Final Project Report: California Wildfires Analysis

Abstract

The main goal of this project is to show the increasing trends of intensiveness, duration, geographical distribution and the main causes of the wildfires in California through a set of independent visualizations on different topics. Each visualization is based on the same dataset that contains California wildfires data between years 1878 and 2017, and is maintained by the Cal Fire. The visualizations are compiled into one easy-to-use web application allowing users to zoom, filter, play animations and perform other interactions with charts. The main finding from each of those visualizations is that the situation becomes worse each year. There is more fire intensity in terms of absolute counts, the area burned, the duration of fires, and the spread of wildfires which covers the whole state of California. The main identified cause is natural (lightning), however the majority of fires have an unidentified cause. Our findings show that the wildfires seasonality is constant, despite other scientific paper findings showing that the seasonality is tending to extend to the fall and winter months.

Introduction

The devastating wildfires became a serious challenge in the past decades. The recent statistics clearly show increasing trends in overall burned area, financial damage to structures and properties, and to human health and lives as well. The main reasons include, but not limited to a warming climate, prolonged droughts, constant population growth, urban area growth, fuel accumulation over the decades. Whether it is a naturally caused fire, arson or a prescribed burn, it becomes more important to use all available tools to keep this dangerous trend under control. This project will use several available dataset attributes to derive several independent visualizations that will cover different topics such as geography, trends, seasonality, causes and duration of California wildfires. Although each visualization covers separate areas, one should be able to combine all the findings and that will lead to one important conclusion: the media and political alerts are not just empty statements. Our concerns are real and they are data driven: the wildfires are getting worse over the years and action is needed now.

Dataset

The California Department of Forestry and Fire Protection (CAL FIRE) maintains historical data about wildfires in California. The dataset contains information on California fires between years 1878 and 2017, providing the following summary information for each identified fire:

YEAR - year of fire
STATE - mainly California, but might include Nevada, Oregon, Arizona
AGENCY - fire agency taking care of the fire
UNIT_ID - fire department taking care of the fire
FIRE_NAME - name of fire
INC_NUM (incident number) - incident ID
ALARM_DATE - start date
CONT_DATE (containment date) - end date
CAUSE - main cause of the fire
COMMENTS - any comments
REPORT_AC - reported number of acres burned
GIS_ACRES - number of acres burned as determined by analysis of GIS data
C_METHOD - data collection method
OBJECTIVE - objective of the fire (suppression, prescribed fire, etc.)

For this project, only the attributes in **bold font** were used.

Data Processing

For each visualization , there was a need to perform different grouping operations of the data to get count, or sum of fires by years, by month and by fire causes. The additional calculated column for the fire duration (in days) was added. For geographical visualization, the forward geofencing was executed using the Radar API (with free version token). Additionally, the best fit line was calculated using the machine learning library for trends charts.

Tasks

The following independent tasks were identified for visualizations:

Problem / Task	Task Type	Audience
Explore wildfire geography over time. Does the activity increase over time? What regions are the most affected?	Explore, compare, show trend	Any fire management agency, general public
What are the wildfire trends in terms of their count and area burned?	Show trend, analyze	Any climate, fire management agency, general public
What are the main causes of wildfires?	Explore, look up, analyze	Any fire management agency, general public
What is the seasonality of wildfires? Can we confirm using this dataset visualization, that the seasonality changes to fall and winter months?	Explore, compare, analyze	Any climate, fire management agency, general public
What is the trend of wildfires' duration over time?	Show trend, analyze	Any fire management agency, general public

Solution

Task 1. Explore wildfire geography over time. Does the activity increase over time? What regions are the most affected?

Visualization design

The task is to explore the geography of the fires, therefore the geographical map with circles (bubbles) for quantitative attributes was chosen.

Implementation

Firstly, the geo coordinates needed to be obtained for each fire unit in the dataset. Secondly, the geo coordinates needed to be mapped to each fire in the dataset. Lastly, the 'group by' operation needed to be applied to get the sum of the area burned for each geographical location and year. After that, the obtained data was passed to the Plotly Express library, which has a built-in function 'scatter_geo' for the bubble map visualizations.

Usage

The user can look at the map and identify the total area burned for a particular geographical location by hovering the mouse on the circles. Additionally, he can switch the year at the bottom of the map or press the 'Play' button to see it animated.

Task 2. What are the wildfire trends in terms of their count and area burned?

Visualization design

The visualization needs to show any trends in counts of wildfires and also in the total area burned over the time. Thus, the line plot idiom was chosen along with the trend line. Since there are two quantitative attributes to explore over the same ordered attribute (Year), the dual axis chart with possible filtering was chosen instead of having two separate charts. In that way, not only the screen space will be saved, but the correlation/causation can be seen as well.

