California WildFire Incidents Report

Introduction

In this project, we have chosen to work on a California wildfire incidents dataset from Kaggle. The original data source is from the California Fire website. We wanted to work with a dataset that contains geographical variables and enough other types of variables so that we could explore the data from different angles. The final dataset we have chosen encapsulates various aspects of 1636 fire incidents that affected California state over the course of 2013 to 2019, which include a total of 40 attributes including 14 numerical variables, 24 categorical variables, and geographic coordinates. After initial exploration, we selected variables most relevant to our project. We have included the selected variables listed below.

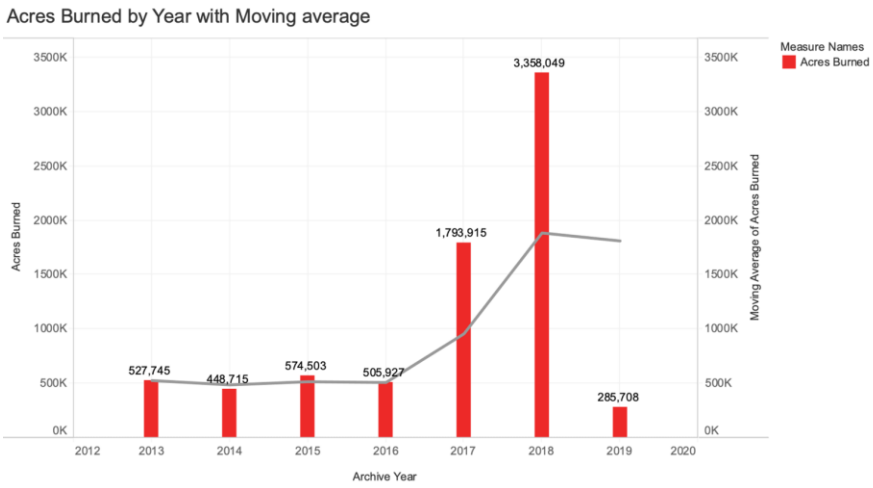
Name	Type	Description
AcresBurned	Numerical	Acres of land affected by wildfires
Archive Year	Categorical	Year of fire incident reported
Counties	Categorical	Counties involved
Latitude, longtitude	Geographical	Geo location of the fire incident
StructuresDamaged	Numerical	Structures damaged by fire incidents
CalFireIncident	Boolean	Is the incident treated as a calfire incident
Fatalities	Numerical	Fatality count
Extinguished	Date time	Extinguished date
Started	Date time	Fire start date

Through various visualization techniques, we aim to demonstrate the impact of the wildfire incidents in the given years, showcase the locations of the fire incidents in various counties and explore any trend or pattern of wildfire incidents if there exists.

Exploratory Analysis

We have only included partial exploratory graphs in our analysis. Firstly, to comprehend the overall trend in fire incidents and identify potential patterns. We employed Tableau to create a chart with ArchiveYear on the x-axis and total AcresBurned on the Y – axis. Graph 1.a below shows a drastic increase in acres burned in the year 2017 and 2018.

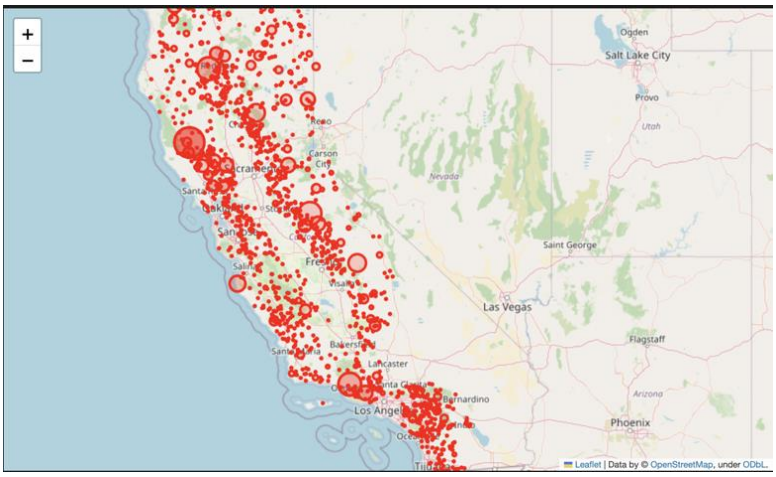
In the initial exploratory graph, we added a moving average line on top of the column chart. However, the moving average line, although typically applied to time series data, is not the best fit here. Instead of time date, we treated Archive year like a factor here. The story is to tell how many acres burned in each year and compare their differences in volume. The line diverts the audience from the story and the plain data it tries to demonstrate. Thence, we removed the line chart in finial visualization.



