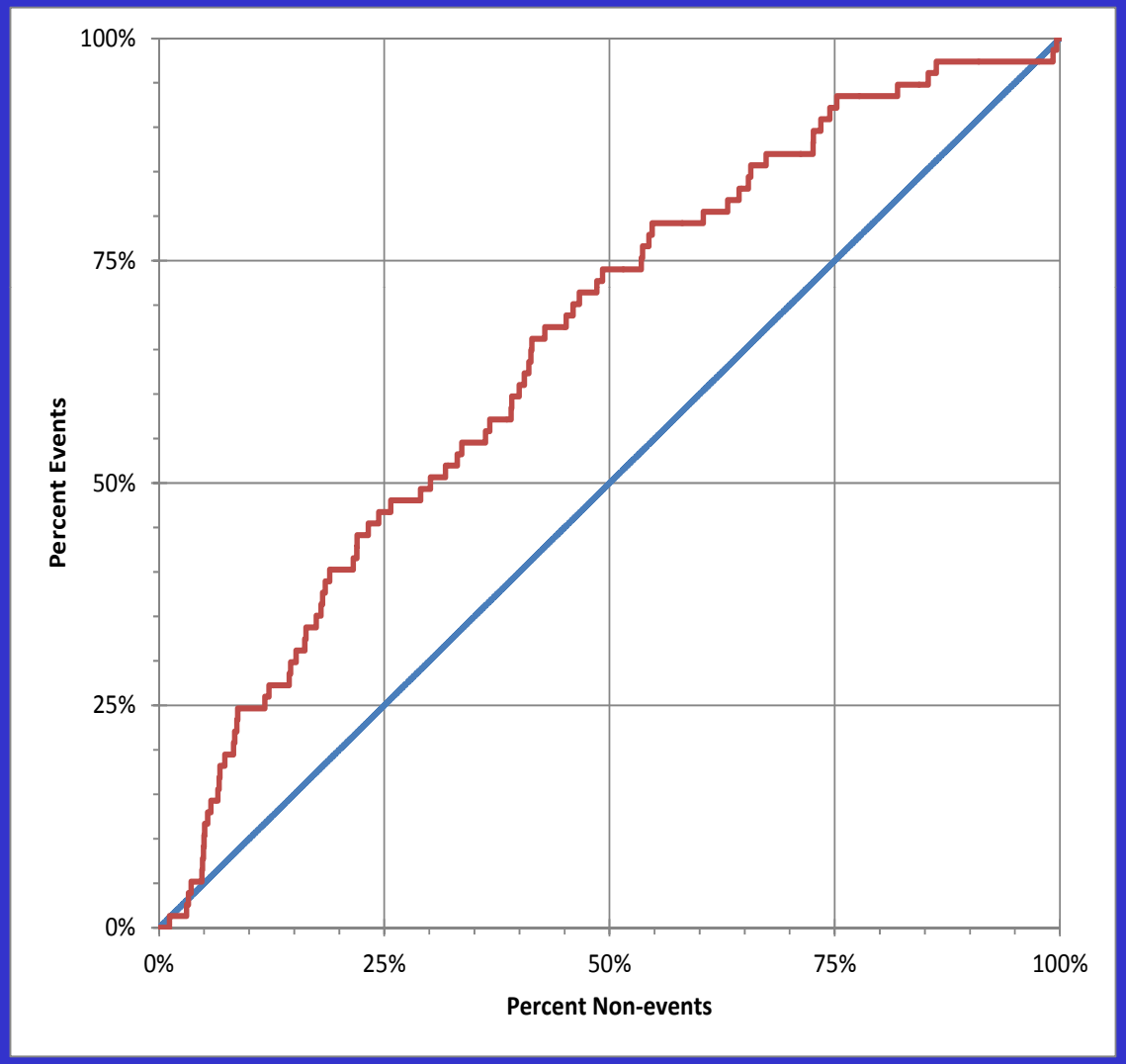
Predicting Homicides and Shootings
and Allocating Police Resources to Prevent Them

- ***THE PROBLEM.*** **The scoring rule using all 21 variables and "trained" on the Eastern YE2010 data doesn't perform as well when applied to the Eastern YE2009 data (A score = 0.655, see ROC curve below).** When the scoring is applied to the YE2009, it doesn't perform as well, although still not too badly →
  - 24.7% of the homicides and shootings occurred in the 10% of the hex-weeks in Category A,
  - 46.8% of the homicides and shootings occurred in the 25% of the hex-weeks in Categories A and B combined, and
  - Only 26% of the homicides and shootings occurred in the 50% of the hex-weeks in Categories D and E combined.

