



# Density Based Crime Mapping

## Misclustered Crime and Unwarranted Punishment

**Katy Weathington**

Department of Mathematics, Statistics & Computer Science,  
Marquette University, Milwaukee, Wisconsin

**Dr. Shion Guha**

Department of Mathematics, Statistics & Computer Science,  
Marquette University, Milwaukee, Wisconsin

### Background

- Crime analysts commonly use 3<sup>rd</sup> party black box software
  - Only the company actually knows specifically how it work
  - Eliminates transparency and accountability
- DBScan is a common density based algorithm that clusters by changes in density
- Merely clusters past data points, not predicting future events
- Therefore, use in future cases must be interpreted by a human

### DBScan Algorithm

Parameters: Density is defined as the number of points, **minPts**, within  $\epsilon$  distance

- For each point, calculate number of points within  $\epsilon$
- If at least minPts are within  $\epsilon$  of a point, label it a core point. Otherwise, label it noise
- Connect all core points along their  $\epsilon$  neighborhood
- Reassign any noise points within  $\epsilon$  of a core point as a border point of that cluster

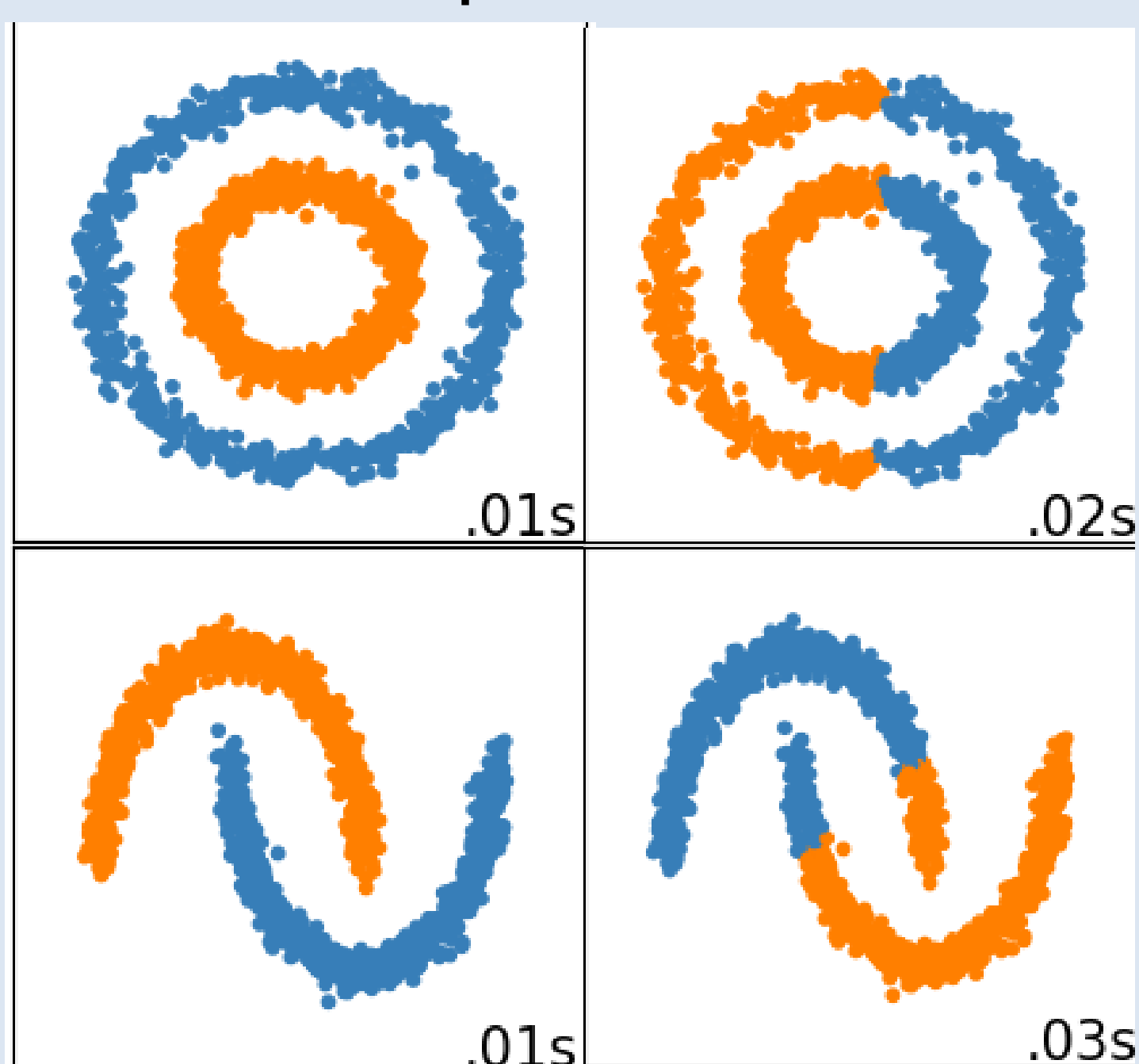


Fig. 1 Comparison of density based(left) and centroid based(right) clustering on toy data. Taken from Scikit Learn documentation

