ISPARK: Interactive Visual Analytics for Fire Incidents and Station Placement

Subhajit Das, Andrea McCarter, Joe Minieri, Nandita Damaraju, Sriram Padmanabhan, Duen Horng (Polo) Chau
Georgia Tech
Atlanta, GA, USA
{das, andream, jminieri, nandita, sriramp, polo}@gatech.edu

ABSTRACT

In support of helping to reduce the response time of firefighters, and thus deaths, injuries, and property loss due to fires, we introduce ISPARK. The ISPARK system determines where fire stations should be located, analyzes the primary causes of fires, the existing infrastructure, and response times, by using visualizations which show the GIS mapping of fire stations on a dashboard. Incidents and response times are shown as additional layers, with clustering of fire incidents to determine predicted fire station locations, forecasting of fire incidents using regression, causal, infrastructure, and personnel analysis, creating an interactive, multi-faceted method for locating fire stations. A comparison of urban and rural fire incident response times is another dimension of this study. We demonstrate ISPARK's usage and benefits using a publicly available dataset describing 300,000 fire incidents in the states of Massachusetts and Maine. ISPARK is generalizable to other geographic areas and domains, such as police stations, schools, hospitals.

Categories and Subject Descriptors

H.1.2 [**User/Machine Systems**]: Human factors; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

General Terms

Visual Analytics, Data Mining, Human-computer Interaction, Design, Human Factors

Keywords

Fire incidents, GIS, clustering, regression, response time, mapping, NFIRS, FEMA, GeoJSON, leaflet, D3

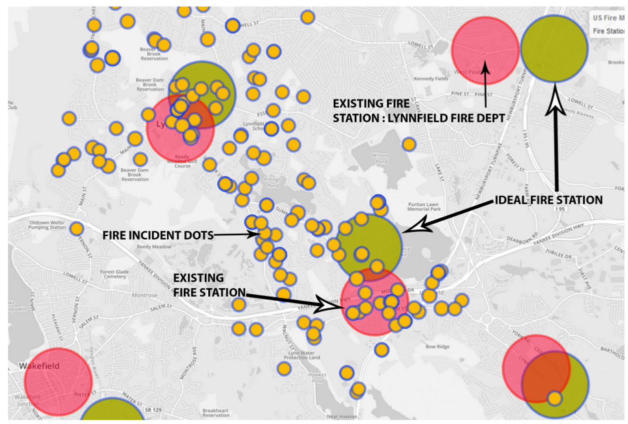

Figure 1: Screenshot of ISPARK showing actual (pink) and predicted (green) fire station locations in Maine determined by our approach, using coordinates with actual driving distances from fire stations to actual fire incidents. Fire incidents are shown as small yellow dots. ISPARK reduces the average driving distance between the fire stations and the fire incidents by about 1/3.

1. INTRODUCTION

In 2013 [14], deaths, injuries, and property losses due to fire were extensive (Figure 2). If the response time can be reduced by just one minute, fewer injuries and deaths should occur, and the cost of reconstruction will be reduced. The goal is to reduce response time for fire stations to aid in a fire, thus reducing injuries, deaths, and property damage from fires. Determining where fire stations should be located to minimize driving distance and response time (Figure 1), analyzing the causes of fires, the existing infrastructure and personnel, and comparison of response times will be beneficial in reaching this goal. There are no federal laws on fire incident response time, but the National Fire Protection Association (NFPA) has detailed standards which most communities use [3]. Response time includes: (1) Dispatch time: 1 min.; (2) Turnout time: 1 min.; (3) Travel time: 4 min.; (4) Setup time: 2 minutes. Since fire grows exponentially in the first 10 min, doubling every second, before flashover, response time is critical.

Past approaches to the problem of fire station location have used GIS [8], optimization, classification, regression, and satellite imagery. GIS, primarily as an historical and descriptive tool, has been used in Oregon [10], Oklahoma City [22], Nevada [9], Moscow [16], and Turkey [18]. Baltimore [7] and the Open Data Institute [12] went beyond

KDD 2015 Workshop on Interactive Data Exploration and Analytics (IDEA'15) August 10th, 2015, Sydney, Australia.

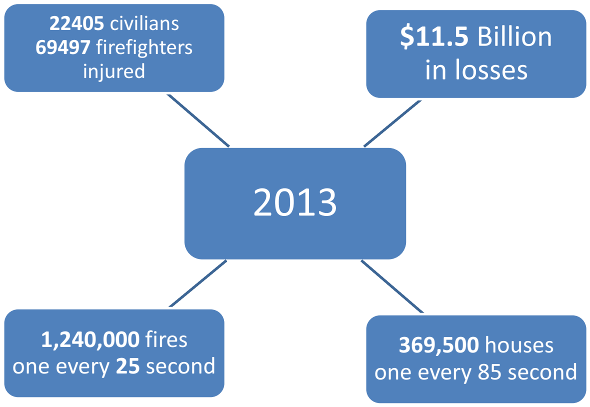

Figure 2: Impact of fire loss for the entire United States in 2013.

