Firewatch

Vikram Sunil Bajaj (vsb259)
Ameya Shanbhag (avs431)
Hitarthi Shah (hus206)

Project goal

With the recent events of forest wildfires happening across the globe, we decided to dive into data to develop some insights. The goal of the project is to analyze the wildfires in the United States over time, what effect does it have on the air pollution and to try to understand the factors causing them, and ultimately increase awareness about climate change.

Data Set

We will be using these two datasets:

1. **1.88 million US Wildfire:**
   Source : https://www.kaggle.com/rtatman/188-million-us-wildfires/
   Description : This dataset is an SQLite database with a table 'Fires' that contains the following information, for the period of 1992-2015 compiled from US federal, state, and local reporting systems:
   Attributes :

- FIRE_NAME = Name of the incident, from the fire report (primary) or ICS-209 report (secondary).(Categorical)
- FIRE_YEAR = Calendar year in which the fire was discovered or confirmed to exist.(Quantitative and Temporal)
- DISCOVERY_DATE = Date on which the fire was discovered or confirmed to exist.(Quantitative and Temporal)
- DISCOVERY_TIME = Time of day that the fire was discovered or confirmed to exist.(Quantitative and Tewmporal)
- CONT_DATE = Date on which the fire was declared contained or otherwise controlled. (Quantitative and Temporal)
- STAT_CAUSE_DESCR = Description of the (statistical) cause of the fire.(Categorical)
- COUNTY = County, or equivalent, in which the fire burned (or originated), based on nominal designation in the fire report.(Categorical and Spatial)
- STATE = Two-letter alphabetic code for the state in which the fire burned (or originated), based on the nominal designation in the fire report.(Categorical and Spatial)

- FIRE_SIZE = Estimate of acres within the final perimeter of the fire.(Quantitative)

2. U.S. Pollution Data

Source: https://www.kaggle.com/sogun3/uspollution

Description: This dataset deals with pollution in the U.S. (Nitrogen Dioxide, Sulphur Dioxide, Carbon Monoxide and Ozone AQI- Air Quality Index) for every day from 2000 - 2016.

Attributes:

- YEAR : Calendar year in which the AQI was recorded (Quantitative and Temporal)
- CO AQI : Quantitative
- NO2 AQI : Quantitative

Analytical Questions and Proxy Tasks

Make a list of questions you want to answer in your project and corresponding proxy tasks (referring to the attributes described in the previous section).

1. What regions are the most and least fire-prone?

Proxy Task: How do you identify 'region'?

Proxy Value: COUNTY/STATE attribute of the *1.88 million US Wildfire* dataset

2. Have wildfires become more or less frequent over time?

Proxy Task: How is time measured?

Proxy Value: FIRE_YEAR attribute of the *1.88 million US Wildfire* dataset

3. Why do most wildfires occur?

Proxy Task: How do you identify 'reason'?

Proxy Value: STAT_CAUSE_DESCR attribute of the *1.88 million US Wildfire* dataset

4. What causes lead to larger wildfires?

Proxy Task: How do you identify causes?

Proxy Value: STAT_CAUSE_DESCR attribute of the *1.88 million US Wildfire* dataset.

5. How long does it usually take to contain a fire?

Proxy Task: How do you measure time?

Proxy Value: CONT_DATE and DISCOVERY_DATE attributes of the *1.88 million US Wildfire* dataset.

6. Does the average fire size vary depending on the average time it took to put out the fire?

Proxy Task: How do you measure the time taken to put out fire?

