

Team BARI  
Brooks, E., Fuss, H., McKenzie, J.  
Final Report  
August 12, 2015

## **Problem Statement**

With our ever-growing amount of real-time data (i.e. tweets, Facebook posts, phone calls, text messages, credit cards, etc.), there is a correlating demand to extract useful information from it. Recently, improvements in statistical and computational algorithms have allowed for such data to provide new insights to a variety of fields previously considered qualitative.<sup>1</sup> For instance, Bollen et al. were able to implement sentiment tracking and machine learning techniques to predict the daily changes in the Dow Jones Industrial Average based on the general public's Twitter data.<sup>2</sup> King et al. used social media data to reverse-engineer the Chinese government's censorship of potential collective action.<sup>3</sup> Through the lens of big data, we can even observe how the constituent behavior of a region may be directly affected by an unexpected emergency or tragedy.

Two years ago on April 15th, 2013, the Boston area experienced a terrorist attack which resulted in a feeling of panic and terror throughout the city and its surrounding areas. On what is called Marathon Monday, a special day for the locals, two pressure cooker bombs exploded at the finish line of the Boston Marathon on Boylston Street, killing 3 people and injuring over 200 others. In the days following the attack, President Obama issued a state of emergency in Massachusetts while the region united behind the phrase "Boston Strong" and witnessed a manhunt which nearly created as much fear as the bombing itself.

Three days after the attack, MIT patrol officer Sean Collier was shot and killed late at night by the two suspects of the bombing, brothers Tamerlan and Dzhokhar Tsarnaev. Shortly after, an armed carjacking led police to engage with the suspects in a firefight in nearby Watertown during the early morning of April 19th. In the frenzy, Tamerlan had been killed while his injured brother had escaped. For the remainder of the day, the greater Boston area was on lockdown while the manhunt continued with door-to-door searches. Later that evening, shortly after the lockdown had been lifted, a local Watertown resident noticed his boat had been tampered with and alerted authorities. This allowed law enforcement to apprehend the wounded Dzhokhar Tsarnaev.

The research question we are working on, proposed by the Boston Area Research Initiative, is focused on this tragic event which forever changed greater Boston. Our goal is to use a collection of real-time, geocoded data to determine the immediate impact of the attack and the manhunt on daily life in Boston and the surrounding region. This is a multi-faceted project with several detailed datasets to analyze. The datasets include records of 911 dispatches from January to June of 2012 and 2013 for the city of Boston, as well as the city's 311 (non-emergency) reports from the years 2010 to 2014. Additionally, we have been given geocoded tweets from April 2nd through 9th, and 12th through 22nd for the year 2013. Using an area we have defined as greater Boston, which includes the entire the Boston Marathon route, we can "clean" the geocoded tweets of any which lie outside this region. Lastly, we have received BARI's collection of census data for the city of Boston, which includes demographic, socioeconomic, and geographic information that will provide further insight into the above datasets. From this pool of data arises several important, primary goals.

The first goal of our research is to track the patterns of information dissemination throughout greater Boston in real-time via the geocoded Twitter data. Secondly, it will be necessary to develop and visualize the flow of sentiments throughout the region the days following the bombing and during the manhunt. Next, the patterns of crime, disorder, and public awareness throughout the city will be mapped with respect to time in order to determine how

daily life may have changed during the crises. This will provide a better understanding of the changes in the city's constituent behavior and allow for us to determine if there are any correlations between a neighborhood's reactions and demographic/socioeconomic conditions. By accomplishing these goals, we will be to develop a new perspective of the crises' effects on the lives of individuals living in the area.

## Methods

For this project we chose to conduct our analyses by using existing Python modules and libraries to write our own modules and scripts which could extract useful information from the given datasets, which were in the form of comma-separated values (csv) files. With the census tract shapefiles we were given, we developed a flexible set of classes that allowed us to easily plot our geocoded data over Boston and its surrounding regions. In order to conduct meaningful analysis of our tweets, we had to first learn more about natural language processing (NLP) and machine learning by watching lectures from Stanford's Coursera course on NLP. After doing so, we developed several text classifiers by writing wrapper classes for a popular Python machine learning library, Scikit-learn.<sup>4</sup> Lastly, we had to train and test our text classifiers by hand-classifying 3600 tweets under the categories we established. Our code and the data we produced for this project can be found at the following GitHub repository:

[https://github.com/brix4dayz/bari\\_reu\\_2015](https://github.com/brix4dayz/bari_reu_2015)

## 911/311 Analyses

Over the past several weeks, we processed the given datasets through the use of Python scripts that parsed the datasets and conducted basic analysis with respect to the varying time scales. As previously stated, the 311 reports are from the years 2010 to 2014, while the 911 dispatches are from 2012 and 2013. Initial analysis included limiting the 311 reports to 2012 and 2013, so that the analysis was comparable to the 911 dispatches. For these datasets, the dates we were particularly interested in investigating were from April 12<sup>th</sup> through 22<sup>nd</sup> for both 2013 and the frame-shifted 2012, so that the days of week matched, for comparison of a "normal" Boston Marathon (control) to the Boston Marathon bombing.

In order to better understand the response of the public to the events of the marathon bombing, we have organized the 911 and 311 data into categories for more detailed analysis. This was achieved by analyzing the latent variables provided in both datasets, which are binary categorizations that describe the nature of the dispatch or report.

The 311 data had the following latent variables associated with each report: public, public denigration, trash, graffiti, housing, private neglect, uncivil use, and big building. Using these latent variables, we have separated the 311 data into three categories: public (public, public denigration, trash, graffiti), private (housing, private neglect, big building), and other. If a report had a '1' in one of the latent variables in a particular category's set of variables, then it was placed under that category. If a report had all '0's for its latent variables it was placed under 'other', which account for nearly half of the 311 reports within the month of April for 2012 and 2013. We also observed a subset of the public category, public denigration, which accounts for graffiti and trash reports and provides some insight into the city's public awareness.

Similarly, the 911 data had the following latent variables: violence, guns, alcohol, social disorder, private neglect, major medical, and youth health. Using the same procedure as we did for the 311 reports, we broke the 911 dispatches into four categories: medical (major medical, youth health), violent (violence, guns), petty (social disorder, private neglect, alcohol), and unsorted. For the month of April in 2012 and 2013, more than 75% of the dispatches fell under the 'unsorted' category. However, by categorizing the datasets into category, we could better

understand how the city's patterns of crime, disorder, and public awareness were influenced during the week of the bombing.

With our map-making module, we were able to plot the density of 311 reports and 911 dispatches for the census tracts of the city of Boston. The density was determined by normalizing the number of reports/dispatches in a given census tracts by its population, so that area with large amounts of activity could be identified. Additionally, we felt it would useful to identify census tracts which although were not dense with activity, had an unusual amount for relative to its average.

Using the map data frame within our map-maker, we calculated each census tract's daily mean, median, and standard deviation of the number of reports/dispatches for each category with all the data we had for 2013. We noticed for every census tract, the median was less than the mean, which was less than the standard deviation. Therefore, we determined a Poisson distribution would be most appropriate for describing our data. We then used the following formula to calculate the Poisson probability,  $P$ , for a given number of reports/dispatches  $n$  for a particular day in census tract  $c$ :

$$P(n) = \frac{\mu_c^n \cdot e^{-\mu_c}}{n!} \quad (1)$$

Where  $\mu_c$  is daily mean number of reports/dispatches for census tract  $c$ . We then used a probability threshold of less than 5%, and determined which census tracts had an "unusual" number of reports/dispatches for each day within our 10-day period.

