TOPIC: Crime and School Violence in New York City

MAIN DATA SET: NYPD Major Felony Incidents,
FORMAT: csv file, 1123465 observations of 20 variables,
LINK: **https://catalog.data.gov/dataset/nypd-7-major-felony-incidents**

SECONDARY DATA SET(s): VARDIS reports on NYC school violence
FORMAT: multiple csv files to be converted to csv. 14730 observations of 48 variables.
LINK(raw): http://www.p12.nysed.gov/irs/school_safety/school_safety_data_reporting.html
LINK (cleaned):
https://docs.google.com/spreadsheets/d/1-S_QbryW4e5jFOqVQZYpgamGVBTZI1htN2yTGhMsZ
0w/edit?usp=sharing

THIRD DATA SET(s): 2000, 2010 NYC Population by borough
FORMAT: csv, 2 observations of 5 variables
LINK: http://www.p12.nysed.gov/irs/school_safety/school_safety_data_reporting.html

PROJECT PROPOSAL:
For our final project submission we plan to use the publicly available datasets to examine crime trends in New York City and their possible connection to incidents of school violence.

We will ask the following questions...
- How does the NYC crime rate vary by day of week?
- What times of day experience the highest crime? How has this changed since 2006?
- Do crime rates go up/down on Major Holidays (Christmas and Thanksgiving vs. Halloween?)
- How do crime rates change by season?
- What are the highest crime areas of NYC? Do different boroughs experience different types of crime?
- Is there a relationship between school violence and violence in the surrounding neighborhood?

Our analysis will include...
- Histograms and plots exploring different perspectives on when/where/what type of crime occurs in NYC.
- Analysis of crime trends by temporal criteria, category of crime and geography.
- A comparison of crimes per capita (per thousand people) and crimes per square mile.
- A plot showing the correlation (or lack of correlation) between school violence and violence in the surrounding neighborhoods
- ...(maybe) A heat map visual of crime by county/borough/school

A First look at the data:

Our primary dataset, (NYPD Felony Dataset) provides information about felonies committed in New York City. Each of the 1,123,465 observations represents a single felony record. The columns of this data set include the type of offense, the time, and the location where it occurred. A complete list of the variables in this dataset is included below (Appendix A). While the dates in the data set range from 1905 to 2015, initial analysis shows that 99% of the data falls within in the time period between 2006 and 2015. We will focus our analysis on that time period.

One of the important views of the dataset come from grouping by "Borough" and by "Offense". The combination view below gives a high-level, detailed insight about which offenses are most prevalent and which boroughs are most heavily-worked. It also speaks to overview questions including: is there consistency in offense proportions between boroughs? which types of felony offenses are most common? and where do the majority of offenses occur?
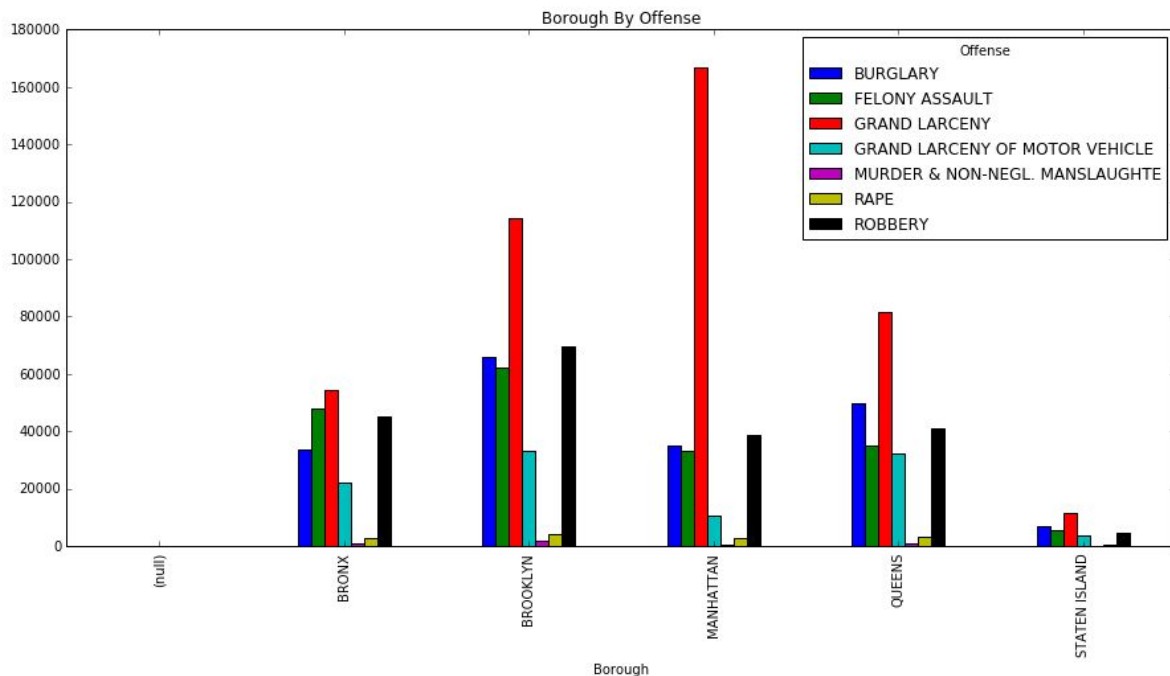

