A Tale of Two Cities

Project by: Pratik Kamath

Objective: "Crime is not spread evenly across maps. It clumps in some areas and is absent in others. People use this knowledge in their daily activities. They avoid some places and seek out others."[1] as quoted by US Department of Justice is in fact quite true and a very important thing to be looked upon. How often have we heard someone pointing out that a particular area is not safe at night because of homicides or other crimes being pertinent and we try to try to avoid it. I intend to do a very detailed analysis on two cities Chicago and St. Louis and want to draw invaluable conclusions which will help find intensity of crimes distributed across these two cities. To keep things interesting, I would like to analyze the two cities I have mentioned above differently.

Data-sets: For Chicago, I have found out a blow-by-blow data-set which happens to include invaluable information such as nature of crime, district, ward, longitude, latitude and therefore I would love to use it for crime heat map analysis and I would love to make important deductions as to where there is occurrence of a particular crime eg. Murder, battery etc. by allotting them different colors on heat map to classify crimes as well.

However, For St. Louis, I would like to do a different analysis. I would like to find out dependency of weather on crime. " Crime occurs all over the world, there are many variables that can play a role in the event of a crime being committed.By researching new variables such as weather, will give law enforcement a better understanding of how the weather affects crime and a chance for authorities to better prepare their departments during weather conditions that influence certain criminal activity." as quoted in a research paper[2] clearly illustrates importance of analyzing weather for the police to predict crimes.

Methodology: I intend to create a detailed crime heat map for Chicago in *R* for plotting crimes with reference to Longitude and Latitude classifying crimes depending on their severity as well. On the other hand, for St. Louis, I plan on doing a Machine Learning analysis using a *python* based back-end employing *scikit-learn* package and using the *Cloudera Hadoop environment*. For static analysis I will be using R, to deduce more observations from the data.

Expected Results: I believe that everyone deserves to be informed about areas so that they can avoid those areas for their safety. Also, I feel that if we can predict the crime rates beforehand, taking into account various factors like temperature, it will be a sterling application of data *analytics* in real life.

References and Motivation:

[1] Mapping Crime: Understanding Hot Spots : U.S. Department of Justice
[2] Effects of Weather on Crime: Rodrigo Murataya and Daniel R. Gutierrez