City Challenge Week(2016) - Team 12 - CUSP Crime Challenge

Performing spatio-temporal analysis and visualization using the NYPD's incident level crime dataset and performing correlational analysis with other publicly available datasets.

The topics addressed for the above mentioned project are as follows:

- Statistics of different felonies across New York City
- Mapping occurrences of crime in NYC at high geo-spatial resolutions
- Exploring correlations between crime and School Statistics
- Exploring correlations between crime and Time Series Crime Analysis

Objective & Goals

- Description of challenges and constraints
  - When, where and why do crimes occur in NYC?
  - What relationships exist between collected datasets?
  - Constraints:
    - Historical Data Collection challenges - only 1 year of complete data
    - Incomplete data - manual addition of data to cover up inconsistencies
- Project objectives and performance metrics
  - Statistics of different felonies by precinct
  - Mapping occurrences of crime by census tract
  - Correlation of school budget and crime incidents - precinct level observations
  - Time series crime analysis at macro level (months) and microlevel (hours, days, weeks)
- What you know and what you don't know
  - What we know
    - Crime Statistics
    - Census Tracts and Statistics
    - School statistics - Budget Allocation
    - Precinct locations - Crime Count per precinct jurisdiction
  - What you don't know
    - Severity of each offense
    - Context of crime
      - Information of victims of crimes
      - Information about perpetrators
      - Motivation and nature of crime

Data & Methods

- Data to be used -- sources and limitations
  - NYPD Seven Felonies Crime Statistics
    - Location (Latitude and Longitude)
    - First 3 quarters of 2015
  - School Statistics (Budget Allocations - Zip Code Level)
    - NYC Open Data Forum
    - Limitations - Not all school budgets' data was collected
  - Precinct Lists and Crime Rates
    - NYC Open Data Forum
    - Limitations - Historical data processing was limited by the time constraints
    - NYPD official website
  - 2014 Census Data (tract level population)
    - FactFinder.Census.gov
- Cleaning and data integration process
  - Merging the census tracts locations with crime locations
  - Reshaping census data by counties and merging with crime data
  - Filtering out crime statistics by precinct and year
- Computational approach
  - Calculated crime incidents
  - Correlation analysis between school budget and crime incidents

Analysis & Results
- Statistics of different felonies by precinct
- Mapping occurrences of crime by census tract
- Correlation of school budget and crime incidents by precinct
- Time series crime analysis at macro level (months) and microlevel (hours, days, weeks)

Next Steps
- Data needs
  - More details around victims: do crimes disproportionately affect different groups?
  - Severity of incident
  - Exact location of crimes (rather than mapped to the middle of the block)
  - Locations and accessibility of police stations
  - Locations of commercial, community, civic, and residential locations
- Implementation constraints
  - Scarcity of tools
  - Lack of fluency
  - Time factors
- Methodological limitations
  - Large Datasets - all the data could not be used whilst processing

School Statistics Analysis

From the maps created in CARTODB (Figure 1) and QGIS (Figure 2), it is evident especially in the map created in QGIS that there exists a correlation between school budget allocation in public schools and crimes incidences by precinct in the year 2015. On Staten Island for example, each of the points representing school locations and their budgets indicate a generally high budget allocation for these schools. These schools are also located in low crime incident precincts shown by the blueness of each of the precincts. However, just across New York City, in east Brooklyn and Queens, there are low public school budgets with high crime counts. This suggests a possible relationship between the two factors, and thus further analysis could be useful, such as looking beyond the scope of this project at historical crime and budget allocation data and cumulatively putting these datasets together on a map to give a more rounded representation of this relationship.
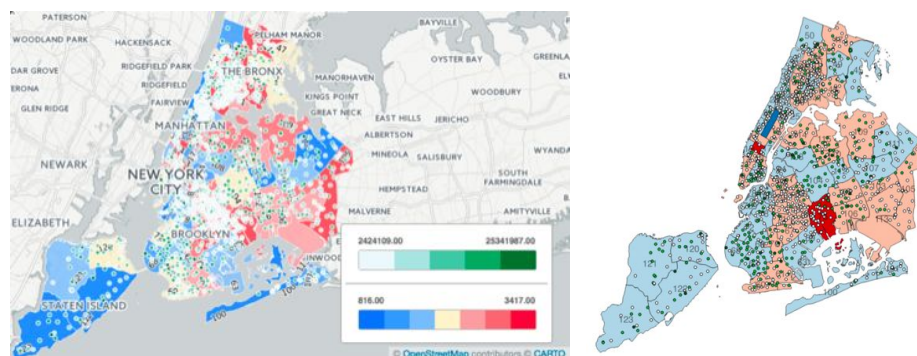

Figure 1 & 2: Map created in CARTODB(left) and QGIS(right) showing the correlation of Crime Incidents in each precinct compared to each public school's budget allocation in the year 2015.

