**Thomas Fire Estimated Housing Damage & Sentiment**

**Team:** Safaa, Shemelis, Shey, Shoujoun

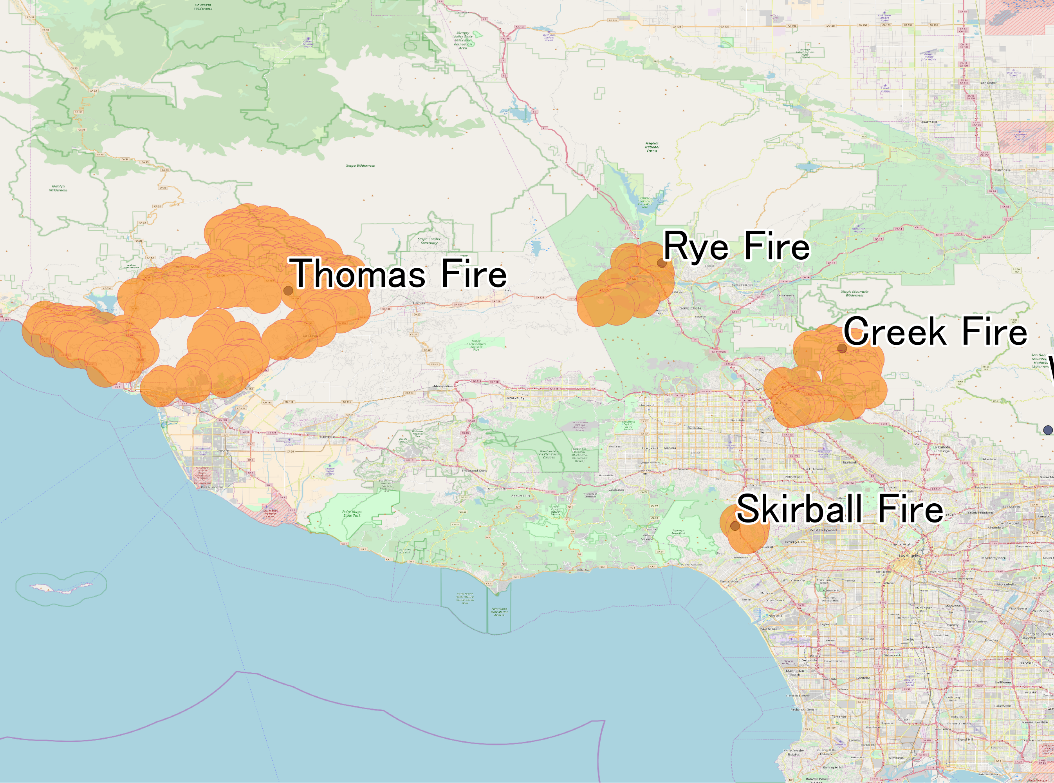
**Abstract**

The aim of this project is to assess the general sentiment of residents in the areas covered by Thomas fire and its surrounding based on their tweets and estimate the housing damage resulting from the fire in these areas.

**Background**

In 2017, there were 56,186 wildfires between January and November compared to 60,236 wildfires in the same period in 2016, according to the National Interagency Fire Center. As of end of November, 2017 ranked higher in number of acres burned compared to the 10-year average. In October, eight counties in Northern California were hit by a devastating outbreak of wildfires which led to at least 23 fatalities, burned 245,000 acres and destroyed over 8,700 structures. The California Department of Insurance reported that insurers have incurred more than $9 billion in claims from the October fires, $8.4 billion in residential claims, $790 million in commercial property, $96 million in personal and commercial auto, and $110 million from other commercial lines. In December, the worst wildfire season in the history of modern California took another bad turn, as five major fires in Southern California destroyed over a thousand homes and buildings. Verisk estimated that 96 percent of homes endangered by the Thomas fire[[1]](#footnote-1) are at high or extreme risk.[[2]](#footnote-2)

The fire started on December 4th and as of the start of this project on December 18th, the fire was not fully contained. The estimated number of acres burned is 270,500 acres where only 45% were contained. An estimate of 18,000 structures threatened, 1,024 structures destroyed, and 250 structures damaged[[3]](#footnote-3).

