

Carolina Data Challenge 2017: New York Police Department

Group: *Petit Larceny* Members: *John Benhart, Joseph Boyle, Kyle Coniker*

Our group seeks to quantify and visualize crime statistics from multiple levels of granularity to gain a better understanding of crime occurring in New York City (NYC). The dataset, produced by the city's Police Department, represents queried crime complaints. The majority of this data, and consequently the subset we employ in our analysis, ranges from the beginning of 2010 to the end of 2015. In addition to the date and time of a given crime, descriptions and locations of the crime are included, in varying levels of generality.

We choose the unit of analysis for this project to be the police precinct. NYC has seventy-seven police precincts, encompassing the entire city, across its five boroughs. We derive a "risk" score for each precinct and for the entire city for each month in the timeframe, from 2015 to 2016. This score for an individual precinct incorporates the severity and frequency of crime for the precinct-month, and divides this by the population and area of the precinct, to produce an informative, unitless quantity to represent the relative, size-adjusted presence of crime in a given precinct. Next, we compare precinct risk to its grand mean across the dataset, mapping the result to visualize the spatio-temporal distribution of crime over the city during the timeframe. We find that the neighborhood bands of Lower to Midtown Manhattan, and East Harlem to the Bronx exhibit uniformly higher rates of crime than do other regions of the city.

Accompanying and complementing this presentation is a mapping of cumulative/city-wide risk by month across the timeframe. (Population and area have already been accounted for, by the construction of risk scores.) Cumulative risk exhibits characteristics of seasonality, declining sharply in the wintertime and peaking in the summertime. No significant increases or decreases in yearly crime appear evident in the dataset.

We present three additional plots. The first is a joyplot depicting frequency of different types of crimes over an average day. This allows comparative analysis of crime to determine the times that certain crimes are more likely to occur. For instance, the vast majority of DUI violations occur between 12am and 4 am, while instances of petit larceny increase throughout the late afternoon and early evening, when people are more likely to be walking around.

The second is a small-multiple representation of a bar chart illustrating reported crimes in each borough, scaled per 100 members of the borough's population. This allows us to draw insights into the prevalence of the most common crimes. For instance, Manhattan has the largest amount of petit larceny, potentially deriving from its status as a tourist and business location.

Finally, we created a series of plots which show the prevalence of certain crimes by proximity to prominent locations in NYC. This provides an intuitive description of the types of threat present in areas around a given landmark, as well as how these threats change by proximity. We observe a high prevalence of petty larceny in close proximity to Times Square, a low prevalence of burglary in Brooklyn, and implicitly conclude a high prevalence of robbery in the Bronx.