CS506 Final Report

Submitted on: 05/1/2018

Department of Defense 1033 Program Analysis

Prepared by:
Joseph O'Donnell
Gautham Ravichandran
Albert Sze
Eric Iren Zhang

Introduction and Problem Formulation:

The topic pursued in this analysis is the use of the Department of Defense (DOD) 1033 program by local Law Enforcement Agencies (LEAs) and how that relates to the demographics of the people protected by those agencies. The 1033 program allows the DOD to donate surplus equipment to LEAs upon request. The analysis focuses on Massachusetts (MA) and Florida (FL) for the city/county level analyses, and expands to the whole country using state level census statistics.

This analysis seeks to determine if there is a correlation between external factors at the local or national level such as demographics or local/national tragedies versus LEA equipment acquisition. Specifically, this analysis seeks to answer the following:

1. What are statistical characteristics of a town/municipality or law enforcement agency that receives a larger than the average amount of weapons?
2. Does education, race, income, or voting pattern have an impact on the types of donations received by a law enforcement agency?
3. Looking at major public safety events in recent history (school shootings, major protests, September 11th, etc.), is there a correlation with the occurrence of an event and donations received immediately following? If there's a positive correlation, how soon after the incident does it happen, and how long does it last, on average? (how long does it take to return to pre-event levels)
4. Is there a similarity between which town demographics in different states drive the acquisition of equipment? Is there a trend between different regions of the United States?

Methodologies:

The primary methodologies used for the analysis are multivariable regression and k-means clustering. Regression is used to identify which demographic trends predict equipment obtainment. The regression model used SKLearn's linear regression implementation to look at which state census features can be used to predict total equipment acquisition. The methodology involved using the metrics of best fit learned in lecture. The state census and categorical acquisition data was split into 80% training and 20% testing and a linear model was created. We looked at various permutations of census data that was included in each state's feature vector during the training and testing phase.

Clustering is used to visually show how demographics are paired with the equipment. The entire database of items was sorted into high level categories described in the data analysis section below. Clustering was focused on items that were either "Weapons", "Vehicles", "and "Tactical Equipment". Items that were classified as "Miscellaneous" were not included, because it could skew the results since each LEA had obtained a large amount of "Miscellaneous" items.

Two sets of feature vectors were then created, where one set of feature vectors contains the quantity of each type of equipment, and another set of feature vectors contains the total dollar acquisition value of each type of equipment. Each item in the feature vectors were then rescaled

between 0-1 so that large values of acquirements would not skew the data. The data was broken down into time buckets of pre-9/11, 9/11/2001-2010, and post 2010. The K-means algorithm was used to create clusters on LEAs that obtained similar amounts of each equipment type. To determine which clustering algorithm is appropriate, three from class were selected: the K-Means , hierarchical , and GMM. To determine the optimal number of clusters, multiple iterations of varying cluster numbers were used, and a silhouette score was computed. Ultimately, the K means method showed the highest silhouette scores, while having error values that were very consistent with the other two clustering methodologies, justifying its selection. K-means clustering was performed at a micro level, by looking at the police stations within MA, and at a macro scale of all US states.

Time-based analysis is done at the aggregate level across MA and the country. This analysis looks at the timing of major public safety events and their correlation to increased equipment acquisition. This technique was mainly guided through visualization of the data and involved research into historical events at the local and national level.

Data Analysis & Results:

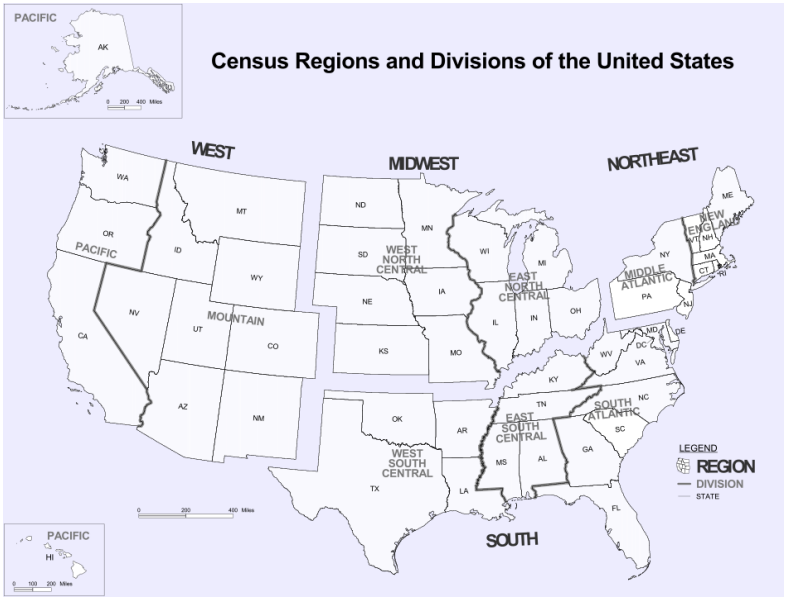
For this analysis two primary datasets were used; the first dataset was provided by the Spark program partner, Nasser Eleroos. The dataset lists the state, LEA, item name, quantity, acquisition date, and acquisition value for 1033 program donations. The items include the various military equipment that was obtained by each LEA. There were over 7000 unique items acquired by various LEAs so to reduce scope, the items were categorized into 4 major groups; weapons, tactical equipment, vehicles and miscellaneous items. The weapons category includes rifles, knives, batons, and pistols. The tactical equipment category includes all weapon attachments, body armor, vehicle attachments, radio and surveillance equipment. The vehicles category includes trucks, vans, mine resistant vehicles, boats, and aircraft. The final category includes any items that do not fit within the previous categories; i.e. medical supplies, tools and wires etc. All items were manually labeled with these identifiers to better categorize state acquiring trends. Key words like “rifle”, “tactical”, “radio”, “surveillance”, “truck” were used to group the equipment list. The outliers which did not fit within the keyword grouping were manually labeled according to the metric outlined above. For the majority of the analysis, the miscellaneous category was dropped. This dataset was then processed to look at total acquisition value of equipment received for each LEA from each state against the categories that were described above.

