Crime in Chicago

2001 - 2017

Ryan Rouleau
Kylee Budai

Questions

1. How much future crime will there be in Chicago ?
2. Where is the most crime in Chicago ?
3. How do the severities of crime in Chicago change over time ?
4. How do severities of crimes spatially correlate ?
5. Where will severe crimes occur in the future in Chicago ?

Project Goal

- Model crime counts and severities for the entire city
- Expand these models to 8 different locations
- Expand those models to even more points
- Generate predictive heat maps

Dataset and Preprocessing

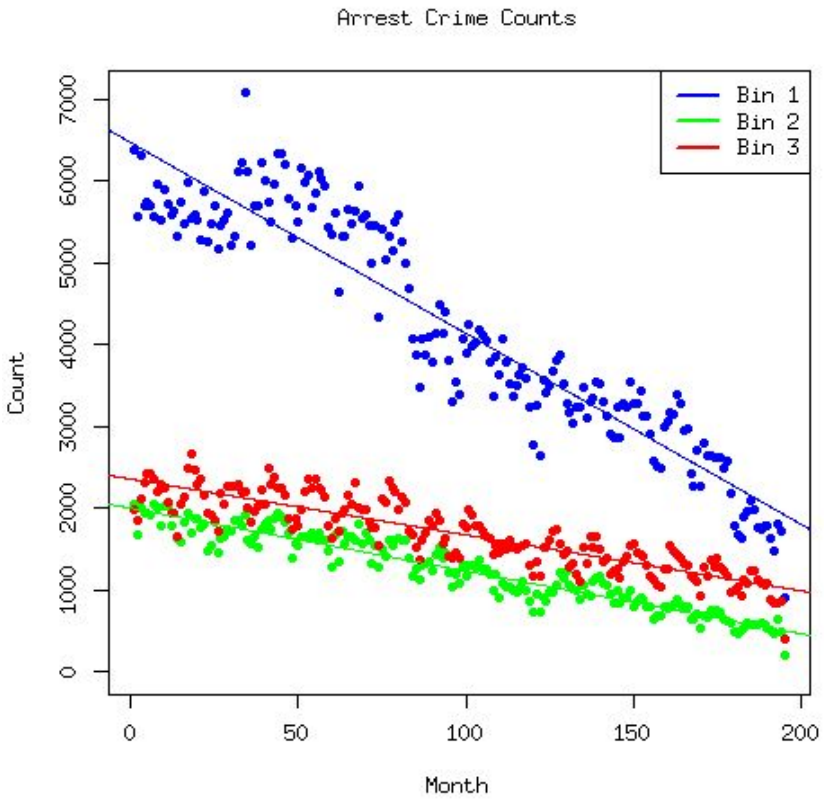
- Remove "OTHER OFFENSE" crime
- Separate data by no arrest/arrest attribute
- Merge data with severity mapping and binning
- Map (Jan 2001, … , Dec 2016) to (1, … ,192)

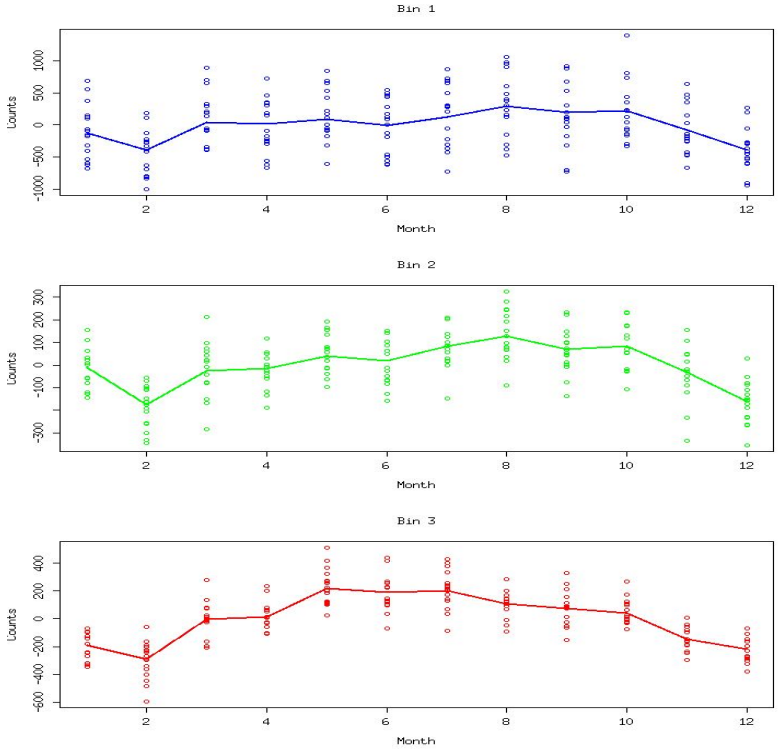
Month	Year	IUCR	Lat	Long	Arrest	Severity	Bin
08	2001	1330	41.896	-87.630	1	1200	3
05	2008	1320	41.699	-87.618	0	700	2
01	2016	0486	41.763	-87.615	0	200	1