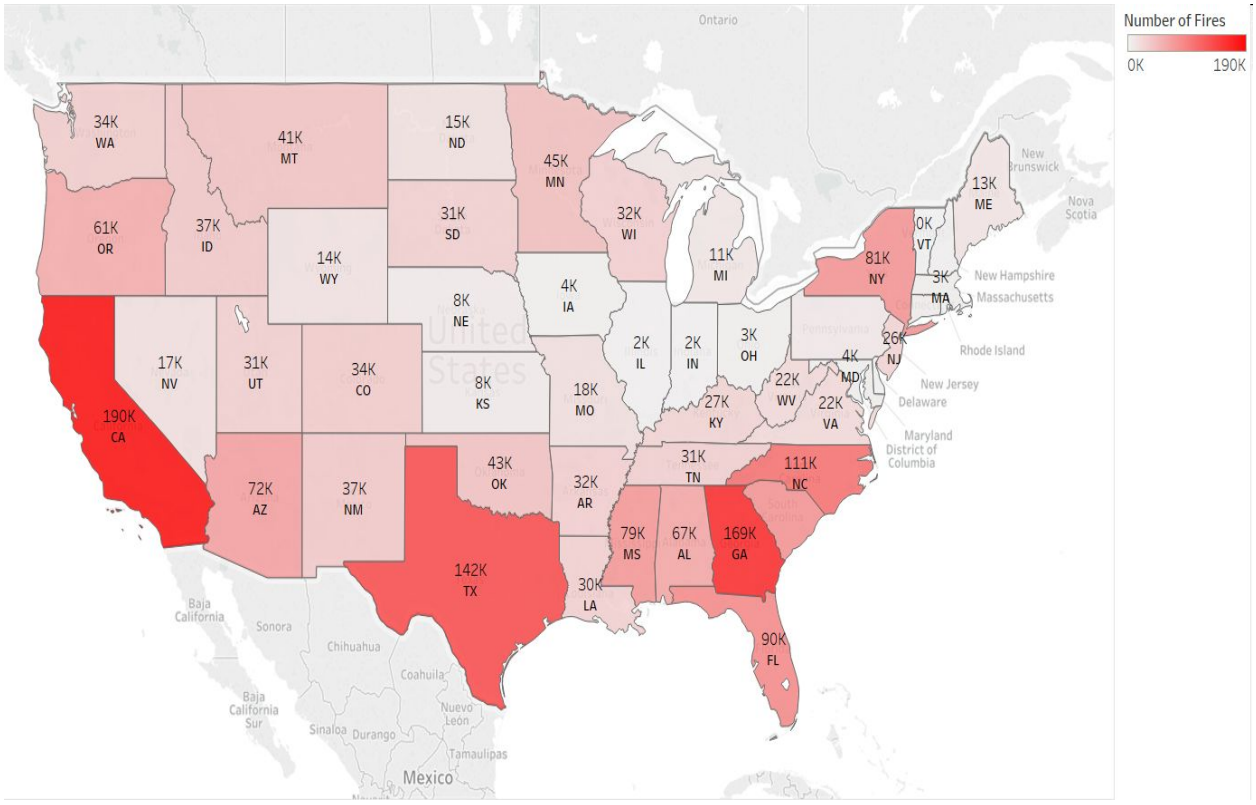
Proxy Value: CONT_DATE and DISCOVERY_DATE attribute of the *1.88 million US Wildfire* dataset.