The second dataset used is census data, obtained via web-scraping, from United States Census Bureau. The census data taken for all US states includes population, population density, racial breakdown, poverty rate, education attainment, foreign born population and violent crime counts. Additionally, local census data for counties in MA and FL was pulled. Massachusetts was chosen because it’s a demographic that is easily relatable with and can be used to represent the Northeast region of the US. FL was chosen because it cumulatively acquired the highest value of military equipment through the 1033 Program. Census data was also obtained for MA cities and towns.

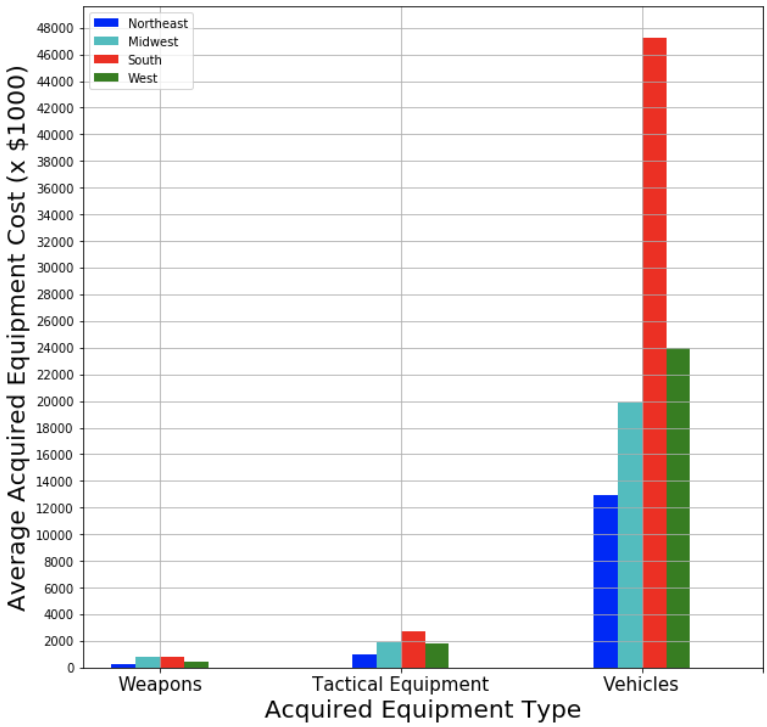
County census data for MA and FL was obtained in CSV format from the Bureau website. The original data had to be cleaned of irrelevant columns and names for easier ingestion in future analysis. Additionally, MA town and city data was obtained the same way, and underwent a similar cleaning process.

Once all the raw data was obtained, each police department in MA and FL was combined with its relevant county and town. The Google Maps API was used to find each LEA and its associated city and county. The methodology failed occasionally due to issues with ambiguous department names forcing manual matching for those delinquant departments. Additionally, county LEAs were dropped when matching departments to cities and towns census data.

Exploratory analysis began at the macro states level. The states were grouped into four regions; Midwest, West, South and Northeast, consistent with the US Census Bureau definition below:

