Analysis of Chicago Crime Trends
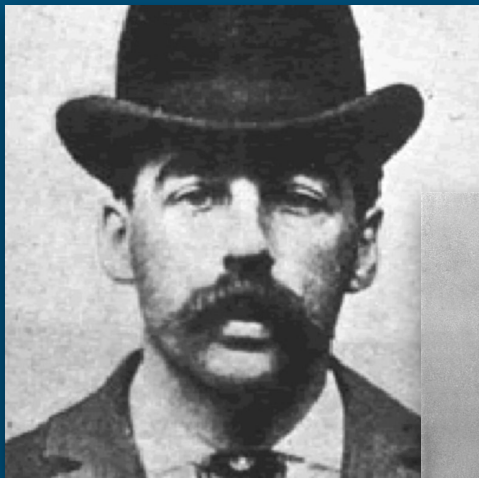
Amanda Killeen, Isaac Hammes, Graham Dominick, Daniel Kingsley

Introduction

Dr. H.H. Holmes
(biography.com)

Al Capone
(FBI.gov)

John Dilliinger
(Wikipedia.com)

Rod Blagojevich
(Wikipedia.com)

Questions

- Is there a correlation between type of crime and location description or district?
- Do arrests occur more often in different locations for the same crime?
- What are the most common crimes for different areas?
- Are there periods when there were waves of certain types of crimes? Trends in crime types?
- Can an accurate model be generated for "arrest made" based on other information about a crime?

Dataset

- Used the City of Chicago's *Crimes 2001-Present Dashboard*
- Contains around 7,000,000 entries of reported crimes.
- Key attributes:
    - Type/Description
    - Area (Address/Neighborhood)
    - Location (e.g. Apartment, Bar, etc.)
    - Date
    - Arrest

Data Preparation work

- **Data cleaning:**
  - Remove unneeded columns
  - Clean string data (ensure uniformity between descriptions, locations, etc.)
- **Preprocessing:**
  - Bin by location
  - Filter by crime type / arrest

Tools

- Python
- Jupyter Notebook
- Sklearn (KNN/Logistic Regression/Naive Bayes/Decision Tree)
- Python mlxtend (Association Rule Mining)
- Numpy
- Pandas
- Tableau

Methods Applied

- Association Rule Mining with the Apriori Algorithm
- Naive Bayes
- Logistic Regression
- Decision Tree
- K-Nearest Neighbors (Failed)
- Partial Data Cube
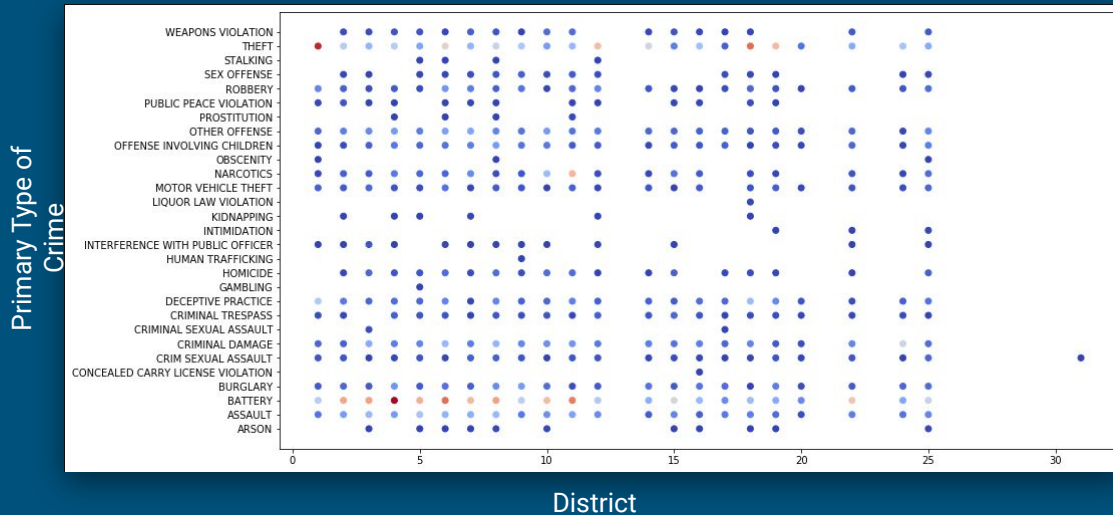- Entropy Measurement (Failed)