### Twitter Analysis

Preliminary analysis of the Twitter data included determining a dictionary of keywords for tweets referencing events of the Boston marathon bombing and manhunt; this will be referred to as relevance. We then familiarized ourselves with natural language processing, sentiment analysis, and basic machine learning techniques for text classification. This allowed us to develop a flexible Python module for tweet classification, given a text file of training data for each possible category or class. These programmed tweet text classifiers have been trained and tested for tweet relevance classification as well as sentiment classification. Sentiment classification allowed us to determine the overall emotion of a given tweet.

Text classification was initially performed using a dictionary of keywords related to the bombing and manhunt, in order to categorize the given Twitter data by relevance to the crises. The dictionary was then tuned by plotting the number relevant tweets as a function of day, where the number of relevant tweets are maximized for days of major events (i.e. bombing, manhunt) and minimized for days preceding the bombing and following the manhunt. Using the developed keyword dictionary and a list the most active and retweeted Twitter users during the crises, we have tracked information dissemination geographically in real-time to observe how predictable the pathways through which the information traveled were.

A second, more sophisticated form of text classification was implemented next to determine the relevance and sentiment of a tweet using machine learning algorithms and a hand-labeled training set we developed. Several classifiers were implemented and trained using Python's Scikit-learn library.<sup>4</sup> We programmed wrappers for traditional multinomial Naive Bayes, support vector machine (SVM) classifiers (which differed in how they determined their margins), logistic regression, maximum entropy, and Bernoulli Naive Bayes. By doing so, it was possible to perform refined sentiment analysis for tracking the emotional response of the constituents of greater Boston.

We first trained our different relevance classifiers by hand-labeling ~1,200 tweets as relevant or irrelevant, and then the sentiment classifiers with ~3,600 tweets labeled as either positive, negative, or neutral. We found approximately 10% of the 1200 tweets were relevant, while roughly one-third of the 3,600 tweets were of each sentiment. The tweets were then “cleaned” of any markup which could have been considered as features by our classifiers. Secondly, the tweets for the sentiment training sets also had any Twitter handles (usernames) replaced with ‘<HANDLE>’, as well as ‘NOT\_’ appended to the front of every word in between a negation (not, n’t) and punctuation.

We then optimized each classifier by training them on a random, large subset of the hand-labeled tweet corpus and testing them on the tweets remaining in the hand-labeled corpus. The relevance classifiers were tested on a development set of 90 irrelevant tweets and 10 relevant tweets, while the sentiment classifiers were tested on set containing 100 tweets for each sentiment. Using a confusion matrix to measure the recall, precision, balanced-F1 (combination of recall and precision), and accuracy we could determine which type of classifier to employ for our different analyses. This is because the smoothing required for the optimized classifier depends on the form of analysis, relevance or sentiment.

After optimization, the best relevance classifier, a quadratic SVM, was roughly 91% accurate with an F1 score of 0.9 (closer to 1.0 the better). However, the best sentiment classifier, a perceptron SVM, was only 59% with an F1 score of about 0.6. We believe this was due to an inconsistency in how we labeled the tweets between the three of us. We each classified a small set of tweets that the other two of us had classified, and determined how much we disagreed. Originally, we classified sentiment tweets for a broader range of emotions (fearful, angry, excited, calm, sad, positive, negative, neutral), however our classifiers were only 40% accurate and we found each disagreed with one another approximately 60% of the time when labeling tweets. We then combined our original sentiment sets into just positive, negative, and neutral and found we improved our classifiers and only disagreed on roughly 20% of the tweets. We also tried training and testing the sentiment classifiers on sets made from just one of us, to hopefully remove discrepancy, however we did not see an improvement in our classifiers. For our Scikit-learn wrapper classes, see the *posnegclassifier* and *relevanceclassifier* modules within the GitHub repository which are documented and commented thoroughly.

Using our relevance and sentiment classifiers as well as the map-making modules, we then attempted to observe how sentiments and information disseminated by the Boston public. In order to represent these movements, we created time-lapse plots of the resulting classification and overlaid them on a map of “greater Boston”, our defined section of the city of Boston and its surrounding region. One limitation of this analysis was the fact that the geocoded tweets only accounted for a small percentage of all the tweets that were posted around the time of the events.

## Map Making

In order to visualize our data geographically, we needed to implement modules which could render maps of the city of Boston, as well as the greater Boston region that we defined. The modules needed to plot data which was geocoded either with longitude and latitude coordinates or census tract identification (ID) number. For the datasets, which were geocoded with longitude and latitude, it was possible to render both scatter and census tract density (number of occurrences per 1000 people) plots over the maps; however, the data geocoded with census tract ID’s could only be plotted by census tract density.

The given BARI census data, not only contained demographic and socioeconomic data for each census tract within Boston, but also contained shapefiles for the boundaries of Boston’s census tracts, block groups, and blocks. Shapefiles use a geospatial vector data format which can be used by geographic information system (GIS) software to visualize maps. Additionally,

shapefiles contain a list of irregular polygons and their coordinates, which make up the boundaries of the areas within the map. Shapefiles may also include other information, such as ID's for counties and census tracts, population, and total area. However, the datasets differed slightly in the areas that they covered and how they were geocoded.

The Twitter data was geocoded with longitude and latitude coordinates and is confined to an area we defined using Google Maps, which encompasses Boston, route 128, and the entire Boston Marathon route, going from Interstate 495 to the coastline. Specifically we used the following latitude and longitude coordinates as the lower left and upper right bounds respectively: (42.1575 ° N, 71.5688 ° W), (42.6296 ° N, 70.6616 ° W).

For plotting the entire Twitter data region, a shapefile was required that encompassed the state of Massachusetts, which we could then use to plot our defined area. The United States Census has a variety of census tract shapefiles for each state. For our Twitter data we used Cartographic Boundary shapefiles<sup>5</sup> which fit to the general shape of the coastlines and could be implemented in conjunction with *mpl\_toolkits.basemap*. Using these shapefiles, it was possible to render several scatter and density plots of the Twitter data for the region of greater Boston with respect to time, allowing us to track how information spread following the bombing.

However, the 911 and 311 datasets were limited to the city of Boston and so we used the provided census shapefiles from BARI. The 311 reports had longitude and latitude coordinates; yet, the 911 reports only included ID numbers which describe the block, block group, and census tract they belong to. Due to these differences, we have found it necessary to develop several distinct map modules which allows us to these datasets over their respective regions given the geocoded information.

Python's *mpl\_toolkits.basemap* module provided a simple interface for reading and rendering such shapefiles. With the combination of *mpl\_toolkits.basemap* and *pandas*, a Python data frame module, and the help of online tutorial for mapmaking in Python<sup>6</sup> we were able to quickly render scatter and density plots of Boston for our datasets. In short, there were four classes from our *bostonmap* module which we ended up using to make our plots: BostonDensity (311), BostonDensityCT (911), GreaterBostonScatter (Twitter), and ColoredGBScatter (Twitter). However, there are several others we implemented which others may find useful. The module which can be found on the GitHub repository is documented with an overview of how the module and each class works, as well as being commented thoroughly for those interested.

