Crime in LA

Madison Razook & David Stone

Project Description

We will be mining a dataset containing information on all crimes reported in LA from 2010 to now. This dataset documents date, time, location, type of crime, victim information, and arrest information. Possible questions:

- How do time and location influence crime?
- Does a delay between the crime and the time it's reported affect the likelihood of arrest?
- Relationship between victim age, sex, etc and type of crime?
- Any interesting patterns with crime and arrest rates over the past 8 years?

Prior Work

- LAPD uses this data to calculate crime statistics for year, area, etc.
    - Can help us decide where to look, i.e. if a certain area has a higher than usual amount of violent crime, we can look into other attributes for those cases to find any predictive factors
    - http://www.lapdonline.org/research

Dataset

- https://catalog.data.gov/dataset/crime-data-from-2010-to-present
- Provided by the City of Los Angeles
- Downloaded on both team members' computers

Proposed Work

- Data cleaning and preprocessing:
  - Some fields are missing
- Analysis:
  - Visualize data
  - Association rule mining
- Data integration
  - May need to integrate data on socioeconomic trends and other topics

List of Tools

- Jupyter Notebook
- Pandas
- Matplotlib

Evaluation

- As we receive results we will check if the results fully answered our question and if the results are trivial
- If our main question is still unanswered, we will reevaluate the question and determine if the results can be used to answer a similar question