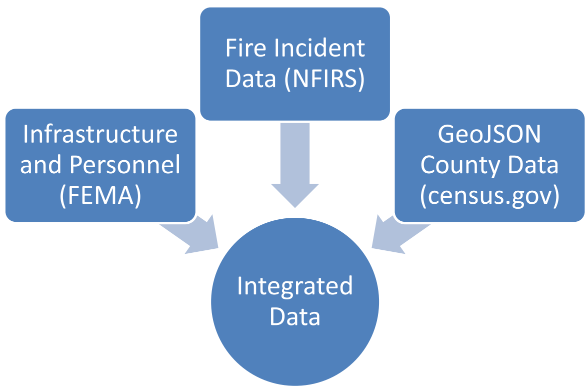

Figure 3: Schematic diagram showing the ISPARK data sources. County and state level detailed infrastructure and personnel data were available from FEMA, fire incident data from NFIRS, and geographic coordinates and population data from the U.S. Census.

these studies to create interactive maps, so the user could click on a fire station, close it down, and see the impact to response times. Malik et al created a system to visualize the impact of closing Coast Guard stations on search and rescue operations [1]. Interactive filtering and linked views were used by Maciejewski et al to visually detect hotspots [19]. Karafyllidis [13], Sen et al [24], and Liu [15] all used optimization combined with GIS to maximize map grid coverage and minimize cost. Sitanggang [20] studied physical data to classify 2693 objects using Naïve Bayes, relating them to fire incidence and location. Hernández-Leal et al [11] developed a fire risk index for forest fires using regression, combining variables such as satellite sensing data. In the city of Boston, Massachusetts (MA) an enterprise GIS system (ESRI based) is available to all staff, but has not yet been applied to locating fire stations, other than showing them on a map [6]. Maine (ME) also has an enterprise GIS system (ESRI) which has not yet been applied to this area [21]. In regard to interactive word clouds, previous work in this area has been done by Viegas [23].

Maine, with a population density of only 43 people per square mile, versus Massachusetts, with a population density of 12,793 people per square mile, were selected [5] so that an urban vs rural comparison could be made. Both MA and ME contribute data to the National Fire Incident Reporting System (NFIRS), a publicly available voluntary database used in our project [17]. One of the members of our project team is a volunteer firefighter in Maine, so his domain knowledge in this area is a significant help to us.

The central theme of our visualization is a GIS-level view of the data, followed by visualizations such as geographic patterns, parallel coordinates for infrastructure, word clouds examining causes, recommended fire station location, response time comparisons in urban and rural areas, and predictions of response times in the future. Previous work lacked interactivity, used expensive tools, was not easily extensible to other areas, and used outdated methods. We introduce ISPARK, which provides the following contributions:

- An interactive, integrated dashboard using open source tools, for implementation at low cost for fire departments across the United States, using ME and MA as the starting points.
- Prediction of the recommended location of fire stations using K-means clustering.
- Prediction of the response times for future years, and comparison of the actual and predicted response times of the firefighters.
- Determination of differences in urban and rural response times.

2. DATA SOURCES AND PREPARATION

Three data sources were used: (1) FEMA, for the infrastructure and personnel; (2) NFIRS for the fire incident data; and (3) the U.S. Census for the GeoJSON county level coordinates and population data (Figure 3). The U.S. Fire Administration collects data via the National Fire Incident Reporting System (NFIRS) system, which is the world's largest national, annual database of fire incident information. NFIRS is a reporting standard that fire departments use to uniformly report on the full range of their activities, from fire to emergency medical services (EMS) to equipment involved in the response. The database comprises about 75 percent of all reported fires that occur annually. Participating fire departments report about 22,000,000 incidents and 1,000,000 fires each year. For this study, approximately 300,000 records and 50 fields were extracted and cleaned for Maine and Massachusetts for 2010 through 2012, which was 500 MB. Two formats were created: csv (comma-separated values) and JSON (JavaScript Object Notation). Most of the data analysis was performed using csv format files, except for shape files and GeoJSON files to overlay shape layers on top of the Leaflet Map.

An additional data source was Microsoft Bing, which was used for geocoding fire station, fire incidents, driving distance (distance from the fire station to the fire incident), and predicted driving duration (time to get from the fire station to the fire incident). The latter was done both for the existing and the recommended fire stations as the originating point. We chose Bing due to their high query threshold, allowing us to issue as many API calls per day as were needed for this project.

