U.S. Residential Fire Deaths in the News: Demographics and Circumstances

Authors:
Yunxuan He - yh7bb
Ben Schenck - bls7cc
Rebecca Stein - rcs8bg
Yiting Tsai - yt9mh

Introduction

Residential fires killed 3,400 Americans in 2017, according to the U.S. Fire Administration. $23 billion was lost to residential fire damage in the same year. Analysing data on the circumstances of residential fires could uncover patterns that would be useful to public sector actors trying to prevent tragic deaths and injuries in their communities, to home builders trying to identify which design features help fire-proof a home, and to insurance companies trying to evaluate the risk of fire related claims on a given property.

For our project, we decided to focus on the impact of residential fires from a public health perspective: we analyzed a dataset of U.S. deaths due to residential fires in 2018 to identify trends incident circumstances and victim demographics.

Our goal was to describe the data in a way that a member of the public with little subject-matter or statistical knowledge could readily interpret. We did not attempt to compute inferential statistics or predictive models because we did not have a sample of victims who survived with which to compare the fatality data. We approached our goal visually, calculating summary statistics and manipulating certain parameter values as an intermediary step to our final product: clear plots and an interactive interface that displays statistics based on user selections.

Data

The data originated from various news sources and was compiled into a dataset by the U.S. Fire Administration ("USFA"), which runs daily internet searches for media reports of residential fire deaths.

The raw data included the following columns: Date, State, City, Age, Age Demographic, Gender, Description, Media URL, Housing Type, Smoke Alarm, Disability (Y/N), and MFI. The first ten are self-explanatory. MFI indexes the deaths for each incident (ex. 1 of 3, 2 of 3, 3 of 3). We removed the Description column because it only contained redundant information, and the Media URL column because it was irrelevant to our analysis. We converted the MFI column to integer values that simply represented the total number of fatalities in the incident that killed a particular victim (for example, where three victims of the same incident would have had MFI string values of "1 of 3", "2 of 3", "3 of 3" in the raw data, they would all have MFI integer values of 3 in our processed data). This enabled us to easily count deaths with aggregate functions provided by the Pandas library. We converted the incident dates, which had been stored as strings, to datetime values, allowing us to plot the number of incidents as a time series. We also converted Age, which was stored as string values, to integer values so that we could compute measures of central tendency.

As mentioned above, USFA pulled this dataset from news sources, and does not update data once entered, so it only includes information available to the media within the first 24-48 hours after a deadly residential fire. Consequently, there are a large number of unknowns, especially under the "Cause" and "Smoke Alarm" parameters. Removing all unknowns was not an option because it would have limited the dataset too severely. Instead, we created new sub-datasets that filtered out unknowns for variables of interest. We had a dataset that only included victims with known ages, a dataset that only included incidents with known causes, and a dataset included incidents with fire alarms working appropriately.

We also used several groupings of the original pre-processed dataset to explore geographic and temporal trends. Using the groupby method in Pandas, we grouped deaths by date, city, and state to count the total number of incidents, and grouped by date and state separately to determine the deadliest days and states with the highest or lowest death counts.