Category	*Likelihood*	*Count of Events*	*Count of Non-evts*	*Percent Events*	*Percent Non-evts*	*Percent Hex-Weeks*	*Odds*
A	0.0205	19	906	24.7%	9.9%	10.0%	2.5
B	0.0122	17	1,372	22.1%	14.9%	15.0%	1.5
C	0.0091	21	2,293	27.3%	25.0%	25.0%	1.1
D	0.0061	14	2,300	18.2%	25.1%	25.0%	0.7
E	0.0026	6	2,308	7.8%	25.1%	25.0%	0.3
Total	***0.0083***	***77***	***9,179***	***100.0%***	***100.0%***	***100.0%***	1.0

Predicting Homicides and Shootings

and Allocating Police Resources to Prevent Them



Predicting Homicides and Shootings
and Allocating Police Resources to Prevent Them

- ***THE CAUSE – OVERFITTING.*** The scoring rule developed using all 21 variables fit the YE2010 data ***too*** well. With only 72 homicides and shooting events the scoring rule fit "noise" as well as "signal."
- ***ONE SOLUTION – VARIABLE REDUCTION.*** The objective is to identify a subset containing only the strongest (most predictive) variables. With fewer "degrees of freedom" the scoring rule is less likely to fit "noise" – the idea is to develop a scoring rule that works almost as well on the YE2010 "training" data, and performs better on the YE2009 "testing" data, as the 21 variable model.
- Of course, there are other possible explanations as well, ***perhaps*** →
  - there is no "signal" – homicides and shooting events are entirely random (unpredictable), in which case the scoring rule was just fit to "noise," even in YE2010,
  - inter-variable correlations (co-linearity) muddied the waters, preventing the "right" variables from receiving enough weight, or
  - the "scarcity" of homicides and shootings prevents the discovery of an effective scoring rule.

Predicting Homicides and Shootings
and Allocating Police Resources to Prevent Them

A three (3) variable scoring rule:

- After reviewing the performance of variables in smaller models, a three variable scoring rule was adopted that performed quite well in both "training" and "testing" scenarios →
  - the number of armed assaults (crimes) occurring in the neighboring hexagons during the previous 52 weeks;
  - the number of aggravated assault calls received from the focal hexagon during the previous 52 weeks (this variable also includes a quadratic term); and,
  - a measure of the change in aggravated assault calls occurring in the focal hexagon in the most recent week versus the preceding three weeks (adjusted by the average number of calls per week during the preceding 52 weeks).
- ***Note:*** the *"testing"* phase used both the weights and variables from the *"training"* model.