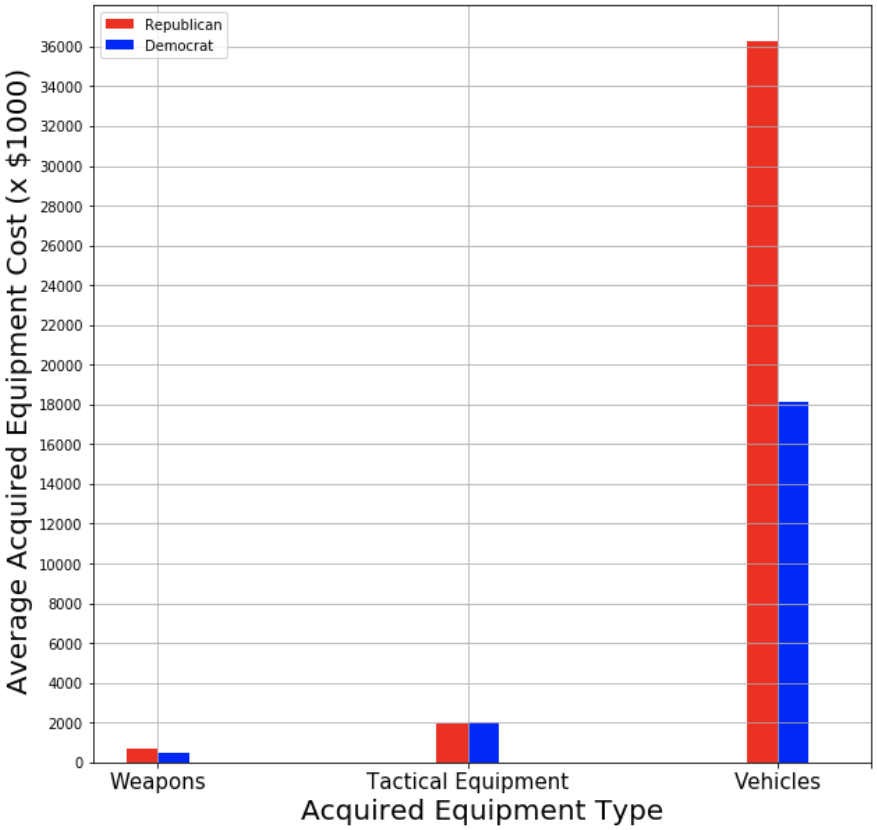


The total acquisition value of categorical equipment received for each region was calculated and plotted in the graph below. The plot was normalized by the number of states in each region in order to get an “average” value for each region.



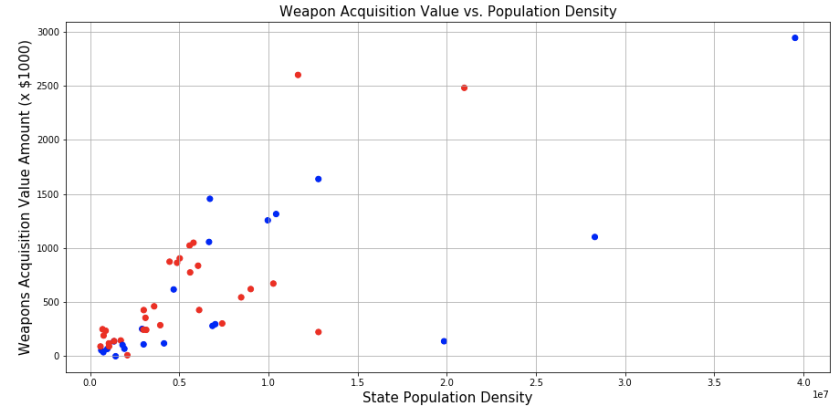
The plot above shows that southern states have acquired more tactical equipment and vehicles than any other regions. It also shows that the northeast region has acquired the least amount of weapons, tactical equipment and vehicles when compared to the other 3 regions. Finally, the plot shows that the overwhelming majority of the equipment cost is from vehicle acquisitions.

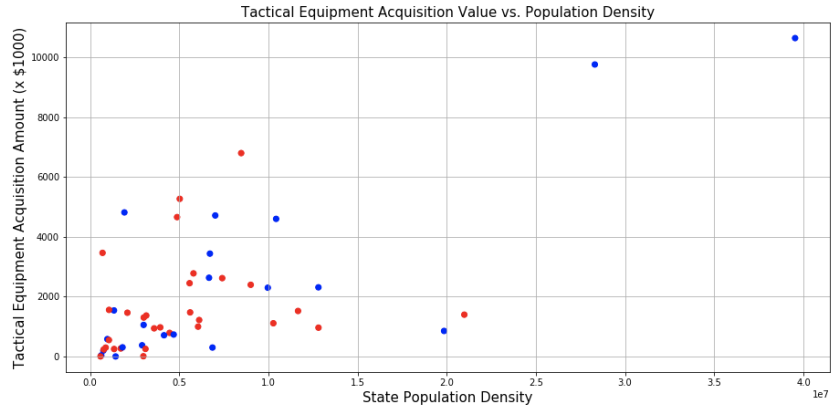
This initial analysis showed a divide in political affiliation between the northeast and the south. Total acquisition value for Republican states and Democratic states were plotted. The data was averaged over the number of states that fall within this binary classification.

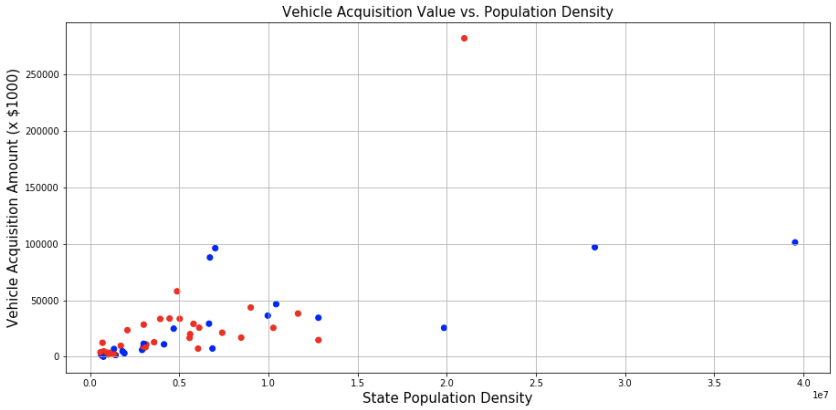


The plot above shows that Republican states have received more weapons and vehicles than their Democratic counterparts, and that southern and Republican states have acquired more equipment than their counterpart states. With this initial finding in mind, the state's census data was used to find correlations between the states and their military equipment acquisition values.