Methods and Analysis

- Computational methods

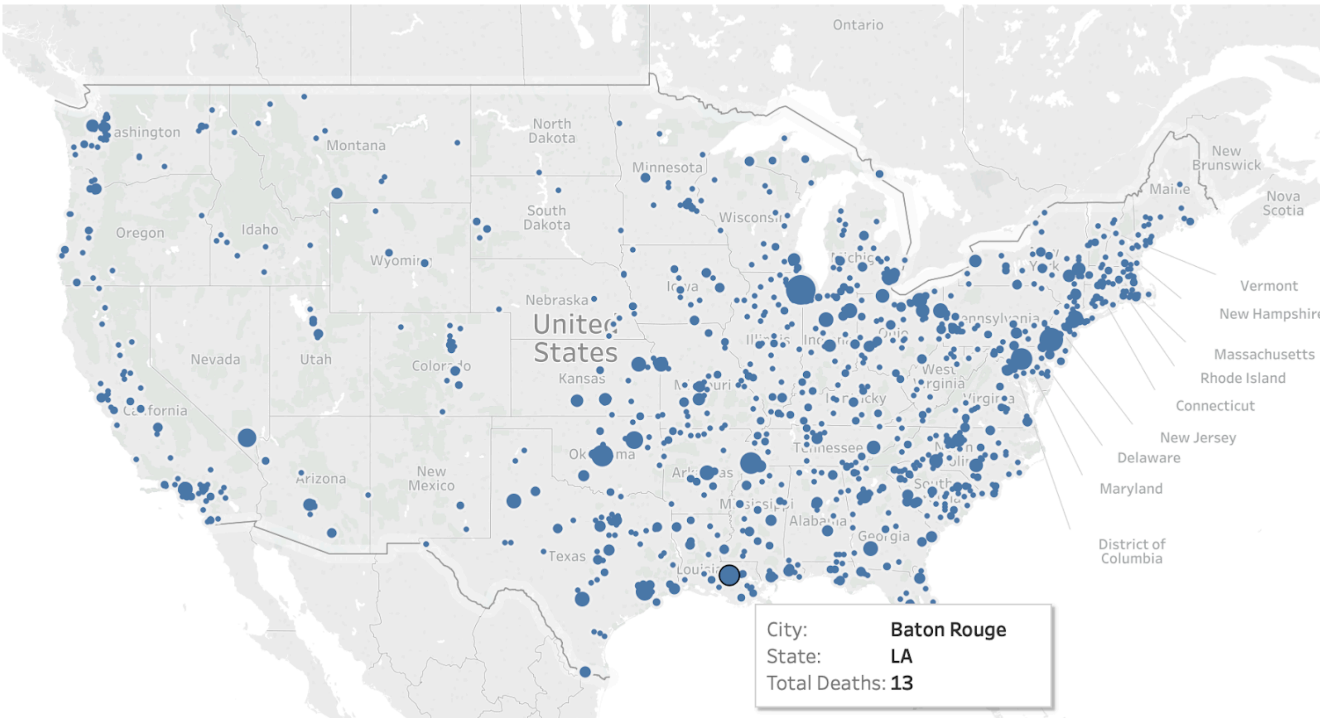
To analyze the number of deaths geographically, we needed a relative measure to account for population size differences. We pulled U.S. Census Bureau population data by state for 2018, grouped the total deaths in our original dataframe by state, then joined the population and death data on the "State" column. We then re-calculated the number of deaths as a proportion of state population.

Using the aggregate functions available in the Python Pandas library, we found the mean age, maximum age, and minimum age of victims by state. For example, we found that New York was the state with the highest number of deaths in absolute terms; however, West Virginia was the deadliest state in relative terms, with .002% of the population perishing in residential fires.

- Graphical methods

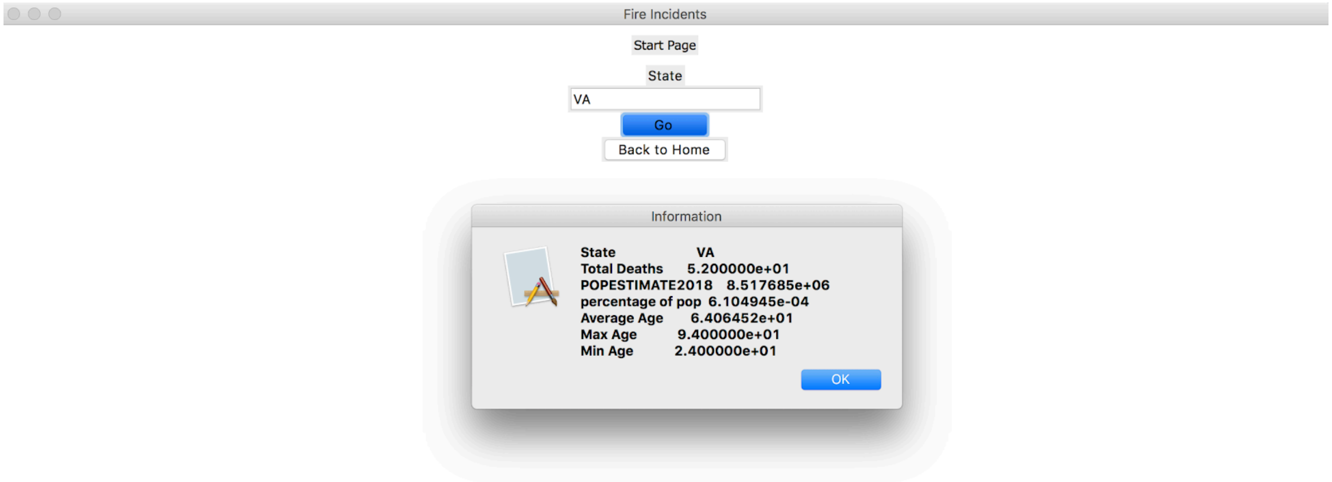
Charts that we generated using Tableau illuminated trends temporally, geographically, and demographically. We plotted the number of nationwide incidents by month and observed an approximate parabolic pattern: with January being the deadliest month, deaths decreasing during the spring and summer, and then creeping up again in the fall. We then isolated incidents by cause and observed that heating related fires largely explained the