Statistics of different felonies across NYC: Crime Type and percentage

To analyze the statistics of felonies across the New York area in year 2015, we first have a look at the crime type for the felonies and their frequencies respectively. In the year 2015, in total 102657 crimes reported, which involves seven types of felonies: 'Murder & Non-Negligence', Manslaughter, Rape, Burglary, Robbery, Grand Larceny, Grand Larceny of Motor vehicle and Felony Assault.

As is shown in the chart, 'Grand Larceny' was reported 41873 in the NYC area, which reached the highest level of 40.79% of total crime committed in 2015, while 'Murder & Non-Negligence'.
Manslaughter' was only committed 336 times, constituting to 0.33% of total crime quantity, which is the least committed crime. Moreover, the crime 'Felony Assault', 'Robbery' and 'Burglary' was also reported fairly high, with each contributed around 15% to 20% to the total crime quantity respectively.
Therefore, we recommend the New York Police Department to emphasis on preventing of 'Grand Larceny' in 2016 and meanwhile give more attention to the crime 'Felony Assault', 'Robbery' and 'Burglary' to maintain the safety of New York City.
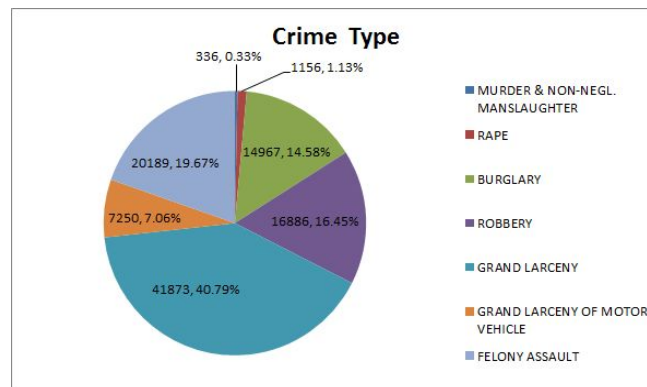

Figure 3: Crime Type in New York City 2015
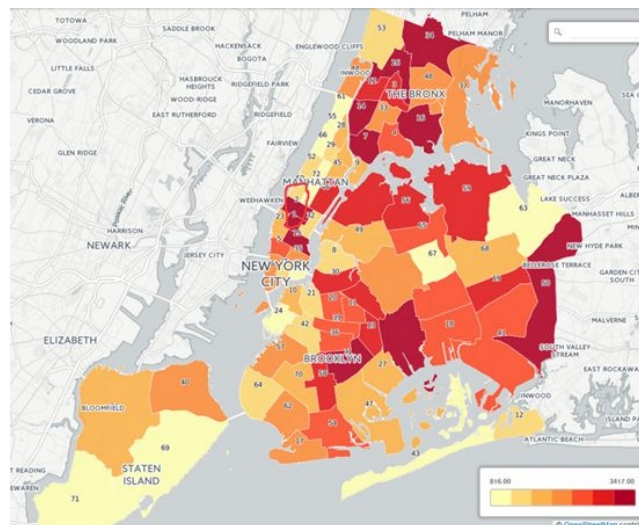
Precinct level crime overview

Figure 4: Rank crime rate per person per precinct

As is shown in the map, in terms of the crime quantity, the precinct 75 has the highest level of the number of crime reported, which is 3417 in 2015 and ranked 1st. Precinct 14 and 44 ranked 2nd and 3rd with crime count as 2728 and 2438 respectively. However, considering the census data and population in each precinct, the crime rate and thus the rank accordingly would be different. To normalize the crime data, we divide the crime quantity per precinct by population per precinct in 2015, and here as is shown in the map, the precinct 14 ranked first in terms of the crime rate with 0.050403, with precinct 18 and 48 followed as 2nd and 3rd. As the precinct 14 ranked 2nd in the crime quantity list and ranked 1st in the crime rate list, it is obvious that the precinct 14 is the precinct that had highest crime committed rate and the New York Police Department should lay great importance in this precinct to prevent the felonies and lower its crime rate in precinct 14.