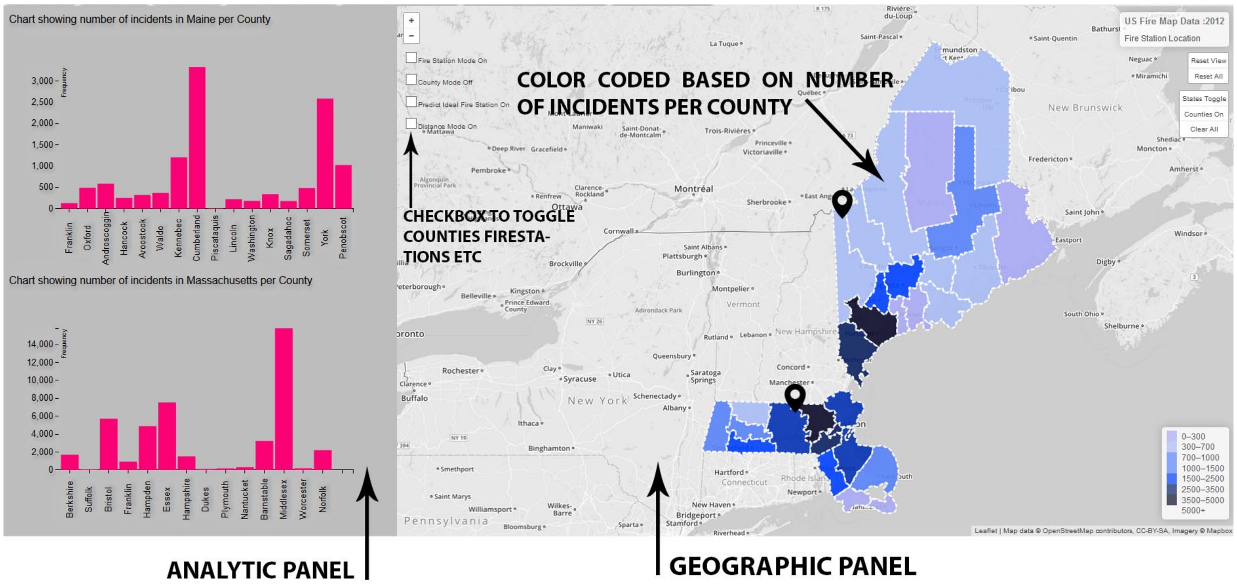

Figure 4: ISPARK's dashboard when opening the application. Left: fire incidents by county in Maine and Massachusetts for 2012 shown as histograms. Right: showing the same fire incident distributions geographically on a map (a darker county means having more fire incidents).

Perl scripts were written to refine the data for Maine and Massachusetts from the NFIRS national database for 2010, 2011, and 2012 in both JSON and csv formats. The latitudes, longitudes, predicted duration and driving distances from the fire station to the fire incident were also obtained from Bing using Perl scripts.

3. ISPARK: DESIGN & CAPABILITIES

3.1 Overview of The ISPARK System

Our approach includes integration of the data, visualization of the data on a dashboard, and various analyses of the data (Figure 4). The dashboard was divided into two sections, an analytics and a geographic panel (Figure 5). The analytics panel, based on D3.js, a JavaScript library for manipulating documents based on data [4], changes depending on the selected mode of the geographic panel. The geographic panel reveals the visual patterns of incidents, fire stations, and recommended fire stations using GIS mapping with Leaflet.js.[2] The analytics information include trends and statistical regressions, infrastructure analytics using D3.js parallel coordinates, K-means clustering to determine the recommended fire station locations, comparison of predicted with actual response times, and word clouds based on the causes of fires.

3.2 Dashboard Summary

The two dashboard panels are: (1) the left analytic panel, in which multiple histograms, word clouds, and other analytics are displayed, including the default 2012 county level data for Maine first, and then Massachusetts underneath; (2) the right geographic panel, which allows the user to zoom to any level desired using the "+" and "-" keys in the upper left section of the panel, with the default display shown as Maine and Massachusetts divided by county and shaded according

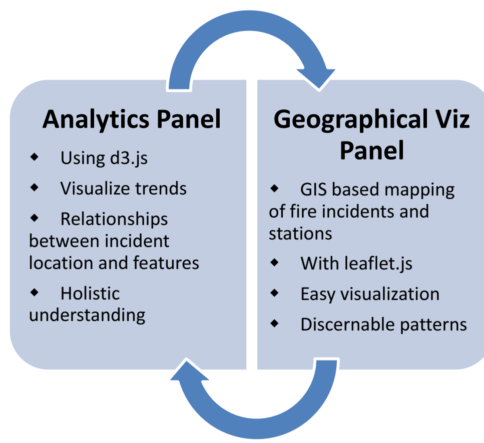

Figure 5: Schematic diagram showing the proposed data visualization toolset UI.

to the 2012 density of fire incidents. The right panel also offers multiple check boxes for "Fire Station Mode", "County Mode Off", "Predict Ideal Fire Station On", and "Distance Mode On". Additionally, prominent teardrop shape markers on each state can be clicked to show state level data. The user can also switch to another year through a menu option offering the years 2012, 2011, or 2010. These features will be described in more depth as we proceed through the design description.