Tools We Used

- R
- Python
- D3.js
- git
- bash

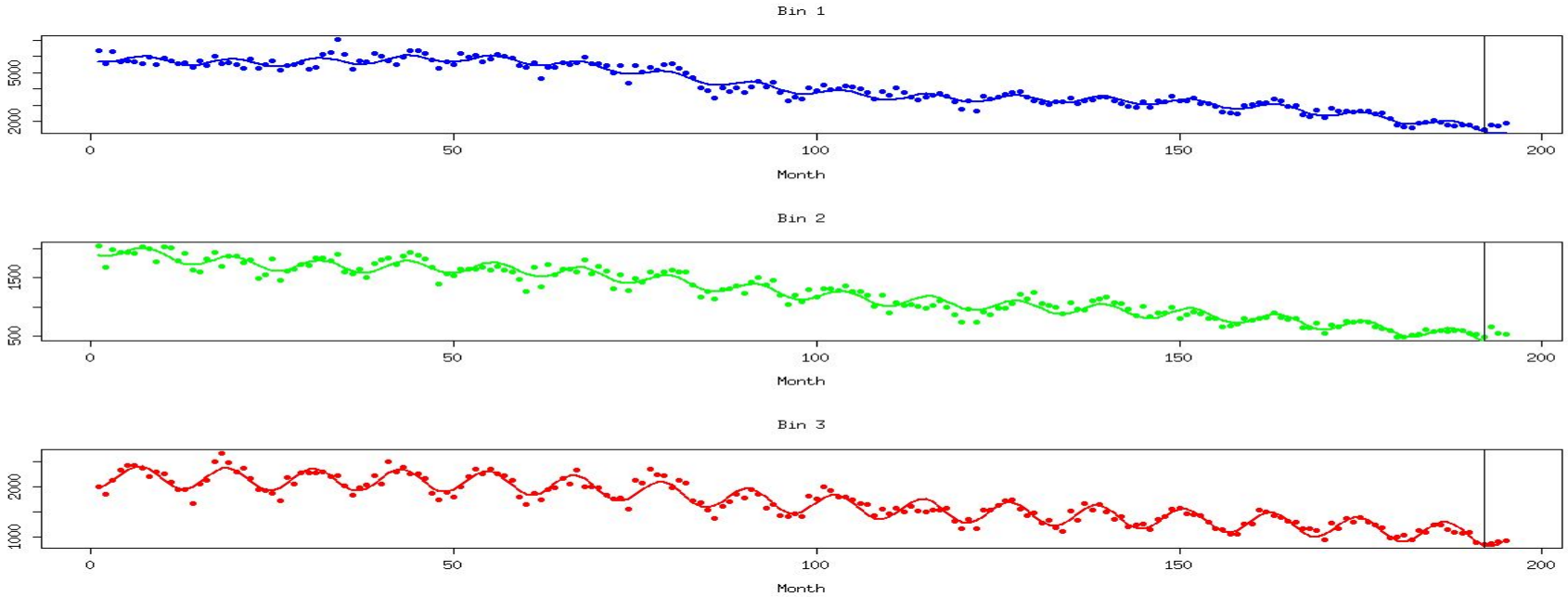
Linear Regression Fits



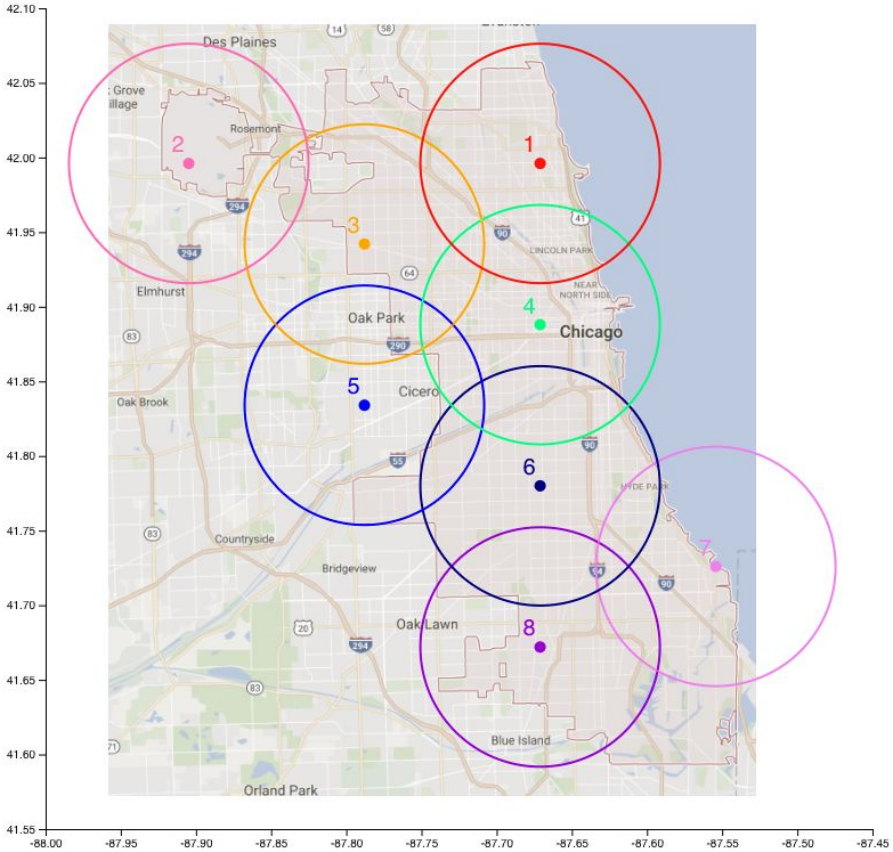


A Better Model for the Data

$$Y = \beta_0 + \beta_1 \cos\left(\frac{2\pi X}{12}\right) + \beta_2 \sin\left(\frac{2\pi X}{12}\right) + f(X)$$