Crime vs Time Overview

We looked at a time series between borough and actual date, but this proved way too noisy to properly analyze what was going on. However, it was apparent that a general increase in crime during the summer months occurred, as well as a much higher crime rate for Manhattan, Bronx, and Brooklyn, compared to Staten Island.

We next looked at a time series between borough and month, and found a much clearer picture showing the differences between borough crime rates (Manhattan & Bronx the highest, Staten Island the lowest). The time series also shows a consistent trend with higher crime rates in the summer months, and lower crime rates in the cooler months. We also looked at the relationship between type of crime and month. The same seasonal trend is reflected here as with the borough versus month investigation, with the warmer months having higher crime rates compared to the cooler months. Grand larceny and felony assault have the highest crime rates We finally looked at the relationship between type of crime and the day of the month. Our hypothesis was that certain crimes associated with transfers of wealth (burglary, larceny, robbery) would occur near bill due dates (the beginning and end of the months). However, looking at the data, this does not seem to be the case. It may be due to a lack of data due to only a select few months having 31 days, but the crime rates for all types of crime seem to tail off at the end of the month. Otherwise, the rates seem constant across the month.
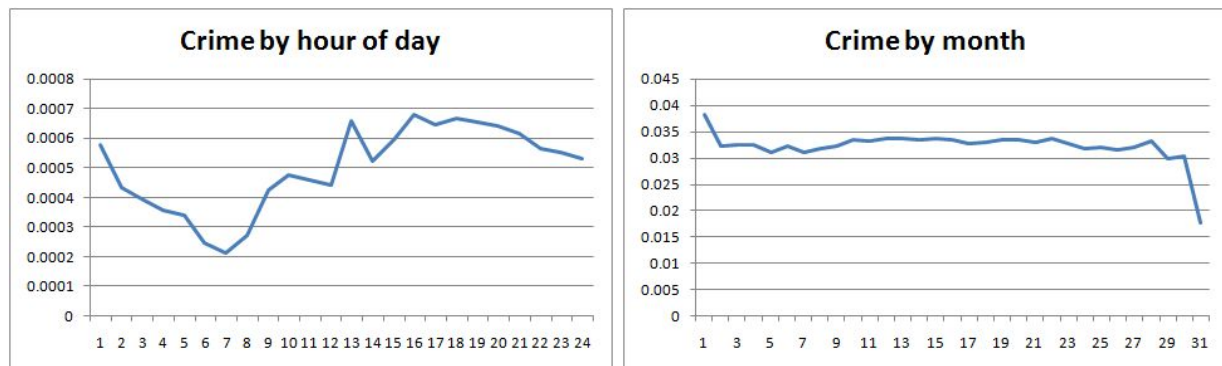

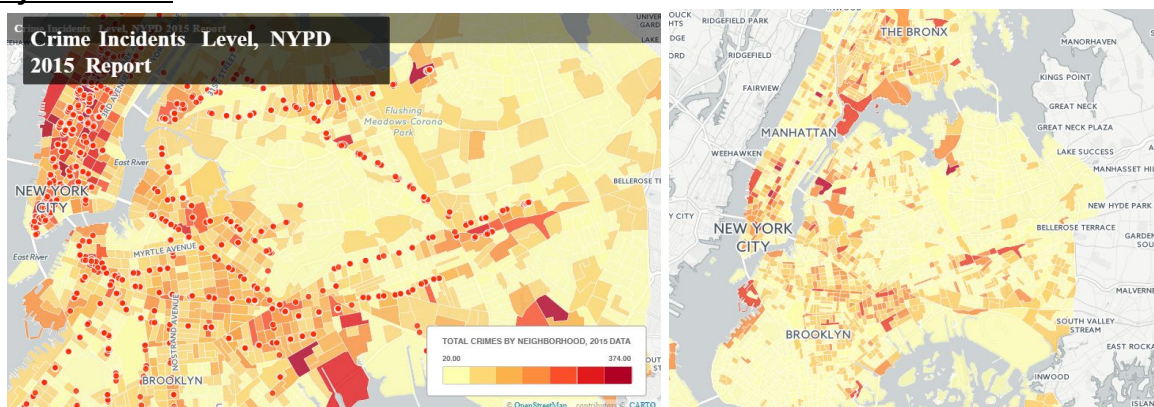**Figure 5:** Graphs of crime by hour and month

Analysis Results

Figure 6: Mapping occurrences of crime by census tract (left). Map created in CARTODB (right) showing the crime to population rate per census tract in the year 2015.