At the state level, total weapons, vehicles and tactical equipment was plotted against key census data points to see if there are any relationships to better create linear models for predicting future acquisitions. There are plots generated below showcasing some of the key census characteristics for states versus their categorical acquisition value. The points are also colored red for Republican and blue for Democrat.







The plot above shows that there is a trend between population density and total value of weapons and tactical equipment acquired. The vehicle acquisition amount does not seem to trend with population density and is relatively flat. Additionally, there does not seem to be any clear distinction between Republican and Democratic states. This basic plotting analysis was completed for other census characteristics including percentage of state population with HS degrees, percentage of state population below state poverty levels, percentage of population by race and per capita income. These features were chosen to answer our initial questions if certain town characteristics correlate to what equipment states acquire through the 1033 program.

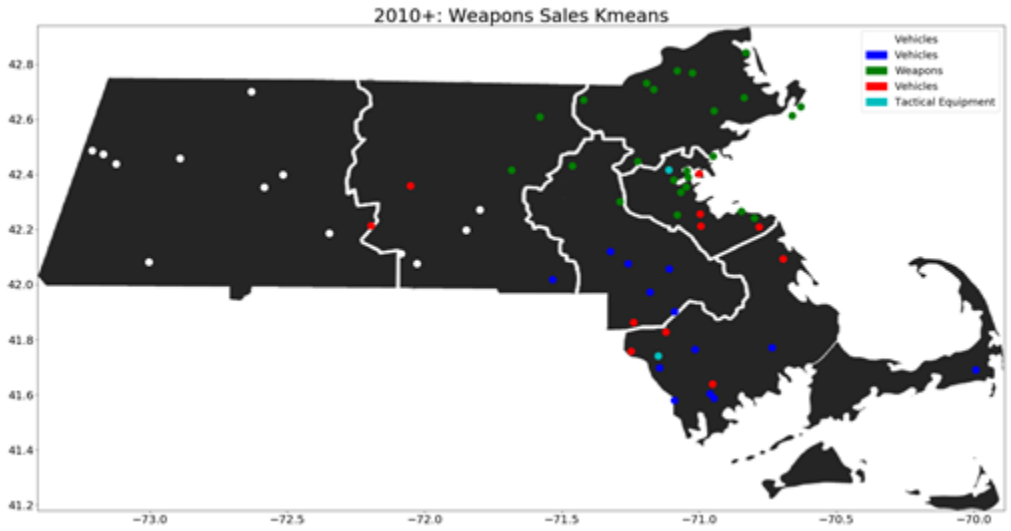
With this exploratory analysis, we found that there was not a strong linear correlation between each state's census features and both the type and value of equipment acquired. The

population density seemed to be the only characteristic found that seemingly had some linear correlation to acquisition type and value.

This same procedure was performed at the county level for both MA and FL in order to see if relationships can be found at the state level. At the state level, only MA counties had a correlation between population density and equipment acquisition similar to the relationship seen at the state level. FL did not have any clear relationship at all. The plot in Appendix A shows that regardless of population density, the value of acquired equipment does not correlate for FL. In all our testing, we could not get an accurate linear model with good R^2 value. The R^2 value was 0 for most configurations and even negative for some particular combinations. An R^2 value of 0 signifies that model explains none of the variability of the response data around its mean.

Clustering Analysis on Quantities of Donations

The K means clustering algorithm was run on feature vectors of equipment quantity to determine which LEAs are similar to each other in the quantities of acquisition. Below illustrates the cluster at an individual LEA level scale for MA.



The cluster graphs show that pre and post 9/11 majority (graphs can be found in appendix B) of the equipment that was obtained were weapons throughout MA, but 2010+ shows more interesting results. 2010+ shows that west and south of MA obtained predominantly vehicles, northeast MA mostly weapons, and Boston mostly tactical equipment. It can also be seen that pre 9/11, most weapons in MA were acquired in the west, and post 9/11, it was central MA. The 2010+ map shows that there are more LEAs acquiring military equipment compared to pre and post 9/11.

