Data Science Capstone

Karl Janich

Introduction

- Crime continues to be a problem in major cities
- Los Angeles is a popular destination for new families with plentiful well-paying jobs
- Families desire to avoid crime-ridden areas
- A tool to avoid a neighborhood at high risk for a major crime is useful to families

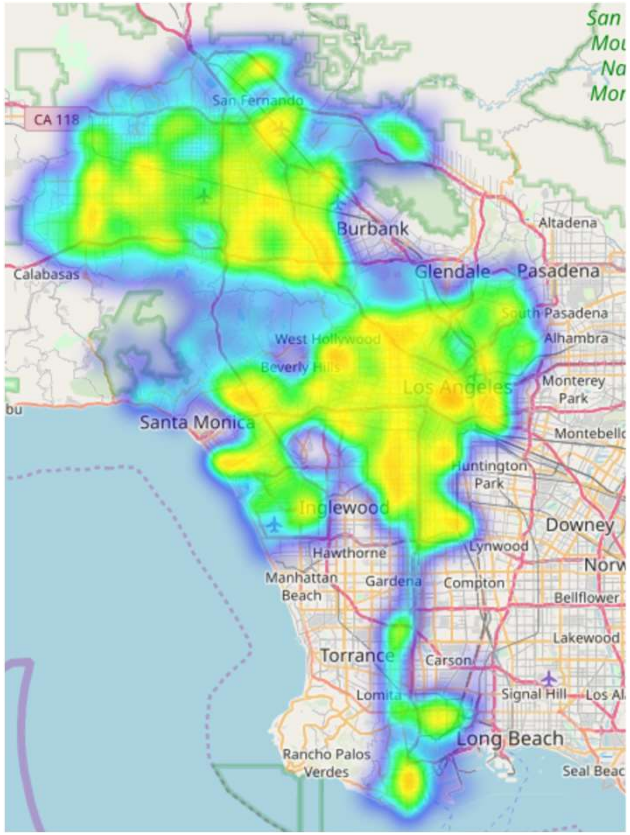
Data Used

- LAPD Crime Statistics Database
  - https://data.lacity.org/A-Safe-City/Crime-Data-from-2010-to-2019/63jg-8b9z
  - Roughly 2 million crimes over 10 years
  - Culled to the past 2 years with roughly 200,000 crimes
    - More meaningful given recent drop in crime and potential for shifting spatial patterns
- LAPD Reporting District GeoJSON
  - https://opendata.arcgis.com/datasets/4398360b1a0242b78904f46b3786ae73_0.geojson

Methodology

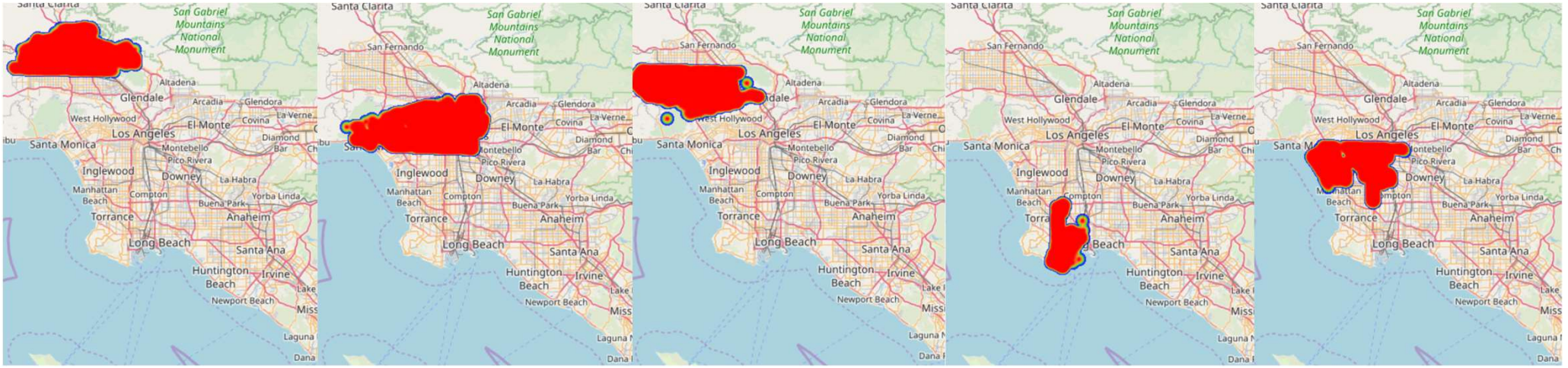
- Data exploration with heatmap
- Further exploration with clustering
- Defining boundaries: Binning with chloropleth map