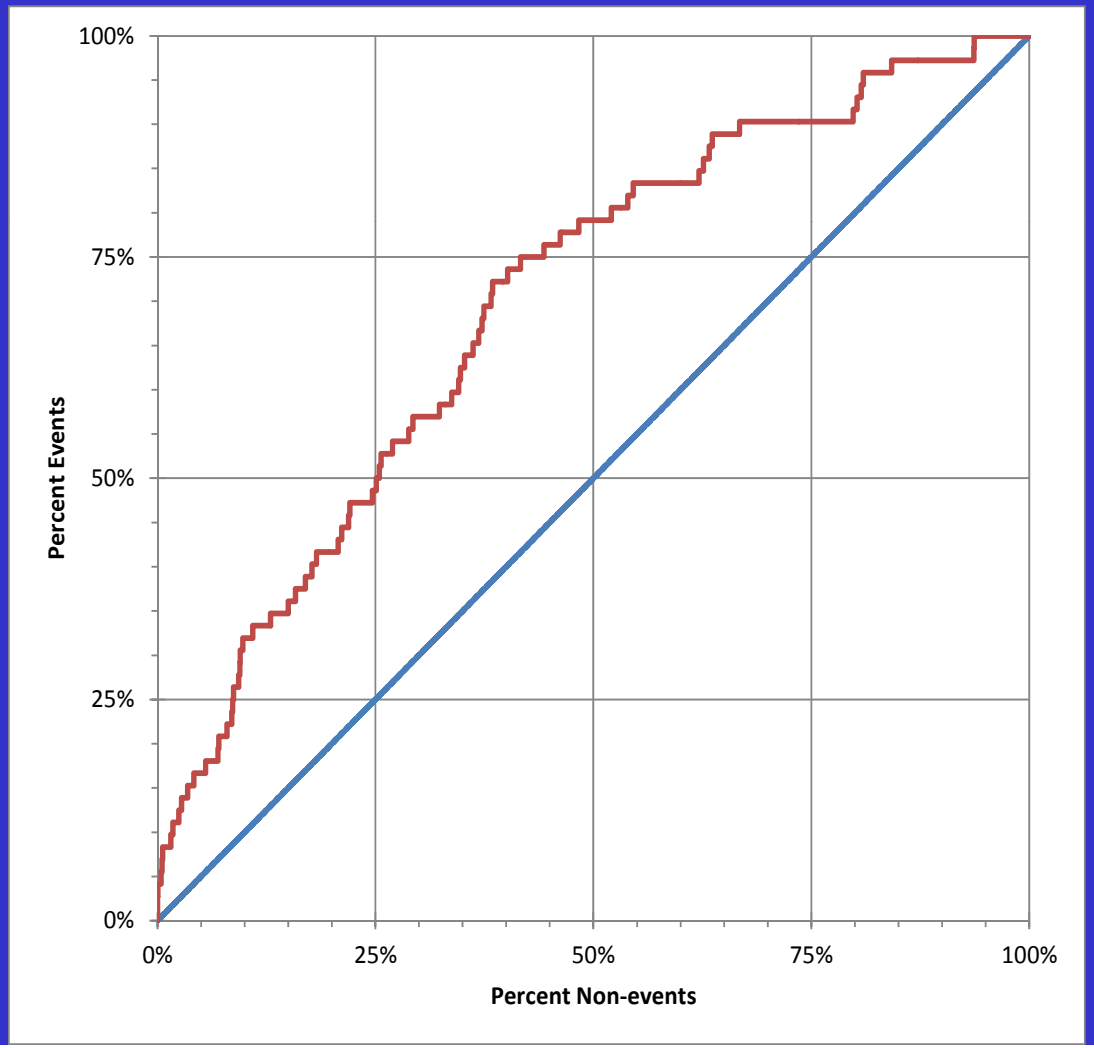
Predicting Homicides and Shootings
and Allocating Police Resources to Prevent Them

- The three variables were used to create a scoring rule *"trained"* on the YE2010 Eastern District data. For AY2010, applying the scoring rule leads to an A-score of 0.697. In other words, like the 21 variable model, the scoring rule accounts for nearly 40% of the uncertainty →
  - the 10% of hex-weeks in Category A include ***31.9%!*** of the 72 homicides and shootings during the year,
  - the 15% of hex-weeks in Category B account for another 16.7% of the homicides and shootings during the year, and
  - only 20.8% of the homicides and shootings occurred in the 50% of the hex-weeks in Categories D and E combined, fewer than in the 10% of hex-weeks in Category A.