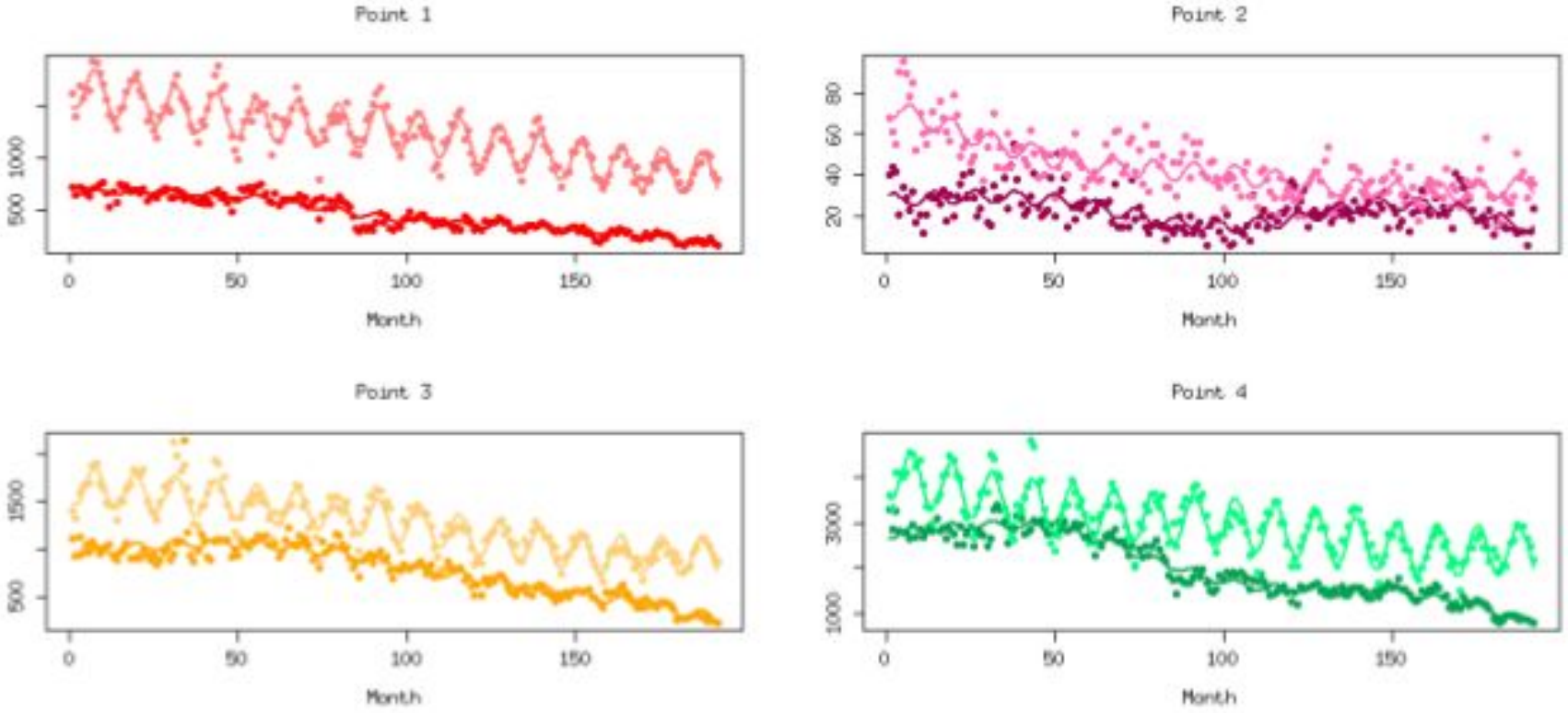
Extending the Model to 8 Points Across Chicago



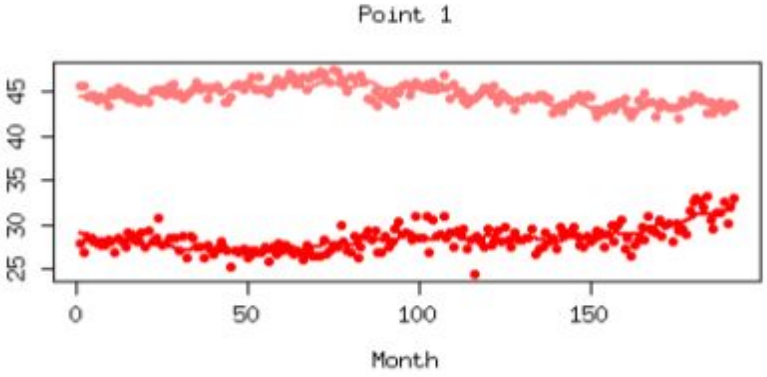
Generate matrices that track number of crimes and average severity of crimes within neighborhoods around points.

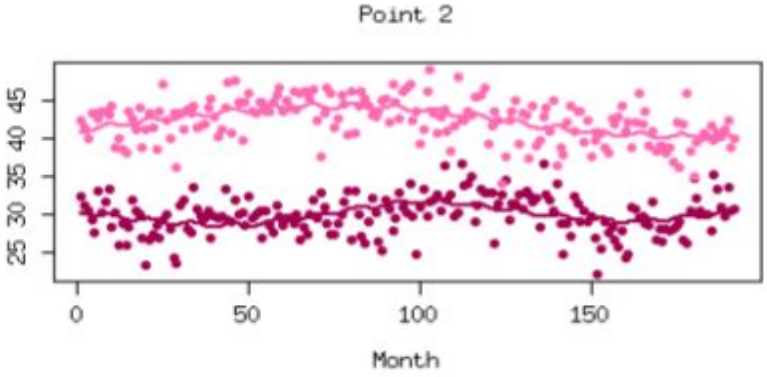
Goal: extend this idea to more points to generate severity heat map.