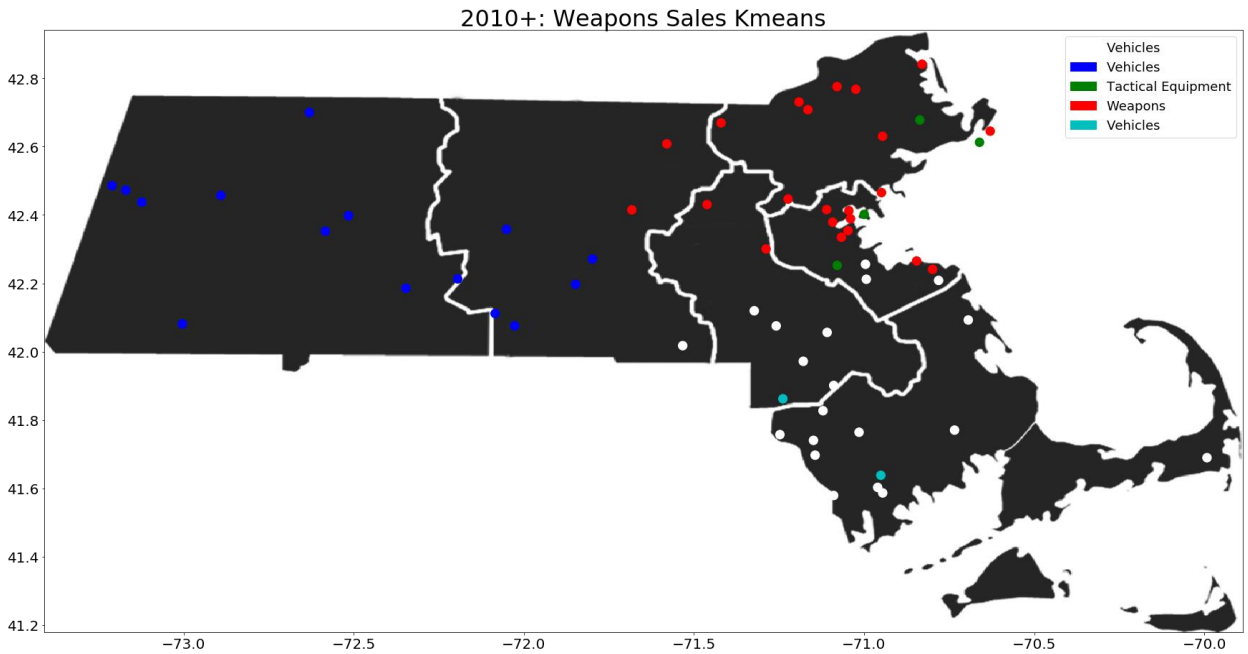
A similar study was then done at the macro level to see if there was any trend across the United States.

The US pre 9/11 and 2010+ cluster graph (shown in appendix B) did not show much. Pre 9/11 shows that the majority of the states acquired weapons. It also shows that Texas, Kansas, Florida, and South Carolina obtained predominately vehicles. The 2010+ clustering shows that most US states obtained predominately vehicles in each state with the exception of Texas acquiring mostly weapons. The US post 9/11 cluster graph, shown above, reveals that the south and southwest of the US acquired mostly vehicles, Texas acquired mostly tactical equipment, and the rest of the country (northeast, northwest, and midwest) received mostly weapons.

An interesting finding from clustering the US at the macro level for pre 9/11, post 9/11, and 2010+, showed that Texas constantly tend to be an anomaly and always differed from the rest of the country.

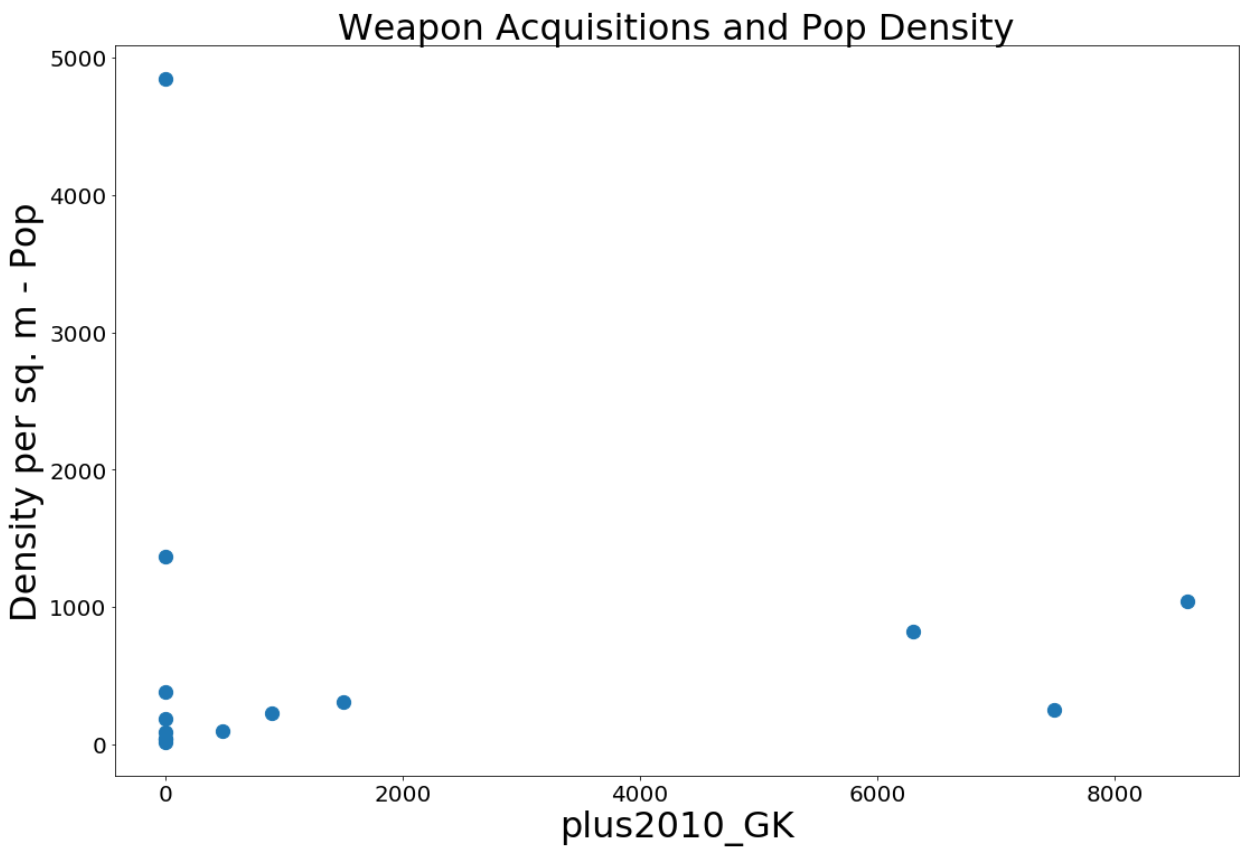
Clustering Analysis on Total Acquisition Value of Donations

