U.S. Fire Administration
National Fire Data Center

National Fire Estimation Using NFIRS Data

White Paper

May 2017

FEMA

National Fire Estimation Using NFIRS Data

White Paper

May 2017

FEMA | U.S. Fire Administration
Working for a fire-safe America

This white paper was produced under contract GS-10F-0195X, order HSFE20-12-F-0240 between 2013 and 2017 by PG Public Services, LLC for the U.S. Fire Administration, Federal Emergency Management Agency.

This paper was peer-reviewed by the following: Professor Charles Jennings, Ph.D., MIFireE, CFO in the Department of Security, Fire and Emergency Management at the John Jay College of Criminal Justice and Professor Susan Simmons, Ph.D., in the Department of Mathematics and Statistics at the University of North Carolina at Wilmington.

Table 1. Meta-analysis of fire factors — continued

Category	Factor	Outdoor	Structure	Structure/ Vehicle	Vehicle	Grand total
Socioeconomic	Age of structure		7			7
	Car ownership		1	1		2
	Education level		7	1		8
	Employment		1			1
	Home ownership		9	1		10
	Home value		1			1
	Household income		29	1	2	32
	Household size		4	3		7
	Housing density		2			2
	Married no children			1		1
	Minority population		8	2		10
	Residential structures		3			3
	Single-parent household		10	1		11
	Single person			1		1
	Telephone ownership		1			1
	Unemployment rate		3			3
	Vacancy		9			9
	White population			1		1
Topography	Altitude	4				4
	Aspect	3				3
	Elevation	2				2
	Slope	6				6
Vehicle	Age of vehicle				1	1
	Number of vehicles				1	1
	Type of vehicle				1	1
Weather	Lightning		1			1
	Precipitation	10	1	1		12
	Relative humidity	6		1		7
	Temperature	17	2	1		20
	Wind speed	6				6
Grand total		**110**	**114**	**17**	**6**	**247**

As an initial foray into creating an estimation methodology, the working assumption is that a ZCTA-based estimation model will be sufficient at the state and national levels, as these differences will cancel out across large geographic areas because of the relatively small geographic differences between ZIP codes and ZCTAs. In addition, using a ZIP code level based area as a preliminary avenue of investigation has the advantage of computation efficiency and the ease to tie directly to the NFIRS data. For the purposes of this initial analysis, ZIP codes and ZCTAs are treated as synonymous.

ZCTAs **may not** be sufficient when dealing with the subparts of the fire problem, such as county or fire cause, but it is believed to be sufficient as an initial avenue of investigation. As geocoding processes are refined and assessed, it may be possible to transition to a census tract based estimation methodology. Census tracts provide more granularity, geographic coverage, and statistical uniformity than ZCTAs and are generally preferred for such analysis. In addition, using census tracts eliminates the translation issues from ZIP codes to ZCTAs.[7] At this time, however, the current geocoding of NFIRS addresses is not sufficiently accurate to use census tracts for national estimates.

Imputation methods

Imputation, a process of replacing missing data with substituted values, is a common statistical technique that is used by many government agencies to adjust for various types of nonresponse in survey data. For example, the U.S. Census Bureau employs “count imputation” for nonrespondents and “characteristic imputation” to generate a complete population count, as well as to impute other demographic data where respondents omit race, age or other personal information. In the 2010 census, count imputation added 1.2 million people, which represented 0.39 percent of the total U.S. population as determined by the U.S. Census Bureau (Cohn, 2011).

In its Current Population Survey, the U.S. Census Bureau also uses several imputation methods for nonresponse (Imputation of Unreported Data Items, 2013). The Bureau of Labor Statistics and the Bureau of Economic Analysis have developed several imputation methods for nonresponse (Eltinge, Kozlow, & Luery, 2003). With respect to fire data, analysts at the Consumer Product Safety Commission use imputation for distributing unknown fire cause entries (Greene, Smith, Levenson, Hiser, & Mah, 2001).

Generally, the U.S. Census Bureau uses averages from similar households within a particular census tract to estimate the nonrespondents. In between censuses, the U.S. Census Bureau uses a combination of birth, death and immigration/emigration records to estimate the current population (Cohn, 2011). A similar technique can be used to fill in the NFIRS data by identifying similar areas.

For this effort, 2010 validated fire incident data from the NFIRS was used as the basis for developing a methodology to estimate the annual overall U.S. fire incidence. There were several reasons for this choice. First, at the time of the initial study in 2012, it was the most recent NFIRS fire incident data available. Second, in using the 2010 NFIRS data, there would be no lag between the population characteristics collected from the 2010 census and the 2010 fire incidents as reported in the NFIRS. Lastly, ACS data at the ZCTA level was available for all ZCTAs in 2010.

The NFIRS fire incident data was adjusted to account for the following: certain additional aid incidents (incidents where one fire department assists another), fire departments and ZIP codes where no fires occurred but other types of incidents were reported, potential missing data, and late submission data.

[7] https://www.census.gov/geo/reference/pdfs/GARM/Ch10GARM.pdf (accessed January 31, 2016).

Fires from the National Fire Incident Reporting System

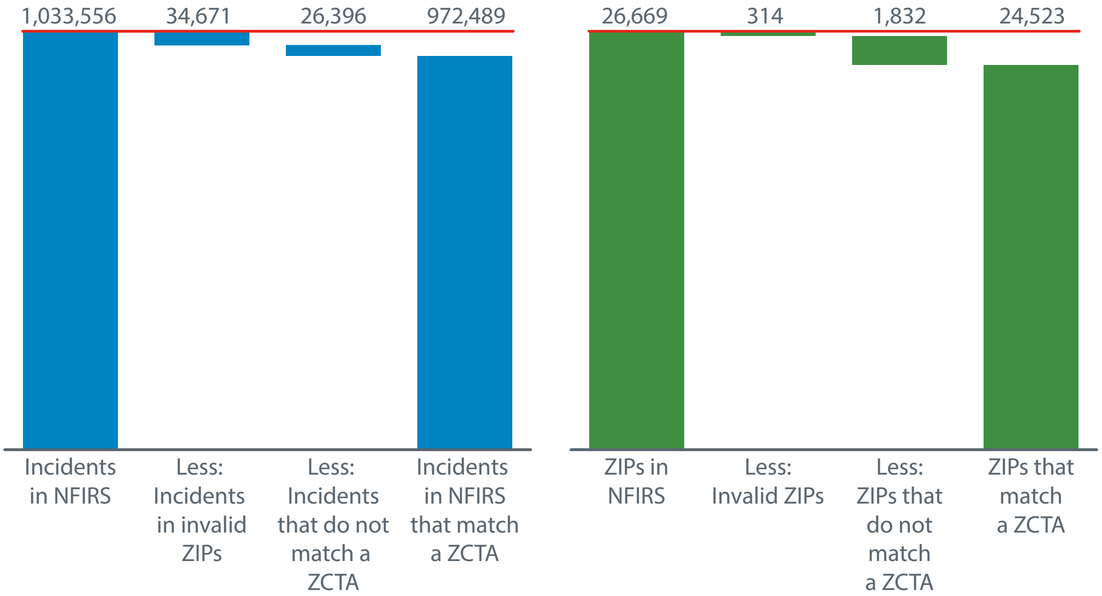
The NFIRS dataset is fluid and constantly changing as fire departments update, upload and submit their data. Once the fire department of record validates its data, it "releases" the data, making it available for analyses at the national level. The most common set of released NFIRS data is the Public Data Release (PDR). This annual subset of NFIRS data is a compilation of a calendar year's incident data which has been validated for release by the fire department of record and submitted to the NFIRS database by July 1 of the following year. The annual PDR is not re-released to reflect subsequent incident submissions after the reporting deadline. The PDR does not contain all validated data submitted by fire departments; it only contains released, validated data.