Graph 1

The second graph we drafted is the glyph map that showcases the locations and severeness of each fire incident, which is measured by the acres burned.

The map below gives an overview of affected areas. The first impression of it is that fire incidences are everywhere, covering the majority of California counties. This graph sets the stage for further exploration of fire incidents' locations in each Archive year. In the later explanatory visualization, we mapped out the fire incidents in year of 2014 – 2019, allowing the audience to compare the areas where fire incidents affected in each year.

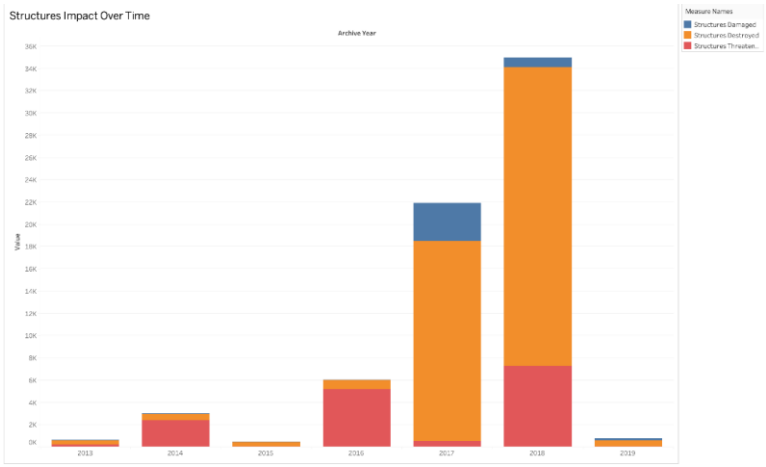


Graph 2

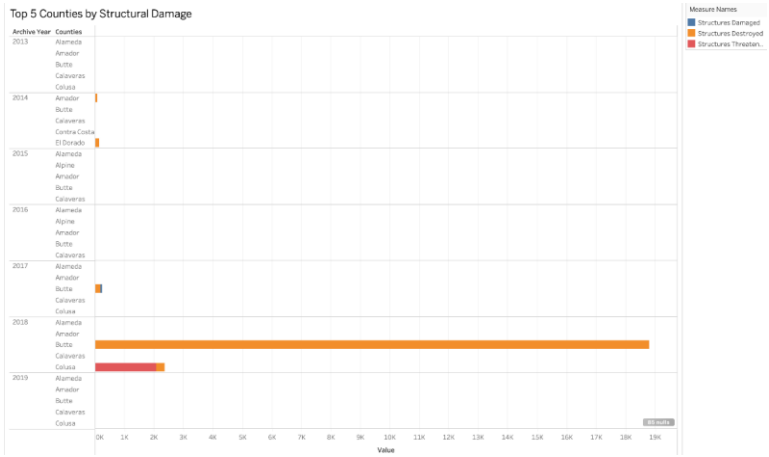
Another topic of interest is the impact of fire incidents, by measure of other variables other than acres burned. We found out that structures destroyed/damaged are another angle we could pursue. However, the column chart Graph 3.a tells a very similar story as Graph 1.b, as we see a much higher number of structures impacted in year 2017 and 2018.