Storyboard

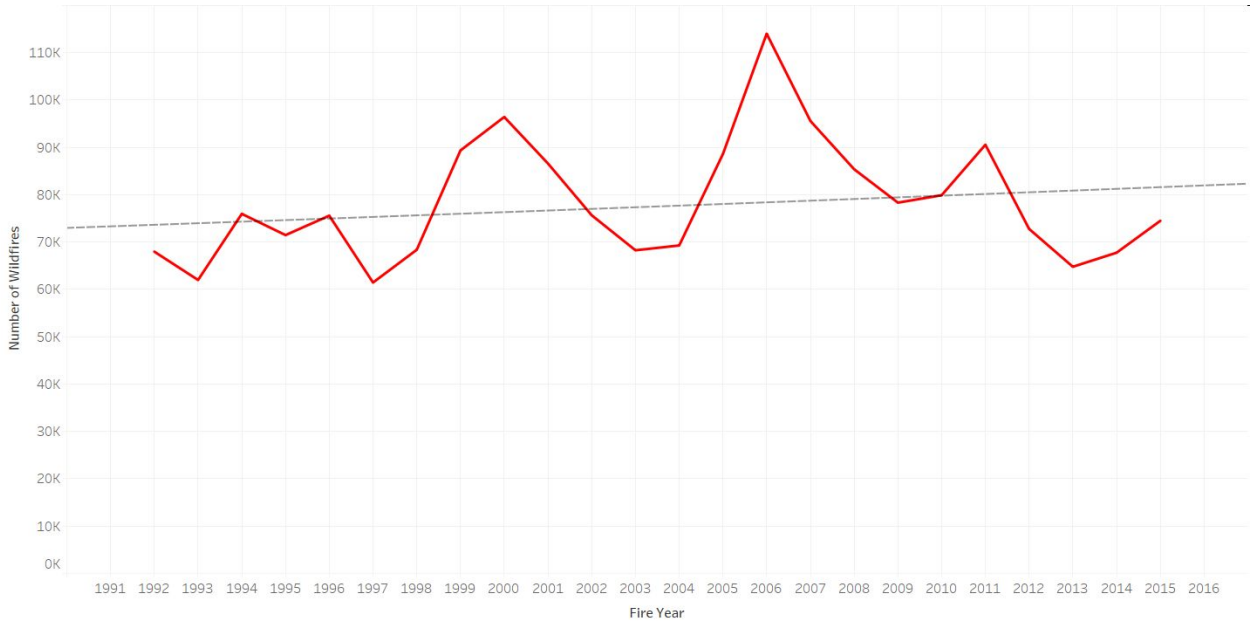
Wildfires are becoming increasingly common nowadays. With a large part of California set ablaze, we couldn't help but think: could this be a cause for global warming? What could have been the reason behind the wildfires? With these questions in mind, we turned to data for the answers. What we found was quite surprising!

First, we obtained data that contained details about wildfires in the US, from 1992-2015.

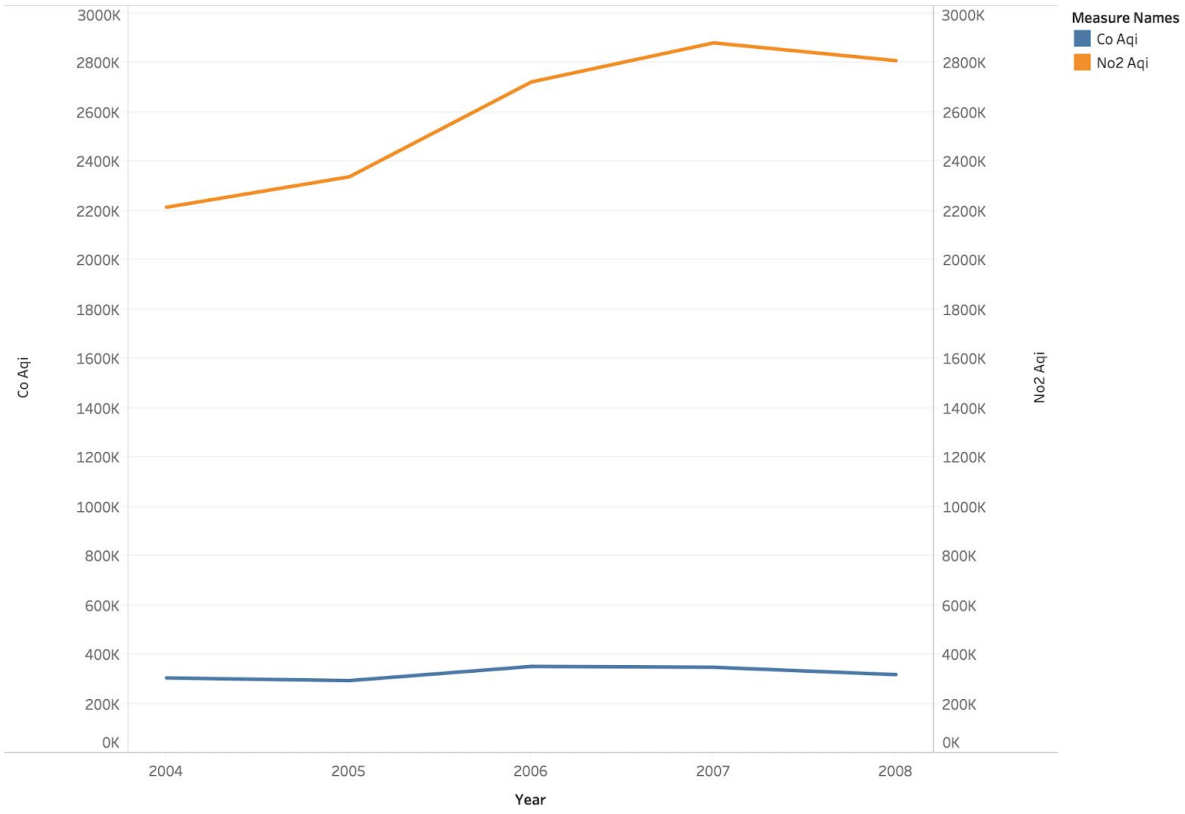
Initially, we wanted to determine which states in the US are most prone to wildfires. California: no surprise there! Georgia is a close second.



