## Intro to Spatial Data Science | Research Question

Julian McClellan October 17, 2016

The University of Chicago Police (UCPD) maintains a searchable database of incident reports spanning July 1st, 2010 to the approximate present, subject to the update rules given in Figure 1. The native format of this database is shown below. For analysis, the database will be scraped and put into CSV format, in addition to a number of other cleaning operations.

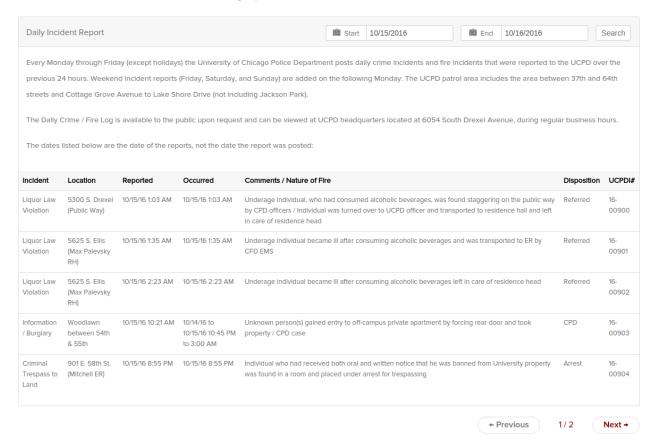


Figure 1:

Looking at the "Location" of these incidents, there are sometimes precise addresses, and other times approximate cross streets given. Thankfully, the Google Maps API can convert all of these into GPS coordinates. The precision of those GPS coordinates can vary somewhat. Looking at the 4th case (from the top) in Figure 1, the Location is given as "Woodlawn between 54th and 55th", hardly a precise address like the rest of the entries, and this is quite common in the dataset. This case in particular poses a unique problem as there are two 54th streets: 54th Street, and 54th place. In this case, the Google Maps API does function, but not with as much precision as precise addresses. Steps may need to be taken to ensure better coordinate accuracy for these types of "Locations". It may be possible to simply hand code the GPS coordinates for any "Location" containing the word "between".

I will be looking to identify clusters and/or spatial heterogeneity in regards to the "Disposition" and/or "Incident" factors of the reports. For example, in Figure 1, one can see that the "Disposition" of the 4th case is "CPD" (Chicago Police Department), meaning that the CPD, rather than the UCPD is handling the case.

Notice also that the "Incident" for this case is "Information / Burglary". With the ~8000 reports contained in the database, one might be able to determine if there is a spatial pattern affecting whether or not a case is handed over to the CPD. Depending on the data itself (which has been scraped but not fully cleaned yet), one might be able to ask this question for a variety of "Incident" attributes and/or other "Dispositions". Figure 1 shows "Referred" and "Arrest" as other possible "Dispositions", but a preliminary pass of the scraped data has shown that there is also "Open" and "Closed" (Cases), and there might be interesting spatial patterns that correspond with these dispositions.

Additionally, the Winter of 2015-2016 occurred during an "El Niño" year, which, in Chicago and the Midwest United States, is characterized by milder winters. It could prove an interesting exercise to analyze spatial patterns of "El Niño" winters against "La Niña" (~normal) winters.