We created a hierarchical bar chart to demonstrate the structural damage over 5 counties in each year. However, there is a low data-ink ratio in the graph, and we dislike the space

underutilized. This exploration sets the stage for us to experiment with other types of visualization. We ended up using a contingency table and star plot to display information that contains multi variables.

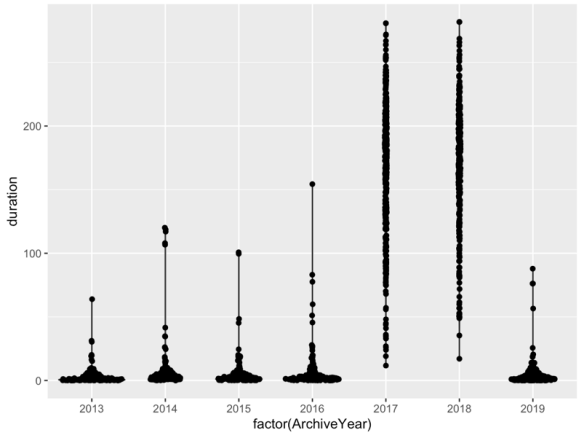


Graph 3.a



Graph 3.b

The last graph we drafted in the earlier stage is a violin and Sina plot of the duration of the fire incidents in each Archive Year. We can tell those fires in 2017 and 2018 lasted longer, and their durations are evener distributed. Fire durations in other years are clustered under 50 days. However, there are a few questions that surfaced based on the visualization. It seems strange that there were no fires that lasted less than one day in 2017 or 2018. This inspired us to dig deeper into our dataset in further analysis.

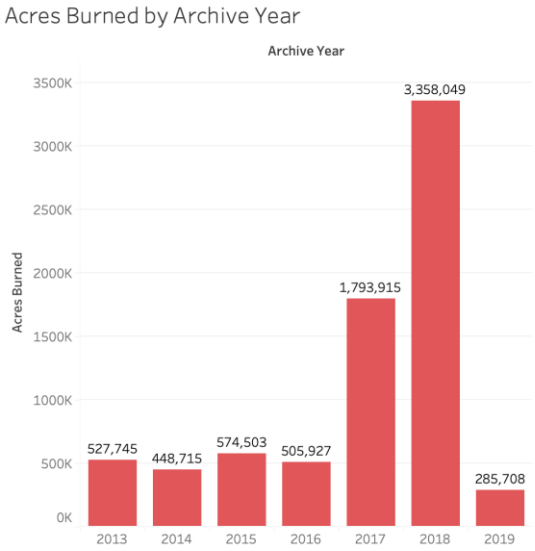


Graph 4

Visualizations

Visualization 1 is a column chart that exhibits the total acres burned in the year 2013 – 2019. The value of the total acre burned is marked on top of each column.