The validated incidents that the PDR does not contain are simply unreleased incidents. Generally, the reason for not releasing data is just an oversight on the part of the state NFIRS program manager. In terms of input to a national estimates methodology, using all validated data — released and unreleased — represents the most complete set of incidents from U.S. fire departments.

There were 1,227,719 validated fire department fire runs (records) reported in the NFIRS in 2010. A run, a record of each time a fire department responds to a service call, is not equivalent to an incident, as more than one department may respond to the same incident. This situation is called "aid." Using the USFA accepted definition of a fire incident, 1,003,556 of these runs were validated fire incidents in 26,669 ZIP codes (Figure 2).[8] In the data, several of the ZIP codes reported were invalid for various reasons. Of those, 314 ZIP codes were invalid because they were not five-digit numbers greater than zero. There were 34,671 fires reported for these 314 ZIP codes (3.3 percent of fires; 1.1 percent of ZIP codes). In addition, since both the scaling and the regression model rely on data from the ACS, only ZIP codes that match ZCTAs could be used in the analysis. This requirement meant that the 26,396 fires reported for the 1,832 ZIP codes that did not match a ZCTA were also removed from the data set (2.6 percent of fires; 6.9 percent of ZIP codes). After removing invalid ZIP codes and matching ZIP codes to ZCTAs, there were 972,489 fires (94.1 percent of all reported fires) in 24,523 ZIP codes/ZCTAs.

[8]The USFA accepted definition of a fire incident only includes records of the primary fire department in whose jurisdiction the incident occurred. This is achieved by excluding records reflecting aid given. The theory is that, in essence, not excluding aid incidents when analyzing incidents will result in double counting those incidents where both the giving and receiving departments report to the NFIRS. See USFA's National Fire Incident Reporting System Version 5.0 Fire Data Analysis Guidelines and Issues.
It is important to note that this definition does not indicate how to count incidents where the aid receiving department does not report to the NFIRS, but the aid giving department does.

Figure 2. Matching NFIRS address data to ZCTAs (2010): Incidents and ZIP codes



Range errors

To identify whether these fires were reported in the correct ZIP code, a novel approach was developed for identifying a "range" error for the ZIP code data. This approach calculated an estimated geographic position of a particular fire department and measured the distance of incident data to check if they appeared to be "out of range." A "weighted center" latitude and longitude for the fire department was calculated by using the latitude and longitude from the ZIP codes covered by the incidents. This "weighted center" was then used to evaluate incident data at a fire department level and to flag incidents that might have been misidentified to a particular ZIP code.

The range error analysis yielded several likely reporting problems. One particularly interesting case occurred in ZIP code 77201. That ZIP code is assigned to a U.S. post office that has no demographic data, but 1,434 incidents were reported there in 2010. Based on the fire department identification, it is believed that the ZIP code should actually be 72201. This kind of pervasive and potentially systematic ZIP code error may result from accidental hard coding of a ZIP code within an entry screen. For purposes of analysis, this ZIP code error was corrected in the data, but other range errors were not corrected.

Aid given data

In general, aid is a voluntary exchange of resources and services between fire departments for mutual benefit or by contractual agreement. In emergency services, there are two basic types of aid fire departments can give or receive: mutual aid and automatic aid. Mutual aid is an agreement among emergency responders to lend assistance across jurisdictional boundaries as a need arises. Automatic aid is assistance dispatched automatically by a pre-arranged contractual agreement between two communities or fire districts.

The NFIRS makes a distinction between primary incident reports, which are filed by the fire department with a jurisdiction over the location of the fire, and "aid given" reports that are filed by neighboring fire departments that may assist in the fire. Aid given reports are valid incidents, but are not counted as fire incidents as they are considered duplicate reports.

Primary incident reports, which include "aid received" reports, are more comprehensive and contain information related to cause, incident type, property use, structure type, and so on. In theory, aid received primary reports and aid given secondary reports can be reconciled and any duplicate reports can be removed from analyses. However, when the primary report is missing, aid given reports serve as the only record of the incident. By current convention,[9] these aid given reports are not included in discussions and counts of incidents reported in the NFIRS. Therefore, aid given reports, for which there is no primary report, provide an additional source of fire incident data that can be explored for use in developing national estimates.

Table 2 gives the number of incidents by type of aid from the 2010 NFIRS valid data set. If aid given reports are excluded under the assumption that all aid given incidents are represented elsewhere in the data, the total number of incidents is 1,003,556. However, the number of aid given incidents is nearly twice the number of aid received incidents. This could be for the following reasons:

- The receiving fire department does not report the aid received; the incidents are reported as no aid.
- Multiple aid given reports are filed for the same aid received incident.
- The receiving fire department does not report the incident because of nonparticipation in the NFIRS or missed reporting. In this case, the aid given report is the only record of the fire.

Table 2. NFIRS fire incidents by type of aid — All valid incidents (2010)

General aid type	Incidents	Percent
Aid received	128,390	10.5
Aid given	224,163	18.3
Other aid	2,091	0.2
No aid	873,075	71.2
Total	1,227,719	100.0

Likely, some portion of the aid given incidents are duplicated in the aid received incidents, but some portion likely is not. A methodology was developed to identify the unduplicated aid given incidents.

The aid given incidents were compared to the aid received incidents to find exact or approximate matches based on the incident identification information. As insufficient NFIRS information was available to identify potential matches from the identification information for aid given incidents, USFA-provided geocoded location data was used to match aid incidents on the geocoded location, date and time of the incident.[10]

[9]Currently, the convention for fire data analysis is to exclude mutual and automatic aid given incidents in counts of fire incidents. The assumption is that these aid given reports are duplicated in the primary incident report. See USFA's National Fire Incident Reporting System Version 5.0 Fire Data Analysis Guidelines and Issues.
[10]While the geocode algorithm used for the 2010 NFIRS data did not yield locations with sufficient precision to enable the use of census tracts as the primary analytic unit, it was sufficiently reliable to use for this de-duplication exercise.

The methodology to identify and match aid given incidents to aid received incidents is based on physical and temporal distances. Physical distance between incidents is achieved using an SQL geography element of geocoded address data for fire records; temporal distance between the incidents is based upon reported incident alarm time. Various versions of de-duplication were attempted using multiple geocode sets, variable distance, variable time, and sensitivity, based on aid type and incident type. An initial baseline set of distances for each major incident type (structures, vehicles, outside/other) was suggested by the USFA. This baseline was a temporal distance of 20 minutes and physical distances of 1, 2 and 5 miles for structure, vehicle and outside fires, respectively (Table 3).

Table 3. Initial match parameters for aid incidents

Match parameter	Criteria
Time	Within a specified date-time range, with aid received alarm time prior to aid given alarm time
Location	Geocoded all addresses and considered to be a match between aid given and aid received incidents if within 1 mile (structure fires), 2 miles (vehicle fires), or 5 miles (outdoor fires)

The first phase of analysis used a broad-ranged matching algorithm to pair each aid record with the closest primary record if there was a match within 10,000 meters that had an alarm time within 120 minutes. This provided a large set of matched records that could be analyzed at a high level to identify patterns in the distribution of matches. Analyses determined that the suggested baseline time and distance radii resulted in an unacceptable number of false positives and that automatic aid (i.e., prearranged) versus mutual aid (i.e., ad hoc) resulted in considerably different distribution of results.