## Results

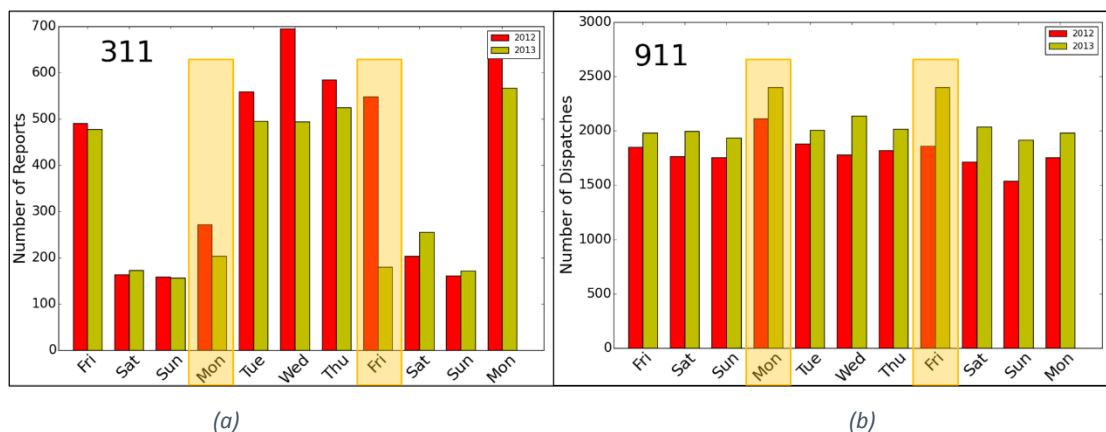


Figure 1. Number of 311 reports (a) and 911 dispatches (b) for each day in our 10-day period for 2012 (red) and 2013 (yellow). The days of the marathon bombing (Mon) and manhunt (Fri) are highlighted in orange.

## 911/311 Analyses

Upon breaking the 311 and 911 data into the categories we previously described, we initially plotted the number of each type of report/dispatch for each day within our frame-shifted (days of the week align) 10-day period of interest, from April 12<sup>th</sup> to 22<sup>nd</sup>, comparing 2013 to 2012. However, the general trends we saw can be demonstrated by just showing the total number of report/dispatches for each day (Fig. 1).

In Fig. 1a, the total number of 311 reports for each day in 2012 (red) and 2013 (yellow) is shown, with the days of the bombing and manhunt highlighted in orange, and it can be seen that we only saw a significant change in 311 activity on the day of the manhunt. This response was particularly strong in both the 'public' and 'public denigration' categories, demonstrating the decrease in public awareness due to the lockdown (see Appendix A, Fig. A1). However, this response was also evident in the 'private' category indicating the public's lack of concern for their own personal space.

In Fig. 1b, it can be seen that there was not a significant difference between 2012 and 2013 in the number of 911 dispatches within our 10-day period (see Appendix A, Fig. A2). Although, the 911 dispatches included a time with the date, and so we broke down the days of the marathon bombing and manhunt by hour in an attempt any sort of immediate effect on 911 activity due to the events.

In Fig. 2, we see the cumulative number of 911 dispatches on the day of the marathon bombing as a function of hour, with 3:00 PM, ten minutes after the bombing occurred, highlighted in orange. Note, 'medical', 'petty', and 'violent' were all normalized so that their axes were of the same scale, but 'unsorted' was not. In all four categories we see the cumulative number of dispatches for both years increase at similar rates; however, at 3:00 PM there is a

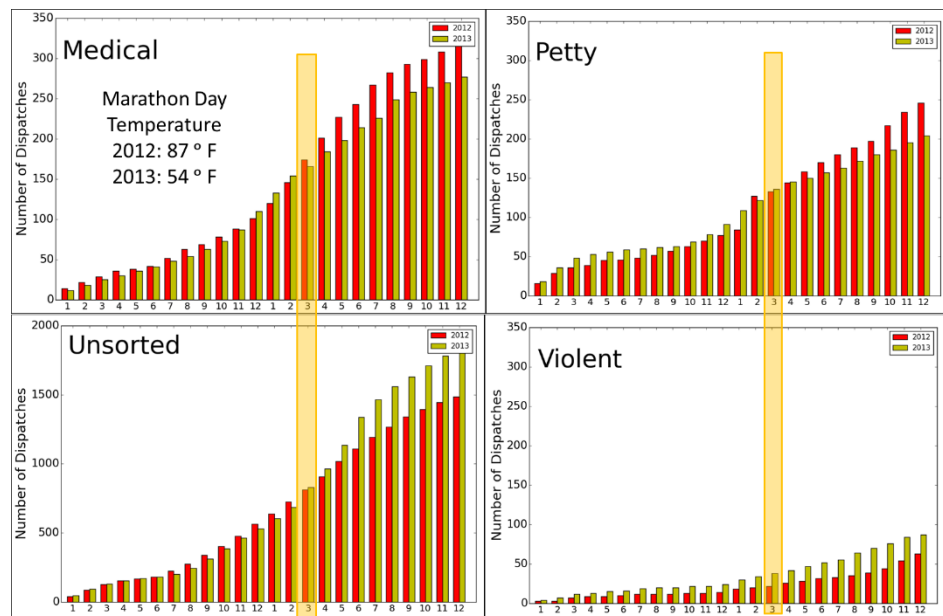


Figure 2. The cumulative number of 911 reports by hour for the day of the marathon bombing for each category. 3:00 PM, the hour closest to the bombing, is highlighted in orange.

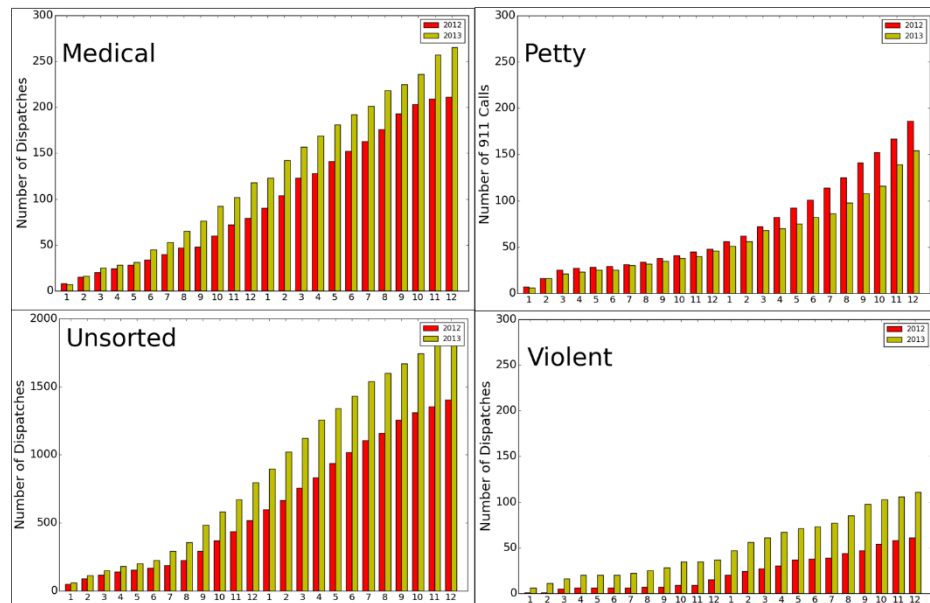


Figure 3. The cumulative number of 911 reports by hour for the day of the manhunt for each category. 3:00 PM, the hour closest to the bombing, is highlighted in orange.



