Group 2: Grace Cote, Hima Gharat, Josh Heisler, Salita Santiago, Marcellis Valentin
UPENN-PHI-DATA-PT-09-2020-U-C
Project 1 Presentation Summary
November 7, 2020

Supplemental Presentation:
https://docs.google.com/presentation/d/1_OGaALunrmiTqD8N2QQZokpoLyl-8ypzU5rTrXwdgGM/edit?usp=sharing

Setting The Data World Ablaze: The Data Stories Behind Wildfires

When initially researching for this project, we wanted to find a topic that was relevant to the present day while also researching something we were genuinely interested in. In hindsight, we should have known 2020 was going to be a dumpster fire when we saw the continent of Australia and their battle with wildfires at the beginning of this year as well as the stateside battle on the west coast, which we are still fighting today. According to FEMA, a wildfire is an unplanned fire that burns naturally (ready.gov). It is important to note that residential fires are not included in this data - this only accounts for wildland fires.

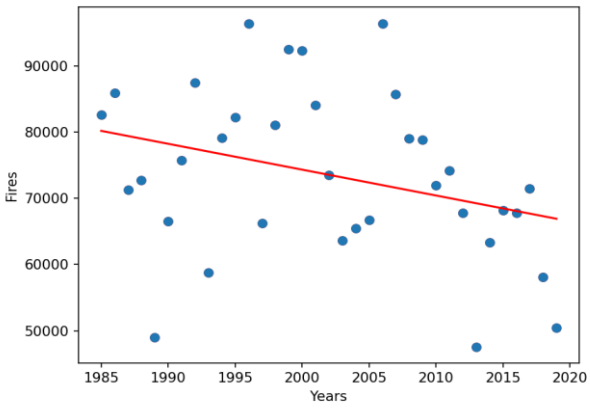
With the recent fires highlighted in the media and the talks of increased wildfires due to climate change, we wanted to see if wildland fires are indeed increasing as much as we think they are. Looking into these wildfires, we wanted to research based on historical, geographical, and financial trends to see if there are noticeable trends or effects. For the project we decided to use data from the National Interagency Fire Center (NFIC), which is a governmental organization that supports and they organize and allocate emergency resources to different federal, state, and local government agencies, such as the National Park Service and Departments of Agriculture and Interior.

Historical:

First, we graphed the number of fires and acres burned since 1985 and found the following information:

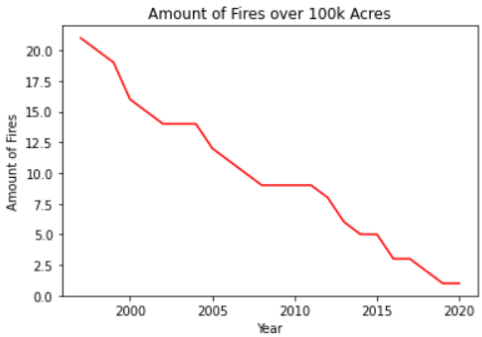
The number of fires fluctuates over the years with a general trend of decreasing as per the below graph.

Graph 1: Fires Since 1985

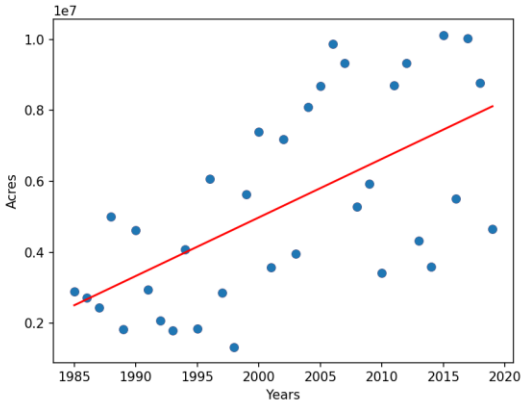


However, the number of acres burned have generally increased, but fluctuate as well. This is most likely due to the increase in intensity of those smaller fires.

Graph 2: Amount of Fires Over 100k Acres



Graph 3: Acres Burned Since 1985

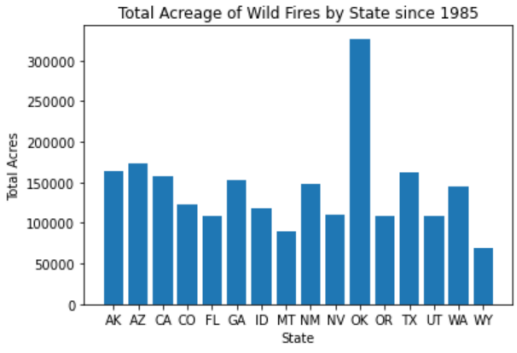


Geographic Data:

After we learned about historical trends, we wanted to learn more about where these fires were happening. The data set used observed 527 wildfires since 1985 that burned over 100,000 acres in one burn, which included the state in which the fire took place.