The second phase implemented a more tightly tailored matching approach, matching only aid given to the primary aid received. Matches were again made by major incident type (structure, vehicle and outside) with the additional filter of aid type (automatic or mutual.) This approach limited the matches to within 2,000 meters and 40 minutes. These data were then spatially graphed in 50-meter physical distance intervals and five-minute temporal distance intervals. A random sample of matches was manually verified. This analysis identified distances and intervals used in this approach. For perspective, a circle with a radius of 50 meters is roughly equivalent to one football field, and one with a radius of 2,000 meters is roughly 5 square miles.

Model fidelity between the analytically derived distances and the initial USFA-suggested distances was analyzed. With the use of the newly proposed match distances, the model improved incrementally. Manual verification of both sets of distances confirmed that using the larger initial distances results in a substantial number of false positives without a substantial increase in successful matches of well-populated records.

This process yielded 49,372 matches between aid given incidents and aid received incidents (Table 4). That means the remaining 174,791 aid given incidents have no match in the aid received incidents and are potentially unique incidents.

Table 4. Matching NFIRS aid given-received incidents, model parameters (2010)

Incident-aid type	Distance (m)	Time (min)	Matches	Unmatched incidents
Structures — Mutual	50	15	19,024	64,596
Structures — Automatic	50	5	17,134	42,510
Vehicle — Mutual	50	5	917	6,729
Vehicle — Automatic	50	5	1,067	5,045
Outside/Other — Mutual	0 union 2,000	40 20	7,535	36,624
Outside/Other — Automatic	50	5	3,695	19,287
Aid given incidents			49,372	174,791

"Other aid" incidents — incidents where a fire department covers and responds to another jurisdiction or locale that has no fire department — were also analyzed. As there were no aid given/received pairs to match, the other aid data was searched for possible duplicates. There were very few; only 122 potential duplicates were identified. For consistency, these incidents were removed from further analyses. A summary of the count of unique fire incidents found in the 2010 NFIRS data is shown in Table 5.

Table 5. Total unique NFIRS fire incidents (2010)

Match group	Incidents
NFIRS valid incidents	1,003,556
Unduplicated aid given fires	174,791
Less duplicate other aid given	-122
Total unique fires	1,178,225

True zeroes

Reporting fire departments with no fires can be identified by reviewing the pattern of reported incidents. If a fire department reported other incidents — such as Emergency Medical Services or hazmat responses — and did not report any fires, it is understood that, while there was activity, there were no fires for the department that year.[11] Using this method, 9,090 ZIP codes were identified that had no reported fires. Of these ZIP codes, 4,373 match ZCTAs.[12] The inclusion of these "true zeroes" to the model should allow for a more accurate estimate by reducing the number of ZCTAs for which incidents need to be estimated.

[11]USFA NFIRS Program Manager.

[12]Reviewing patterns of fire reporting does not identify departments that had no reported incidents of any type. If a department participates in the NFIRS, there is a mechanism that allows fire departments to submit a monthly "report of no activity" if the department experiences no calls of any type during the month. These "reports of no activity" were analyzed for 2010 with no material effect on the national estimate: 46 "true zero" ZIP codes were added, reducing the estimate by 327 incidents.

Missing data and late data

Fire departments that faithfully report incidents to the NFIRS may fail to report a particular incident due to a variety of circumstances. Additionally, it is known that the annual data is incomplete. The NFIRS has a closing date of July 1 post year, but data continues to arrive to the NFIRS after this date. The identification of this missing and late reporting data is critical in calculating an accurate estimate of the fire problem. There are known issues of underreporting; overreporting of fire incidents is rare, if it exists at all.[13]

A conservative approach to identifying and adding missing data to the model was adopted, and a model was developed of fire incidence for each fire department that reported data to the NFIRS between 2006 and 2011. To account for seasonality of data submission, the fire incident data was analyzed on a monthly basis. A monthly average and standard deviation for each fire department was calculated. Using a 97.5 percent confidence band, if the reported number of incidents was less than two standard deviations below the average number of incidents for that month (suspected underreporting) then the average was substituted for the reported value.[14] For the base analysis year of 2010, this method adds 85,097 fires to the total.

The modeling of late data represents a particularly hard challenge to estimation. After the submission period, there can be additions, changes and deletions of incident data. Three years of data (2010 to 2012) were examined to understand the extent of change to the data after the close of the submission period.[15] The average number of additional late incidents was 0.8 percent of the total number of incidents for the year — a non-negligible change.

It was theorized that the existing missing data model could explain some of this change. An analysis at the fire department level was developed to compare the number of validated, released incidents as reported in 2010 (the 2010 PDR); the total number of validated incidents for 2010 reported as of 2014 (the model input data set); and the predicted number from the missing data model.

The analysis showed that most fire departments make no change to their data after the close of the submission date (including releasing previously unreleased data) — only 7.2 percent of fire departments in the 2010 PDR changed their total valid incidents by 2014. Comparing fire departments that had increases in their valid incident counts with the incident counts predicted by the missing data model showed 97.2 percent of the increase in incidents could be predicted by using the existing missing data analysis. Given the overall robustness of the missing data model, additional factors to account for late data were not added to the model.

Cumulative effect of data adjustments

In 2010, there were 1,003,556 reported fires for 26,669 ZIP codes. Applying additional data correction techniques increased the number of fires and the number of ZIP codes for which data were available. Table 6 shows the various correction techniques used and the estimates each respective technique produced.

[13]USFA NFIRS Program Manager.
[14]The analysis was conducted using a one-sided confidence interval because the incidence of overreporting is uncommon.
[15]Because the NFIRS fire incident data for the models contains all validated 2010 incidents as of 2014, and the PDR contains only released validated incidents as of 2010, an exact comparison on the effect of late submitted data could not be made, but a bound on the size of the late data increase could be determined.

Table 6. Summary of NFIRS incident census (2010)

Category	NFIRS valid incident data	NFIRS valid incident data matched to ZCTAs	NFIRS valid incident data with unduplicated aid	NFIRS valid incident data with unduplicated aid matched to ZCTAs[16]	NFIRS valid incident data with unduplicated aid and "true zeroes" matched to ZCTAs	NFIRS valid incident data with unduplicated aid, "true zeroes" and missing data added matched to ZCTAs
Fires	1,003,556	972,489	1,178,225	1,178,225	1,178,225	1,263,322
ZIP codes	26,669		30,410			
ZCTAs		24,523		24,330	28,703	28,703

Scaling model

The original paper by Hall and Harwood on national estimates (1989) proposed a method for creating a national estimate of fire incidence by relating the population protected for each of the fire departments in the NFIRS to the total U.S. population and then scaling up the reported incidents.

Historically, the population protected data in the NFIRS was inaccurate, and the data is no longer collected. Since the NFIRS collects the ZIP code for each incident, it is possible (assuming a ZIP code — ZCTA equivalence) to estimate the total population of ZIP codes for areas where fires were reported and to then compare that portion to the total population. This approach is similar to that described by Hall and Harwood but does not require an estimate of population protected.

A novel approach at estimating the population associated with each fire department was undertaken within the NFIRS. NFIRS data was combined with ACS data to generate an estimate of the total population covered.