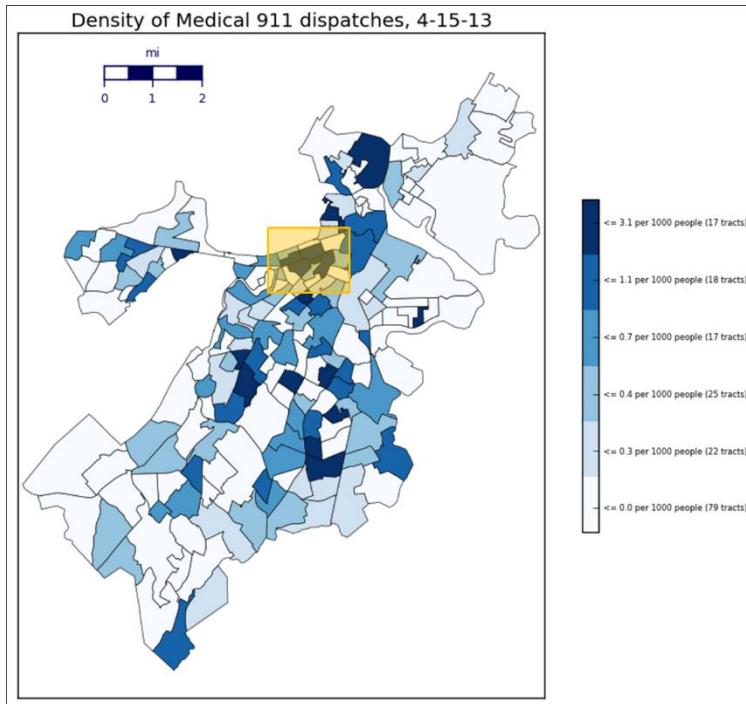


Figure 4. The density of 911 medical dispatches for each census tract in Boston on the day of the marathon bombing. The area where the bombing occurred is highlighted in orange.

response. The ‘unsorted’ and ‘violent’ number of dispatches increases more rapidly in 2013 following the bombings, while the ‘medical’ and ‘petty’ dispatches increases more rapidly in 2012. We note there was a significant difference in temperature between the two Marathon Monday’s, which could account for the change in ‘medical’ dispatches as the day approached the hottest part of the afternoon. It also important to note, that we saw a large number of ‘bomb threat’ dispatches within the ‘unsorted’ category for 2013 on this day, which does not come as a surprise and most likely attributes to the increase in ‘unsorted’ dispatches.

In Fig. 3, we show plots similar to Fig. 2, however for the day of the manhunt. On the day of the manhunt, we see both ‘unsorted’ and ‘medical’ dispatches “jump” in 2013 around 7:00 AM. As with the day of the bombing, the ‘violent’ dispatches showed a similar to trend the ‘unsorted’ dispatches, but sees its “jump” later in the day around noon. However, ‘petty’ followed the same

trend it did on the day of the bombing, with the number of reports in 2012 increasing more rapidly but doing so around 1:00 PM instead.

Lastly, we made density (number of reports/dispatches per 1000 people) plots for our 311 and 911 categories over the census tracts of Boston for each day in our period of interest for 2013. For example, we the density of ‘medical’ dispatches for the day of the bombing, with the area where the bombing occurred highlighted in orange (Fig. 4). It can be seen that this area was one of the densest that day, however there were other census tracts throughout the city which also saw a large amount of activity. Although we can visualize where there were responses, it is more difficult to determine if there is any correlation between areas that responded. Going forward, we hope to plot census tract density for a particular day as a function of different demographic and econometric variables we have within our census data, in order to determine correlations between an areas which seemed to have a response.

## Twitter Analysis

After developing our keyword dictionary (see Appendix B, Fig. B1) we applied the keyword dictionary to the given tweets to find the keyword-relevant tweets, which we then used to determine the most unique “tweeted-at” relevant Twitter users within our 10-day period (Fig. B2). This provides us with users, such as city institutions and news organizations, through which we could simply track information dissemination.

Once we trained and optimized our relevance classifier, we determined the frequency of relevant geocoded tweets in the morning (AM) and evening (PM) for our 10-day period, April 12<sup>th</sup> through 22<sup>nd</sup> of 2013, the week of the Boston Marathon bombing, using both our keyword dictionary and the relevance classifier. We also did so for the previous week, April 2<sup>nd</sup> through 9<sup>th</sup> of 2013, and compared the performance of the two methods for both weeks (Fig. 5, note difference in scales). Theoretically, there should not be any relevant tweets during the previous

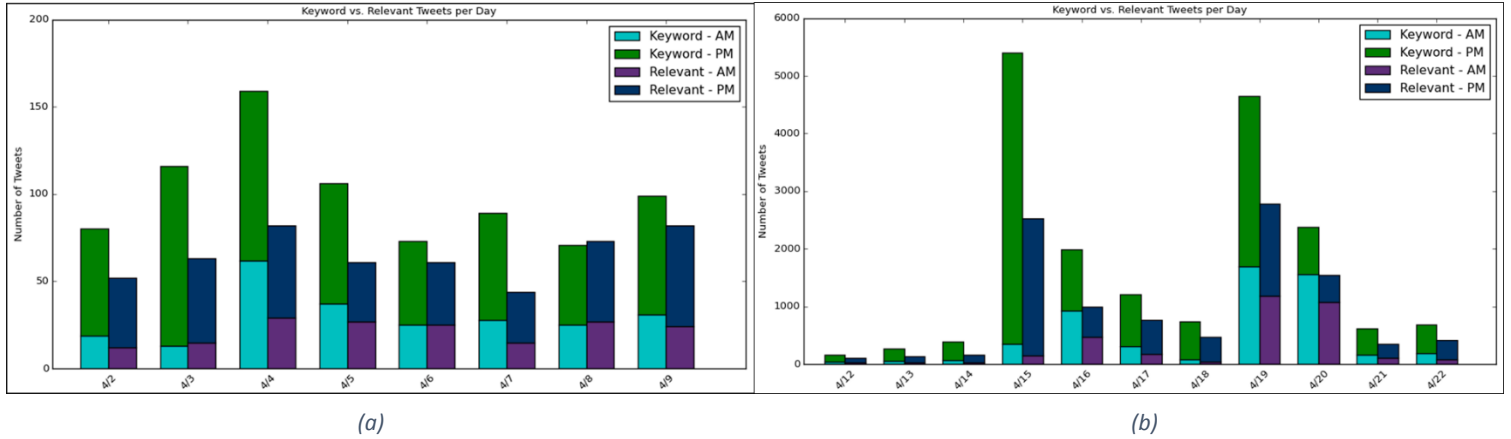


Figure 5. The number of keyword-relevant (green/cyan) tweets vs. the number of relevant-classified (blue/purple) tweets each day for the previous week (a) and the week of the marathon bombing (b).

week (Fig. 5a), as the bombing and manhunt had not yet occurred; tweets marked as relevant on the days before the significant events occurred shall be referred to as outliers. This analysis gives an idea as to the number of false positives for the two developed relevance analysis methods.