From this visual, we can see that "Grand Larceny" is the most offensive felony in general. The proportions of offense follow roughly the same pattern except for Manhattan, which has an extremely large percentage of Grand Larceny but a low count of everything else. This could suggest that the Manhattan borough has more grand larceny offenses, or it has a more general definition of Grand Larceny, or maybe that Grand Larceny is used more of as a "catch all" type of offense in Manhattan. Another possibility is that since Manhattan is the financial center of New York, its high rate of Grand Larceny may actually reflect incidents of embezzlement. Wikipedia suggests that these two types of crime can be hard to distinguish from each other. If this is the case, then some of Manhattan's Grand Larceny may actually by categorically different than the Grand Larceny that occurs in other boroughs… making our direct comparison misleading. Further investigation of the locations of these crimes may help shed some light on this issue.

Another significant approach to the felony data is as a time series. The daily view at right suggests a cyclical pattern to offenses over time. This could be related to seasons, or the flux of arrests to releases over time, or it could be related to constantly renewed efforts to minimize crime and then relaxing those efforts repeatedly. Also note that while the offense counts are cyclical over time, the count seems to be generally declining. Further analysis will focus on whether this cyclical pattern is true for all types of felonies and whether crime rates are also cyclical by time of day or day of week.
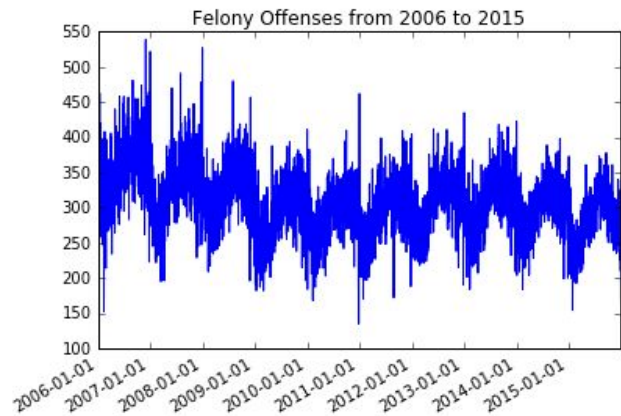