We also looked at total value of acquired equipment to see if there were trends in the value of the equipment acquired. The acquired military equipment data was broken into the same buckets as the clustering analysis on the quantities. However, instead of the feature vector storing quantities of each item, it stores the total acquisition value of each bucket for that particular LEA. The plots for the pre 9/11 and post 9/11 time frame were inconclusive and were not included, however, the plot below shows the acquisitions that have occurred after 2010.



From this, we saw that the upper right hand corner of MA had numerous LEAs that obtained a similar total value of equipment, with the dominant group being weapons. Southern and Western MA had large total value of vehicles acquired. LEAs in the Greater Boston Area acquired similar total value of Tactical Equipment. This looks very similar to the results when clustering on quantities of acquisitions.

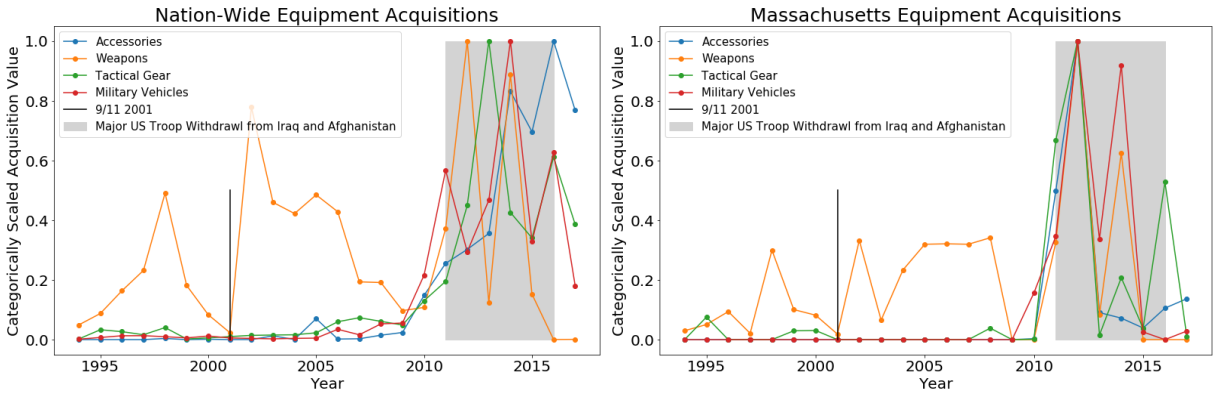
Now that we have the clusters defined, we tried to determine the key census characteristics that tie each LEA in their respective clusters together, as the LEAs in each cluster acquired similar amounts of equipment. Using the findings from clustering the data (both at the micro and macro level), an attempt to correlate the clusters for 2010+ and the census data was done. However, it was determined that there is no correlation between the quantities of equipment being acquired and the demographic of that town/state. One example of such a finding is shown below, the LEAs in the red “Weapons” cluster are plotted with total weapon acquisition value on the X axis, and population density on the Y axis.



From this, the only conclusion that can be made, is that population density within a town is not a good indicator for weapon acquisitions, whether it be in quantity or total acquisition value. Plots like this are consistent across all town demographics that are provided by the US Census data, leading to the conclusion that town demographics are not good parameters to use for drawing conclusions on the equipment acquired by its corresponding LEA.

Time-Based Analysis

The time-based analysis of acquisition value shows that the largest factor in acquisitions may simply be availability of equipment. There was a nation-wide increase in acquisitions immediately following the start of the first major withdrawal of US troops from Iraq in 2011 through the end of the major withdrawal of troops in Afghanistan in 2016. 2011 is when vehicles (category 3) and tactical equipment (category 2) became widely available to LEAs, and this is when the total value of acquisitions spikes.