Through this comparison we found that the relevance classifier performed fairly well, finding only a limited number of outliers in the previous week and reporting significant peaks on the day of the bombing and manhunt in our 10-day period (Fig. 5b). On the other hand, our keyword dictionary provided us with a larger number of relevant tweets but it discovered a much greater number of outliers. Considering our relevance classifier had an accuracy on 90% on our development set and did better at recognizing irrelevant tweets as opposed to relevant tweets, the actual number of relevant tweets for each day after the bombing was probably somewhere in between the two methods.

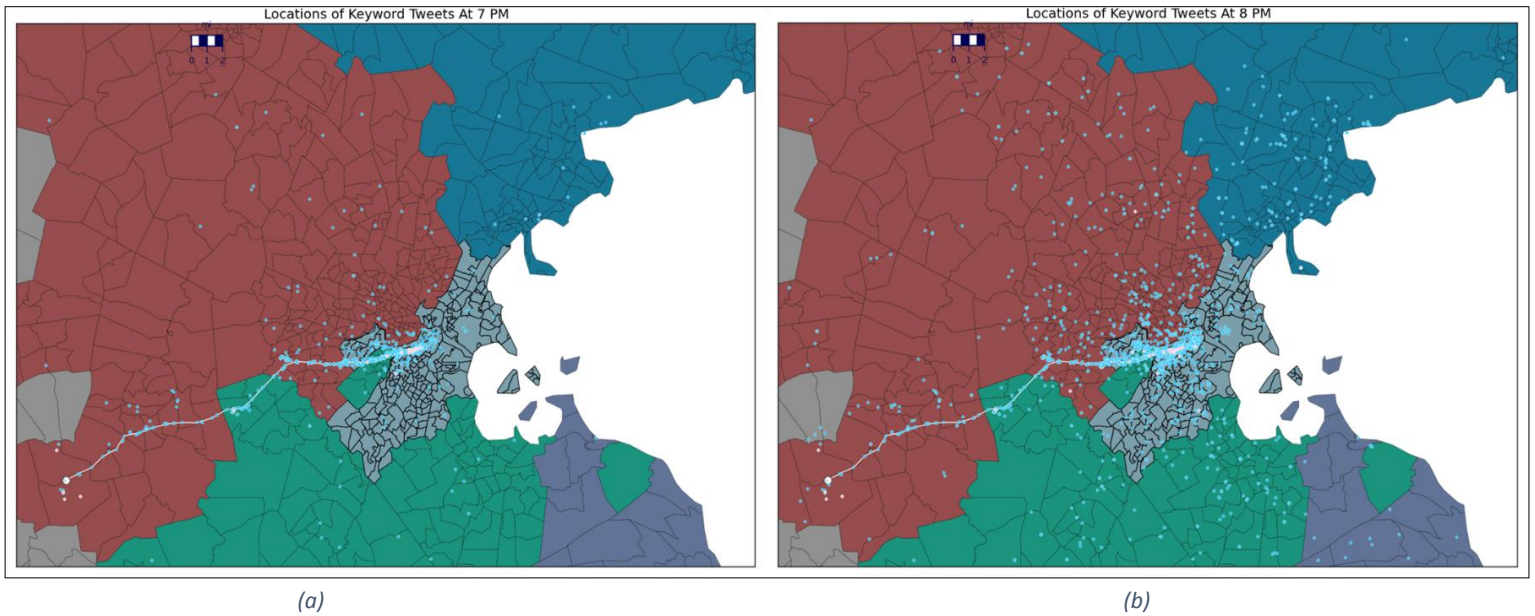


Figure 6. The location of keyword-relevant tweets (blue) vs. informative tweets (light red) on the day of the marathon bombing at 7:00 and 8:00 PM demonstrating a spike in relevant Twitter activity.



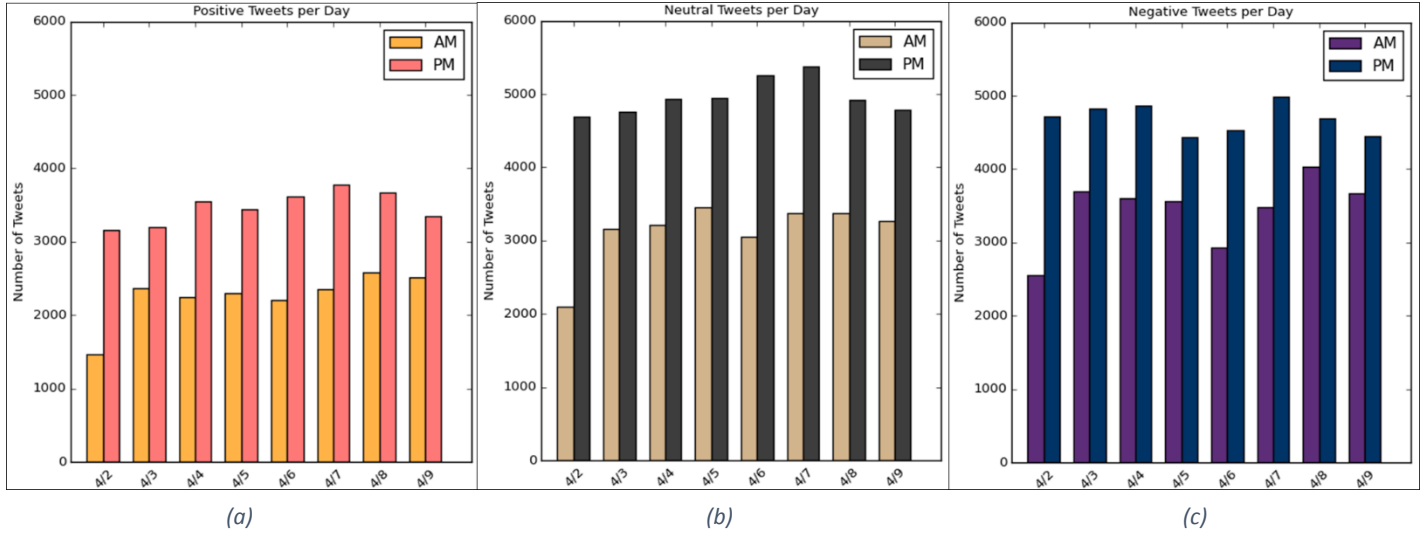


Figure 7. The number of positive (pink/orange), neutral (grey/beige), and negative (blue/purple) tweets for each day during the previous week.

We broke down the day of the bombing by hour, and plotted the location of keyword-relevant tweets, as well as relevant-classified tweets, over a map of greater Boston (see Appendix B for example of greater Boston, Fig. B3) which provided us with a rough sketch of how news about the bombing spread throughout the region. Improving upon this, we then plotted the location of relevant tweets versus the location of relevant tweets containing one of most “tweeted-at” Twitter usernames, what will be referred to as informative tweets, by hour for the day of the bombing to see if we could not better visualize how information may have traveled through the social network.