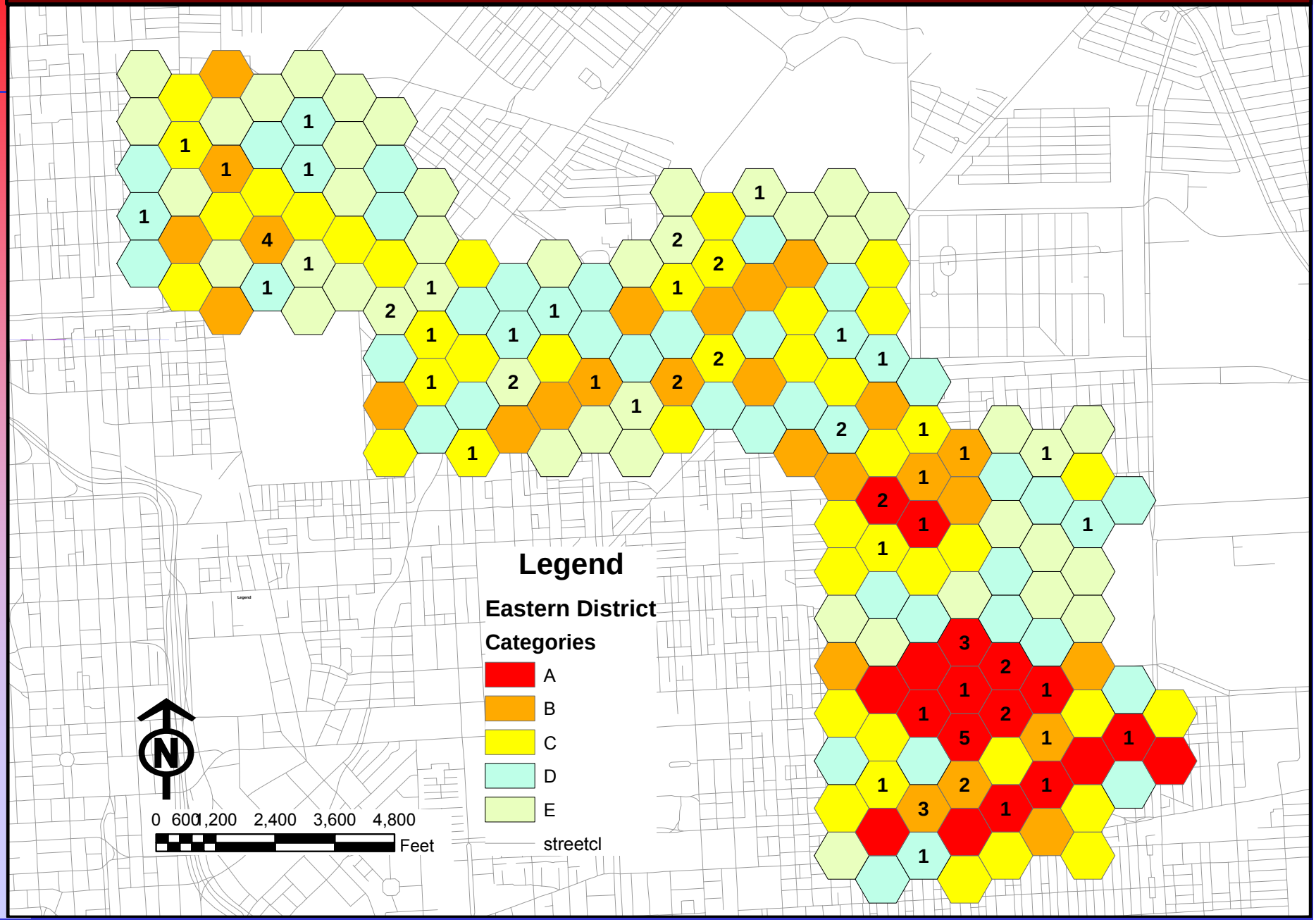
Category	*Likelihood*	*Count of Events*	*Count of Non-evts*	*Percent Events*	*Percent Non-evts*	*Percent Hex-Weeks*	*Odds*
A	0.0249	23	902	31.9%	9.8%	10.0%	3.2
B	0.0086	12	1,377	16.7%	15.0%	15.0%	1.1
C	0.0095	22	2,292	30.6%	25.0%	25.0%	1.2
D	0.0035	8	2,306	11.1%	25.1%	25.0%	0.4
E	0.0030	7	2,307	9.7%	25.1%	25.0%	0.4
Total	***0.0078***	***72***	***9,184***	***100.0%***	***100.0%***	***100.0%***	1.0

Predicting Homicides and Shootings
and Allocating Police Resources to Prevent Them

