STOR 565 Project Proposal

Mackenzie Bogiages, Kevin Chen, Austin Dunn, Tony Lieu, Alicia Ar

For our project, we want to learn about crimes that occur in the city of Durham. We found a dataset that has approximately 120,000 crime reports that occured in Durham. The excel file has a list of all reported crimes that happened since 2009. Various information about the crime was recorded such as: The longitude/latitude where the crime occured, time/date of occurence, type of crime, and zone.

From quantitative analysis of the data, we hope to answer the following questions:

1. Is there a relationship between time of the day and type of crime? In other words, can we predict certain types of crime (such as domestic violence or robbery) to happen at certain hours of the day?
2. What time of the day do crimes happen the most, in general?
3. Where in Durham, location-wise, are crimes most likely to occur?
4. The dataset shows both time the crime occurred and time the crime was reported. We can see that in some cases, the time reported was several hours or even several days after the crime occurred. What type of crimes have large time gaps between reported and occurred?
5. How can we use our analysis to make Durham law enforcement and citizens more aware of potential crime?

There are several regression and classification methods that we should consider when answering these questions. Here is a rough outline of our process:

1. Split the data into a training and test set by year. Use earlier years as training data and fit a model. Then, see how well the model performs on “future” years.
2. Consider a variety of regression and classification methods
   1. Hour vs. crime counts: Linear model, piecewise polynomial
   2. Types of crime at different times of the day: Logistic Regression, KNN, LDA, QDA, support vector machines. Use validation methods to choose appropriate complexities.
   3. Location of crime: Possibly split Durham into square plots of land, and see which areas are most prone to crime. Would likely have to implement GIS software to our project.
3. Put together findings in a clean and aesthetic presentation

Necessary software for this project include:

* R
* QGIS

This project was inspired by recent efforts of law enforcement to implement machine learning algorithms and artificial intelligence to predict time and locations of potential crimes in large cities. This technology creates a safer living space for citizens.