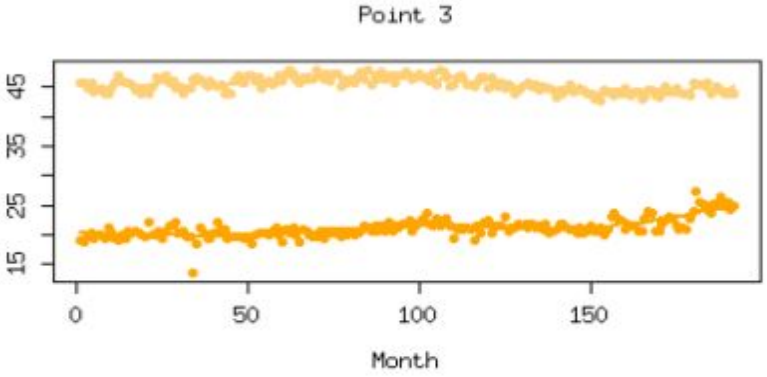
Use same model as before to fit counts at each point.

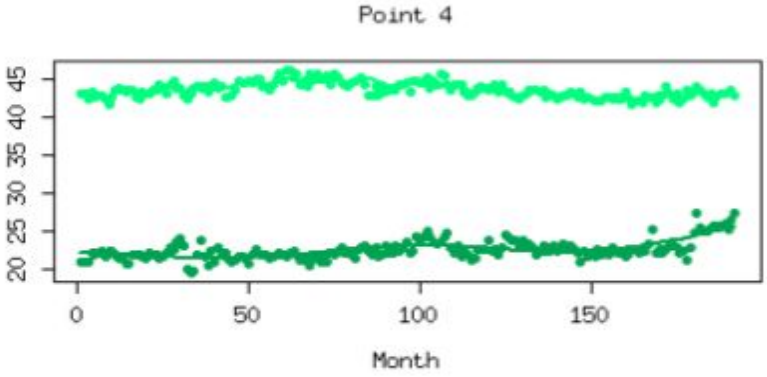


Fit average severities with same model

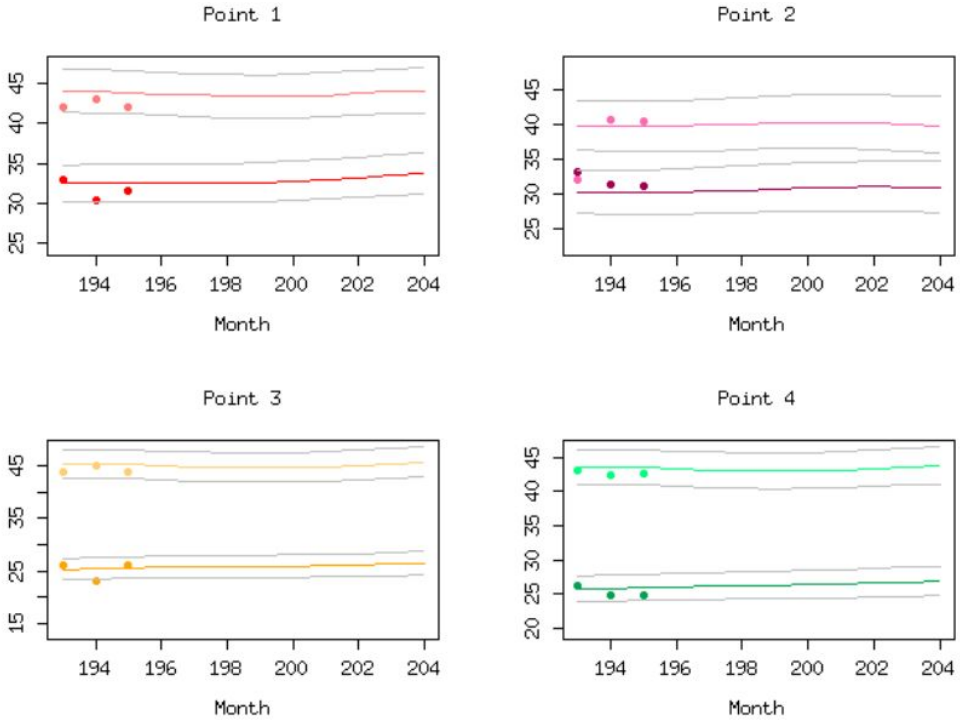








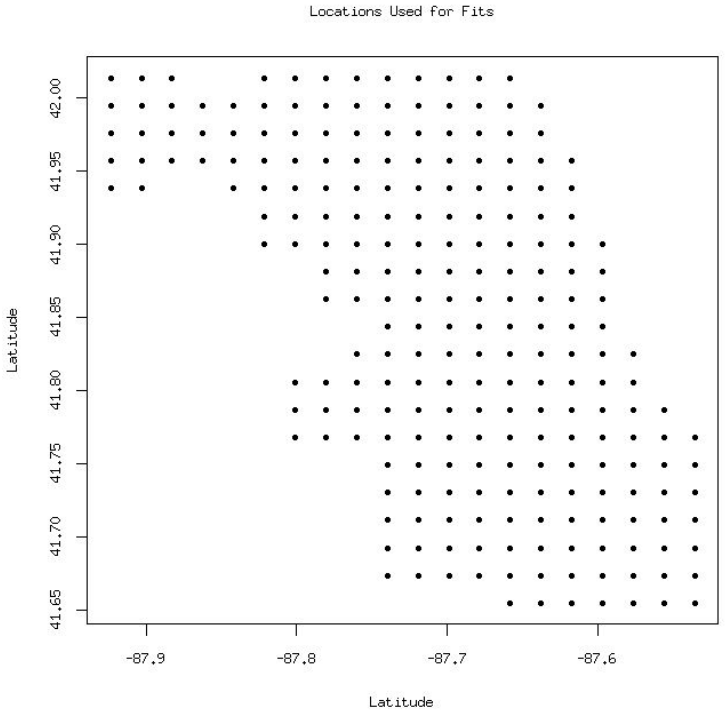