In Fig. 6, we see the locations of all of the keyword-relevant tweets (blue) versus informative tweets (light red) for the day of the bombing at the hours of 7:00 and 8:00 PM (Fig. 6a and 6b respectively). The marathon route can be seen in white, with Suffolk county (Boston) being the grey region with thicker borders. Up until 8:00 PM, relevant tweets seem to spread along geographical pathways, however as shown in Fig. 6, there is a significant increase in the

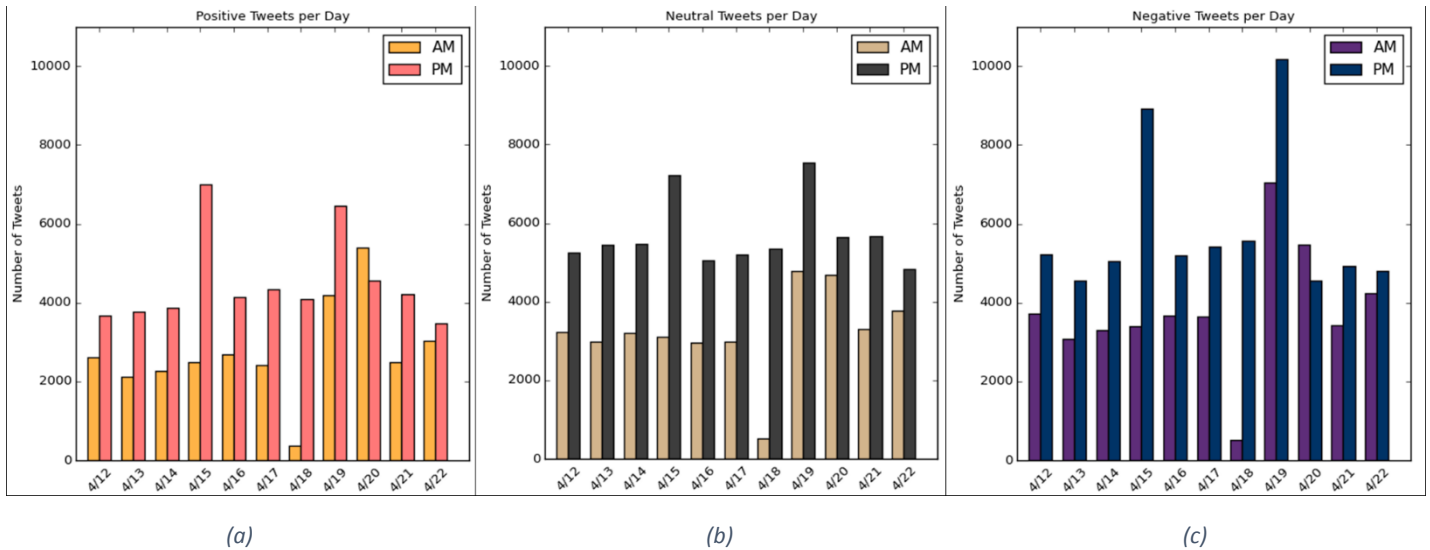


Figure 8. The number of positive (pink/orange), neutral (grey/beige), and negative (blue/purple) tweets for each day during the week of the marathon bombing.

number of relevant tweets throughout the whole region. We made similar plots for other days, and did not see this spike in activity between these two hours. However, there is not a large increase in the number informative tweets between these two hours, and so we believe this spike may have been result of people getting home at the end of the day and the evening news, and not necessarily due to information being spread through Twitter. Although it is important to keep mind that our geocoded tweets only account for a small percentage for the total tweets that day.

Upon training and optimizing our sentiment classifier we made plots similar to ones we made for determining tweet relevance. We ran our classifier on both the previous week and our 10-day period to observe the number of positive (pink/orange), negative (blue/purple), and neutral (grey/beige) tweets for typical week (Fig. 7) in comparison to the week of the marathon bombing (Fig 8).

In Fig. 7, we see that the number of tweets for each sentiment remains relatively constant throughout the entire week. While in Fig. 8, we see large spikes of Twitter activity on the day of the bombing and manhunt, with each sentiment having significant peaks in the evening of those days. Particularly in Fig. 8c, we a large number of negative tweets in both the morning and evening of the day of the manhunt indicating the public's frustration and fear associated with the lockdown. The day following the manhunt, a Saturday, we then see a large number of positive tweets in the morning demonstrating the change in the public's general sentiment with the manhunt finally ending and the region beginning to return to normalcy. Note, each sentiment had a small number of tweets on the morning of the 18<sup>th</sup> which we were unsure as to why. That day had a particularly small number of geocoded tweets overall and we speculate it could be because location services on Twitter were down that morning.

Lastly, we plotted the locations of the sentiment tweets over the greater Boston region by hour for both the manhunt and the previous Friday, April 12<sup>th</sup> (Fig. 9). The negative tweets (blue) dominated the positive (orange) and neutral (green) tweets overall, as was shown in Fig. 8. By what was an indicator that our classifier was doing a decent job despite its low accuracy was shown by the regions highlighted in orange, meant to show the Watertown area. Watertown was where the firefight took place early in the morning on the day of the manhunt, causing the