Prior to 2011 there are other factors which can be seen more readily. Although only correlation can be shown and causation cannot be proved, the acquisition value of small arms does spike in nearly every state across the country one year after 9/11, 2001, and one year after the 1997 North Hollywood shootout event. In MA, the Boston Marathon Bombing in 2013 also preceded a spike in weapons obtained. This spike is also seen across the country suggesting that the bombing may have prompted other states to request more weapons.

Showing these trends and outlining certain historical factors has shown that equipment acquisition impact by local public safety events may not be as big a driver as equipment availability appears to be for equipment acquisition.

Preliminary Conclusions:

The largest driver in equipment acquisitions is availability of equipment. Availability spikes following the withdrawal of US troops from foreign combat. This is shown most clearly in 2011-2016 as troops were being withdrawn from Iraq and Afghanistan at increased rates and equipment acquisition spiked across the country (specifically a spike in vehicle acquisition). The secondary driver for acquisition impacts mainly weapons and is the perceived need for militarization by police forces due to major crises. The most clear correlation can be seen by a spike in weapon acquisition post Sept. 11 2001. In addition a spike in military obtain can be seen in 1998 after the infamous North Hollywood shootout, which is considered as the turning point for police militarization for the United States.

Our linear regression models showed that there is no relationship between each state's census data and its equipment acquisition value. One of our key questions was if there were certain traits of towns/ states which could be used to predict equipment acquisitions. The only classification that had any distinct patterns was Republican versus Democratic states. Republicans states have on average, acquired more weapons and vehicles than their democratic counterparts. This finding is further corroborated when looking at the region specific equipment acquisitions. The southern states which are all Republicans states, acquired more tactical equipment and vehicles than any other region.

Based off the clustering analysis, it was shown that clustering of quantity and total value of military equipment acquisitions had similar results. Utilizing the findings from clustering,

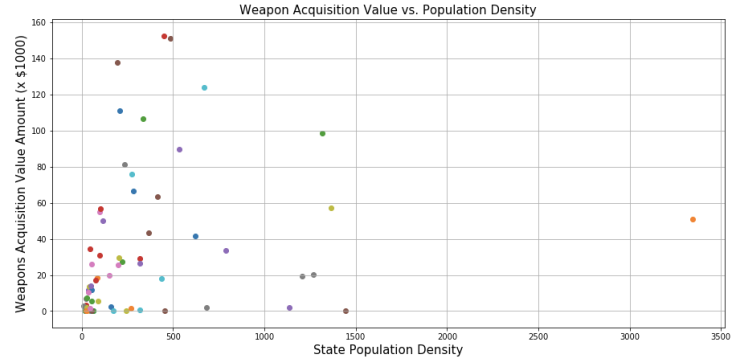
correlation between military items and census data was attempted. However, the analysis showed no strong correlation between weapon, tactical equipment, and vehicle acquisition vs town/state demographic.

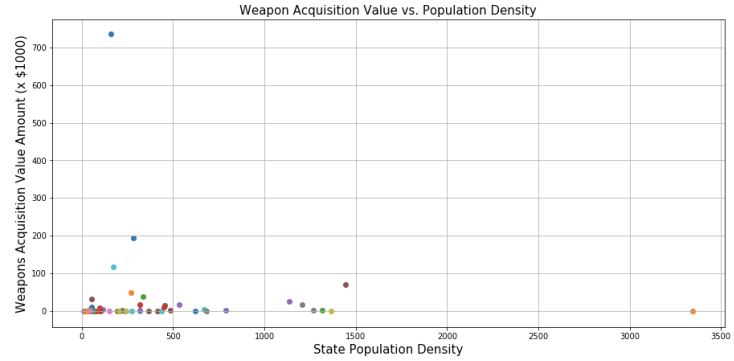
Further Analysis

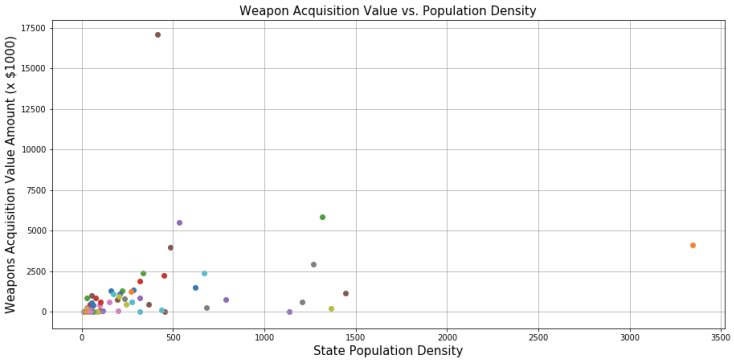
Further look into this study would include contacting LEAs which showed specifically high spikes in acquisitions around certain events to ask what their drivers were, as well as a deeper look into US and world history around these times, potentially contacting the BU History Department. Another dataset that can be analyzed is US terrorism index and its effect on LEA equipment acquisition.