We also confirmed that wildfires are becoming more frequent over time! This is definitely cause for concern.

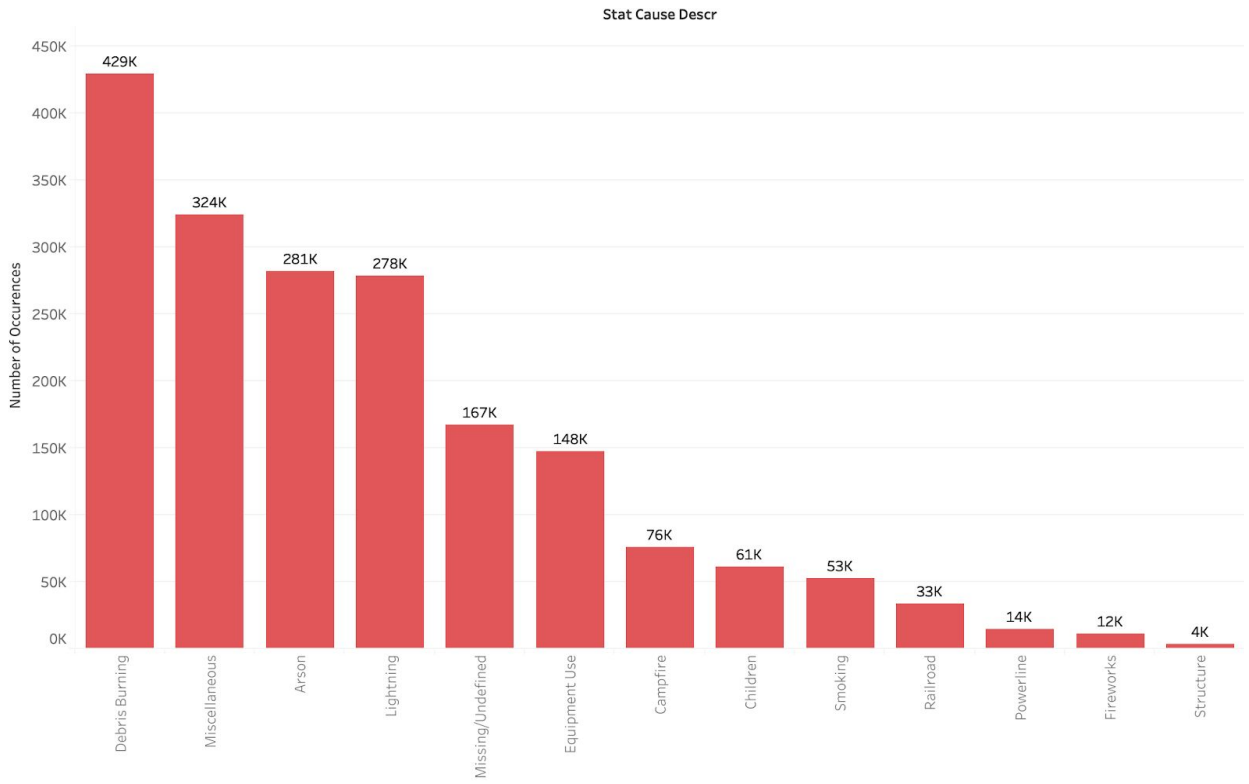


Although there is an upward trend in the number of wildfires per year, the sudden increase in 2004 followed by the sudden drop in 2007 led us to investigate a little further on the impact the wildfires had during that period on the Air Quality Index.