To assign the population to a particular fire department and ZIP code combination, an analysis was used for each ZIP code to identify all fire departments and the associated U.S. census population for that ZIP code. If multiple fire departments had incidents in a ZIP code, each fire department was allocated an equal share of the total incidents for that ZIP code. Future work could allocate the population on a pro-rated basis, but this was not undertaken for this analysis.

The resulting model produces a simple population coverage estimate for each fire department — one that, in concept, is mutually exclusive of any other fire department estimate.[17]

[16]Excludes erroneous ZCTA matches to areas other than the 50 states and District of Columbia.
[17]This process can be further simplified. If fires and population by ZIP code are available, there is no need to do any allocation to fire departments to calculate protected population. A simple proportion is possible as well.

The known incidents, including "true zeroes" where applicable, represent fires in areas where a known number of people live. According to the ACS, the 2010 population for the 50 states and the District of Columbia was 309,122,569. A scaling estimate for the total incidence rate was developed by relating the population coverage of reporting fire departments to the total population and multiplying that by the incidents as reported. National estimates derived using the scaling method are summarized in Table 7.

Table 7. National estimates using scaling (2010)

Category	NFIRS valid incident data matched to ZCTAs	NFIRS valid incident data with unduplicated aid, true zeroes and missing data added matched to ZCTAs
Fires	972,489	1,263,322
Population in ZCTAs with data	280,323,251	295,964,706
Estimated fires	1,072,399	1,319,486

Regression model

A regression model is a statistical approach to forecast values of a dependent variable based on change in one or more independent variables. However, relationships depicted in a regression analysis are associative only, and any cause-effect or causal inference is purely subjective. As with the scaling model, this model is focused on using a ZCTA as a preliminary avenue of investigation for computation efficiency and the ease to tie directly to the NFIRS data.

Research efforts identified factors that might help predict the total fire incidence: demographic, socioeconomic, weather, business and vehicle crash factors (Table 8). Each of these factors was collected from either the ACS five-year data (2007 to 2011), the 2010 Census County Business Patterns (CBP), the Wildland Fire Assessment System (WFAS), or the Fatality Analysis Reporting System (FARS).

Table 8. Summary of data factors and sources

Category	Factor	Data source
Business	Business population	CBP — Number of establishments
Population	Age	ACS S0101 — Total estimate by age[18]
	Population density	ACS S0101 — Total population[19]
Socioeconomic	Age of structure	ACS DP04 — Year structure built, various[20]
	Car ownership	ACS DP04 — Calculated. Occupied housing units minus occupied housing units with no vehicles available.
	Children	ACS DP02 — Families with children under 18
	Education level	ACS DP02 — Households with at least a high school graduate
	Home ownership	ACS DP04 — Owner occupied units (not found significant in this analysis)
	Home value	ACS DP04 — Estimated median value
	Household income	ACS DP03 — Median household income
	Household size	ACS DP02 — Average household size
	Housing density	ACS DP04 — Total housing units
	Married, no children	ACS DP02 — Calculated. All married couples minus married couples with children under 18.
	Minority population	ACS DP05 — Calculated. Total population minus non-Hispanic white population.
	Residential structures	ACS DP04 — Year structure built, various
	Single-parent household	ACS DP02 — Calculated. All male-headed households without a spouse with children, plus all female-headed households without a spouse with children.
	Single person	ACS DP02 — All householders living alone
	Unemployment rate	ACS DP03 — Calculated. One minus (total employed population divided by population 16 or over).
	Vacancy	ACS DP04 — Vacant housing units
	White population	ACS DP05 — Non-Hispanic white population

[18]Older age population was calculated by adding the total population by age statistics for age bands of 65 and above for a given ZCTA.
[19]Population density is calculated by using the total population for a ZCTA and dividing it by the total land area for a given ZCTA.
[20]Age of structure is the total number of housing units constructed in each of the time bands collected by census for a given ZCTA.

Table 8. Summary of data factors and sources — continued

Category	Factor	Data source
Vehicle	Fatal crashes	DOT FARS — Fatal crash location
Weather	Precipitation	WFAS — Fire danger index data
	Relative humidity	WFAS — Fire danger index data
	Temperature	WFAS — Fire danger index data

In addition, the research efforts identified data factors that could not be sourced with national-based data. These factors are detailed in Table 9.

Table 9. Additional factors not analyzed

Category	Factor
Behavior	Alcohol use
	Playing with fire
	Smoke alarm presence
	Smoking
Fuel	Biomass
	Soil moisture
	Stand age
	Time from last fire
Land use	Distance to campsite
	Distance to farmland
	Distance to road
	Distance to town
	Industrial
	Livestock units
	Park
	Paths
	Privately owned
	Road density
	Secondary housing
	Urban
Topography	Altitude
	Aspect
	Elevation
	Slope
Vehicle	Age of vehicle
	Number of vehicles
	Type of vehicle
Weather	Lightning
	Wind speed

Relating demographic and socioeconomic data

Data from the ACS, collected by the U.S. Census Bureau, can be used to complement the fire data from the NFIRS. The U.S. Census is taken every 10 years, but the ACS data is an ongoing, confidential survey in which results are released every year. Over two million surveys are conducted each year, and socioeconomic data is collected at differing geographic levels, including age, sex, race, income, insurance and housing.

ACS data consists of "period estimates." Essentially, it represents population and housing characteristics over a specific data collection period. For this data, the U.S. Census Bureau releases estimates for five years, three years, and a single year, depending on the population thresholds for the specific geographic areas. For geographic areas with populations of 65,000 or more, the collected data produces one-year estimates; for geographic areas with smaller populations (20,000 or more), three years of data are necessary to produce estimates; for the smallest population areas, five years of data are necessary. These are called "one-year," "three-year," and "five-year" estimates (Guidance for Data Users Main, 2013).

The ZCTA estimates that were selected are five-year estimates. This ACS data at the ZCTA level provides a rich set of demographic data that can be used to analyze and predict fire incidence.

Relating weather data, land use, fuel and topography

As reviewed in the background research, there has been extensive research linking weather data, land use, fuel and topography with fire risk and wildfires.

The U.S. Forest Service (USFS) Wildland Fire Assessment Program collects a number of these factors to calculate a "fire danger rating" that includes "current and antecedent weather, fuel types, and both live and dead fuel moisture."[21] This data is collected in their Geographic Information System called the WFAS. Working with a USFS researcher, shape-files from the U.S. Census Bureau and the WFAS were combined to develop ZCTA-based estimates for average annual temperature, precipitation and humidity.[22]

Land use, fuel or topography factors were not analyzed due to a lack of availability of this data on a ZIP/ZCTA basis on a national level. These factors may have additional explanatory power if they could be gathered.

Relating business population data

The general thesis for fire incidence estimation is that fire incidence is largely related to the presence of people. Demographic and socioeconomic data have been analyzed with respect to residential fire incidence, but additional factors are typically included when analyzing outdoor or wildfire incidence. For example, many of the land use factors identified, such as distance to towns, campsites, parks or farmland, are essentially identifying the locations where people work or play.