Another extension of this research would be analysis into the budgets of each LEA, and how that affected the donations of surplus military equipment. This would require more time because the data on each LEA and their budget is not readily available. To further complicate things, 1033 program record information is not subject to review, thus, finding said data would be very difficult, but doable if given ample time for research. Methods of collecting data could include surveying police officers to determine why a certain LEA would feel the need to request certain types of equipment. Surveying DOD sources could lend more insight into what really demonstrates need, and how decisions are made as to which LEA gets what equipment.

Appendix

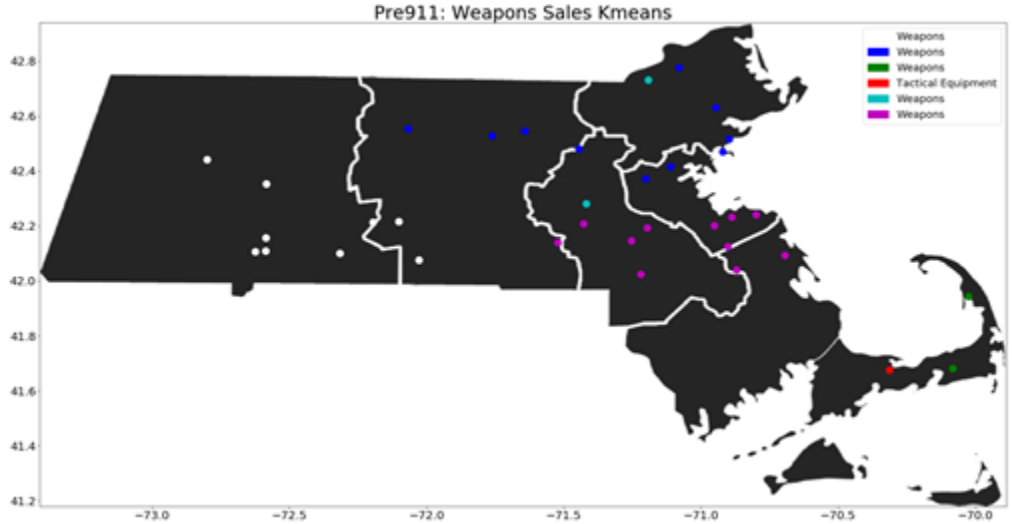


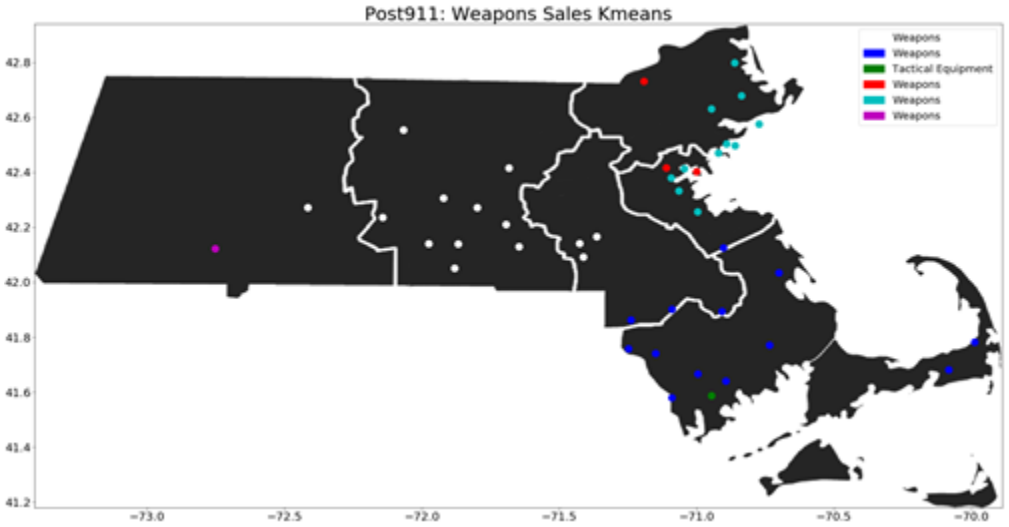




Appendix A - FL County Equipment Acquisition vs. Population Density

Appendix B


Pre911: Weapons Sales Kmeans
Weapons
Weapons
Weapons
Tactical Equipment
Weapons
Weapons
42.8
42.6
42.4
42.2
42.0
41.8
41.6
41.4
41.2
−73.0
−72.5
−72.0
−71.5
−71.0
−70.5
−70.0


Post911: Weapons Sales Kmeans
Weapons
Weapons
Tactical Equipment
Weapons
Weapons
Weapons
42.8
42.6
42.4
42.2
42.0
41.8
41.6
41.4
41.2
−73.0
−72.5
−72.0
−71.5
−71.0
−70.5
−70.0


Pre911: Weapons Sales Kmeans
Weapons
Weapons
Vehicles
45
40
35
30
25
−120
−110
−100
−90
−80
−70


2010+: Weapons Sales Kmeans
45
40
35
30
25
−120
−110
−100
−90
−80
−70
Vehicles
Vehicles
Weapons
Vehicles
Vehicles
Vehicles
Vehicles
Vehicles
Vehicles