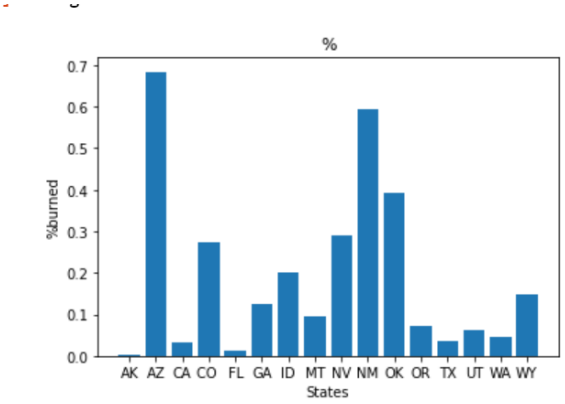
As you can see on the graph, everyone looks relatively the same, with the exception of Oklahoma which has a very high number of acres of land burned within the state. Oklahoma only had 6 fires, but all 6 fires burned over 275K each.

Graph 4: Total Acreage of Wildfires by State



We also acknowledged that it is hard to compare the burned acreage of a small state like Rhode Island to a large state like Alaska. To solve this, we took the burned acreage and divided that by the state's landmass to find a percentage burned, as per the below:

Graph 5: Percentage Acres Burned by State



Arizona and New Mexico had the highest wildfire acreage relative to their landmass. It is also important to note that Oklahoma is not as large of an outlier as the other states, when comparing it relative to its landmass.

As per the graph, there is a wide range of percentage burned by state, so we wanted to run a chi square test to see if there is statistical significance to the higher percentage burn states

Taking the average of all the states' percentage burn, we got our expected value. When comparing the 16 states at a 95% confidence, the statistic is less than our critical value, and therefore, we reject our null hypothesis and say that at least one of the states we researched had a different percent burned than all the others.

Graph 6: Chi Squared Statistics

	Percentage Burned	Expected Percentage
0	0.002714	0.19151
1	0.684243	0.19151
2	0.030981	0.19151
3	0.273945	0.19151
4	0.013982	0.19151
5	0.124636	0.19151
6	0.200042	0.19151
7	0.094127	0.19151
8	0.291501	0.19151
9	0.593038	0.19151
10	0.391785	0.19151
11	0.070678	0.19151
12	0.034501	0.19151
13	0.061793	0.19151
14	0.046882	0.19151
15	0.149306	0.19151

Financials:

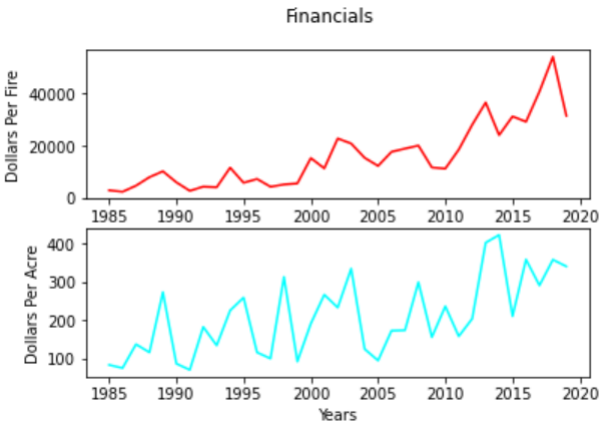
On the financial side, we looked at the suppression costs of extinguishing the fire. The dollars spent per fire noticeably increased, but this is likely from inflation. The dollars per acre did not have a true pattern most likely because the number of acres burned fluctuated as well.

As a summary of what we have found, this research can help us to educate and prevent future fires from occurring, while also informing consumers of potential risks of living in certain areas. The pattern of historical wildfires show they are decreasing in number but fluctuate in size.

From our research, the greatest number of wildfires occurred on the West Coast and 4 corners region or Alaska. On the financial side, the cost of suppression is increasing with inflation, but this could also be due to the possible increase in technology being used to fight the fires.

Next steps include looking into the causes of why these fires happened, whether they were natural or human error. In addition, we want to look further into the geographical research because we only investigated the larger fires. And then we also want to research if there were other financial costs involved and what these costs are actually spent on. Was there an increase in technology? Was there an increase in labor costs? That information could be crucial to finding out what works when fighting these

Graph 7: Financials Subplot:



Conclusion:

Upon further reflection, we had tons of growth opportunities. For example, when creating the chi squared table, we ran into difficulty as we did not originally calculate an Expected Value. We had to reassess our table and calculate this value to eventually conclude the rejection of the null hypothesis (that at least one of the states differed from the others in terms of Percentage Burned)

Overall, this research is a work in progress, but learning their patterns using this data and others like it can help us to prevent them in the future.