The user will see the opening visualization with the map of the United States on the right, and the defaults shown as indicated (Figure 4). By clicking the buttons for each year, the county incident histograms on the left panel and density of shading in the counties in the right panel will change

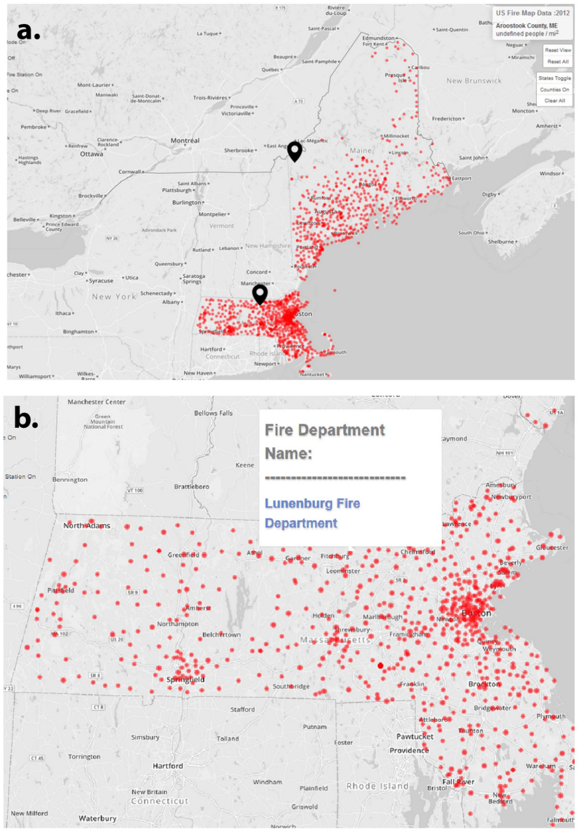

Figure 6: (a) Screenshot showing the fire stations (red dots) mapped over the states of MA and MN; (b) Fire stations zoomed in over the state of MA, and pop-up of fire department name, Lunenberg, when hovering over fire station.

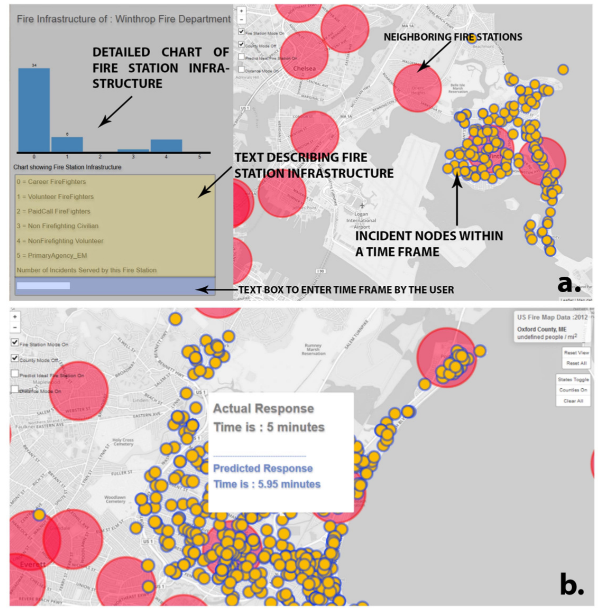

Figure 7: (a) Fire stations (red discs) zoomed in with left panel showing infrastructure for Winthrop fire station being hovered over; (b) Fire incidents (yellow dots) within a certain response time interval (e.g., 5 min) for the Winthrop fire station.

appropriately. Since the fire incidents are too numerous to show individually, the counties are shaded according to the density of fire incidents, thus avoiding overplotting .

One of the options on the right panel offered to the user is to go into the "Fire Station Mode" (Figure 6a). Checking this box on the right panel will add the fire station location overlays on the states of MA and ME. By also checking the "County Mode Off" box, the fire station display will be cleaner looking. By clicking the "+" sign in the upper left of the screen, the user can zoom in as far as desired, and then, by hovering over a fire station node (red), see the name of any of the individual fire stations (Figure 6b). Note that the infrastructure, i.e., the number of various types of staff is shown (career, volunteer, paid per call, and so on) is also shown for that station.

3.3 Mapping Fire Incidents: Techniques & Design

Open source web technologies have been used, with a two panel interface, the right panel containing the Leaflet map with layers, and the left panel containing a JavaScript enabled interface, with multiple D3 visualizations. These two panels are designed to communicate with each other, so that the related charts show up as the user selects various sets of information options on the visualization.

This map showcases incident data from 2010, 2011 and 2012, and provides a basis for future fire station locations for the entire country. The longitude and latitude values are used to retrieve the counties for each incident, and then shade them based on the number of incidents per year. The dashboard includes popups which load up as soon as the user clicks on any of the fire incidents or fire stations, providing more detailed information about the fire department, its capacity, and average response time. For the fire incidents, popups provide the actual response time compared to the predicted response time for that incident.