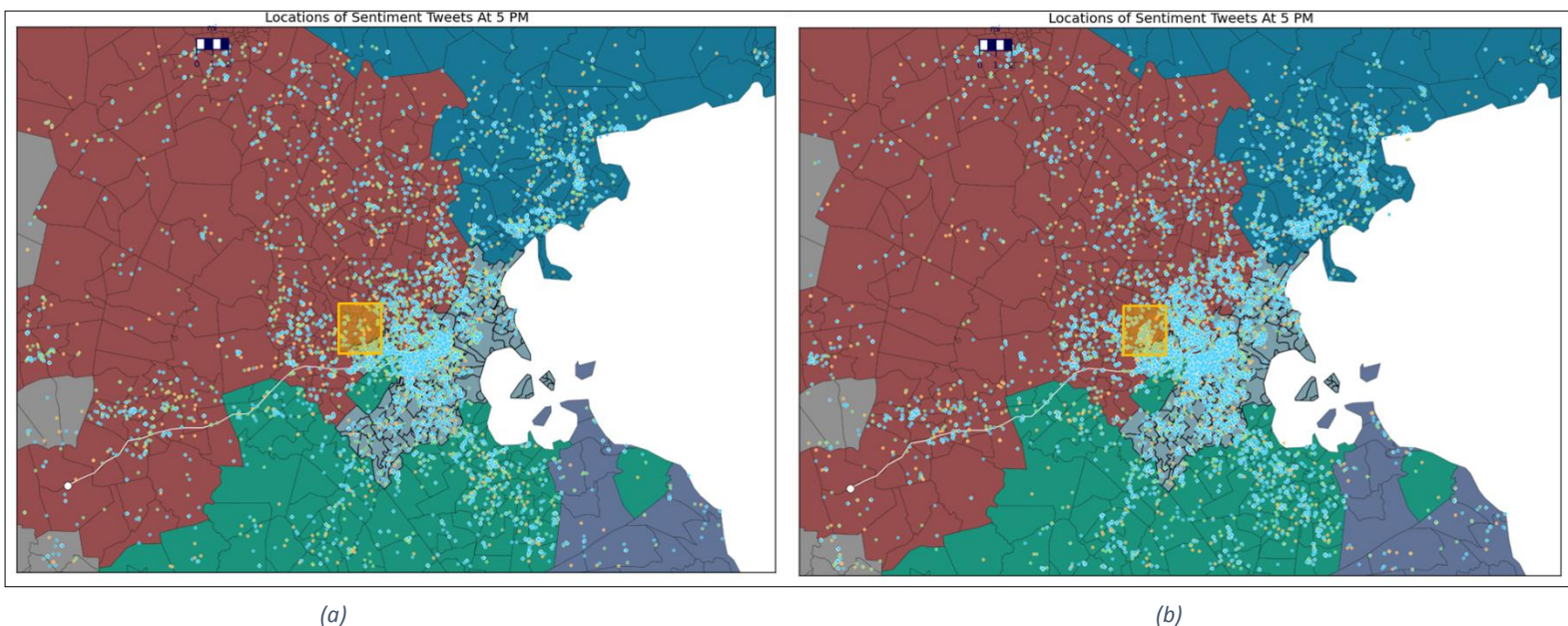


Figure 9. The locations of positive (orange), neutral (green), and negative (blue) tweets at 5:00 PM on a "typical" Friday, April 12<sup>th</sup>, (a) and the day of the manhunt, April 19<sup>th</sup> (b).

lockdown. Fig. 9 shows a large concentration of negative tweets in that area on the day of the manhunt, but a dilute number of tweets on the previous Friday which intuitively makes sense. However, despite this our sentiment classifier has room for improvement and can be applied further to for a more in-depth analysis of the flow of sentiments during the week of the marathon bombing.

## Suggestions for Future Work

Although we took a deep dive into the datasets we were given, we felt there was still a lot that could be done in order to find more meaningful results. For the 911/311 analyses, we discussed how the census tract density required further attention. In our Methods section we also mentioned how we were able to measure the Poisson probability of a census tract having a certain number of reports/dispatches for a certain day. We think a good place to start for someone continuing this project would be to plot the census tract density and Poisson probabilities as functions of the different econometric and demographic variables given in census dataset, such as proportion of census tract's population that is a certain race or ethnicity, or the Z-score for one of the 911 latent variables. We speculate that a typical day will not have clear correlation between the Poisson probability and demographic variable, however on the days of the bombing or manhunt there could be a general trend that could provide some insight to the conditions of the neighborhoods which had significant responses.

For the Twitter analysis, developing a larger, and more consistent training set for the sentiment classifiers could greatly improve their overall performance. Additionally, experimenting with different methods of tokenization and "tweet cleaning" could also optimize the sentiment classifiers' accuracy, precision, and recall. Lastly, trying other sophisticated forms of information and sentiment tracking could reveal more significant results, such as using word networks for information dissemination and semi-supervised learning lexicons for sentiment analysis.

## References

1. Shaw, J. (2014). Why "Big Data" Is a Big Deal. Harvard Magazine.  
<<http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal#article-images>>
2. Bollen, J., Mao, H., Zang, X. (2011). Twitter Mood Predicts the Stock Market. Journal of Computational Science.  
<<http://www.sciencedirect.com/science/article/pii/S187775031100007X>>
3. King, G., Pan, J., Roberts, M. (2014). Reverse-engineering censorship in China: Randomized experimentation and participant observation. Science Magazine.  
<[http://gking.harvard.edu/files/gking/files/experiment\\_0.pdf](http://gking.harvard.edu/files/gking/files/experiment_0.pdf)>
4. Pedregosa, *et al.* (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. <<http://scikit-learn.org>>,  
<<http://jmlr.org/papers/v12/pedregosa11a.html>>
5. United States Census Bureau. (2014). Cartographic Boundary Shapefiles. Geography.  
<[https://www.census.gov/geo/maps-data/data/cbf/cbf\\_tracts.html](https://www.census.gov/geo/maps-data/data/cbf/cbf_tracts.html)>

6. Hegel, S. (2013). So You'd Like To Make a Map Using Python. <[http://sensitivecities.com/so-you-d-like-to-make-a-map-using-python-EN.html#.VZGG\\_VyZ65Y](http://sensitivecities.com/so-you-d-like-to-make-a-map-using-python-EN.html#.VZGG_VyZ65Y)>

## Appendices

### Appendix A: Additional 911/311 Figures

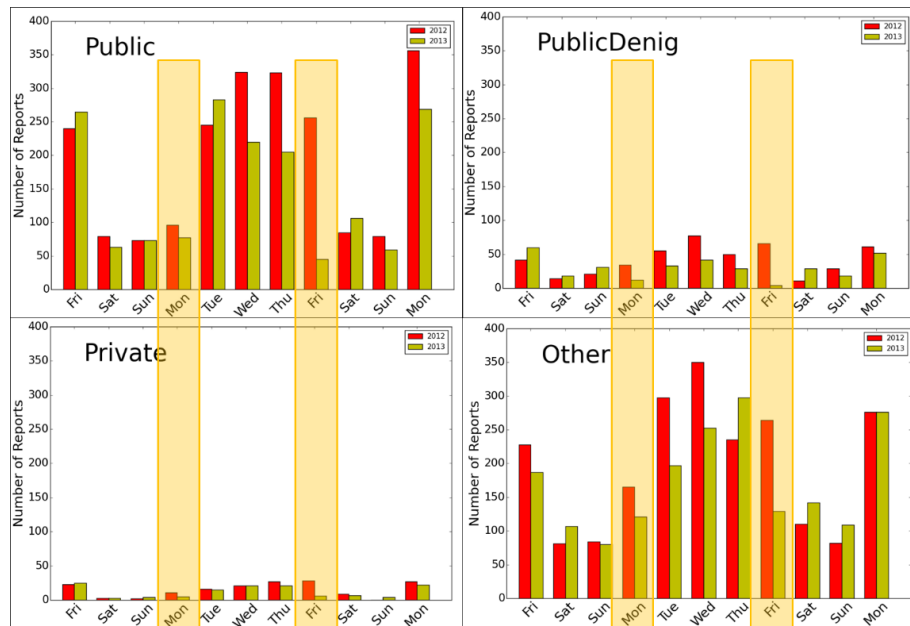


Figure A1. The number of 311 reports each day for each category for the week of the 10-day period. The day of the marathon bombing (Mon) and manhunt (Fri) are highlighted in orange.