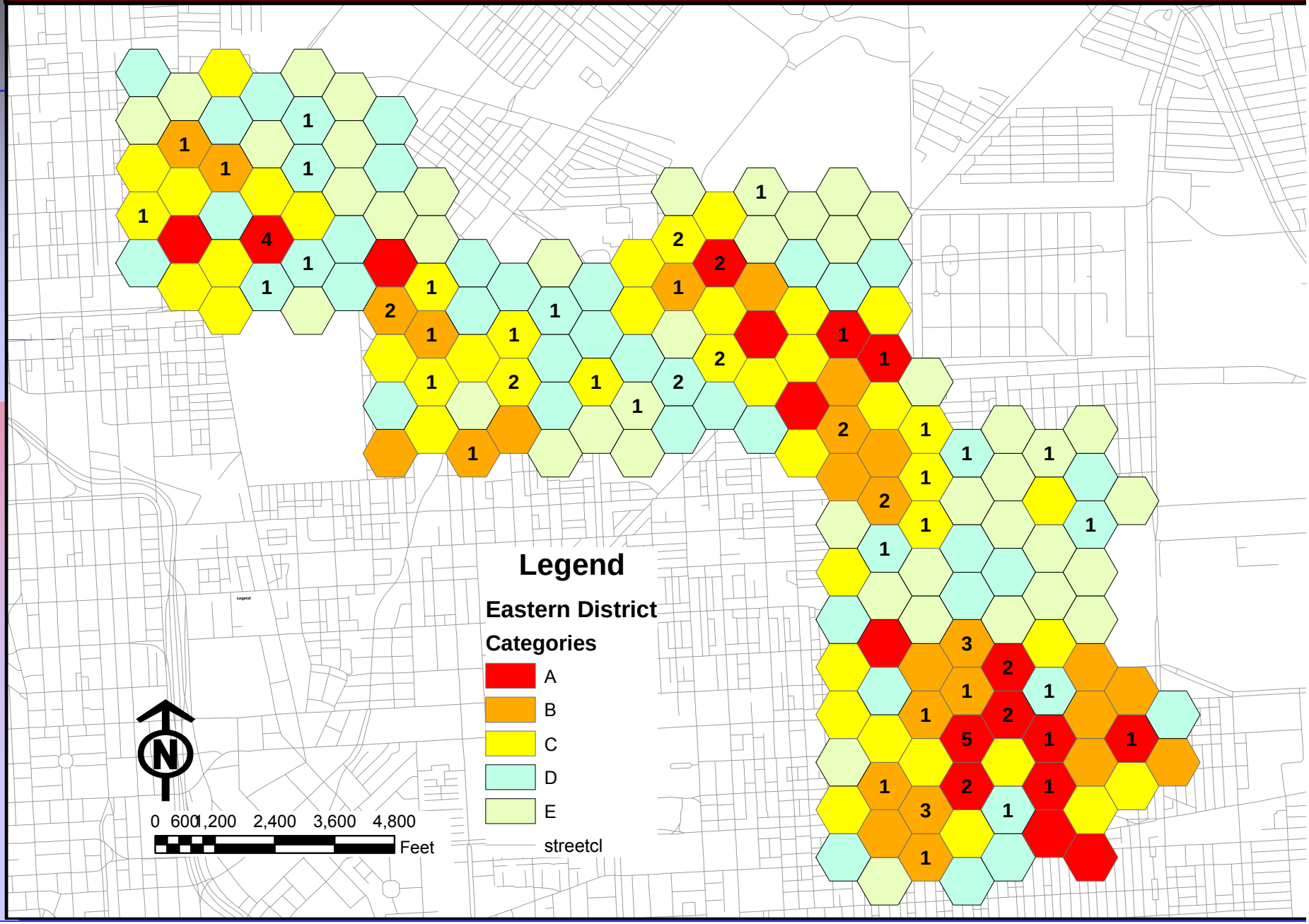
- The ROC curve below results from a scoring rule using only three variables *"trained"* on the Eastern YE2010 data (A score = 0.697).



3 Variable Scoring Rule, Eastern District Effort Allocation Categories Week of Sept 24, 2009 with Homicide and Shooting Count (YE 6/30/10)



Three Variable Scoring Rule, Eastern District Effort Allocation Categories
Week of March 25, 2010 with Homicide and Shooting Count (YE 6/30/10)