Extrapolate fits to predict average severity in January, February, and March 2017



Percentage of actual values that fell within confidence intervals:

	Arrest	No Arrest
Bin 1	87.5%	95.8%
Bin 2	58.3%	95.8%
Bin 3	66.6%	87.5%

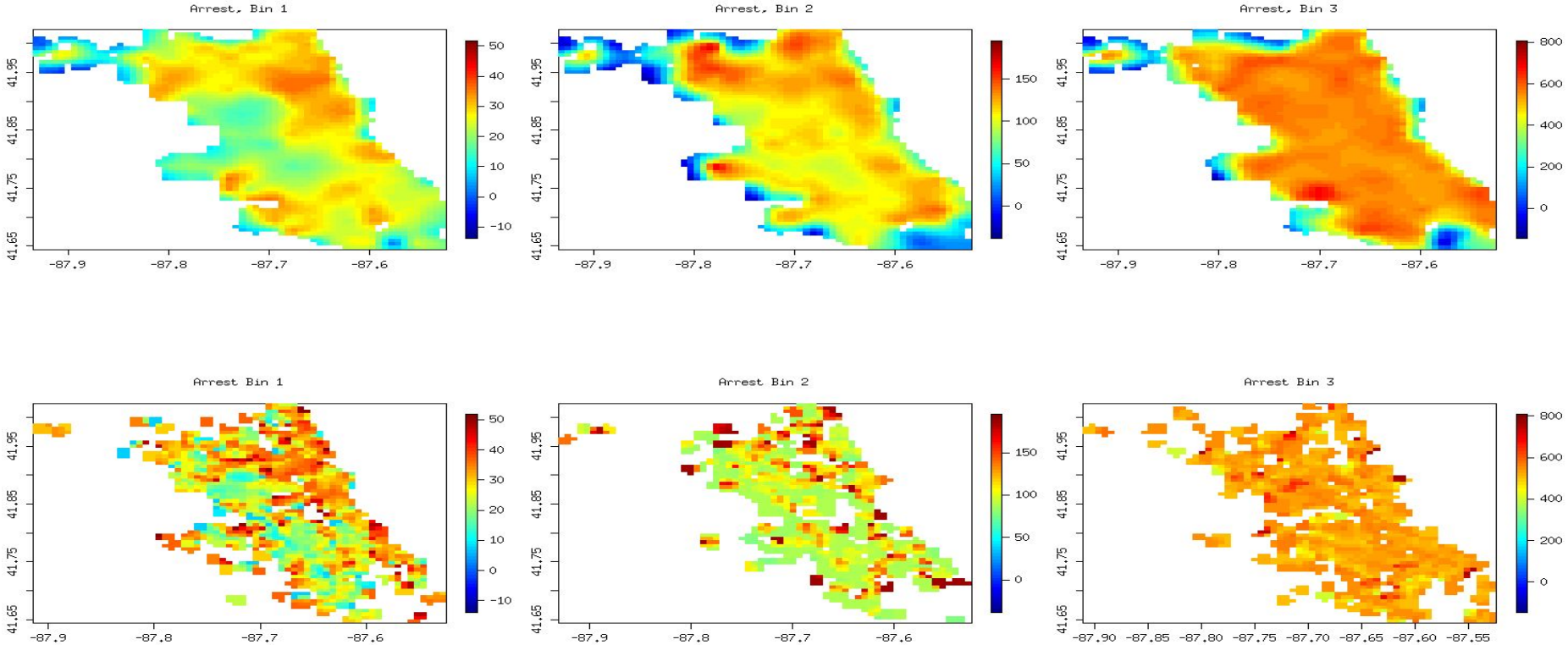
Extending to ~200 points, using same model on average severities, and extrapolating