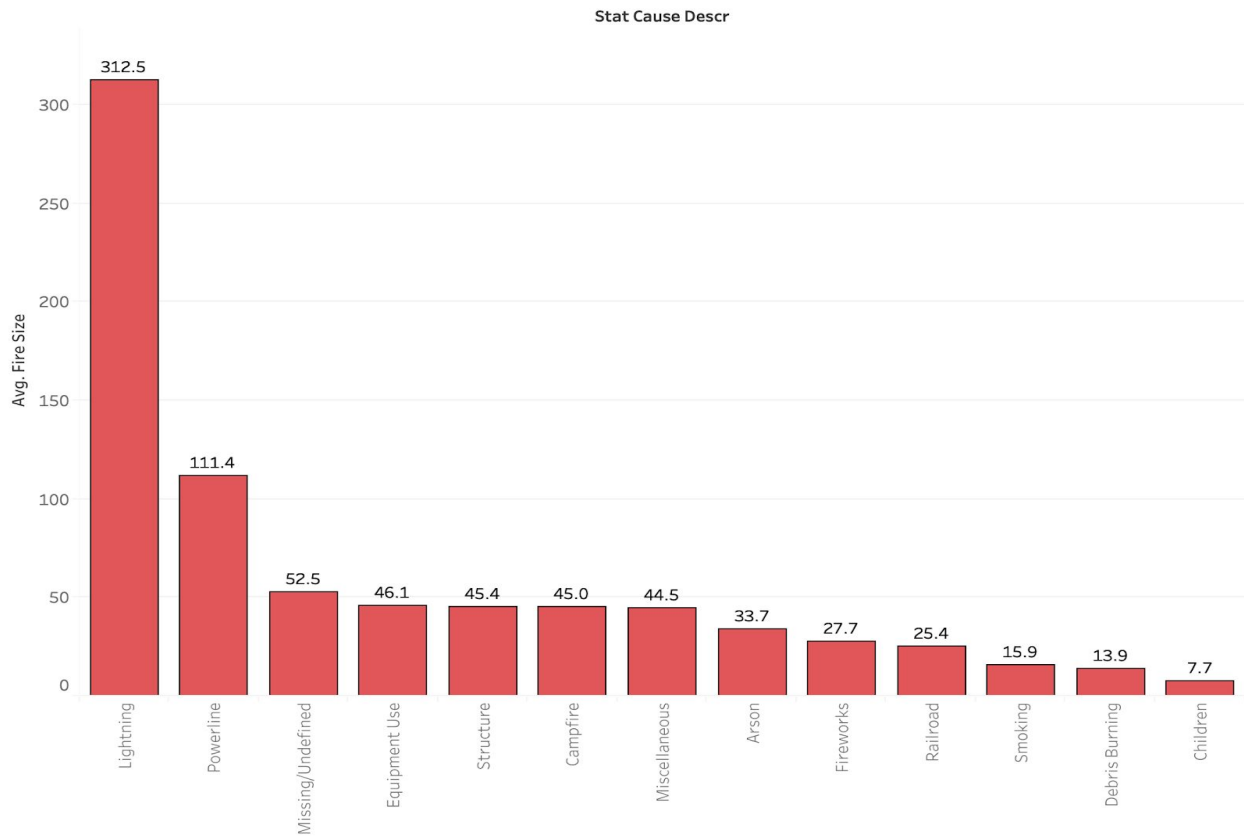


As we can see that there is sudden increase in the CO and NO2 levels from 2004-2006, the same time period where the number of wildfires had increased followed by decrease in the CO and NO2 levels similar to decrease in the number of wildfires from 2006-2008. The large amount of smoke caused by the wildfires could be responsible for the change in the CO and NO2 levels.