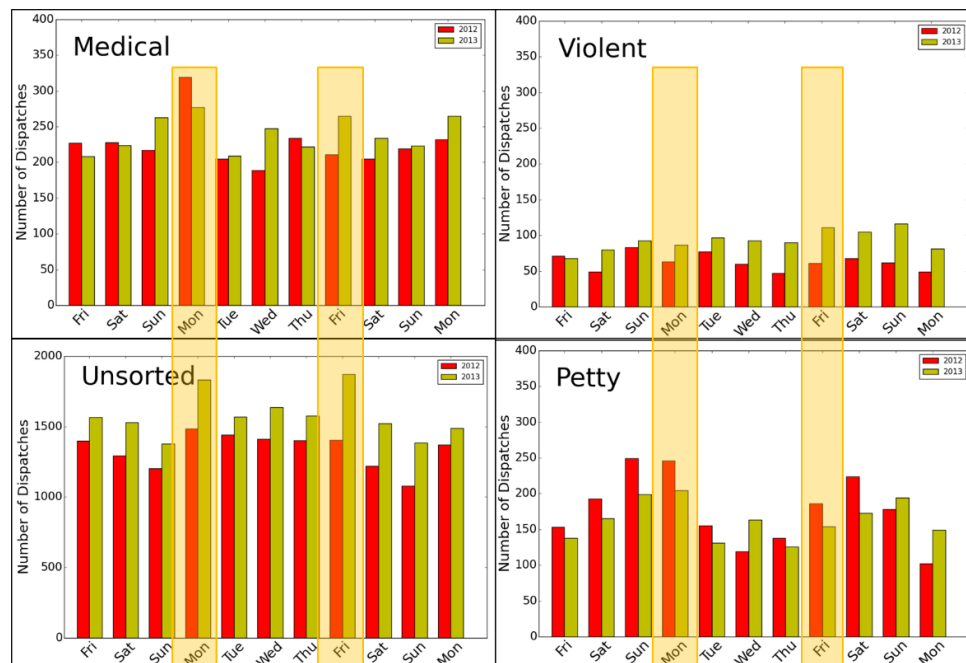


Figure A2. The number of 911 dispatches each day for each category for the week of the 10-day period. The day of the marathon bombing (Mon) and manhunt (Fri) are highlighted in orange.

## Appendix B: Additional Twitter Figures

Frequencies of Keywords		
Keyword	Frequency	# Tweets
marathon	3604	3450
bomb	3558	3406
#bostonstrong	3443	3433
#bostonmarathon	2090	2072
pray	2061	1939
watertown	1923	1830
#prayforboston	1824	1798
explosion	1220	1193
lockdown	860	846
boylston	643	598
finish line	533	528
manhunt	527	518
#marathonmonday	363	363
tsarnaev	259	250
dzhokhar	202	197
muslim	155	142
#bostonpride	112	111
collier	107	105
explode	96	94
#pray4bos	69	69
#patriotsday	55	55
tamerlan	39	37

Figure B1. The keyword dictionary we developed for determining a tweet's relevance to the bombing and/or manhunt, with the frequency occurred and number of tweets used in for each word during our 10-day period.

Top 25 Most 'Tweeted-At' Accounts	
Twitter Handles	Frequency
boston_police	69
bostonglobe	61
bostonmarathon	49
redsox	48
barackobama	37
bostondotcom	30
stoolpresidente	29
nhlbruins	27
mlb	26
cnnbrk	24
ap	23
7news	22
fox25news	21
youranonnews	20
cnn	20
middlebrooks	18
dzhokhar_a	18
wcvb	18
cbsboston	18
universalhub	18
tdgarden	15
massstatepolice	14
j_tsar	12
huffingtonpost	12
wbcsays	11

Figure B2. The 25 most uniquely "tweeted-at" Twitter accounts within keyword relevant tweets for our 10-day period.

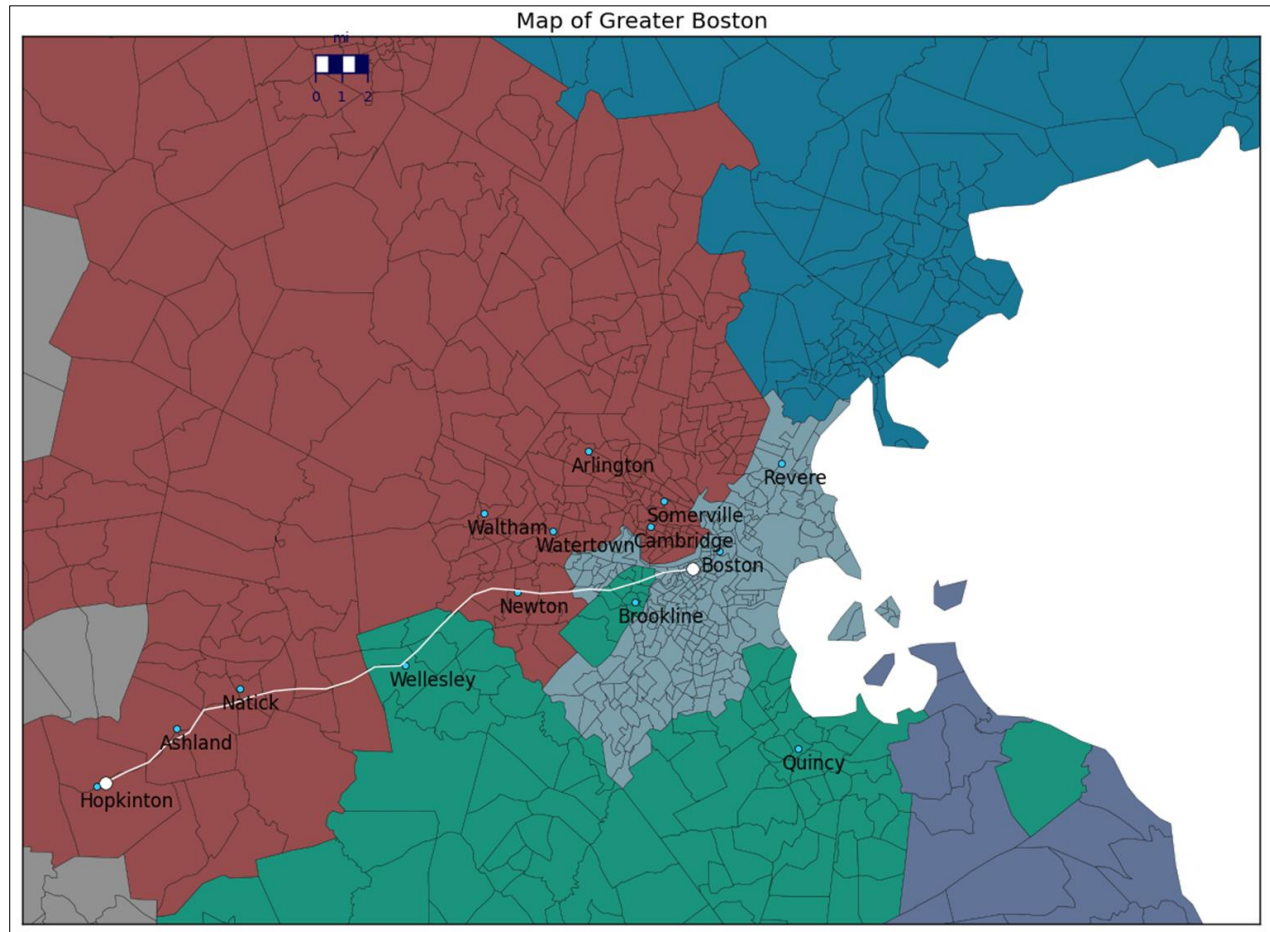


Figure B3. A map of the region we defined as greater Boston with the marathon route in white and the different counties color-coded. We used the following latitude and longitude coordinates as the lower left and upper right bounds respectively: (42.1575 ° N, 71.5688 ° W), (42.6296 ° N, 70.6616 ° W).