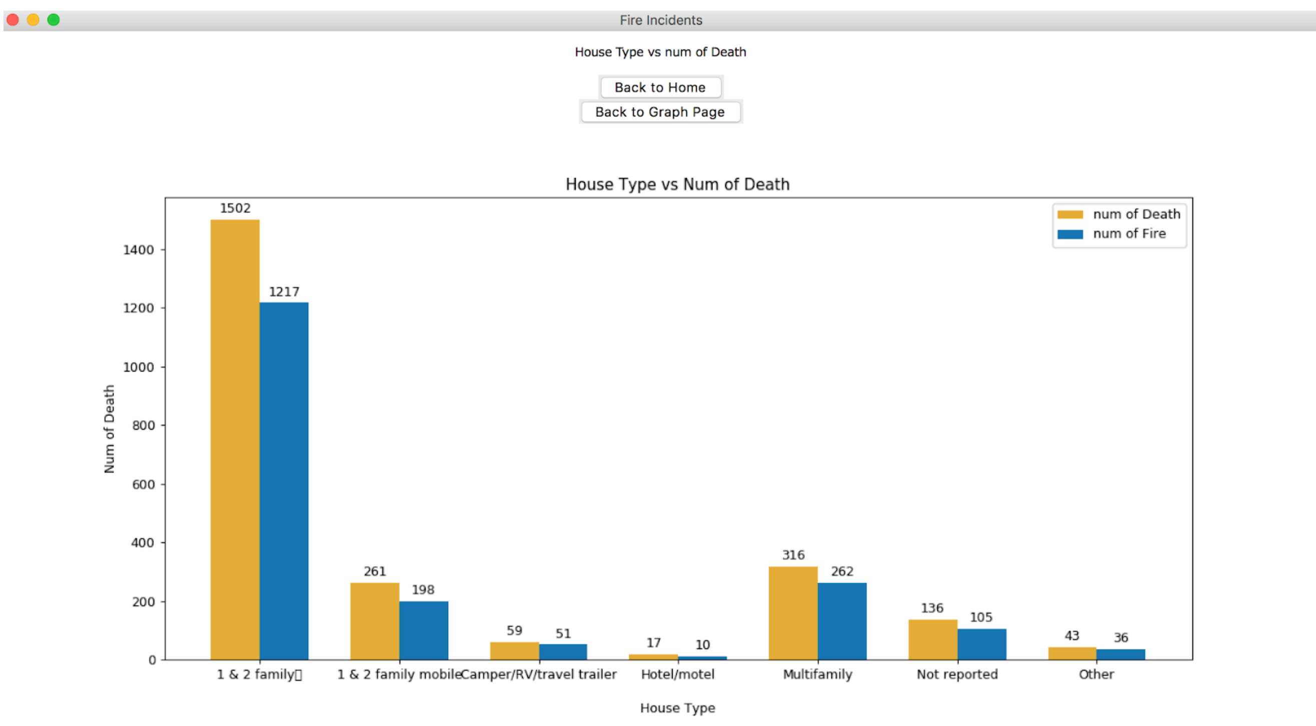
January and December spikes. Tableau includes an automatic geocoding and mapping feature, so we were able to plot the incidents as points on a map of the United States. To further visualize the data geographically, we added number of deaths per city as a parameter translating to point size, so cities with more deaths are plotted with larger points. An example with Baton Rouge, LA highlighted can be seen below.



