Capstone project 1 proposal

1. What is the problem you want to solve?

I will use crime data of Chicago and the question I am trying to answer is: where do crimes happen and what affect them? This is an important question to many people in Chicago.

2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?

Regarding crime analysis, the clients should include law enforcement, people living in Chicago and maybe even the government. For law enforcement, the pattern of different crimes in the area can better help them assign their resources to the most important spots instead of randomly patrolling the entire area. For average people, maybe they can try to avoid areas (or the time) where certain types of crimes are high. For the government, the reason some places have high crime rates may be related to racial, economical, and even educational reasons. Trying to find ways to mitigate the situation could be an important task for the government.

3. What data are you going to use for this? How will you acquire this data?

First the most important data file, the crime data can be acquired from Chicago data portal. I will also be using weather data (from NOAA website), US Census data regarding social, education, household income and poverty status (from US census website). From the same Census website, the tracts data can also be downloaded to use as location to visualize all the data on the map. Chicago transportation data will be used as well to analyze the relationship between public transportation lines and crime locations.

4. In brief, outline your approach to solving this problem (knowing that this might change later).

To solve this problem, I will first clean the crime data and other data tables and use the date time as an index to explore time series. Other data tables can be linked to the crime data by the date time index. Visualization on the map can be accomplished using location data in the crime data set as well as in other data sets. Especially for the Census data, time and location can be used to link to the crime data set.

5. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

My deliverables will include all my code and analyses. Most of my conclusions will be supported by visualization on a map. Different aspects (income, demographics, unemployment, racial) can overlay on top of different crime data on the same map to make better effects. A brief paper will also be written up to accompany the analysis and as a final report.