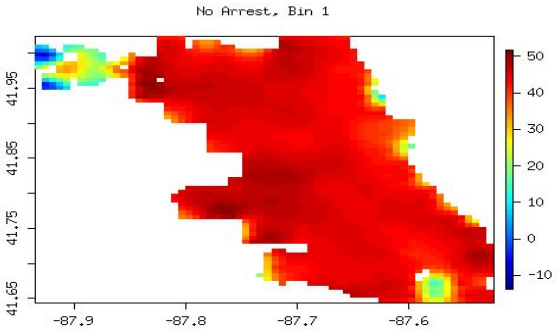
Percentage of actual points that fell within 95% confidence intervals after extrapolating.

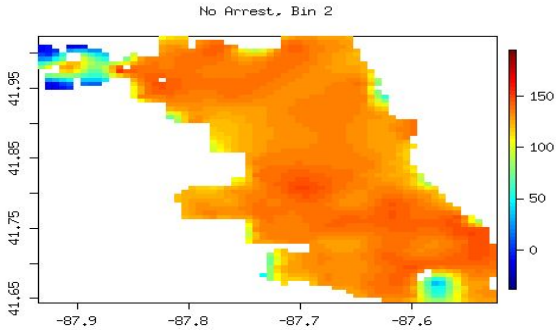
	Arrest	No Arrest
Bin 1	70.51%	62.45%
Bin 2	68.80%	63.03%
Bin 3	71.24%	60.59%

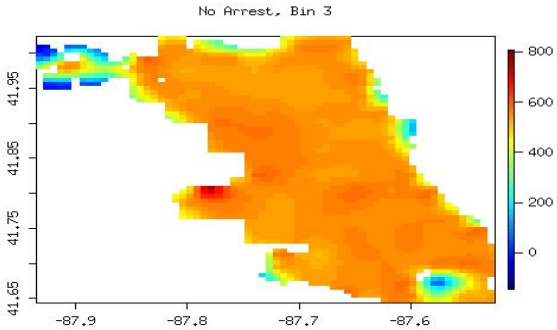
Generated predictive severity heat maps for arrest data using ordinary kriging on predicted values

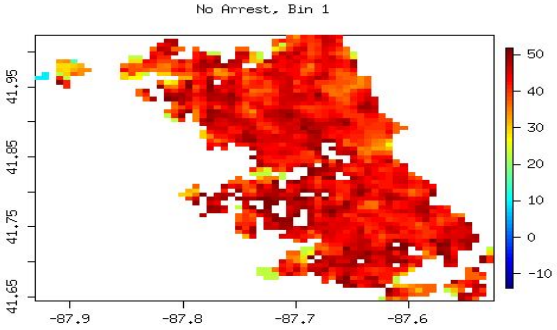


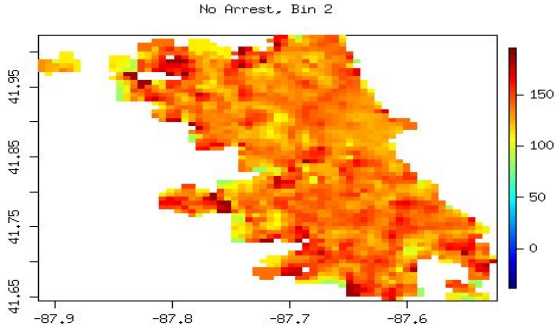
Predictive severity map for no arrest data