Multiple modes are provided, including the fire station and the incident modes. The fire station mode allows the user to zoom in to see the locations of all of the fire stations, and to see the number of different types of personnel (professional, volunteer, EMS staff) for each fire station. Selection buttons at the top by year provide the data for those time frames. Markers on top of each state provide various types of state level data when clicked.

3.4 Visualizing & Predicting Response Times

In Figure 7a, one of the fire stations has been selected by the user, and a response interval entered in the left panel, with the fire incidents served by that fire station within the specified time interval (5 min.) shown as yellow nodes. If the user now hovers over an incident, both the actual and the predicted response time will be shown (Figure 7b). Most often, the actual response time will be faster than the predicted response time, as in the example below, since the fire

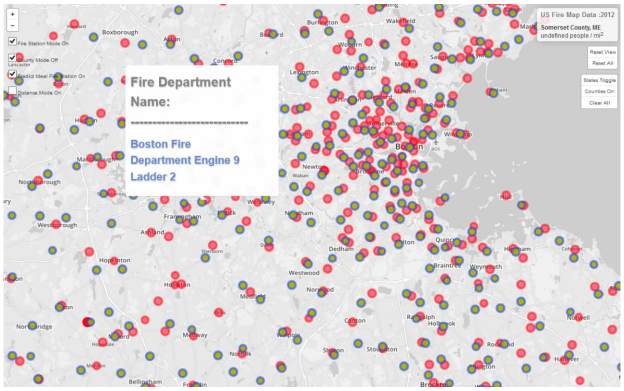

Figure 8: View showing predicted fire station dots in green color against existing fire station dots in red color.

truck can speed to the fire.

Statistical regressions using SAS were used to explore the relationships between variables. The fire location arrival times for each incident were subtracted from the time of the fire alarm. 3-5 percent of the records had either 0, nothing, or were negative, so they were removed from the analysis. Unrealistically long response times (over one hour) were also removed. Unrealistically long distances (over 30 miles) between the fire station and the incidents were also removed. The times and distances retained as realistic were selected based on the domain knowledge of the volunteer firefighter from Maine on our team.

K-means clustering, using Python, was performed on the fire incident data to obtain recommended fire station locations based on the coordinates of fire incident data. K-means clustering partitions n observations (fire incidents) into k clusters (fire stations), with each observation belonging to the cluster with the closest mean. This method provides the centroids of all the k clusters, i.e., predicted locations for fire stations.

Clicking on “Reset all” and checking the box “Predict Ideal Fire Station On” will show all recommended fire stations, as shown in (Figure 8). For ME, the recommended fire stations are shown in green, whereas in MA they are shown as blue. The driving distances for each of the recommended fire station locations to the incidents were calculated, and, on the average, the distance to the incidents was reduced by one-third compared to the actual fire station locations. These recommended locations would result in a significant reduction in response times. However, these locations are not located on existing streets, and expecting local governments to move their fire stations based on this data is unrealistic, which will be addressed in the discussion section.

3.5 Comparing and Predicting Firefighter Response Times across States and Years

By pressing the “Reset All” button on the right panel, the state level option menu is shown (Figure 9), including fire station response times, response times by year, actual versus log response times, driving distance comparisons, and the prediction of response times.

Clicking on the third menu item, “Actual vs Log Response Times” shows that the distribution of response times is a highly skewed distribution (Figure 9a). In order to use linear regression, all response times were converted to the log of response time, which was a more normal distribution. Clicking on the fourth menu item, the “Driving Distance Comparison” option shows that it is consistently about a mile further in ME (3 miles) compared to MA (2 miles) to get to the fire incidents, since the rural areas are more spread out (Figure 9b).

The average actual response times in both MA and ME are shorter than those predicted ($p < .0001$), with MA showing an actual response time about three min faster than the predicted time (Figure 9d). ME also showed faster actual than predicted times, but only by one minute. This is to be expected, given that the fire trucks can go faster than regular vehicles. For ME, increasing their speed is made more difficult by the poor road conditions and often older equipment.

Clicking on the fifth menu item, the “Prediction of Response Times” option shows that the annual predicted response time for both MA and ME and the actual response times for 2010, 2011, and 2012 show very little change from year to year, although there are significant differences in response times between the two states, with ME showing longer response times ($p < .0001$) by about two minutes every year. No changes in future years response times for either state are predicted.

An historical regression over a three year period showed significant ($p < .0001$) relationships between the log of the response time and the log of the driving distance to the fire incident, the state (with Massachusetts showing lower response times), and the month/year of the fire incident. The overall strength of the relationship, however, was very low (r squared=.071). Due to the small amount of change in response time by year, any reduction in response time will need to be from variables such as the relocation of the fire stations. Basically, the firefighters are getting to the fire incident fast enough, they just need to be located closer to the incidents.

