



# MA CSS Research Topic

## Benjamin Rothschild



# Research Question

How accurate can I create a model to predict locations of crimes in Chicago?



# Background

Crime research spans academia, government, and industry.

In Los Angeles a model was *1.4 - 2.2* times better at predicting crime compared to a dedicated crime analyst and with this improved performance *7.4%* of crime was reduced in a controlled experiment that used a model to predict hotspots.

In this study the prediction algorithm used historical crime data to predict future crime hotspots (did not rely on census or demographic data)

Randomized Controlled Field Trials for Predictive Policing

G. O. Mohler, M.B. Short, Sean Malinowski, Mark Johnson, et al.



# Background

Recent social science research has taken advantage of “Big Data” sources to achieve more accurate predictions and measurements of real-life phenomena.

For example Google Street View images were successfully used to measure income in New York City and Boston.

Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life

Edward L. Glaeser, Scott Duke Kominers, Michael Luca, Mikhil Naik



# Datasets

- [Plenary.io](https://plenary.io)
  - 311 Service Requests
  - Business Licenses
  - Environmental Inspections
  - 311 Vacant Buildings
  - Crimes
  - Food Inspections
  - Red Light Tickets
  - Divvy Trips
  - Building Violations
- [Zillow/Craigslist](https://www.zillow.com/craigslist/)
  - Rental Price Data
- [Twitter](https://twitter.com/)
  - Geo-located tweets



# Methods

- Spatial Regression
- PySAL Python Library



# Evaluation Methods

- I expect the output of my model to give a score for specific areas (blocks/radius/census tracts, etc) and I will try to correlate these scores with the data from the Chicago Crime Dataset.
- Evaluation Criteria:
  - Error Rate
  - Proportional Reduction in Error (PRE)
  - Area Under the Curve (AUC)





# Consideration and Goals

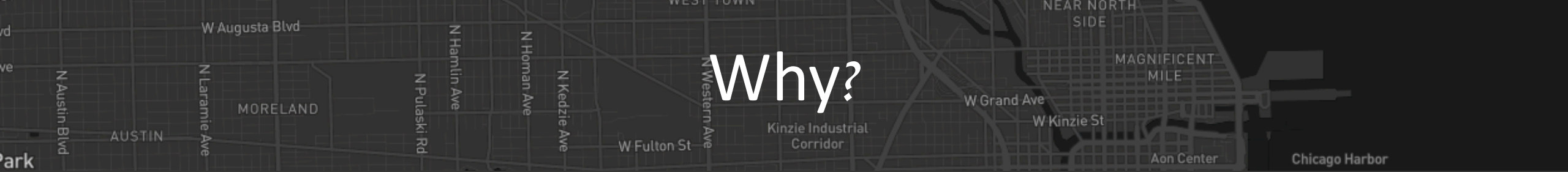
- I want to try to make my geographic units of analysis as small as possible
  - How does my error rate changes with different boundary sizes
- Similarly I want to make my unit of time analysis as small as possible
  - How often can I make a prediction? How does this affect the error rate?
- What datasets are most informative in helping my prediction?
  - How does adding datasets affect the Error rate?





# Possible Conclusions

- I can predict there will be a crime on a .5 mile radius and 5 hour timespan with at 15% success rate
- I can predict there will be a crime on a 2 block radius and 10 hour timespan at 30% success rate
- I can predict nothing 😞



# Why?

- A model like this could help with resource allocation with police department or with community investment initiatives
- Analysis like this can lead to further questions about what environments lead to crime
- By seeing how the model improves or not when data sources are added I can hypothesize about what other datasets would be helpful in improving my model



# References and Resources?

- Datasets

- <https://censusreporter.org/profiles/86000US60657-60657/>
- <https://data.cityofchicago.org/Service-Requests/311-Service-Requests-Vacant-and-Abandoned-Building/7nii-7srd/data>
- <https://data.cityofchicago.org/view/5cd6-ry5g>
- <https://www.zillow.com/research/zillow-rent-index-methodology-2393/>
- <https://dev.twitter.com/rest/public/search-by-place>

- Articles

- Glaeser, Edward L., Scott Duke Kominers, Michael Luca, and Nikhil Naik. "[Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life.](#)" (pdf) *Economic Inquiry* (forthcoming). [View Details](#)

- Websites

- <https://www.civicscape.com>

A dark-themed map of Chicago, showing the city's grid of streets and major landmarks. The map is rendered in shades of gray, with the city's name 'Chicago' prominently displayed in the center-right. The word 'Questions?' is overlaid on the left side of the map in a large, white, sans-serif font. The map includes labels for various neighborhoods such as West Town, Near North Side, Loop, South Loop, Near West Side, Greektown, North Lawndale, South Lawndale, Cicero, Crawford, Bohemian California, Lower West Side, South Branch Addition, Armour Square, South Commons, Douglas, Burnham Park, Oakland, Fuller Park, Grand, Canaryville, South Brighton, Brighton Park, Corwith, Sleepy, and Stickney. It also shows major roads like W Augusta Blvd, W Madison St, W Jackson Blvd, W Roosevelt Rd, W 16th St, W Cermak Rd, W 26th St, W 31st St, W 35th St, W Pershing Rd, W 43rd St, W Washington Blvd, W Washington St, W Van Buren St, W Taylor St, W Grand Ave, W Kinzie St, and W Fulton St. Other landmarks include Garfield Park, East Garfield Park, Douglas Park, Soldier Field, University of Illinois at Chicago, Aon Center, and the Stickney Water Reclamation Plant. The word 'Questions?' is positioned over the Cicero and North Lawndale areas.

# Questions?