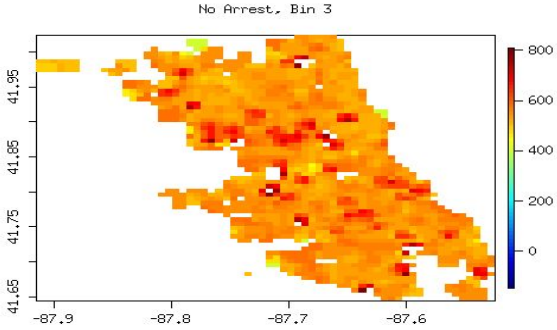








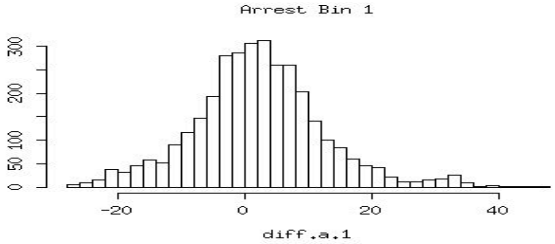


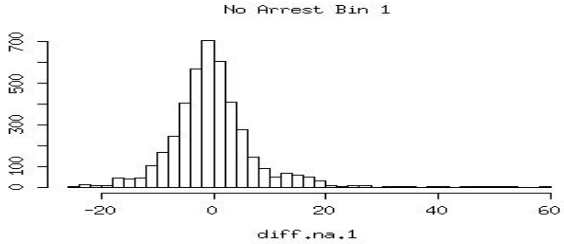


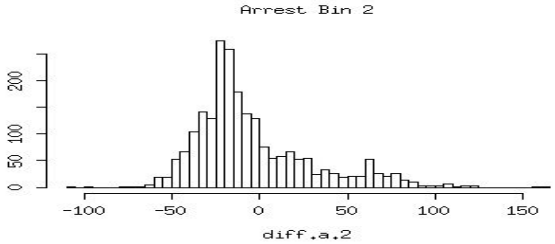
Accuracy of severity heat maps

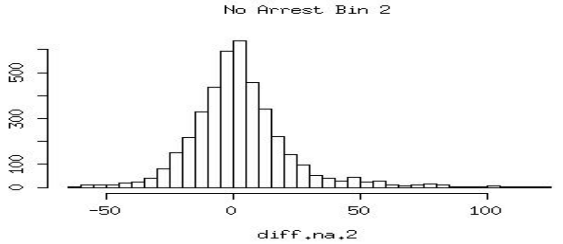
Were not able to generate confidence intervals on plots to determine accuracy.

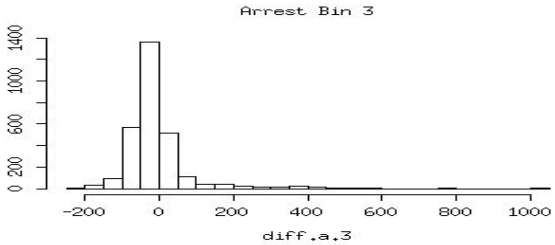
Severity maps tended to overestimate actual severity values

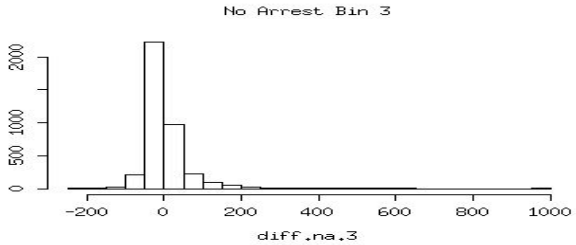












Law enforcement will be able to more efficiently allocate resources