Preliminary Investigation into Fire Service Survey Responses

David Marx

2/3/2017

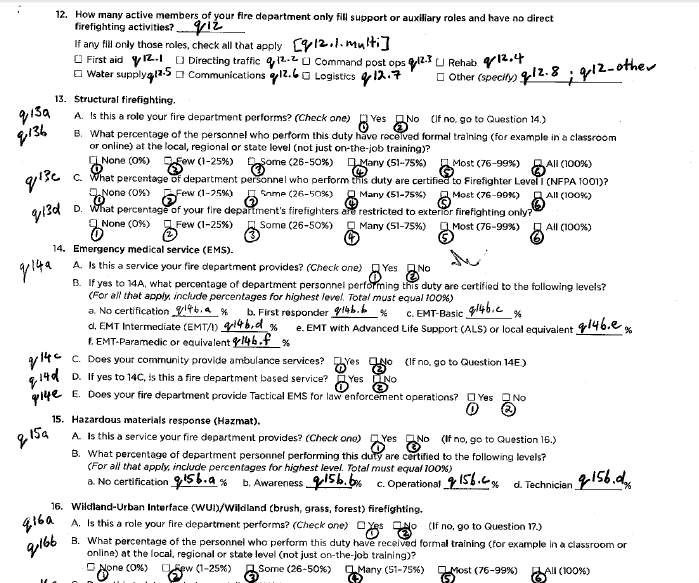
## Executive Summary

The Fire Service Survey is an annual survey administered by the NFPA. One of the main functions of the survey is to perform a needs assessment of departments in the US fire service. A preliminary investigation of a sample of survey responses identified that:

* 18% of departments experienced a decling in paid/FTE firefighters since 2011
* Over 54% of departments are running apparatus that is 15+ years old
* 21% of volunteer departments had at least 14% female staff, whereas only 1% of paid departments did.

## The Survey

Survey results were provided to the analytics team with very little context. A data dictionary was provided in the form of an annotated version of the survey, where the questions asked by departments were matched to the corresponding column names in the data.



Data Dictionary

It is unknown what year these results are from, but it is assumed they are all from the same year. Additionally, Part II, Question 5 asks:

"Has your department had a reduction or increase in total firefighter positions of full time equivalents (FTE) since 2011?"

so it can therefore be presumed that the year of the data was from 2012 or later.

Departments from all across the country were represented, with 5458 departmetns responding in total.

## Suggestions for Improvement to survey

### Survey Administration

The provided sample survey (the data dictionary) illustrates that the survey is being administered via pen and paper, completed by hand and then results are collected by human data entry. This process is time consuming for all involved, error-prone, and unnecessarily expensive. Associated costs include:

* Printing the survey
* Mailing the survey
* Delivering responses
* Human data entry of responses

All of these costs are relieved by administering future surveys using a digital platform. Because the contents of this survey are not sensitive, future surveys could be administered using free platforms such as Google Forms or SurveyMonkey.

Possibly more important than mitigating costs associated with the current survey system, migrating to a digital platform would increase the quality of the data collected in the survey.

* Someone performing data entry may misread or mistype the value of a response
* Digital forms can enforce that only valid entries are submitted and perform data integrity checks as respondents complete the survey
* E.g. questions 14b asks for percentages that need to add up to 100, question 34 asks for several values that need to add up to the value of a human entered "total" field
* Digital forms can enforce that required answers are alwys completed as a prerequisite for allowing submission

### Survey Questions

Less problematic than the mechanism by which the survey is administered, the survey prompts for responses on certain questions in a way that unnecessarily reduces data granularity. Many questions on the survey ask for the respondent to provide a value, but then invite the respondent to communicate the value as a range. For example, consider question 16, which asks about Wildland Firefighter Training:

"What percentage of personnel who perform this duty have received formal training?" A. None (0%) B. Few (1-25%) C. Some (26-50%) D. Many (51-75%) E. Most (76-99%) F. All (100%)

Contrast this with question 14b, which when asking about EMS training, requires the respondent to provide (or at least aproximate) specific values of percentages for each EMS training level in their department.