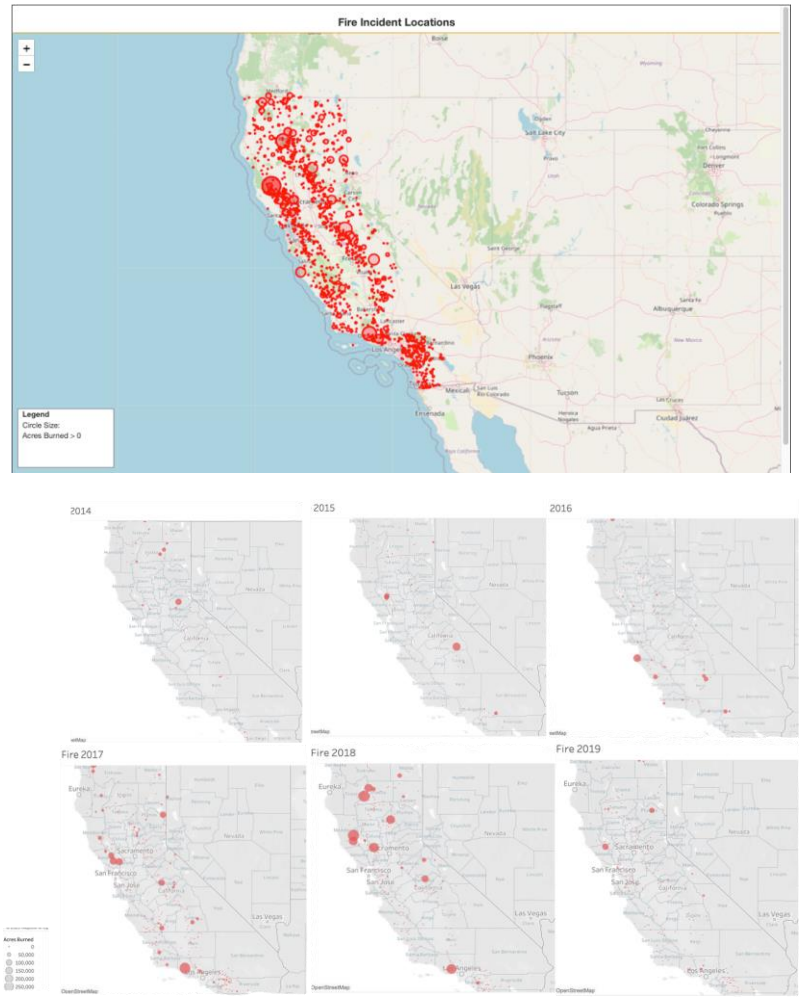
We removed the moving average line from the draft graph. We also made sure to display the Archive Year as a discrete variable instead of a continuous one. The graph helps the audience see the value of the acres burned in each archive year and compare their differences.



Visualization 1

Visualization 2 is a glyph map and small multiples of the same kind, which demonstrates the locations of the fire incidents. Additionally, the size of the dot corresponds to the acres burned. The small multiples are the breakdown maps of the fire incidents in the year 2014 – 2019. The purpose of including small multiples in our visualization is to give a detailed overview

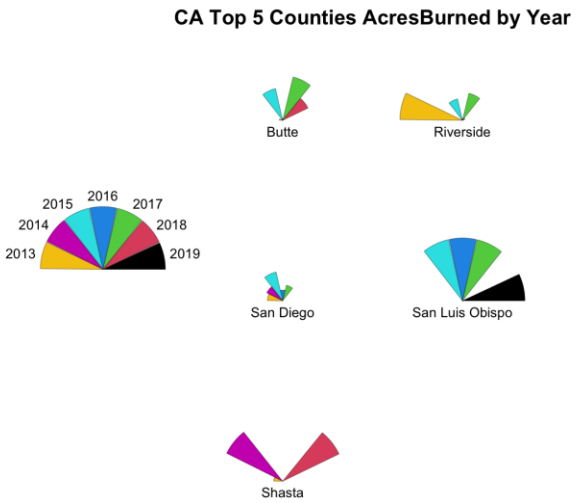
of the fire locations each year. The message we want the audience to receive is where the fire incidents occurred in each year and the severity of the fires.



Visualization 2

Visualization 3 is a star plot that shows the top 5 CA counties and their acres burned in the year 2013 –2019. The star plot makes it easy to compare the impact of fire incidents on the counties as well as comparing the difference of impact each year per county. The counties selected are the counties affected the most over the year 2013 – 2019. We observe from the graph that there are different patterns for each county. For exmaple, San Luis Obispo was affected the most overall but zero or close to zero fire incidents in 2013, 2014 and 2018.

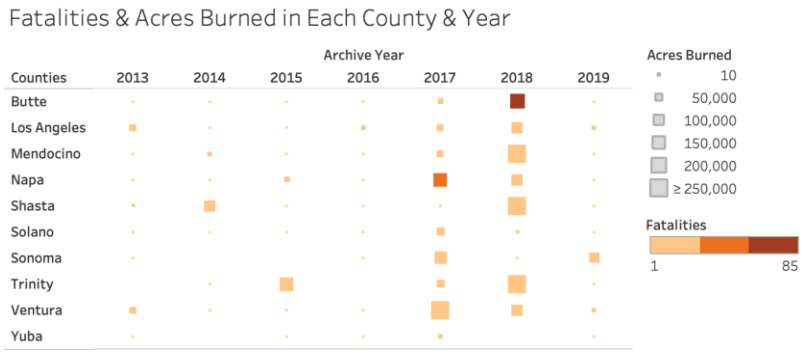
Taken feedback from class, we have relocated the legend and reversed the year sequence.