### Data

- Gathered raw court data from City of Milwaukee Municipal Court and Circuit Court.
- Sanitized data into API readable format
- Ran data through Google Maps API to achieve latitude and longitude geocodes
- Multiple charges often originated from a single event
  - For clustering accuracy, we only plotted an event once

### Crime Clustered by DBScan

These maps were generated using the Sci-Kit Learn library for Python

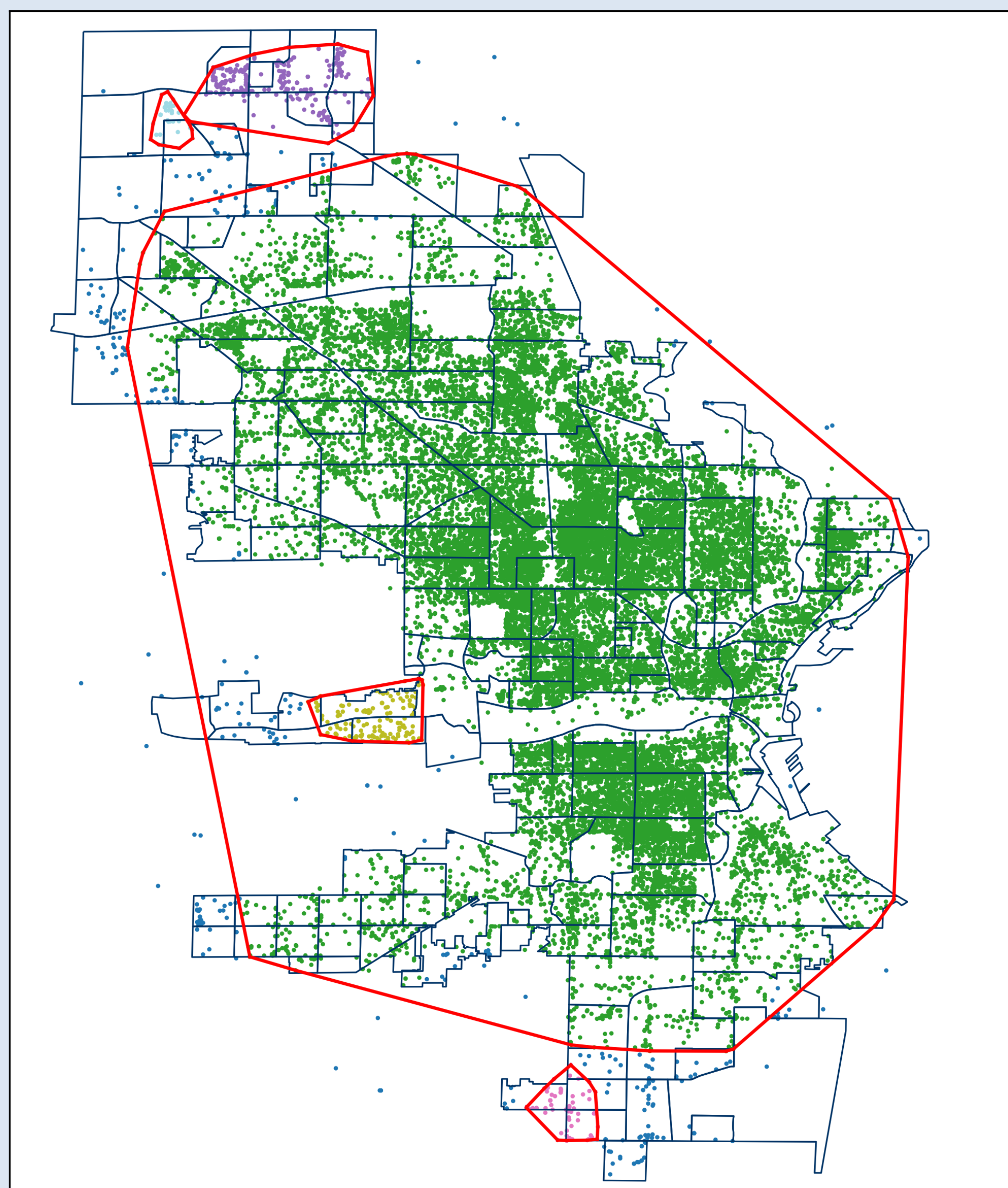


Fig. 2 Example of DBScan results. This particular example uses only guilty cases, and has  $\epsilon=.0075$  and  $\text{minPts}=45$ . Unique clusters are outlined in red.