Below, a more granular breakout by felony offense type reveals that the trend in crime levels is different depending on the type of crime in question. The only offense type not declining over time is "Felony Assault". The most notably consistent decrease in crime rate is in auto theft ("Grand Larceny - Motor). There is also a marked drop in both "Grand Larceny" and "Robbery" that coincides with the onset of the recent recession (2008-9). Given our earlier questions about the murky distinction between "Grand Larceny" and embezzlement, one wonders whether this drop is primarily reflective of Manhattan's data (as a proxy for the financial industry)... or is the drop visible in all boroughs?
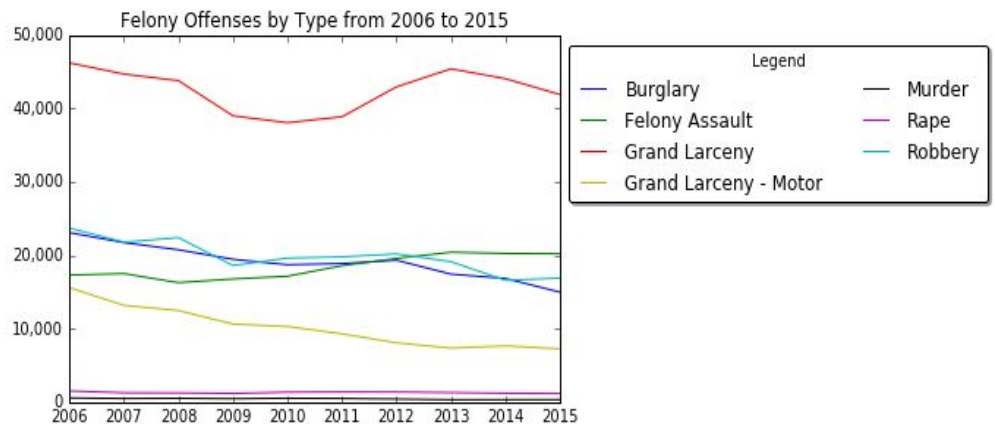

The final part of our project will examine possible connections between crime trends and incidents of violence and crime in New York Schools. For this portion of our project we will draw upon a secondary dataset consisting of information from the Violent and Disruptive Incident Reports (VADIR) that are filed annually by all New York Public Schools and whose results are made available to the public. We have pulled and collated the available data for the same time period as our NYPD dataset (2006-2015). Each of the 14,730 entries in this second dataset represent a single school's incidents in a single year. The categories of recorded incidents range from bullying and alcohol possession to kidnapping, assault and homicide. The data also include information about the district and number of students for each school. A full list of the variables for this second dataset is available below (Appendix B).

A significant portion of our work thus far has gone into simply cleaning and collating this data set. As to be expected when working with real data, the format of the report has changed over time. For example the reports from 2009, 2010 listed schools by county instead of borough. Other years were missing the district category all together. Joining the raw data required us to resolve these and other issues like inconsistent column names and inconsistent capitalization in data entries. At the beginning of this document we've included links to both the raw (csv) files and our (semi)cleaned version. More work will need to be done to process this data, however the charts below provide an initial snapshot of trends in NYC school crime.

The graph at right shows that since 2006 there has been a consistent increase in the number of school incidents. It is not clear if this trend is reflective of a true increase or may be due to better reporting and more buy in (the data is self reported by schools to the government). Further analysis will look at whether this trend reflects all types of school incidents or is primarily due to certain types of school crimes. Below, we can see that while incidents without weapons don't appear to have a strong trend upwards or downwards, incidents with weapons appear to be rising on the rise.
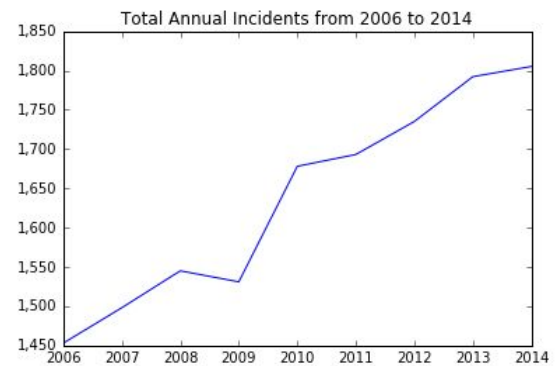
The next steps for our project include completing the felony crime analysis suggested above and linking these two datasets geographically to examine a possible relationship between crime in the community and crime in schools. We have two ideas about how this could be done. One possibility would be to get GPS coordinates for the schools in the second data set and use a distance formula calculation to identify nearby crimes from the first dataset. If identifying GPS coordinates for the schools proves challenging our alternate plan will be to categorize the crimes by school district instead of
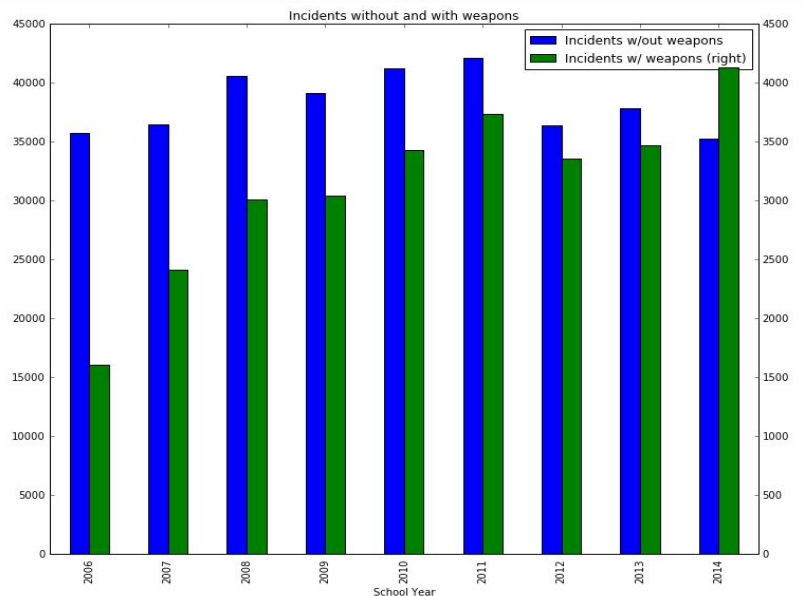
proximity to a specific school (we have an ESRI file detailing the boundaries of the school districts and GPS coordinates for the crimes). A final backup if both of these prove beyond the scope of our programming abilities would to to simply look at a possible correlation in trends by borough. Time permitting, we are interested in attempting to create a visual "heat map" which the location and frequency of crime incidents (in schools and in the community) over time.