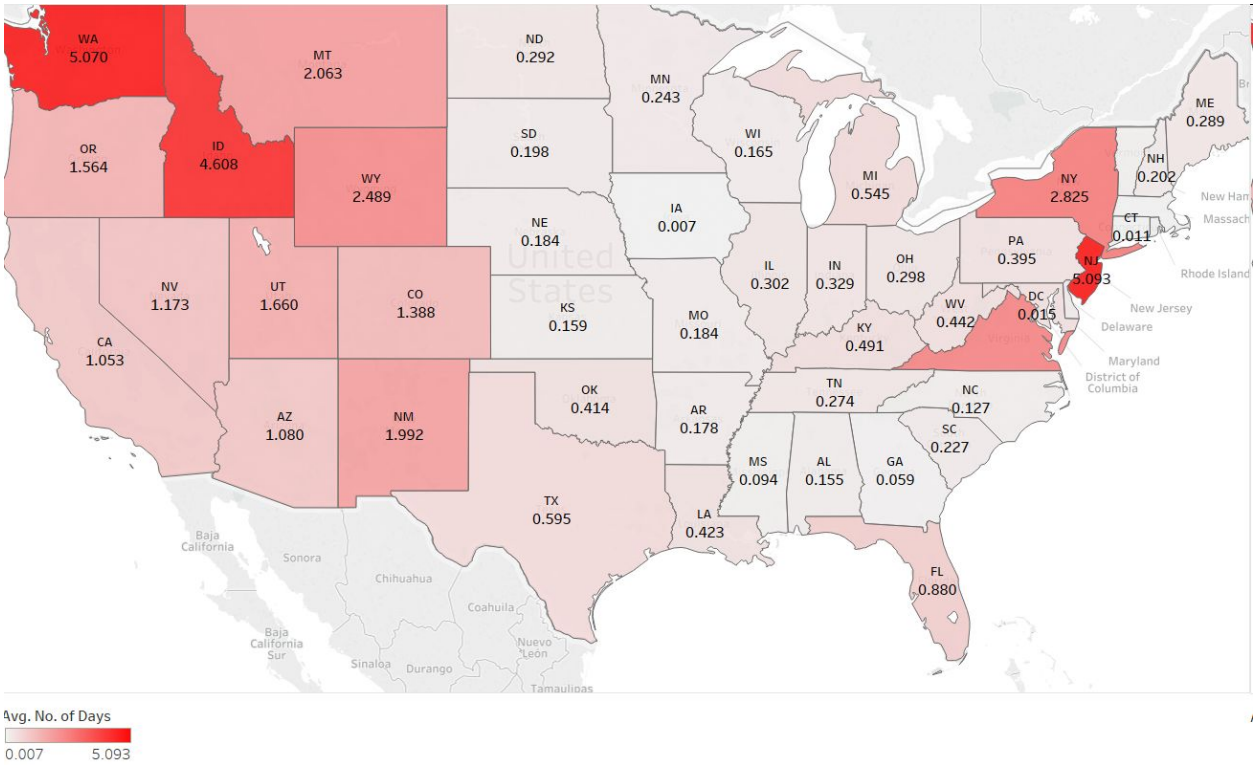
But why are wildfires happening? Data had the answer yet again: most wildfires were caused due to burning debris, who'd have thought?