Visualization 3

Visualization 4 is a contingency table that illustrates the fatalities and acres burned data of the top 10 counties that have the highest fatalities in each Archive Year. Rows represent counties, and columns display the Archive Year. Each box represents fatalities and acres burned information corresponding to a county in a year.

Instead of using a hierarchical bar chart, which can quickly run out of space, the contingency table packs the information neatly. It takes advantage of both area size and color encodings. The visualization captures how wildfire incidents impact the environment and human lives. The audience can tell instantly which counties are affected the most in either environmental or fatality aspects. Apparently, Napa and Butte County suffered the most fatalities, while the severeness of acres burned differed in counties and years.

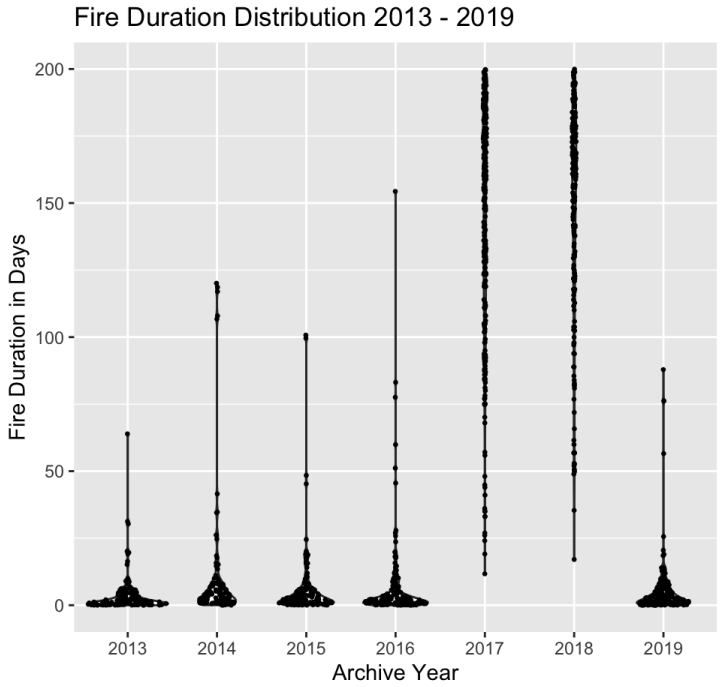


Visualization 4

Visualization 5 is a violin Sina plot that displays the fire duration distribution over the years. After the initial graph, we rechecked our dataset and the source website and realized the dataset was missing some records. For example, according to the state website, there was a fire

recorded in April 2017 that was put out within a day, but we could not find a corresponding record in our dataset. Although we do miss some data points, the violin and Sina plot tells a compelling story that fires in 2017 and 2018 are in larger scales and generally took longer to extinguish.

In addition to rechecking the data source, we added a title and adjusted the axis labels and resized the dots, so they are clearer to understand.



Visualization 5

Conclusion

Our visual narrative unfolds on multiple fronts, beginning with a column chart placing Archive Year against Total Acres Burned. This chart allows the audience to see a year-on-year comparison of acres burned. The glyph maps and small multiple maps detail the locations and severity of fire incidents. The sizes of each dot convey the magnitude of each incident. The star plot puts the year-over-year impact for each county in direct comparison. The contingency plot utilized visual encodings to convey several information in a compact space. Finally, the fire duration looks at the fire impact through another lens and emphasizes again that the year of 2017 and 2018 stand out in terms of the severeness of wildfire incidents and its impact.