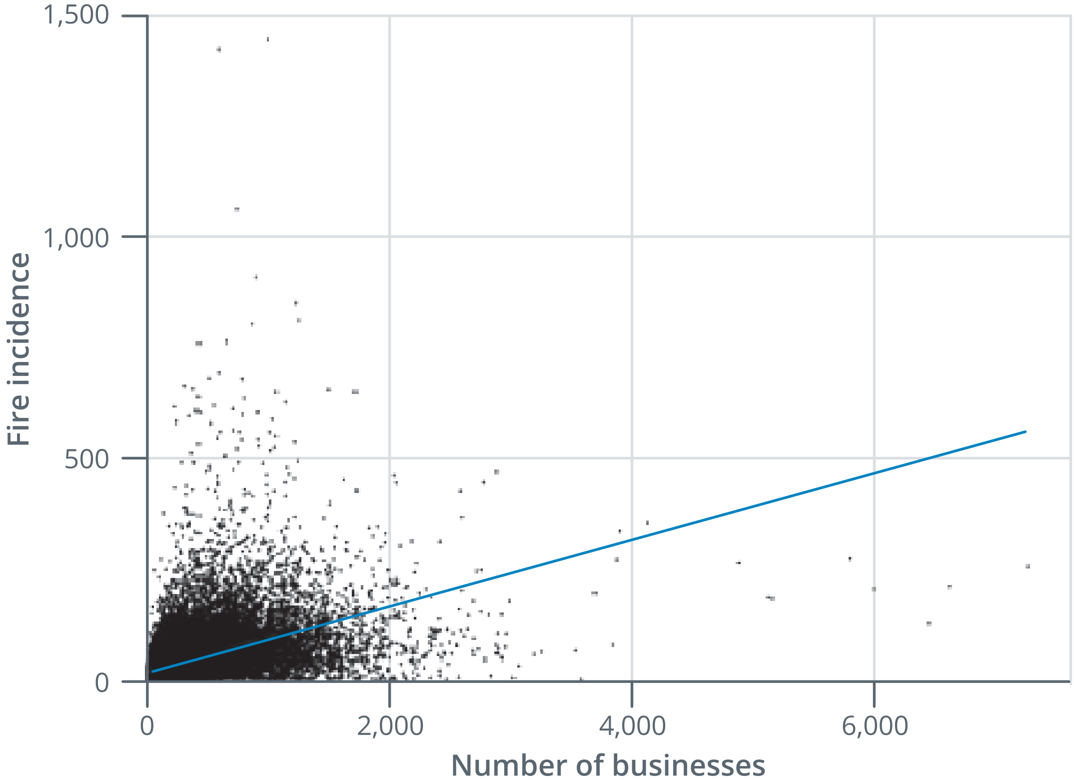
Unfortunately, 'daytime population' is not accurately captured on a national basis at a ZIP code level. It was postulated that the number of businesses could be a potential proxy for this information. The U.S. Census Bureau's CBP collects the total number of businesses

[21]http://www.wfas.net/index.php/fire-danger-rating-fire-potential--danger-32.
[22]W. Matt Jolly, Ph.D., Research Ecologist, Project Administrator, WFAS, U.S. Department of Agriculture, USFS. February 2015.

at a ZIP code level by the North American Industry Classification System (NAICS) codes. NAICS was developed as the standard for use by federal statistical agencies. It is used to classify business establishments for the collection, tabulation, presentation and analysis of statistical data describing the U.S. economy. The U.S. Census Bureau assigns and maintains only one NAICS code for each establishment, based on its primary activity. This can be used as a direct measure for the number of businesses within a geographic area.[23]

An analysis was conducted to compare the total number of businesses located within a ZCTA and the fire incidence for that ZCTA. A scatter plot of those two factors is found in Figure 3. The simple linear regression analysis showed an R-squared (R^2) value of 0.2478. Since roughly 25 percent of the data is accounted for by the model, business activity was included in the pool of variables tagged for further analysis.[24]

Figure 3. Number of businesses and fire incidence



Relating vehicle crash data

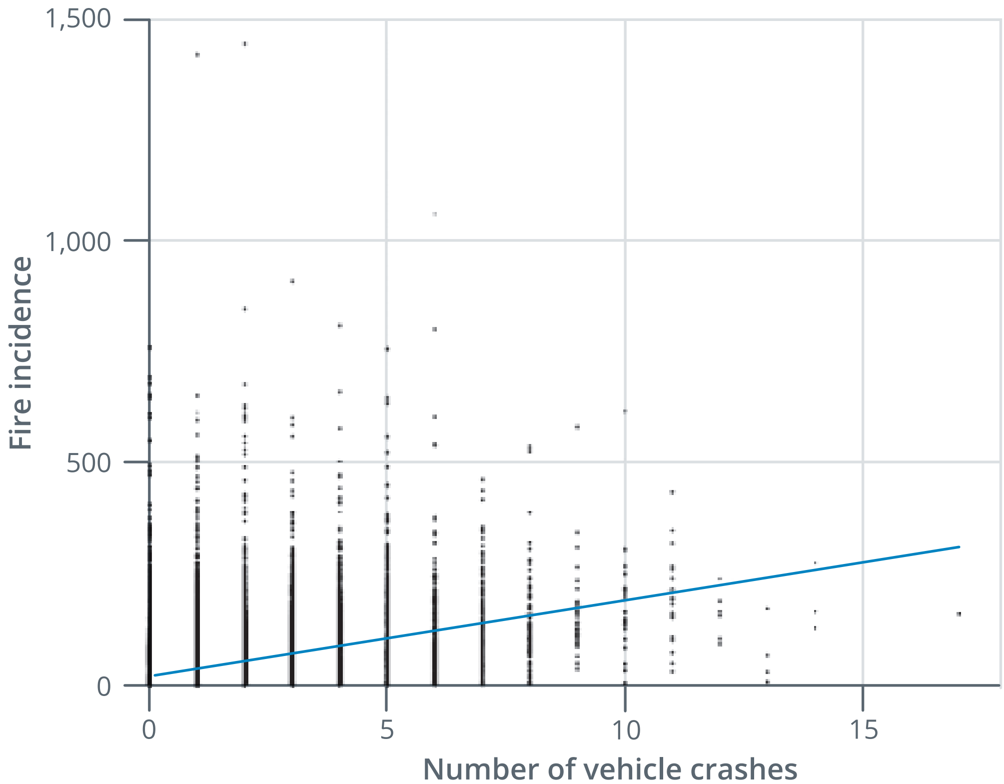
One of the most challenging aspects of identifying the location and volume of fire incidence is accounting for vehicle fires. Vehicle crash data was leveraged to identify locations that might have substantial travel volume that were separate and distinct from where people work or live. The Department of Transportation's (DOT's) National Highway Traffic Safety Administration collects fatal vehicle crash data on an incident basis in the FARS.[25] There

[23]In addition to the number of establishments, the U.S. Census Bureau collects employment populations as part of the CBP, but this data was not available in 23.5 percent of ZIP codes and was excluded from further consideration.
[24]A log transform on fire incidence was considered, but it was discarded as it did not improve the model when including the other variables.
[25]http://www.nhtsa.gov/FARS.

are approximately 30,000 fatal car crashes per year in the U.S. One of the reported fields for fatal crashes is the crash location, which is recorded with a latitude and a longitude. A commercial geo-location service was used to transform the latitude-longitude into a ZIP code for each incident and to summarize the number of incidents by ZIP code. A ZIP code was identified for 99 percent of the latitude-longitude pairs.