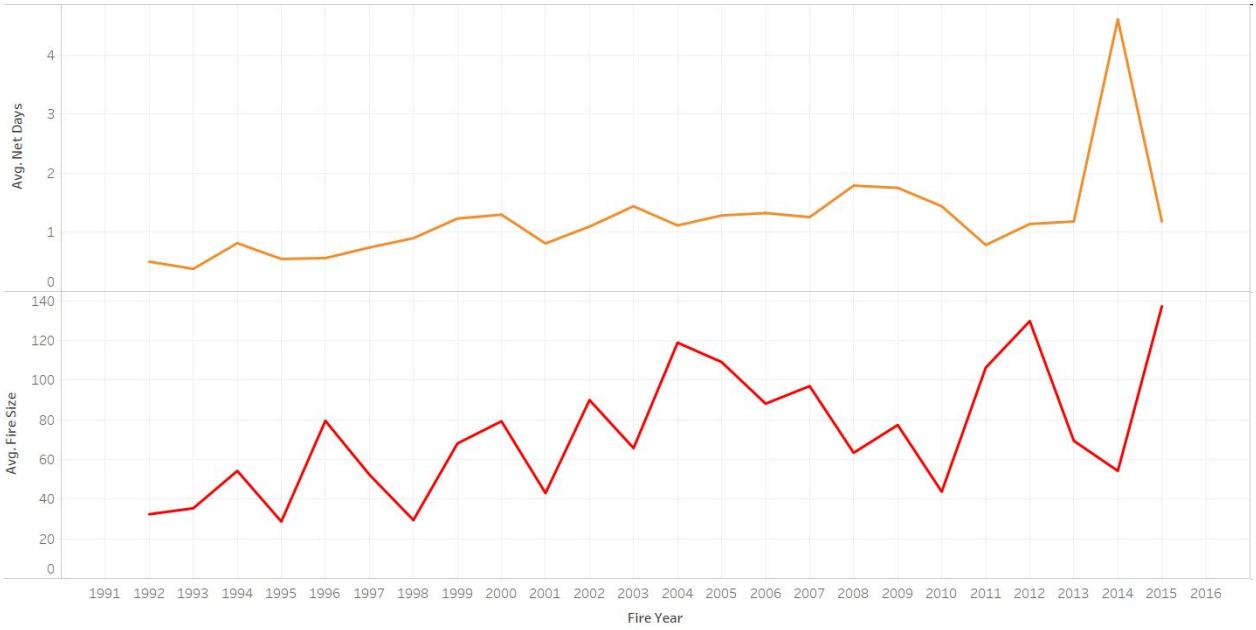
After observing the chart, we can see that the wildfires because of lightning were less than those due to debris burning. This is what surprised us, so we wanted to determine not the cause for the most wildfires, but the cause for the largest wildfires. Turns out, although burning debris caused the most wildfires, it mostly caused really small fires. Lightning, however, was the cause for the largest wildfires! This is, unfortunately, not something that humans can control. Other causes, such as power line failures, equipment use, campfires etc., however, can and should definitely be controlled.



The second most important part, after identifying the real cause of wildfire is how long does it take to put out the wildfire. This is important because wildfires take away homes, vegetation and wildlife. The soil in the area of wildfire is destroyed because as soon as the wildfire hits the soil, all its nutrients are gone. Ash and smoke produced from the wildfire causes serious health problems to people having allergies. Hence the faster the fire is put down, less is the impact of wildfire on the ecosystem.



Clearly, some fires took longer to put out than others. But does this mean that fires which took longer to put out became more widespread? Did a delay in putting out the fires cause them to grow larger? This is surely a cause for concern, if true.



On comparing the average fire size with the average number of days it took to put the fire out, across the years, a correlation does seem to exist. This correlation has, however, disappeared in recent years.

Changelog

- Removed the sea ice level dataset and the correlation between the wildfire dataset and sea ice extent.
- Changed the title from ‘Fire and Ice’ to ‘Firewatch’
- Amended the goal of the project to reflect the removal of the sea ice extent dataset.
- Added the air pollution dataset to observe the correlation between the wildfires and their impact on air pollution.
- Added a new question to determine the average number of days it took to put out the wildfires in each state.
- Added a new question to check if wildfires that took longer to put out grew larger.

Items Already Implemented

- D3 visualizations for 4 questions (2 interactive maps implemented - index_1.html, 2 bar charts implemented - index_2.html)
- Incorporated changes suggested by Professor

Items to be Implemented

- 2 line chart visualizations
- Flipping the bar charts from horizontal to vertical
- Combining visualizations in a single HTML file
- Refining the story and completing the webpage