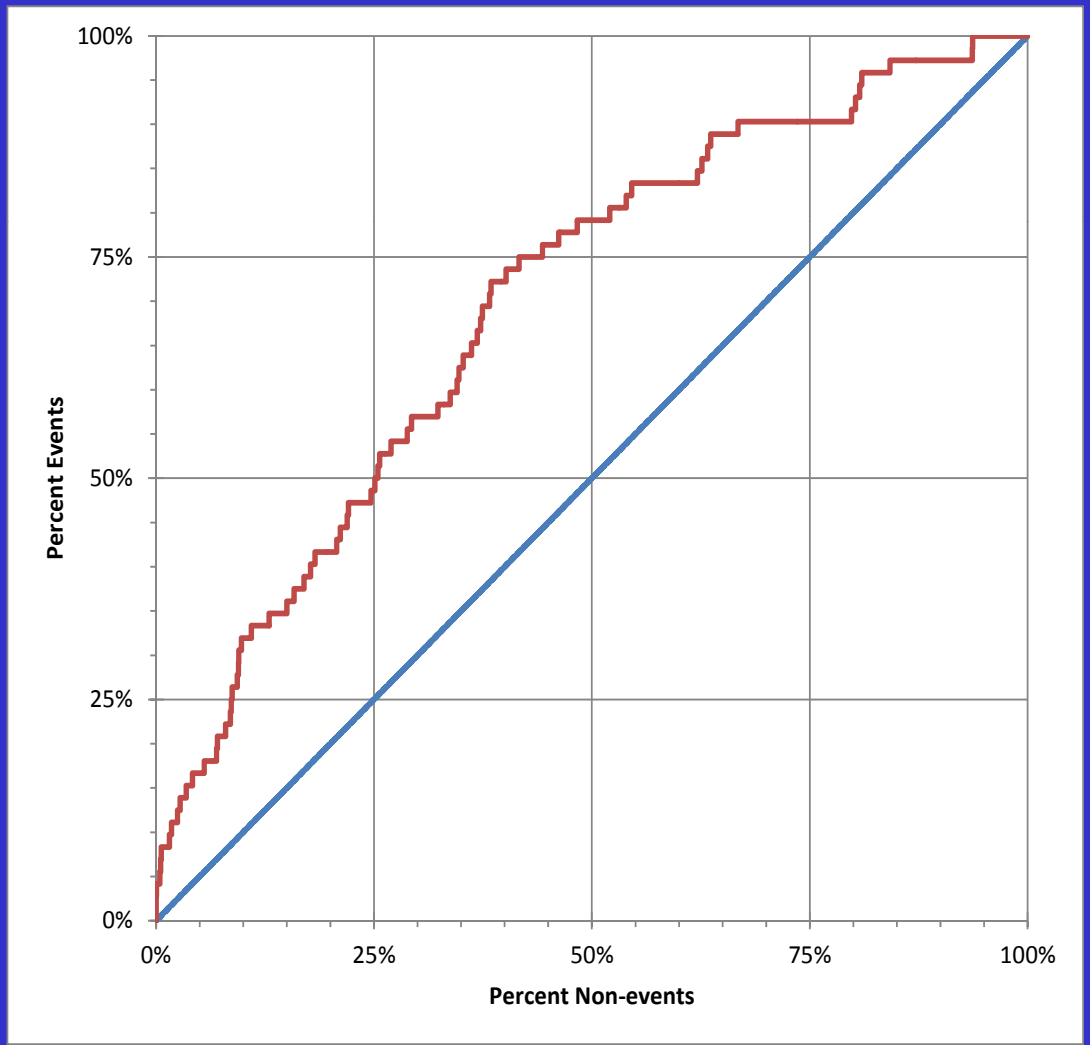
Predicting Homicides and Shootings
and Allocating Police Resources to Prevent Them

- When the three variable scoring rule "trained" on the YE2010 data is applied to the YE2009, it perform almost exactly as well as it did on the YE2010 training data (A score = 0.680) →
  - 22.1% of the homicides and shootings occurred in the 10% of the hex-weeks in Category A,
  - 52.0% of the homicides and shootings occurred in the 25% of the hex-weeks in Categories A and B combined, and
  - only 22.1% of the homicides and shootings occurred in the 50% of the hex-weeks in Categories D and E combined.