3.6 Causality Analysis

Word clouds were created using D3 and JavaScript to visualize the causes of the fires. This is a high level view based on unigrams created from the fire cause data, with noise words removed. Each term is given a font size proportional to its frequency. This provides a quick, intuitive idea of the distribution of causes for incidents by county. Clicking on each word causes the incidents with that cause to be highlighted on the map.

By clicking “Reset All”, and then clicking on an incident with a particular cause, in this example an ”accident”, highlighting of all incidents caused by that particular type of event in cyan will occur (Figure 10). Causes of fires can also be highlighted from the main maps for both states by clicking on one of the words on the left, such as accident, and then all of the incidents will appear. Zooming in, after selecting a cause from the word cloud, allows the user to see the actual nodes for each fire incident. If a particular county is then selected, then the distribution of fire causes for that county can be seen in the word cloud on the left.

3.7 Infrastructure and Personnel Analysis

The fire station infrastructure visualizations (Figure 11) were developed using a ”Parallel Coordinates Graph” from

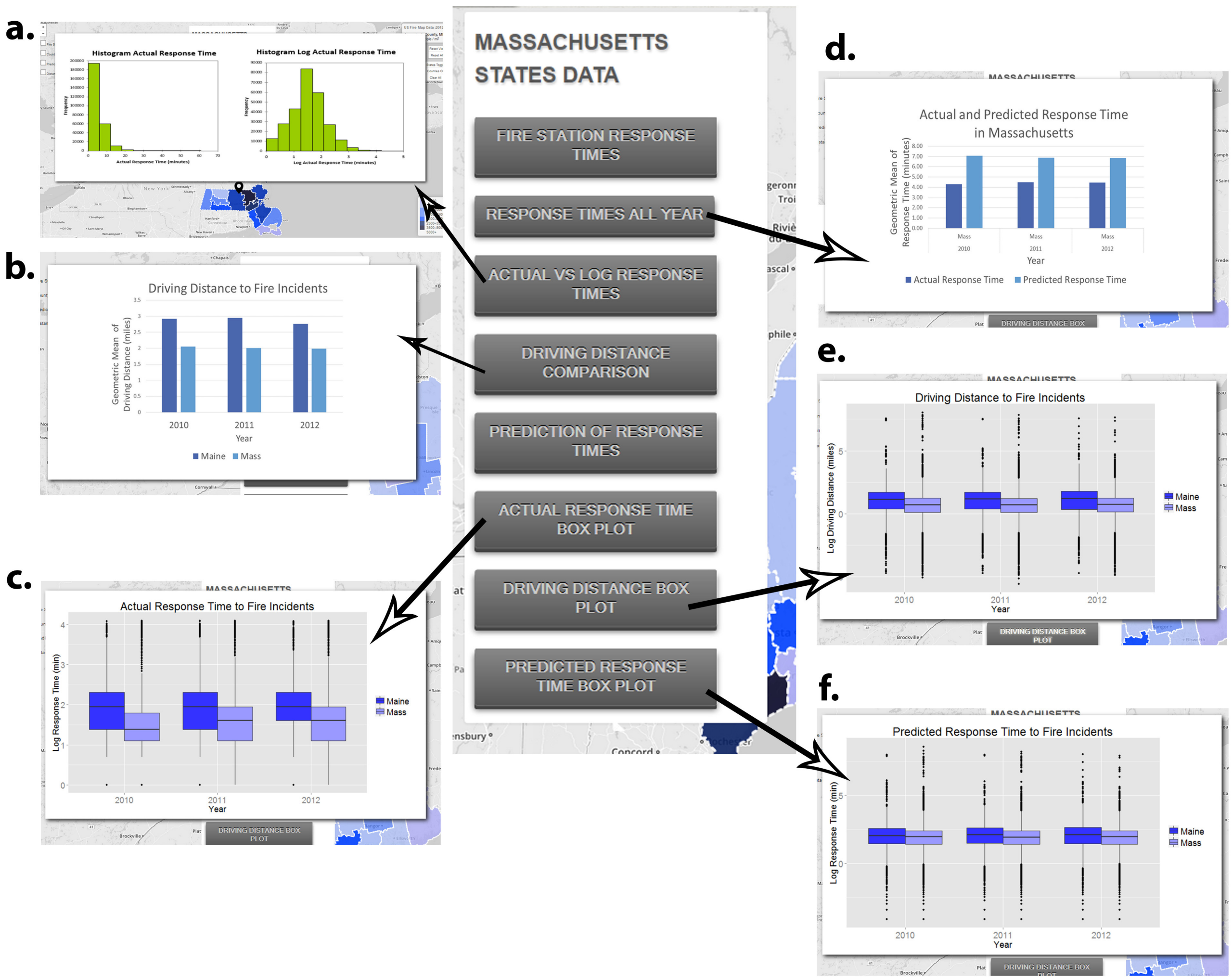

Figure 9: State level menu : (a) Linear and log distributions of response times. (b) Actual driving distance to fire incidents by year and state. (c) Actual Response Time to Fire Incidents Boxplot (d) Actual and predicted response times.(e) Driving distance to fire incidents (f) Predicted Response Time to Fire Incident