Additionally, there are some questions in which the binning does not span the full range of values, requiring an "other" field, creating a mixed variable where some respondents provide precise values, but others respond with a category. Concretely, questions 22d and 22f, asking about the frequency of physical exams and fitness assessments, respectively, present the following options for responses:

A. New Firefighters Only B. Every 6 Months or Annually C. Every 2 years D. Every 3 years E. Other

It is likely that these options will capture the majority of departments, but there's really nothing gained using this format relative to simply asking for a concrete number of months or "other" for responses requiring explanation.

Asking for precise values in one place but ranges in others is inconsistent and produces data that is higher granularity for some variables than other. To mitigate this, respondents should be asked to provide (or approximate) precise values whereever possible.

### Data representation

The data was delivered for analysis in CSV format. It is unclear whether this was the original raw format of the aggregated data, or if this was an export from a database. In any event, there were several issues identified in this data representation that would be trivial to mitigate for future administrations of the survey, which would reduce preprocessing overhead for analytics.

First and foremost, null data was represented in multiple ways, primarily as either empty cells (199,102 occurrences), the text value '#NULL!' (188,236 occurrences) or white space (56,520 occurrences), suggesting the data was likely copied from another excel spreadsheet. It was additionally determined that in many instances, null data was not intended communicate a null response, but likely a zero-value response (e.g. in fields prompting for counts). Migrating to digital survey administration would eliminate this ambiguity and enforce a consistent representation for null data and ensure zero values are represented explicitly.

Additionally, numeric values were often represented as comma-delimited numbers, i.e. text of the form "xx,xxx". Representing numbers as texts adds an unnecessary preprocessing step when working with the data and can result in mistakes if analysts, attempting to work with the full set of variables, do not notice the incorrect data type for such variables. Again, this is a problem that would be completely rectified by using a digital form to administer the survey.

Finally, several columns in the data were out of order relative to their ordering in the survey. This is a relatively small problem, but at first glance it makes it appear as though a quarter of the variables are missing and introduces unnecessary risk of error if an analyst attempts to use numeric indexing to get to the variables they want rather than accessing them by name. When generating flat text files of this kind, in the future measures should be taken to preserve meaningful column ordering.

## Data preprocessing

### Cleaning Nulls

For the purpose of this analysis, it is assumed that null entries on questions asking for numeric values are equivalent to a reponse of "0". Cleaning in this fashion had the additional effect of creating a new "null" category on questions associated with categorical/ordinal variables, since categories are otherwise 1-indexed in the data. Handling nulls in broad strokes such as this is generally inadvisable, but was determined to be suitable for the purposes of performing a quick preliminary exploration of this data.

### Cleaning numbers stored as text

Because of the width of this data set (227 columns), instead of identifying columns in which numbers were stored as text individually and fixing their data as needed, the following automated approach was utilized to attempt to clean all such data in a single pass. For each column:

1. Count the maximum number of characters for any value in the column. If this value is higher than 10 (i.e. if it is a number, it is on the order of 10,000,000, which is greater than the US population) assume the column is not numeric.
2. Remove any commas appearing in entries in the column.
3. Attempt to coerce entries in the column to numeric.

This method was successful for all numeric data assessed in this analysis.

## FINDINGS

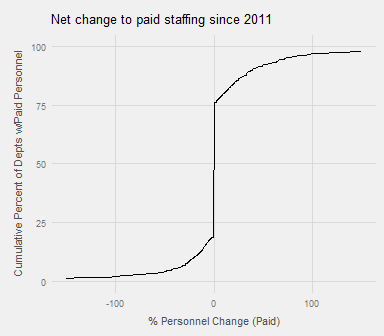
## Stability of staffing levels

The preliminary investigation found that staffing levels of paid firefighters are stable relative to their 2011 values.

52% of departments have no full-time paid personnel. Of those that have paid personnel, relative to 2011 (and excluding departments that reported zero paid personnel in 2011):

* 18% experienced a staffing decline
* 25% experienced a staffing increase
* 57% experienced no net change to staffing

The following chart provides the empirical distribution function for the absolute value of the net percent change for each department relative to their 2011 paid staffing.