An analysis was conducted to compare the total number of vehicle crashes located within a ZCTA and the fire incidences for that ZCTA. Figure 4 shows a scatter plot of these two factors. The simple linear regression analysis showed an R^2 value of 0.1946 (Figure 4). As nearly 20 percent of the data is accounted for by the model, the variable was included in the pool of variables tagged for further analysis.[26]

Figure 4. Number of vehicle crashes and fire incidence



Estimating fire incidents — Regression model

The goal of the model is to estimate fire incidence in areas for which no incident data are available, but demographic, socioeconomic, business, vehicle crash, and climate data are available. Other models estimate specific incidence rates (Jennings, 1999) (Corcoran et al., 2011); however, this research and model look at the entire fire problem, including residential, vehicle and outside/wildfire incidents for the whole country. This model, therefore, contains elements that help predict each type of fire.

A multivariate regression model was developed using incident data, including demographic and socioeconomic variables (population density, older population size, housing stock age, vehicles present at homes, owner occupancy, housing vacancy, presence of children, single

[26]A log transform on fire incidence was considered, but it was discarded as it did not improve the model when including the other variables.

parent households rate, married with no children, minority population, unemployment rate), environmental variables (precipitation, temperature, relative humidity), business variables (business population and density), and vehicle crash data as shown earlier in Table 8. The incident data included the effects of all of the data adjustment techniques, including accounting for aid, range errors, “true zeroes,” and missing data.

This linear regression produced an estimate for any ZIP code where there was no data. Of the 32,989 ZCTAs in the 50 states and District of Columbia, almost 87 percent have fire incident data present and, under the model, would not need to be estimated.

Lastly, a variable was added that was specifically needed because of the ZCTA based analysis. ZCTAs vary widely by the area of land that they cover. For example, ZCTA 99557, an area close to the center of Alaska, has a land area of over 13,400 square miles and is over 10 times larger than the state of Rhode Island. To properly account for any size discrepancy in ZCTAs, the area of land for each ZCTA was included in the model. The full set of regression variables is shown in Table 10.

Table 10. Variables and variable definitions used in regression model

Variable	Definition
Population density	The total number of persons residing in the ZCTA divided by the total land area in the ZCTA.
Older population	Total number of the population over the age of 65.
Total housing units built after 2010	Total number of housing units in the area built after 2010.
Total housing units built 2000 to 2009	Total number of housing units in the area built between 2000 and 2009.
Total housing units built 1990 to 1999	Total number of housing units in the area built between 1990 and 1999.
Total housing units built 1980 to 1989	Total number of housing units in the area built between 1980 and 1989.
Total housing units built 1970 to 1979	Total number of housing units in the area built between 1970 and 1979.
Total housing units built 1960 to 1969	Total number of housing units in the area built between 1960 and 1969.
Total housing units built 1950 to 1959	Total number of housing units in the area built between 1950 and 1959.
Total housing units built 1940 to 1949	Total number of housing units in the area built between 1940 and 1949.
Total housing units built before 1939	Total number of housing units in the area built before 1939.
Houses with vehicles	The total number of occupied housing units with at least one vehicle available.
Owner occupied units	The number of occupied housing units that are occupied by the owner.
Vacant unit	The percentage of housing units that are vacant.

Table 10. Variables and variable definitions used in regression model — continued

Variable	Definition
Children present	Number of households with their own children under 18 years of age.
Education level	Number of households with at least a high school degree.
Household size	Average household size.
Married, no children	The total number of households that are married, but who do not have children.
Single-parent household	The sum of the percentage of single-mother households with children, and the percentage of single-father households with children.
Minority population	Total population that is not white (not-Hispanic/ Latino).
Household income	Median household income.
Unemployment	Total population over 16, minus the population that is employed.
Precipitation	The yearly total precipitation for the closest weather station to that ZCTA.
Temperature	The yearly average of mean daily temperatures for the closest weather station to that ZCTA.
Relative humidity	The log of yearly average relative humidity, as measured by the closest weather station to the ZCTA.
Total businesses	Total number of businesses registered by NAICS within the ZIP code.
Fatal crashes	Number of fatal vehicle crashes within a ZIP code.
Total land area	The total land area contained within that ZCTA.

As shown in Table 11, each of the independent variables included in the model show high "t" statistics for their partial regression coefficients, and they are significant at the 0.001 level. The exception is "household income," which was significant at the 0.01 level. "Household size" and "married, no children" were found to not be significant and were removed from subsequent analysis. A variation inflation factor test was performed and no multicollinearity was detected for the variables in the model. A Breusch-Pagan test was performed, and heteroscedasticity is not present in the model.

Table 11. Multivariate regression coefficients

Coefficient	Standardized coefficients beta	Standard error	"t"
Intercept	(186.7000)	19.5300	(9.5600)
Population density	(0.0013)	0.0001	(13.3570)
Older population	(0.0118)	0.0006	(18.8310)
Total housing units built 2000 to 2009	0.0079	0.0004	17.8910
Total housing units built 1990 to 1999	0.0101	0.0006	16.3780
Total housing units built 1980 to 1989	0.0039	0.0006	6.3680
Total housing units built 1970 to 1979	0.0072	0.0006	11.7020
Total housing units built 1960 to 1969	0.0135	0.0007	18.1120
Total housing units built 1950 to 1959	0.0035	0.0007	4.7900
Total housing units built 1940 to 1949	0.0103	0.0010	10.5900
Total housing units built before 1939	0.0125	0.0004	32.8050
Houses with vehicles	(0.0139)	0.0005	(29.6570)
Owner occupied units	0.0072	0.0005	14.7940
Vacant unit	(19.9700)	2.0910	(9.5520)
Children present	(0.0087)	0.0009	(10.0260)
Education level	0.0062	0.0003	17.8220
Household size	(0.0614)	0.7409	(0.0830)
Married, no children	0.0020	0.0013	1.5340
Single-parent household	0.0493	0.0015	32.8070
Minority population	(0.0013)	0.0001	(16.4450)
Unemployment	0.0065	0.0003	21.4400
Household income	(0.0000)	0.0000	(3.2310)
Precipitation	0.0129	0.0013	9.6730
Temperature	0.5334	0.0749	7.1180
Relative humidity	11.0900	1.4930	7.4290
Total businesses	0.0233	0.0012	19.8530
Fatal crashes	4.2610	0.2068	20.6030
Total land area	0.0141	0.0017	8.2140

The R^2 value calculated for the model is 0.5849, indicating that 58.5 percent of the total variation in the incidence of fires has been explained by the independent variables. The remaining 41.5 percent has not been explained. The adjusted R^2 value for the model was 0.5844. An R^2 value greater than 0.35 is considered meaningful for the social sciences (Jost, n.d.). Additionally, this model compares favorably to other fire rate estimation models, as shown in Table 12 (Chhetri, 2010).

Table 12. Comparison to other models

Model	R^2	Adjusted R^2	Residual standard error
P. Chhetri (2010)	.451	.390	52.1
USFA	.585	.584	41.2

Multiyear analysis

The same factors were analyzed across multiple years to examine the general robustness of the model. For each of the three years, demographic and socioeconomic data were retrieved from the ACS, weather data from the WFAS, business data from the CBP, and crash data from the FARS. The Gazetteer ZCTA data was the same over the three years, as the boundaries of ZCTAs were consistent through the examined period.

Overall, the model was very robust, with the average adjusted R^2 value remaining over 0.57 across the three years analyzed (Table 13).