User Interaction Features

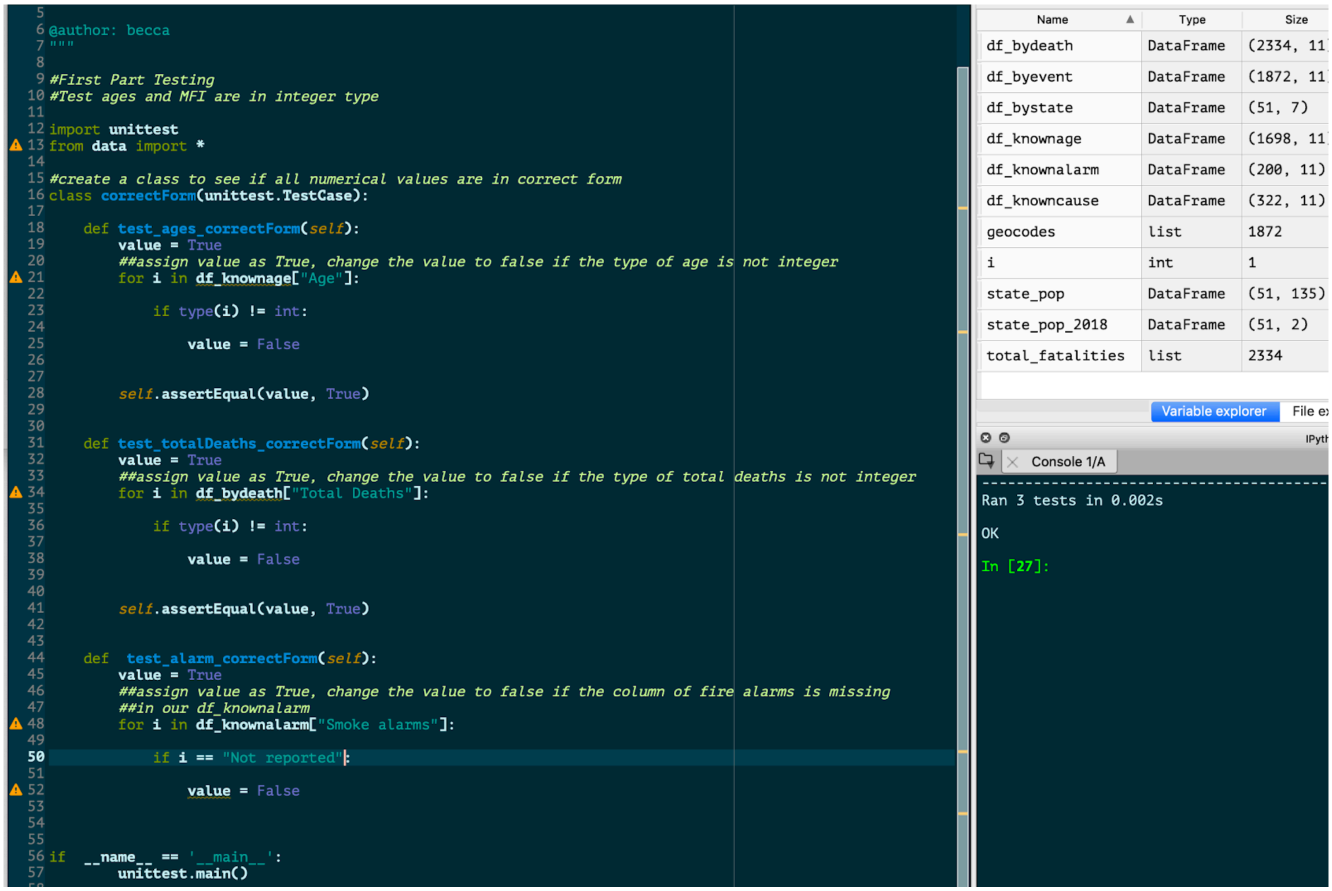
Using the tkinter library, we built a simple graphical user interface that allows users to view summary statistics from selected states and analysis graphs. The program works by importing the csv files generated as output of the pre-processing and analysis modules rather than referencing those modules directly. With this approach, it would take less editing to apply the GUI to different datasets if one chose to do so. In addition, we can obtain the graphs that we need immediately. An example of VA statistics and graphic interface is as below.