Exploratory Heatmap



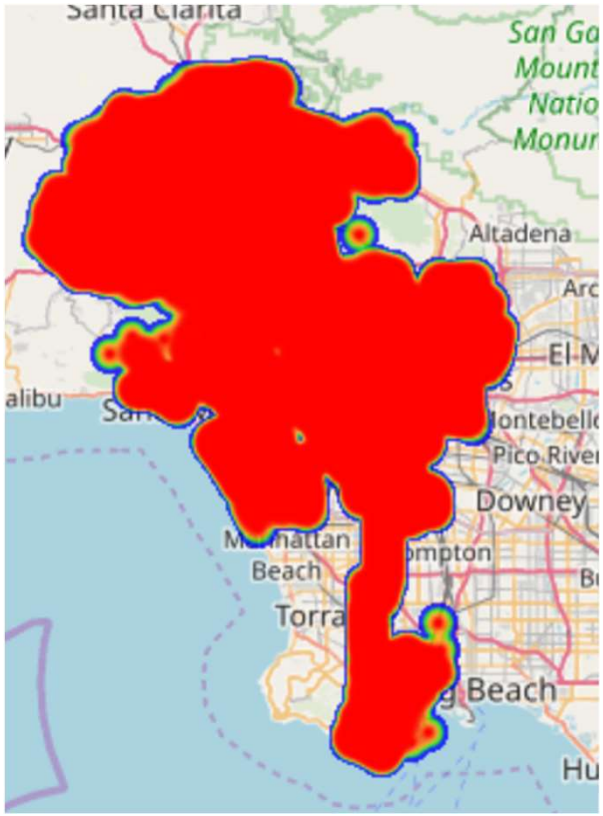
- Can see a few clusters that are merged
- Some heated areas centrally especially
- Poorly defined borders

K-Means Clustering



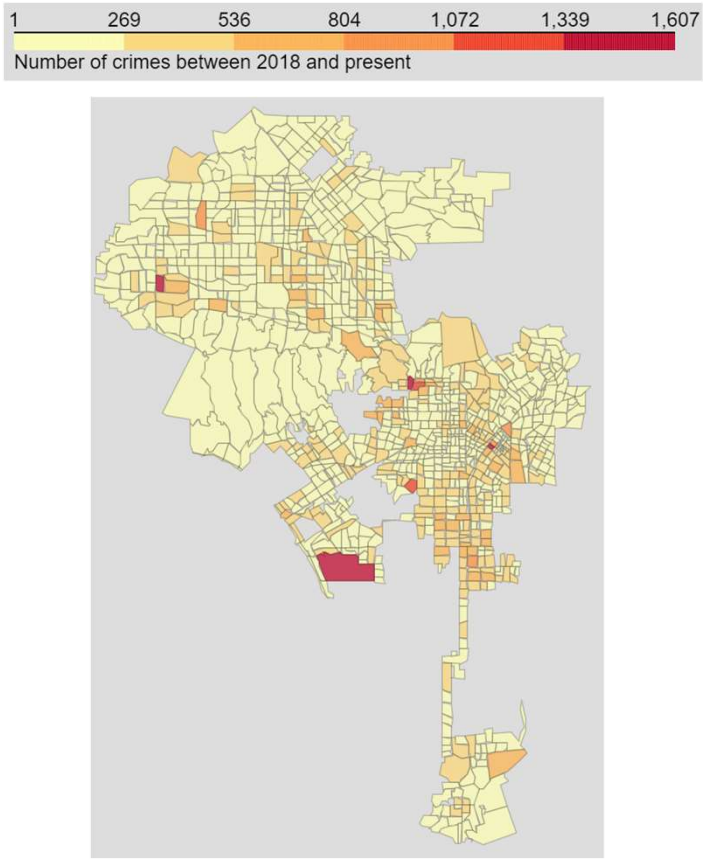
- Can better-delineate the clusters
- Does not add much to better define areas at this point
- No significant changes with new random start points

DBSCAN



- DBSCAN will create one cluster unless radii are set unreasonably low.
- DBSCAN also is suboptimal given the memory and runtime requirements for the size of our dataset.

Chloropleth Map



- Can see well-defined districts demonstrating locations with very high concentrations of crimes.
- The light yellow regions provide optimal regions with lower probabilities of being affected by crime.

Discussion and Conclusions

- Mapping of crime data is viable with pubically available data
- We can use maps to determine where foci of crime exist
- Ethical issues are brought up by this information
  - Do maps like these facilitate “redlining”?
  - How can we help these problem areas?