Table 13. Comparison of model across years (2009-2011)

Model	2009	2010	2011
Adjusted R^2	0.5725	0.5844	0.5669
Residual standard error	39.6	41.2	42.0

Estimation model

An estimation model was developed using the results of the regression analysis. The base regression model was used to estimate incidents in ZCTAs where no incident data was reported in the NFIRS and all associated predictive factors were available. If a negative estimate was predicted, a zero was substituted for the estimate of that ZCTA. There were 751 ZCTAs where not all of the multivariate data was present. In these ZCTAs, an estimate was calculated by using a linear regression based on population only. A range was estimated based on a 95 percent confidence interval for each estimate.[27]

Using either the multivariate or the population-only models, estimates were calculated for a total of 3,399 ZCTAs. The estimates were then added to the actual responses. In 2010, the total national estimate using the multivariate linear regression (MLR) model was 1,278,263. This compares to using the scaling estimate of 1,319,486 and the NFPA estimate of 1,331,500. This process was repeated for 2009 and 2011, and the results are summarized in Table 14.

[27]The margin of error for an individual ZCTA estimate was calculated using the standard error multiplied by the appropriate z-score (1.96 for 95 percent). The total range for the forecast was calculated by multiplying the margin of error by the total number of estimated ZCTAs, and dividing by the square root of the sample size.

Table 14. Summary of national estimates by different methods

Model	2009	2010	2011
NFIRS valid incidents	948,089	1,003,556	999,457
NFIRS unduplicated aid	1,124,038	1,178,225	1,181,115
National estimate — scaling	1,283,989	1,319,486	1,326,666
National estimate — MLR	1,233,667	1,278,263	1,287,174
National estimate (MLR) — range	± 4,519	± 4,709	± 4,666
NFPA	1,348,500	1,331,500	1,389,500

Limitations

There are some limitations to the current model, in particular with missing data and the application of estimates. The current model, based on the ZCTA, is capable of producing estimates for state and county level fire incidence, but the estimates will have higher margins of error because of trending and aid given effects.

The missing data model substitutes an average if the monthly reported number for that fire department is below two standard deviations from the mean. Currently, the model does not include any yearly trend factors for fire incidence; therefore, a ZCTA where population is changing quickly may be underestimated or overestimated. It is believed that this factor is not significant at the national level, but this may create errors at the county or state level.

The estimation model substitutes an estimate for a ZCTA if there is no data for that ZCTA. If a ZCTA has a high fire incidence and the fire department primarily responsible for that area does not participate in the NFIRS, the model result could be underestimated for that ZCTA. If the high fire incidence ZCTA has aid incidents reported by participating fire departments, then those aid incidents will be presumed to be the incidence for the entire ZCTA. A preliminary analysis was conducted to correct this underprediction. The analysis compared the actual numbers and estimates for each ZCTA, and substituted the prediction if the ZCTA was underestimated by two standard deviations from the mean. The analysis identified only 248 ZCTAs out of 33,120 where this post-estimation fix would be applied. This is within natural probability of such an error. While it is not believed that this factor is significant at the national level, it may be a factor for state or county level estimates.

Conclusions and Further Work

It is possible to develop a method of generating an annual estimate of fire incidence for the U.S. based on fire incident data from the NFIRS, in conjunction with other publicly available data. Two models were examined: a simple scaling model using population protected and a more sophisticated multivariate regression model using a combination of demographic, socioeconomic, weather, business, and vehicle use data. Both models produced similar estimates, within two percent of each other.

During the course of analysis, two major improvements were made relative to existing NFIRS data: identifying additional incidents within the NFIRS where aid is given using advanced geolocation techniques and imputing missing values with more reliable data from a prior year. The accurate identification of aid incidents reveals an additional 14 to 18 percent of unique incidents. The imputation of missing data further identifies an additional 7 to 9 percent of incremental incidents. This total of 21 to 27 percent of additional incidents is a very substantial increase and significantly adds to the overall number of reported U.S. fires.

Population appears to be the most important predictive variable in estimating fire incidents; however, a greatly improved estimation model can be built by leveraging a combination of additional socioeconomic, climate, business and vehicle data. Identifying the location of population at home, work, and recreational areas, as well as their movement between these places, is important to accurately estimating the location of incidents at a ZCTA level.

If they could be gathered, other factors, including land use, fuel or topography, could add additional explanatory power to the model. Additionally, interaction and higher-order terms may provide additional explanatory power to the regression model.

As the estimates are made at the ZCTA level, it may be possible to create estimates at the state or county level by aggregating the ZCTAs appropriately. Because NFIRS participation is not necessarily uniform over small geographic areas and may lead to large estimation errors, further work may be necessary to examine the robustness of this approach.

References

Ahrens, M. (2005). *U.S. vehicle fire trends and patterns.* Quincy, MA: National Fire Protection Association.

Amatulli, G., Rodrigues, M., Trombetti, M., & Lovreglio, R. (2006). Assessing long-term fire risk at local scale by means of decision tree technique. *J. Geophys. Res., 111*(G4). doi:10.1029/2005JG000133

Asgary, A., Ghaffari, A., & Levy, J. (2010). Spatial and temporal analyses of structural fire incidents and their causes: A case of Toronto, Canada. *Fire Safety Journal, 45*(1), 44–57.

Brotons, L., Aquilué, N., de Cáceres, M., Fortin, M., & Fall, A. (2013). How fire history, fire suppression practices and climate change affect wildfire regimes in mediterranean landscapes. *PLoS ONE, 8*(5). doi:10.1371/journal.pone.0062392

Bunn, T., Slavova, S., & Robertson, M. (2012). Crash and burn? Vehicle, collision, and driver factors that influence motor vehicle collision fires. *Accident Analysis and Prevention,* 140–145. doi:10.1016/j.aap.2011.10.008

Chhetri, P., Corcoran, J., Stimson, R., & Inbakaran, R. (2010). Modelling potential socioeconomic determinants of building fires in south east Queensland. *Geographical Research, 48*(1), 75–85. doi:10.1111/j.1745-5871.2009.00587.x

Clare, J., Len, G., Plecas, D., & Jennings, C. (2012). Reduced frequency and severity of residential fires following delivery of fire prevention education by on-duty fire fighters: Cluster randomized controlled study. *Journal of Safety Research, 43*(2), 123–128. doi:10.1016/j.jsr.2012.03.003

Cohn, D. (2011, May 4). *Imputation: Adding people to the census.* Retrieved from http://www.pewsocialtrends.org/2011/05/04/imputation-adding-people-to-the-census/.

Corcoran, J., Higgs, G., Brunsdon, C., Ware, A., & Norman, P. (2007). The use of spatial analytical techniques to explore patterns of fire incidence: A South Wales case study. *Computers, Environment and Urban Systems, 31*(6), 623–647.

Corcoran, J., Higgs, G., & Higginson, A. (2011). Fire incidence in metropolitan areas: A comparative study of Brisbane (Australia) and Cardiff (United Kingdom). *Applied Geography, 31*(1), 65–75. doi:10.1016/j.apgeog.2010.02.003

Cortez, P., & Morais, A. (2007). A data mining approach to predict forest fires using meteorological data.

Díaz-Delgado, R., Lloret, F., & Pons, X. (2004). Spatial patterns of fire occurrence in Catalonia, NE, Spain. *Landscape Ecology, 19*(7), 731–745.

Dutta, R., Aryal, J., Das, A., & Kirkpatrick, J. (2013). Deep cognitive imaging systems enable estimation of continental-scale fire incidence from climate data. *Nature*, 3188. doi:10.1038/srep03188

Eltinge, J. L., Kozlow, R. A., & Luery, D. M. (2003). *Imputation in three federal statistical agencies.* Retrieved from https://www.bls.gov/advisory/fesacp1101703.pdf.