the D3 Library. Clicking the "Infrastructure" button at top of screen reveals data obtained from US Census Data provided though the Federal Emergency Management Agency (FEMA). For ME, *career firefighters* are the smallest group, followed by *volunteer* and *paid on call* firefighters. For MA, the biggest group is the career firefighters. Also we made the parallel coordinates graph interactive by enabling black boxes as brushes on top of each vertical axis. The user can drag them vertically and also can enlarge their sizes. Their role is to help the user to subset range of values to be displayed in bold colors, while other feature values would be ghosted on the back. This helps the user to selectively comprehend the flow and direction of the data.

4. DISCUSSION

Regarding the prediction of the recommended location of fire stations using K-means clustering, while the re-location of the fire stations based on this plan would reduce the distance to the fire stations to the fire incidents, such a plan is unrealistic. One improvement would be to treat the problem as a facility location analysis, with a set of potential realistic locations, but still using a clustering methodology. Another approach would be to provide interactive opening and closure of selected fire stations to observe the impact on the response times of the remaining fire stations. Since regression showed that the response times are not expected to increase in the future and the firefighters are either close to meeting or exceeding the required response times, this approach would make even more sense.

We observed that rural firefighters, on average, travel about one mile further to get to the fire incidents than those in an urban area, and their response times are about two minutes slower than in urban areas. Possible explanations for these findings include: (1) rural firefighters may need to travel

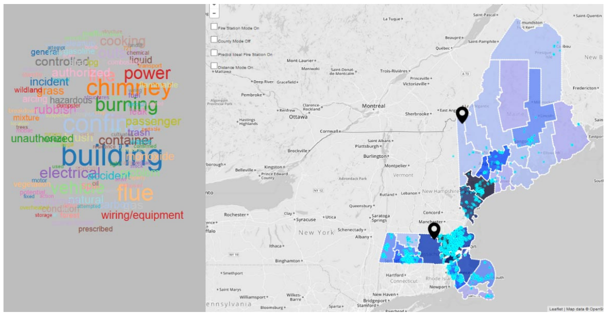

Figure 10: View showing all those fire incident dots in cyan related to a specific fire cause word ("accident") when clicked in the left panel.

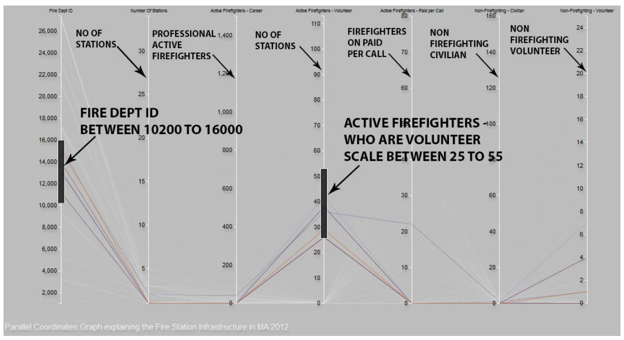

Figure 11: Infrastructure distribution for a county in Massachusetts in 2012.

on gravel and dirt roads, which slows them down; (2) firefighting equipment in rural areas is older, slower, and less specialized; (3) rural firefighters are more likely to be volunteers than career firefighters, so they may need to first travel to the fire station (e.g., from home) before going to the incident location; (4) rural fire incidents are more spread out geographically.

ISPARK and our approaches may easily work with data from any fire departments in the U.S. or other countries. However, the expertise to maintain the application will need to be available to the fire department. ISPARK may also work with data from other related domains, such as police, hospitals, and schools, so they could use similar technology.

It is noteworthy to mention that the word cloud we presented as one of the core features of visual analytics, shows a wide array of fire causes per county. Like wise to fit them in a restricted space, many of the words are overlapping. But to make it easier to be able to select each and any of those words by the user, we implement opacity changing mechanism, when the user's pointer hovers over them. This way the user in real time gets to know which word tag he is going to pick to see further analytics.

Our future work includes revising the methodology for determining recommended fire station locations, adding options for interactively closing and opening existing fire stations to observe the consequences on response times, extending ISPARK to the rest of the United States, and exploring more factors, including the spatiotemporal variables that would influence the fire station workload, such as time of day, day of the week, and month of the year. Additional studies should add traffic conditions to the predicted response time for increased accuracy. Also, wildland fires likely have unique characteristics which should be explored.

5. CONCLUSIONS

Our project has led us to the following four conclusions. First, an interactive, integrated analytical and geographic dashboard can be developed using entirely open source tools. Second, geographic mapping of the recommended fire stations, existing fire stations, and the fire incidents served by the fire stations can be done, but the recommended location methodology needs to be further refined. Third, the incident response times over the period 2010 through 2013 have been stable, and firefighters are either close to meeting or exceeding the standard response times now. Last, rural and urban areas do show differences in fire response times.