[[4]](#footnote-4)

Our goal is to use publicly available data to do the following:

1. Run a sentiment analysis contrasting the sentiments of residents directly impacted by the fire and other residents who were not directly impacted by the fire through,
2. Estimate the property damage in some of the same areas
3. Run an analysis to study the correlation between residents sentiment, damage, and a wealth index based on the housing prices in the targeted locations

**Methodology**

Since our goal is to contract sentiment across different groups of residents based on the extent of their exposure to the fire, we decided to look at three groups of cities:

Group 1: Cities where the fire has not been contained yet – San Fernando and Ojai

Group 2: Cities where the fire has been contained – Beverly Hills/Bel Air and La Crescenta/Crescenta Highlands

Group 3: Cities which there was no fire – Long Beach and Santa Barabara

The selection of the cities was purposive due to the first spread and efforts to contain it and based on familiarity with these areas in California. However, in order to represent these cities, our data was sampled randomly within these cities.

In order to achieve representation of the targeted cities, we identified the coordinates of the city through Google maps and selected a random sample of 20 coordinates with 0.5 miles radius around each of these points to assess sentiments of residents of those areas.

**Data**

In order to conduct the targeted analysis, several data sources and APIs were used. These include:

1. Twitter data and tweepy
2. Zillow data and Zillow API
3. Google maps API
4. Census API

**Data Management**

**Analysis**

**Limitations**

The following are limitations imposed by data access and availability:

1. Twitter data is only available for the past week, thus we couldn’t conduct analysis on a random sample of all cities impacted by the fire.
2. Although we can request tweets from specific areas, we cannot save the location of the tweets, thus the tweets sample can only be linked to the city level
3. We only used tweets with “Thomas Fire” hashtags to avoid contaminating the data with data on other fires (e.g. domestic fire incidences)
4. Coverage at the housing estimates at the zip code and city level to link to tweeter data since exact location of tweets cannot be saved

**Division of Tasks:**

* Safaa: Documentation, data analysis, and visualization
* Shey: Zillow data
* Shoujun and Shemilis: Google maps and twitter data

**Timeline**

|  |  |  |
| --- | --- | --- |
| Date | Tasks | Comments |
| Saturday – 12/16/2017 | Develop proposal/research questions  Identify targeted locations | Team |
| Monday – 12/18/2017 | Explore Zillow data API and start data extraction  Write Documentation  Work on Google maps and twitter data extract | Shey  Safaa  Shoujun and Shemelis |
| Wednesday – 12/20/2017 | Data review, cleaning, and merging | Team |
| Saturday – 12/23/2017 |  |  |
| Monday – 12/25/2017 |  |  |
| Wednesday – 12/27/2017 |  |  |
| Saturday – 12/30/2017 |  |  |
| Monday – 01/01/2018 |  |  |
| Wednesday – 01/03/2018 | Data analysis and visualization | Safaa |
| Saturday – 01/06/2018 | Develop PowerPoint presentation | Team |
| Monday – 01/08/2018 | In class presentation | Team |

1. https://en.wikipedia.org/wiki/Thomas\_Fire [↑](#footnote-ref-1)
2. https://www.iii.org/fact-statistic/facts-statistics-wildfires [↑](#footnote-ref-2)
3. http://cdfdata.fire.ca.gov/incidents/incidents\_details\_info?incident\_id=1922 [↑](#footnote-ref-3)
4. By Phoenix7777 - Own workData source: MODIS Active Fire Detections for CONUS - 11/29/2017 through 12/06/2017 1700 MST, Remote sensing application center, USDA Forest Service modis\_fire\_last7\_2017\_340\_conus\_shapefile.zip, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=64633073 [↑](#footnote-ref-4)