APPENDIX A:

Complete list of variables in the NYPD data set.

ColumnName	Type
Identifier	object
Occurrence Date	object
Day of Week	object
Occurrence Month	object
Occurrence Day	float64
Occurrence Year	float64
Occurrence Hour	float64
CompStat Month	float64
CompStat Day	float64
CompStat Year	float64
Offense	object
Offense Classification	object
Sector	object
Precinct	float64
Borough	object
Jurisdiction	object
XCoordinate	float64
YCoordinate	float64
Location 1	object

APPENDIX B

Complete list of variables in the VADIR_incidents data set.

*Note: columns ending with "nw" means "no weapon" and "ww" means "with weapon".

ColumnName	Type
Alcohol Possesion	float64
Alcohol Possession	float64
Arson	float64
Assault With Serious Physical Injury_nw	float64
Assault With Serious Physical Injury_ww	float64
Assault with Physical Injury_nw	float64
Assault with Physical Injury_ww	float64
BEDS Code	float64
Bomb Threat	float64
Burglary_nw	float64
Burglary_ww	float64
County	object
Criminal Mischief_nw	float64
Criminal Mischief_ww	float64
District	float64
Drug Possesion	float64
Drug Possession	float64
Enrollment	float64
False Alarm	float64
Forcible Sex Offenses_nw	float64

Forcible Sex Offenses_ww	float64
Grade Organization	object
Homicide_nw	float64
Homicide_ww	float64
Intimidation, Harassment, Menacing, or Bullying_nw	float64
Intimidation, Harassment, Menacing, or Bullying_ww	float64
Kidnapping_nw	float64
Kidnapping_ww	float64
Larceny, or Other Theft_nw	float64
Larceny, or Other Theft_ww	float64
Minor Altercations_nw	float64
Minor Altercations_ww	float64
Need/Resource Category	object
Other Disruptive	float64
Other Disruptive Incidents	float64
Other Sex Offenses_nw	float64
Other Sex Offenses_ww	float64
Other Sex offenses_nw	float64
Other Sex offenses_ww	float64
Reckless Endangerment_nw	float64
Reckless Endangerment_ww	float64
Riot_nw	float64
Riot_ww	float64

Robbery_nw	float64
Robbery_ww	float64
School Name	object
School Type	object
School Year	float64
Use Possession or Sale of Alcohol	float64
Use Possession or Sale of Drugs	float64
Weapon Possession_nw	float64
Weapon Possession_ww	float64