Knowledge Gained

- Classification model predicted arrests with an ~85% accuracy.
- Property Crime had the lowest rate of arrests being made, with the highest likelihood of not being arrested being the item set {Residential, Not Domestic, Property Crime}
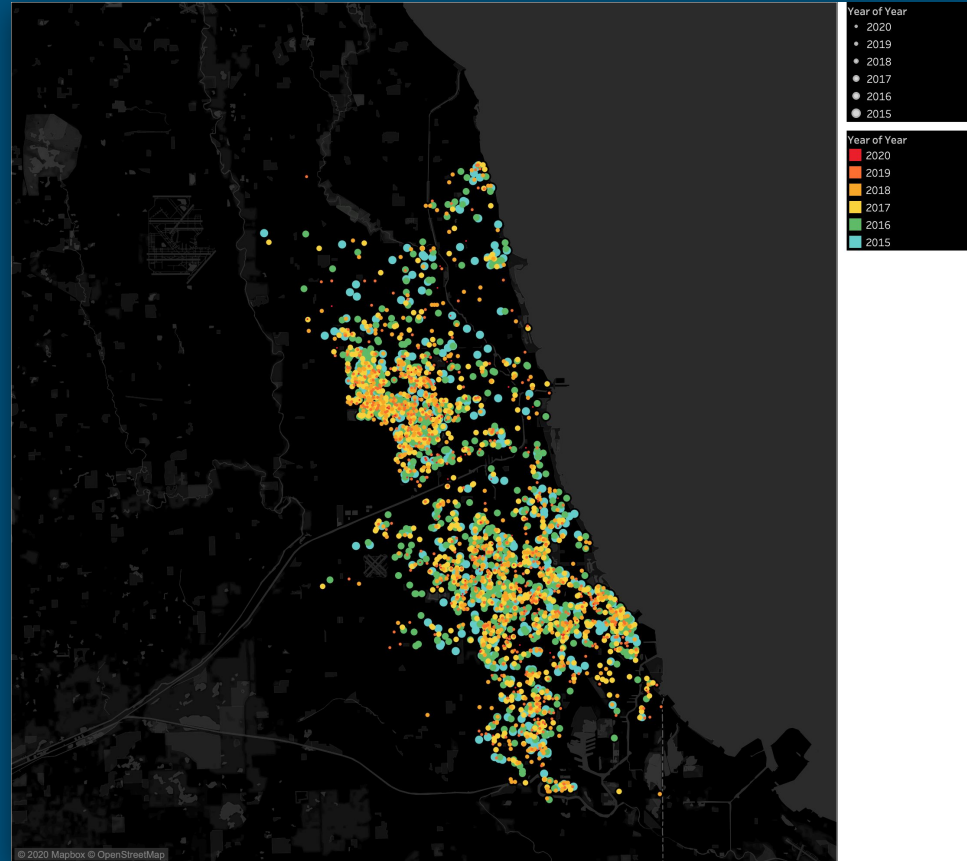
Knowledge Gained

- Increased frequency of certain types of crimes by district
- Based on a sample, concentration of "Battery" accounts in districts 2 through 11. We also see a concentration of "Theft" accounts in districts 1, 18 and 19.

Knowledge Gained

- Some areas of Chicago have a history of homicide, with homicide repeatedly occurring in the same approximate location year over year.

Knowledge (not) Gained

- Question: Do clustered patterns of killings indicate gang activity?
- Use entropy (from information theory) to see if there is a pattern in killing
- Didn't work, the difference in total killings is too large to compare patterns between districts

Applications

- Law enforcement can use results based on location to increase presence and/or implement neighborhood watch, in particular residential areas due to high rate of no arrest.
- Potential for researchers to investigate districts with higher rates of battery and theft to see if there is a correlation between these crimes and other factors (socioeconomic, education, gang presence)
- Use arrest prediction to see if there is bias in crime investigation. i.e. do some districts have higher arrest rates than others because of less police attention?