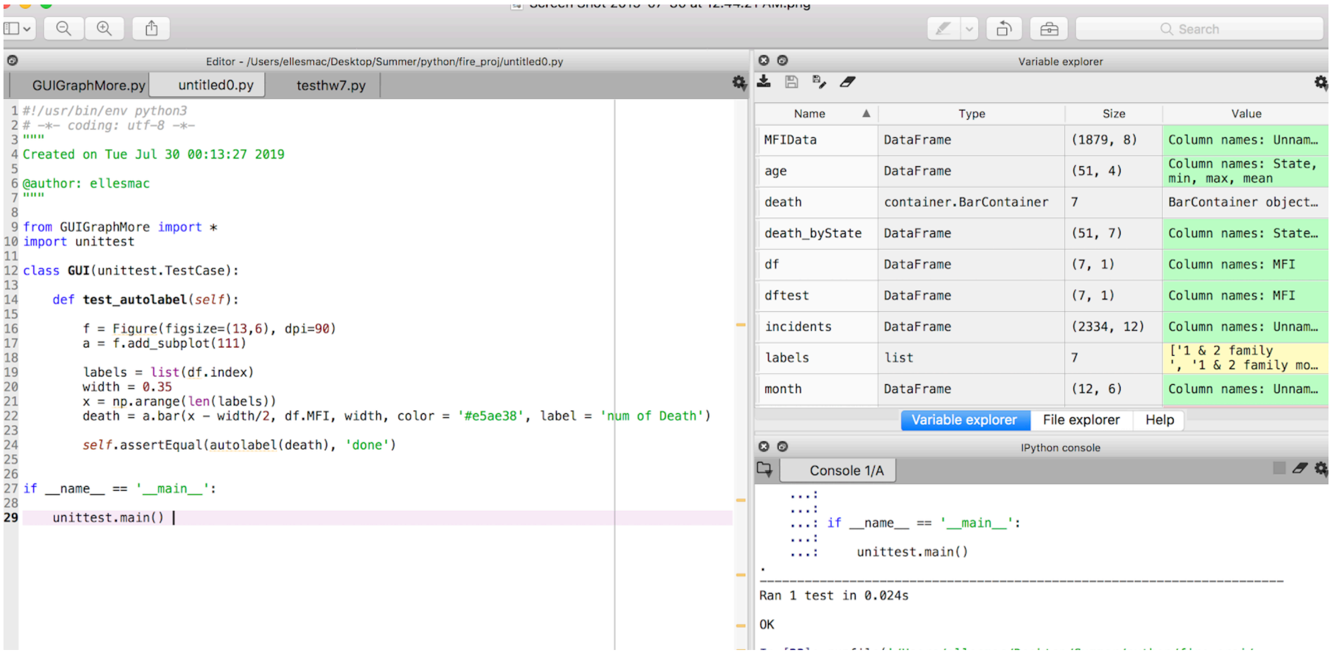


Testing

We created and ran two testing modules for this project. One test is for our pre-processing and analysis program and the other for our user interaction program, which depended on correct output from the pre-processing module. Using the unittest library, we tested that our preprocessing code a) correctly converted all known age values to integers; b) correctly reformatted all "Total Death" values and converted them to integers, and c) successfully removed observations where smoke alarm presence was unknown for the df_knownalarm dataframe. The output for this test is below.

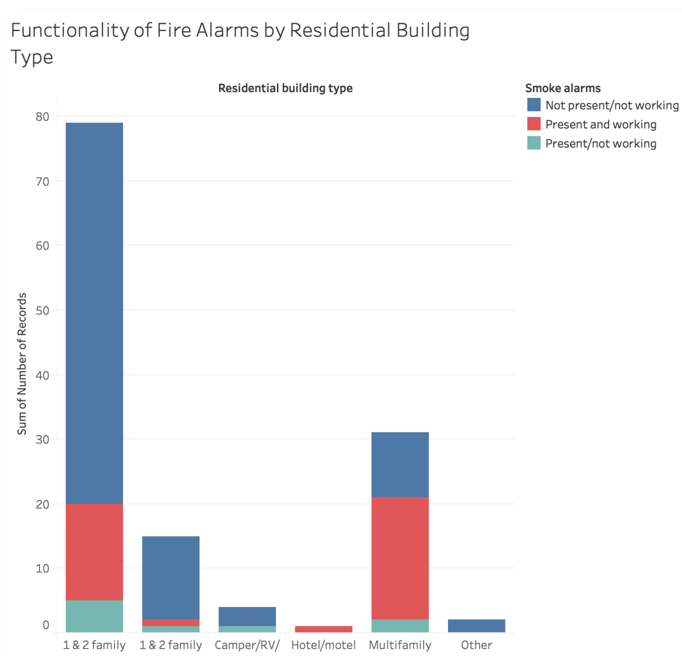


For our user interaction module, we tested whether GUI correctly labeled the "Graph Page" section. The output for this test is shown below:



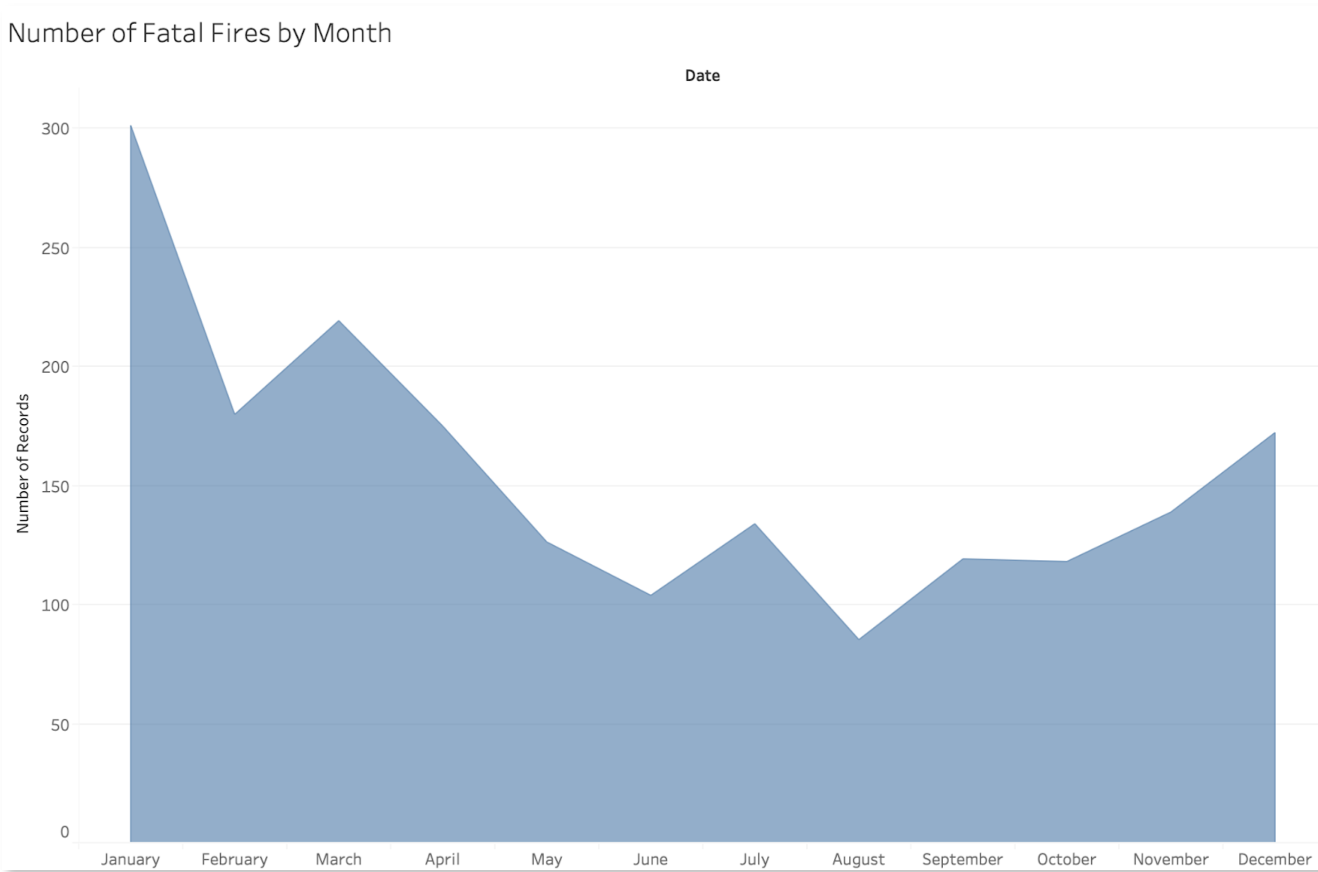
Conclusions

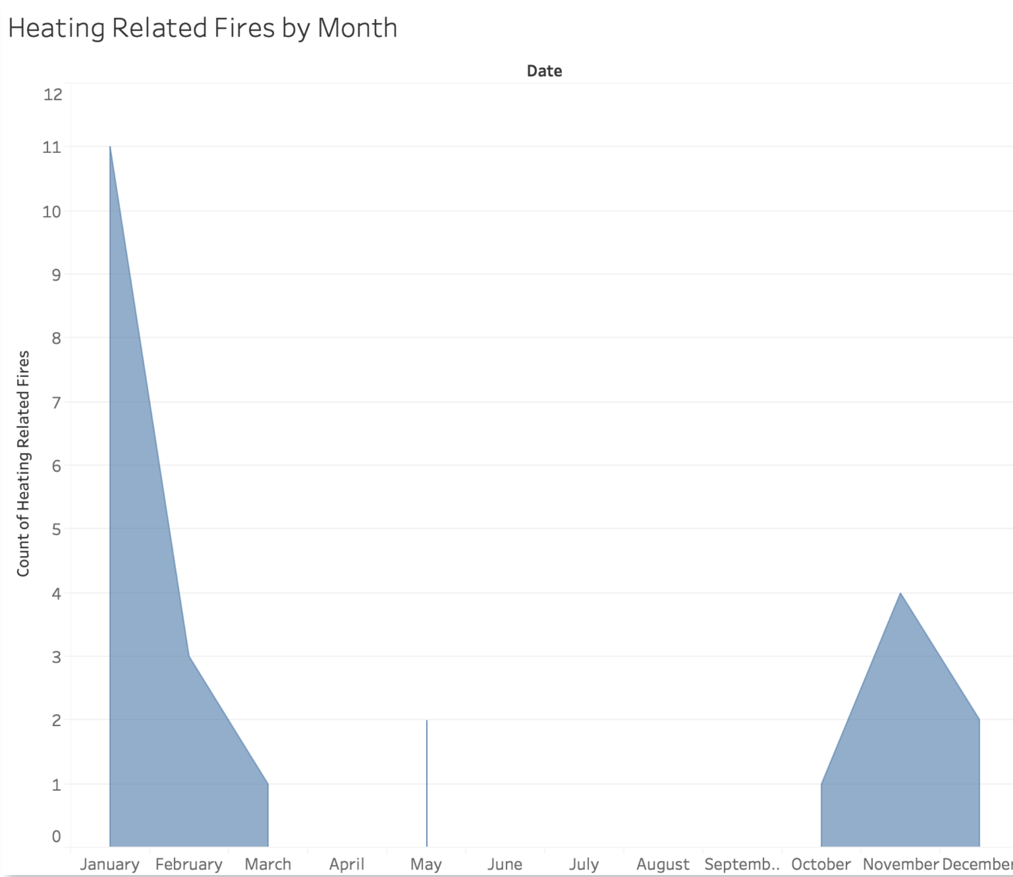
It appears that a larger proportion of incidents involved residences without fire alarms present at all, except in the case of multifamily residences (see: *Functionality of Fire Alarms by Residential Building* below). Approximately 81% of incidents pertaining to single or double family residences involved an absent or dysfunctional fire alarm. On the other hand, only



about 39% of incidents pertaining to multi-family homes involved an absent or dysfunctional fire alarm.

From the chart showing the total number of fatal fires by month, it appears that there is a higher frequency in the winter months. While more research should be conducted to determine causality, this is likely due to a combination of drier air as well as increased use of space heaters and electric blankets. As further evidence of this, in the graph of heating related fires by month, it appears that these are almost exclusively occurring in the winter months.





Further Study

There are many directions that can be taken to further the study of fire-related deaths. In our analysis, we looked at records of fire fatalities as reported in the news. Moving forward, it would be interesting to acquire data on all fire-related fatalities in the country and compare the analysis to our data. From this, we could gather insights as to which factors (ages, genders, causes, etc.) are correlated with being reported on the news. Also, similar to how we found trends in the data based on the month, by obtaining more records going back several years it may be possible to detect more trends over a longer timespan. Furthermore, acquiring data on non-fatal fires would provide new insights into the analysis that we performed. With this data, research could be conducted into predictive analysis, figuring out if we could predict whether or not certain factors are likely to produce a fatality.