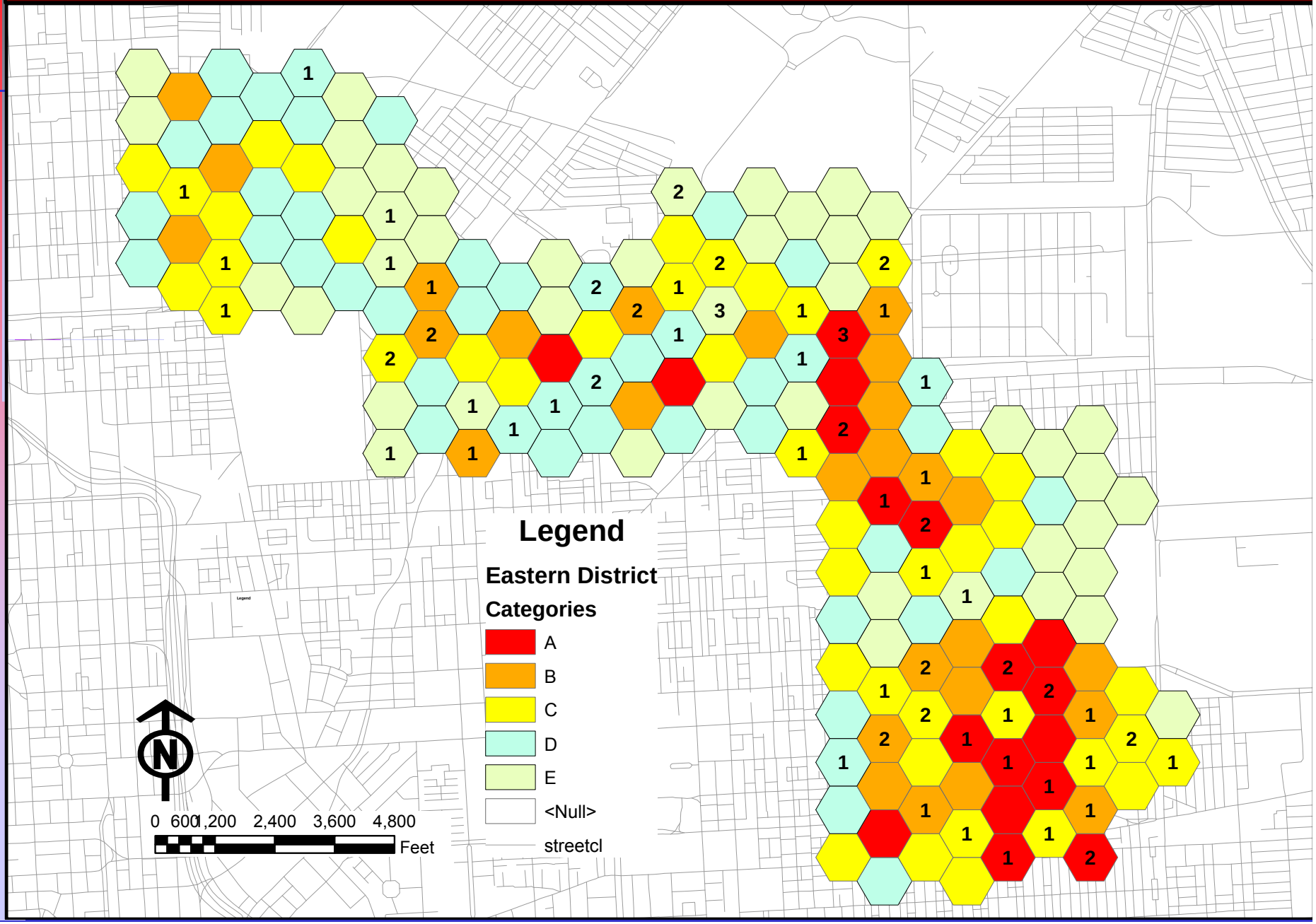
Category	*Likelihood*	*Count of Events*	*Count of Non-evts*	*Percent Events*	*Percent Non-evts*	*Percent Hex-Weeks*	*Odds*
A	0.0184	17	908	22.1%	9.9%	10.0%	2.2
B	0.0166	23	1,366	29.9%	14.9%	15.0%	2.0
C	0.0086	20	2,293	26.0%	25.0%	25.0%	1.0
D	0.0035	8	2,307	10.4%	25.1%	25.0%	0.4
E	0.0039	9	2,305	11.7%	25.1%	25.0%	0.5
Total	***0.0083***	***77***	***9,179***	***100.0%***	***100.0%***	***100.0%***	1.0

Predicting Homicides and Shootings
and Allocating Police Resources to Prevent Them