## Representation of Women

Controlling for department size, women have measurably higher representation in volunteer departments. This was determined by treating reported counts of part-time/volunteer vs. paid/FTE as separate "departments" (so for this analysis, "volunteer department" is used to mean "volunteer *side*" for departments that employ both paid and volunteer firefighters), then fitting a logistic regression to the department type against the predictors of total firefighter count and female firefighter count for the department. Concretely, using the model formula:

gave the resulting model:

mod\_summary = as.data.frame(summary(mod\_prob)$coefficients)  
mod\_summary[,'Pr(>|z|)'] = '<2e-16'  
  
kable(mod\_summary)

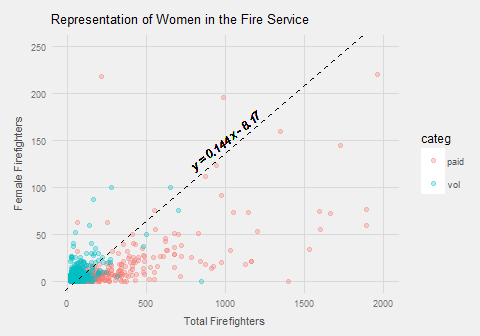
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | z value | Pr(>|z|) |
| (Intercept) | 1.6414057 | 0.0612496 | 26.79865 | <2e-16 |
| female\_ff | 0.2008580 | 0.0126106 | 15.92773 | <2e-16 |
| total\_ff | -0.0289682 | 0.0013922 | -20.80738 | <2e-16 |

#mod\_summary

Because volunteer departments are generally small, we should expect the size of the department to have a strong effect on our ability to predict, in the absence of other information, if that department is paid or volunteer. From the coefficients of this model, we can see that adding a single female firefighter to a department increases the odds of that department being volunteer approximately seven times more than if we remove a random firefighter.

The above model provides us with a decision boundary given by the line

In other words, if we observe that a department has over about 14% female firefighters, it's extremely likely (77% accuracy) that department is volunteer. We can see this phenomenon graphically:



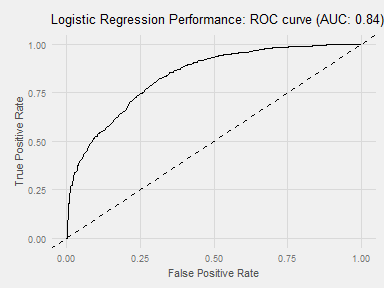
Concretely, 21% of volunteer departments had at least 14% female staff, whereas only 1% of paid departments did.

### Model Validation

To confirm this analysis, the accuracy statistic was investigated via permutation testing. To ensure the model was not "overfit" to the data, performance on out-of-sample data was evaluated using the bootstrap technique. A value of 77.0% was found for the bootstrapped accuracy with 2000 replications, with a 95% confidence interval of 75.3% - 78.7%. This confirms that the model was not fitting noist

To ensure that the model performs better than random (i.e. that the accuracy was not a consequence of class imbalance), significance testing was accomplished by performing a permutation test (aka "target shuffling") with 200 replications over the bootstrapped accuracy statistic (with 200 replications, which was found to be sufficient to match the results of the 2000 replications bootstrapped accuracy), producing an estimated p-value of p < .005.

Finally, the overall model performance was evaluated using a Reciever Operator Characteristic (ROC) curve. The curve illustrated that high probability assignments capture most of the positive class in the data (i.e. the model assigns high probability appropriately). Additionally, the area under the ROC curve (AUC) was found to be , providing additionally evidence that the model performs well (AUC close to 1 is good).



## Aging Fleet

Many departments have units in service that are at least 15 years old. Percentage of departments with at least one in-service suppresion unit of the following types, that is at least 15 years old:

#t(data.frame(Engine='51%', Ladder='41%', Tanker='54%'))  
data.frame(Unit=c('Engine','Ladder','Tanker'), 'Percent'=c('51%', '41%', '54%'))

## Unit Percent  
## 1 Engine 51%  
## 2 Ladder 41%  
## 3 Tanker 54%

That 54% of departments are operating old tankers only gives us a lower bound on the estimated percent of departments that are running any kind of suppression apparatus that is at least 15 years old. The figure of how many departments genearlly are running old units was not investigated in this analysis.