6. REFERENCES

[1] Malik A, Maciejewski R, Maule B, and Ebert D.S. 2011. A visual analytics process for maritime resource allocation and risk assessment. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on* (23-28). VAST, 221–230.

[2] Vladimir Agafonkin. 2011. Leaflet: An open-source JavaScript library for mobile-friendly interactive maps. (May 2011). `http://leafletjs.com/`

[3] National Fire Protection Association. 2015. National Fire Protection Association Website. `http://www.nfpa.org/`. (March 2015).

[4] Michael Bostock. 2011. A JavaScript visualization library for HTML and SVG. (February 2011). `http://d3js.org`

[5] US Census Bureau. 2015. Quick Facts. U.S. Census Bureau, U.S. Department of Commerce. `http://www.census.gov/quickfacts/table/PST045214/00`. (February 2015).

[6] Fire Department City of Boston Official Website. 2015. City of Boston, Official Website. `http://www.cityofboston.gov/fire/`. (February 2015).

[7] ESRI. 2011. Baltimore City (MD) Fire Department Uses GIS Technology to Optimize Resources. (September 2011).

[8] ESRI. 2015. GIS for Fire Station Locations and Response Protocol(2011). `http://www.esri.com/library/whitepapers/pdfs/gis-for-fire.pdf`. (March 2015).

[9] William Finley. 2007. Fire Station Location Master Plan. (January 2007). `http://www.usfa.fema.gov/pdf/efop/efo40119.pdf`

[10] David K. Hard. 2006. Analysis of Factors and Recognized Standards Utilized to Determine Fire Station Locations. *Klamath County Fire District No 1, Klamath Falls, Oregon* (June 2006).

[11] Pedro A Hernández-Leal, Alejandro González-Calvo, Manuel Arbelo, Africa Barreto, and Alfonso Alonso-Benito. 2008. Synergy of GIS and Remote Sensing Data in Forest Fire Danger Modeling. (December 2008).

[12] Open Data Institute. 2015. London Fire Stations. `http://london-fire.labs.theodi.org`. (March 2015).

[13] Ioannis Karafyllidis and Adonios Thanailakis. 1997. A model for predicting forest fire spreading using cellular automata, Vol. 99. Ecological Modelling, 87–97. Issue 1.

[14] Michael J. Karter Jr. 2014. Fire Loss in the United States during 2013. `http://www.nfpa.org/newsandpublications/nfpa-journal/2014/september-october-2014/features/2013-fire-los`. (September-October 2014).

[15] Nan Liu, Bo Huang, and Magesh Chandramouli. 2006. Optimal Siting of Fire Stations Using GIS and ANT Algorithm, Vol. 20. Journal of Computing in Civil Engineering, 361–369. Issue 5.

[16] Sergey Mescherin, Igor Kirillov, and Stanislav Klimenko. 2014. Optimizing and Visualizing Fire Dispatcher Activity. (October 2014).

[17] Federal Emergency Management Agency US Fire Administration National Fire Incident Reporting System, Department of Homeland Security. 2015. NFIRS 5.0 Reference Guide (2013). `https://www.nfirs.fema.gov/documentation/reference/`. (March 2015).

[18] Recep Nisanci, Volkan Yildirim, and Yasar Selcuk Erbas. 2007. *Fire Analysis and Production of Fire Risk Maps: The Trabzon Experience, Risk Management for the Future - Theory and Cases. Chapter 10.*

[19] Maciejewski R, Rudolph S, Hafen R, Abusalah A, Yakout M, Ouzzani M, Cleveland W S, Grannis S J, and Ebert D S. 2010. A Visual Analytics Approach to Understanding Spatiotemporal Hotspots. In *Visualization and Computer Graphics, IEEE Transactions on* (23-28). VAST, 205–220.

[20] Imas Sitanggang, Yaakob Sukaesih, Mustapha Razali, A N Norwati, and Ainuddin. 2012. Application of Classification Algorithms in Data Mining for Hotspots Occurrence Prediction in Riau Province Indonesia, Vol. 43. Journal of Theoretical and Applied Information Technology, 361–369.

[21] Fire Marshall State of Maine Official Website. 2015. State of Maine Official Website. `http://maine.gov/dps/fmo/index.html`. (February 2015).

[22] Division of System Planning Corporation TriData. 2006. Oklahoma City Fire Department Fire Station Location Study.

[23] Fernanda B Viégas, Martin Wattenberg, and Jonathan Feinberg. 2009. Participatory Visualization with Wordle. In *IEEE Transactions on Visualization and Computer Graphics* (23-28), Vol. 15. 1137–1144.

[24] A Şen, T Önden, C Gökgöza, and C Sen. 2011. A GIS Approach to Fire Station Location Selection. Inter al Soceity for Photogrammetry and Remote Sensing.