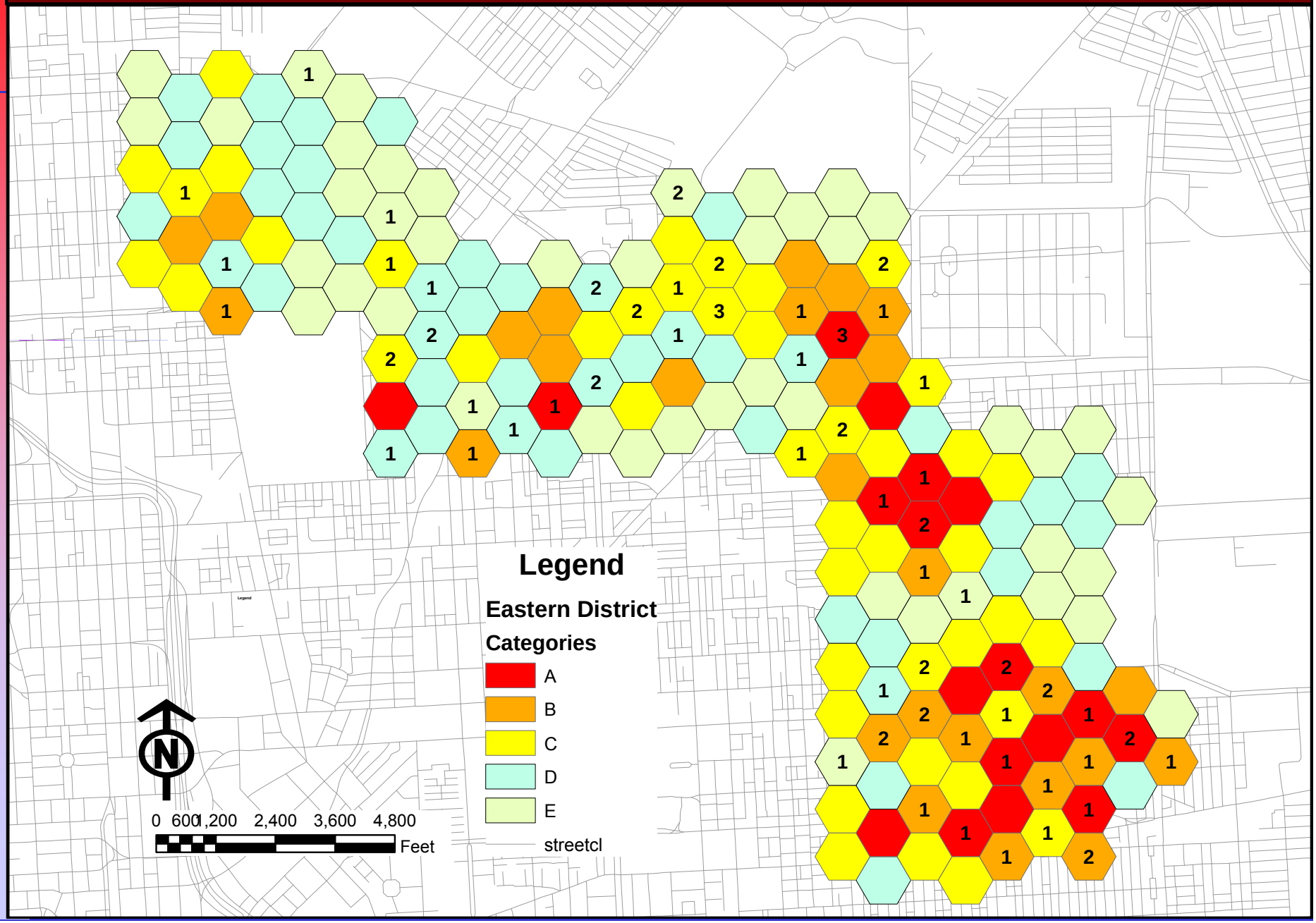
- The ROC curve below shows the results of applying the 3 variable scoring rule "trained" on YE2010 to YE2009 (A score = 0.680).



3 Variable Scoring Rule, Eastern District Effort Allocation Categories Week of Sept. 25, 2008 with Homicide and Shooting Count (YE 6/30/09)



3 Variable Scoring Rule, Eastern District Effort Allocation Categories Week of March 26, 2009 with Homicide and Shooting Count (YE 6/30/09)



Predicting Homicides and Shootings
and Allocating Police Resources to Prevent Them

Using the model to allocate police resources:

- Each week the scoring rule should be applied to updated data.
  - Category A, consisting of the highest scoring 10% of the district's hexagons should receive 25% of available resources;
  - Category B, consisting of the next highest scoring 15% of the hexagons also should received 25% of available resources;
  - Category C, the next highest scoring 25% of the hexagons also should received 25% of available resources; and,
  - Categories D and E, collectively consisting of the *lowest* scoring ***50%*** of the hexagons also should received 25% of available resources, with Category D hexagons typically assigned somewhat more than half of the allocation.
- This will result in a 2x–3x increase in police resources per homicide versus an even distribution of resources across all the hexagons.