Fahy, R., & Norton, A. (1989, January/February). How being poor affects fire risk. *Fire Journal, 83*(1), 28–36.

Federal Emergency Management Agency, U.S. Fire Administration, & National Fire Data Center. (1997). *Socioeconomic Factors and the Incidence of Fire.* Research Report. Emmitsburg, MD. Retrieved from https://www.usfa.fema.gov/downloads/pdf/statistics/socio.pdf.

Federal Emergency Management Agency, U.S. Fire Administration, & National Fire Data Center. (1998). *An NFIRS analysis: Investigating city characteristics and residential fire rates.* Research Report. Emmitsburg, MD.

Fox, D., Martin, N., Carrega, P., Andrieu, J., Adnés, C., Emsellem, K., . . . Fox, E. (2015). Increases in fire risk due to warmer summer temperatures and wildland urban interface changes do not necessarily lead to more fires. *Applied Geography,* 1–12. doi:10.1016/j.apgeog.2014.10.001

Garcia, C., Woodward, P., Titus, S., Adamowicz, W., & Lee, B. (1995). A logit model for predicting the daily occurrence of human caused forest fires. *International Journal of Wildland Fire, 5*(2), 101–111.

Greene, M. A., Smith, L. E., Levenson, M. S., Hiser, S., & Mah, J. C. (2001). *Raking fire data. Proceedings of 2001 FCSM research conference.* Washington, DC: Federal Committee on Statistical Methodology, Office of Management and Budget. Retrieved from https://fcsm.sites.usa.gov/files/2014/05/2001FCSM_Greene.pdf.

Guidance for data users main. (2013). Retrieved from https://www.census.gov/acs/www/guidance_for_data_users/guidance_main/ (accessed January 31, 2016).

Gunther, P. (1981). Fire-cause patterns for different socioeconomic neighborhoods in Toledo, OH. *Fire Journal, 75*(3), 52–58.

Gunther, P. (1982). Rural fire deaths: The role of climate and poverty. *Fire Journal.*

Guyette, R., Stambaugh, M., Dey, D., & Muzika, R. (2012). Predicting fire frequency with chemistry and climate. *Ecosystems,* 322–335.

Hall, J. (1993). *The U.S. fire problem overview through 1992: Leading causes and other patterns and trends.* Quincy, MA: National Fire Protection Association.

Hall, J. R., & Harwood, B. (1989, May). The national estimates approach to U.S. fire statistics. *Fire Technology, 25*(2), 99–113. doi:10.1007/BF01041420

Harris, S. A. (2012). The relationship between fire behaviour measures and community loss: An exploratory analysis for developing a bushfire severity scale. *Natural Hazards, 63*(2), 391–415.

Harvey, P. A. (2004). Strategies to increase smoke alarm use in high-risk households. *Journal of Community Health, 29*(5), 375–386.

Imputation of unreported data items. (2013, April). Retrieved from https://www.census.gov/cps/methodology/unreported.html.

Jennings, C. (1996). *Urban residential fires: An empirical analysis of building stock and socioeconomic characteristics for Memphis, Tennessee.* Cornell University.

Jennings, C. (1999). Socioeconomic characteristics and their relationship to fire incidence: A review of the literature. *Fire Technology, 35*(1), 7–34.

Jost, Steve. (n.d.). *Linear correlation.* Retrieved from http://condor.depaul.edu/sjost/it223/documents/correlation.htm.

Karter, J., & Donner A. (1978, January). The effect of demographics on fire rates. *Fire Journal, 72*(1), 53–65.

Lizhong, Y., Heng, C., Yong, Y., & Tingyong, F. (2005). The effect of socioeconomic factors on fire in China. *Journal of Fire Sciences, 23*(6), 451–467.

Lizhong, Y., Xiaodong, Z., Zhihua, D., & Weicheng, F. (2002). Fire situation and fire characteristic analysis based on fire statistics of China. *Fire Safety Journal, 37*(8), 785–802.

Mandallaz, D., & Ye, R. (1997). Prediction of forest fires with Poisson models. *Canadian Journal of Forest Research, 27*(10), 1685–1694.

Munson, M. (1975). *Urban neighborhoods and the fear of fire: Towards the specification of a measure of the quality of life*. Princeton, NJ: Princeton University.

Munson, M., & Oates, W. (1983). Community characteristics and the incidence of fire: An empirical analysis. *The Social and Economic Consequences of Residential Fires.* Lexington, MA: D.C. Heath and Co.

Nicolopoulos, N. (1997, June). *Socio-economic characteristics of communities and fires.* New South Wales Fire Brigades.

Roberts, I. (1996). Smoke alarm use: Prevalence and household predictors. *Injury prevention, 2*(4), 263–265.

Romero-Calcerrada, R., Barrio-Parra, F., Millington, J., & Novillo, C. (2010). Spatial modelling of socioeconomic data to understand patterns of human-caused wildfire ignition risk in the SW of Madrid (central Spain). *Ecological Modelling, 221*(1), 34–45. doi:10.1016.j.ecolmodel.2009.08.008

Runyan, C., Bangdiwala, S., Linzer, M., Sacks, J., & Butts, J. (1993). Risk factors for fatal residential fires. *Fire Technology, 29*(2), 183–193.

Runyan, C., Johnson, R., Yang, J., Waller, A., Perkins, D., Marshall, S., . . . McGee, K. (2005). Risk and protective factors for fires, burns, and carbon monoxide poisoning in U.S. households. *American Journal of Preventive Medicine, 28*(1), 102–108.

Schaenman, H. J. (1977). *Procedures for improving the measurement of local fire protection effectiveness.*

Serra, L., Saez, M., Juan, P., Varga, D., & Mateu, J. (2013). A spatio-temporal Poisson hurdle point process to model wildfires. *Stochastic Environmental Research and Risk Assessment,* 1671–1684. doi:10.1007/s00477-013-0823-x

Shai, D. (2006, March-April). Income, housing, and fire injuries: A census tract analysis. *Public Health Reports, 121*(2), 149–156.

United States Census Bureau. (n.d.). *When to use 1-year, 3-year, or 5-year estimates.* Retrieved from https://www.census.gov/acs/www/guidance_for_data_users/estimates/ (accessed January 31, 2016).

Wotton, B., Martell, D., & Logan, K. (2003). Climatic change and people-caused forest fire occurrence in Ontario. *Climate Change*, 275–295.

Xin, J., & Huang, C. (2013). Fire risk analysis of residential buildings based on scenario clusters and its application in fire risk management. *Fire Safety Journal,* 72–78.

Xu, D., Guofan, S., Dai, L., Hao, Z., Tang, L., & Wang, H. (2006). Mapping forest fire risk zones with spatial data and principal component analysis. *Science in China, Series E: Technological Sciences, 49*(1), 140–149.

Acronyms

ACS	American Community Survey
CBP	County Business Patterns
DOT	Department of Transportation
FARS	Fatality Analysis Reporting System
MLR	multivariate linear regression
NAICS	North American Industry Classification System
NFDC	National Fire Data Center
NFIRS	National Fire Incident Reporting System
NFPA	National Fire Protection Association
PDR	Public Data Release
R^2	R-squared
USFA	U.S. Fire Administration
USFS	U.S. Forest Service
WFAS	Wildland Fire Assessment System
ZCTA	ZIP Code Tabulation Area

U.S. Fire Administration
Working for a fire-safe America

16825 South Seton Ave.
Emmitsburg, MD 21727
www.usfa.fema.gov

May 2017