When mapped by census tract, the crime data shows a concentration of crime occurrences in Midtown Manhattan, Northern brooklyn, and the Bronx. There are also large areas of relatively low crime occurrences, most specifically in Southern Brooklyn,central Queens, and Staten Island.

Incidentally, those large areas of relative safety appear to correlate visually to the MTA Subway map, showing that areas that are less accessible by train show fewer crimes. This could be an effect of income whereby wealthier families are less reliant on subway lines, suggesting a link between subways, income, and crime occurrences. The number of crimes within each census tract can not tell us the whole story. These tracts have different characteristics: some are big areas, some are populous while some are deserted.

Their population densities are not the same. Hence, we take a step further by including the population into our analysis. We only consider the areas with more than 1000 residents, which accounts for 99.5% of all the areas covered. Figure 3 shows the crimes per 100 residents by census tract, providing us a different picture on the safety level of certain areas. While Bushwick South and Vinegar Hill in Brooklyn are the highlighted ones with high numbers of crimes in the borough, they are relatively safe using the new metric, crimes per 100 residents. The same situation happens to Carroll Garden in Brooklyn , which is no long safe areas when we take the population into account.

The neighborhoods such as Flushing in Queens, Midtown South in Manhattan, East New York in Brooklyn, Stapleton-Rosebank in Staten Island are still the unsafe areas based on either number of crimes or the ratio of crimes to population.

The median income may be another factor that influence the crime probability in the area since we observe that in Queens, the middle class neighborhoods such as Richmond Hill and Jamaica have higher ratios of crime to population than working-class neighborhoods like Springfield Gardens North. Specifically, Grand Larceny and Felony Assault have show different spatial trends. Grand Larceny has the highest level in midtown manhattan, downtown brooklyn, and specific neighborhoods of East New York, Flushing, and West concourse. The high occurrence of larceny in midtown manhattan could be attributed to the high levels of commercial, business, and tourist activity as well as the large number of people frequenting those areas.

The part of East New York that has a high frequency of larceny is largely a shopping area. Felony Assault has the highest occurrences in Northern Brooklyn, Harlem, and the South Bronx, area that have fewer office buildings, and more retail and residences. These are important distinctions when considering the original map of crime incidents.

Although there is a lot of crime happening in midtown manhattan, it larceny, as opposed to crime happening in brooklyn and the bronx, which is felony assault. Larceny is the theft of property without the use of force, which is considered robbery. Larceny suggests that there is no confrontation between the victim and the perpetrator. Felony Assault is by definition confrontational and results in physical injury to the victim. So although, midtown manhattan shows a high incidents of crime, that crime is less dangerous than the crime happening in the outer boroughs, which is more violent and threatening to physical safety.
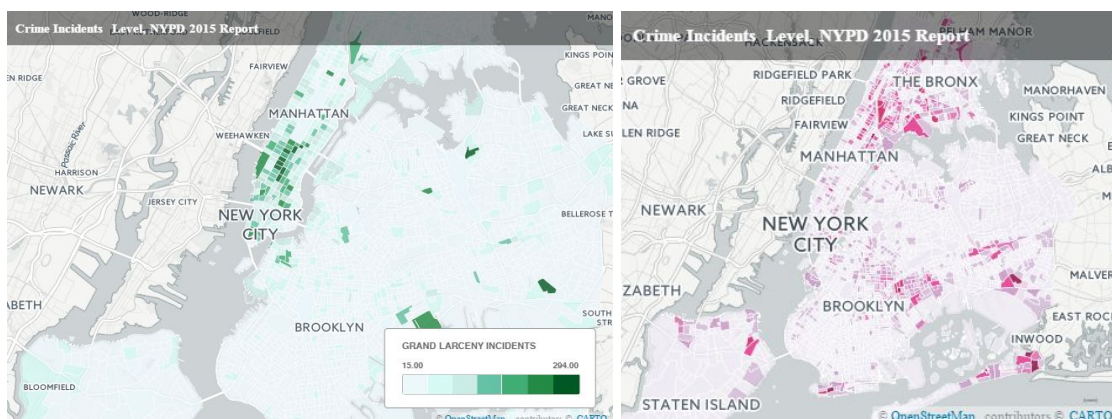

Figure 7: (left) Incidents of grand larceny compared to incidents of felony assault (right).