Some clusters can exist within other clusters. This implies an enclave which may require extra attention and consideration

### Parameter Estimation

- By finding the “elbow” of a K-Nearest Neighbors graph you can more accurately choose an  $\epsilon$  for a given minPts
- You can run this for the full set of points, to get an  $\epsilon$ 
  - minPts can then be determined using the above result and an inverted K-Nearest Neighbors

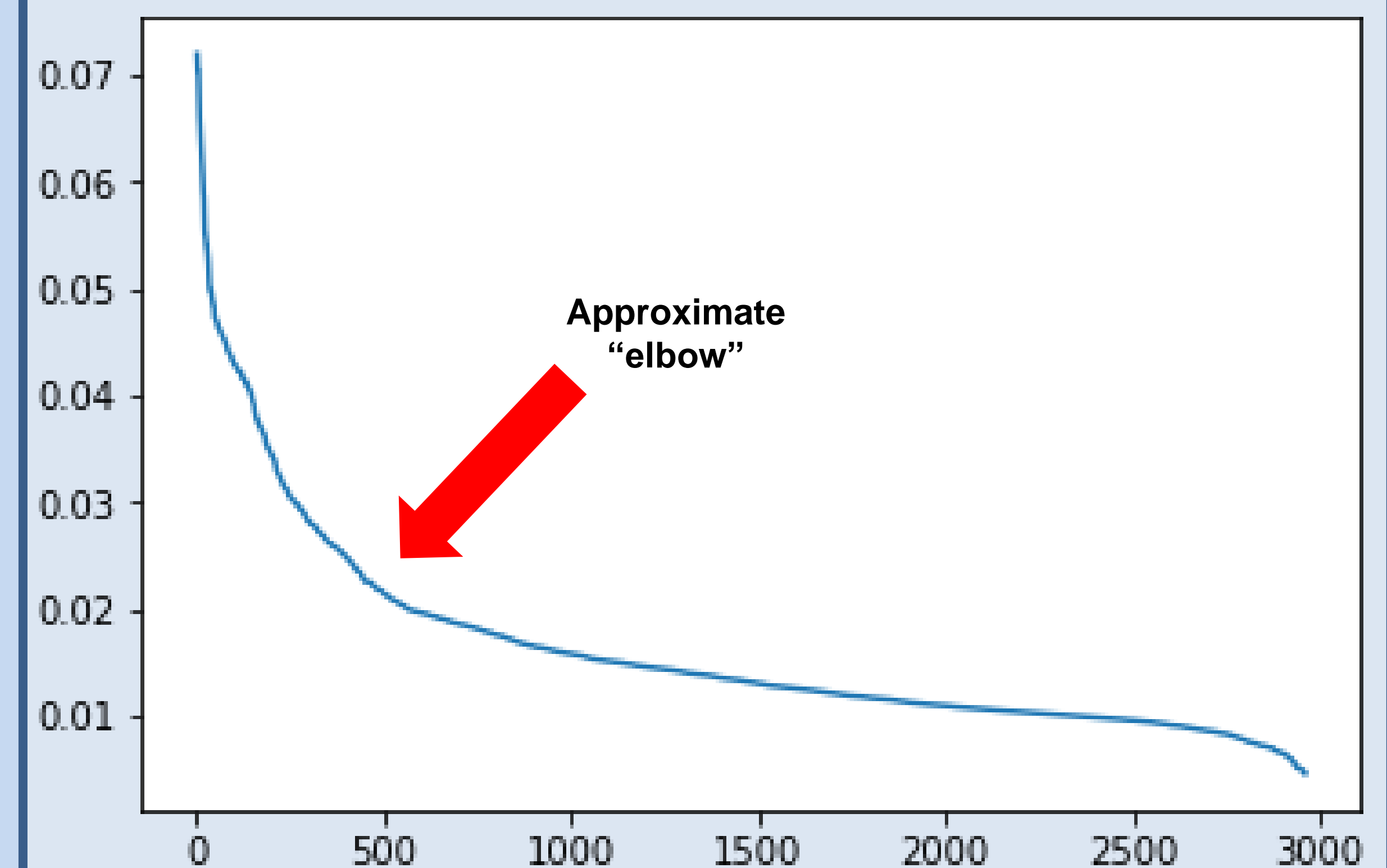


Fig. 3 Example of a K-Nearest Neighbors plot

### Takeaways

- Small variance in parameters have a large impact on outcome
- DBScan does not create many meaningful crime clusters on a macro scale
  - Typically outputs one large cluster that covers the entire city
  - Small clusters can help highlight or identify a location associated with increased crime
  - Smaller clusters can be difficult for user to distinguish and therefore utilize