Implementation

Firstly, the 'group by' operation needed to be applied to get the sum of area burned and the count of fires for each year. Secondly, the trend lines data needed to be calculated for both charts. After that, the obtained data was passed to the Plotly library, which has a go.Scatter() function to generate custom scatter plots.

Usage

The user can see the main line plots and the trend lines which indicate the changes over time. He also can filter/disable unwanted plots by clicking on their corresponding items in the legend box.

Task 3. What are the main causes of wildfires?

Visualization design

The stacked bar chart was chosen because it allows one to visualize one quantitative (Count), one qualitative (Cause) and one ordered attribute for analysis and comparison.

Implementation

Firstly, the 'group by' operation needed to be applied to get the count of fires for each year and each wildfire cause. After that, the obtained data was passed to the Plotly Express library, which has a built-in function for bar charts with the parametrization for making the bars stacked by categories.

Usage

The user can explore, compare and analyze the bars by their length and clearly see the most affecting fire causes. He can use a filtering feature to hide the unwanted categories.

Task 4. What is the seasonality of wildfires? Can we confirm using this dataset visualization, that the seasonality changes to fall and winter months?

Visualization design

The task was divided into three subtasks to help finding any hidden seasonality insights:

a. Overall count of fires by months
b. Categorical division of counts by month
c. Overall annual count of fires by month - animation by year

The bar chart was chosen for each subtask, because it allows to visualize one quantitative (Count) and one ordered attribute for analysis and comparison.

Implementation

Firstly, the ‘group by’ operation needed to be applied to get the count of fires for each year, for each month and each fire cause. After that, the obtained data was passed to the Plotly Express library, which has a built-in function for bar charts with the parametrization for making it divided with facets.

Usage

The user can explore, compare and analyze the bars and clearly see the months where the most fires occur. He can use the faceted view to see the seasonality by categories. Also he can use the animation view in order to get an idea about the shift/change in seasonality.

Task 5. What is the trend of wildfires' duration over time?

Visualization design

The area chart is good for this task because one quantitative attribute (Fire Duration) and one ordered attribute (Year) are being used and the goal is to show trends and make analysis.

Implementation

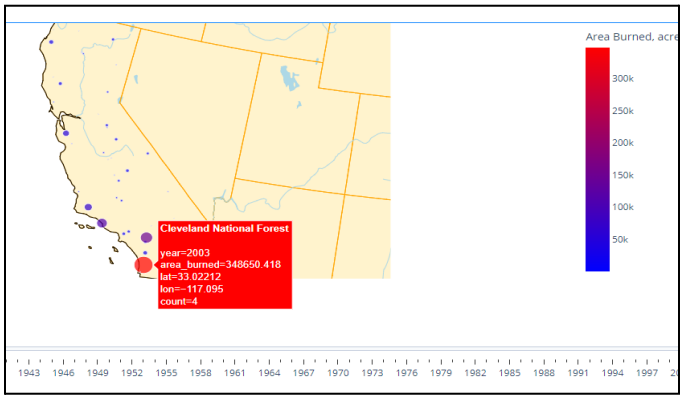
One additional column for fire duration needed to be calculated and then ‘group by’ operation of average duration by each year needed to be executed. After that, the obtained data was passed to the Plotly Express library, which has a built-in function for area charts.

Usage

Area charts are simple to use and analyze. The user can see the growing area over time and make conclusions out of it.

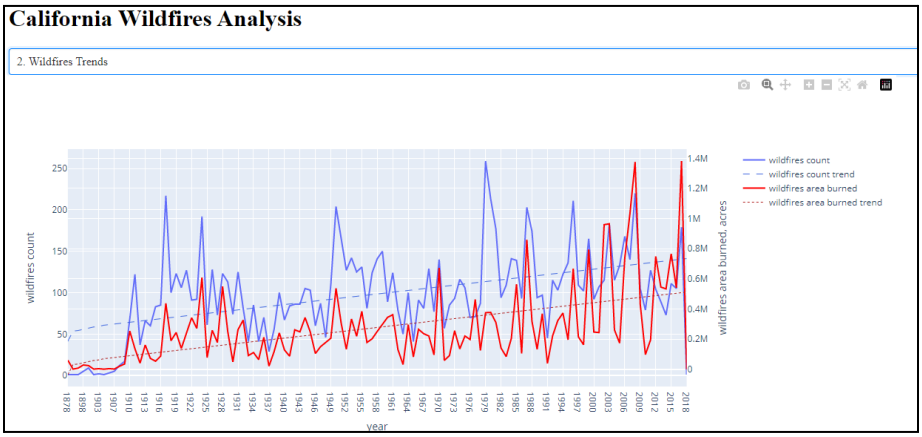
Results

Task 1. Explore wildfire geography over time. Does the activity increase over time? What regions are the most affected?



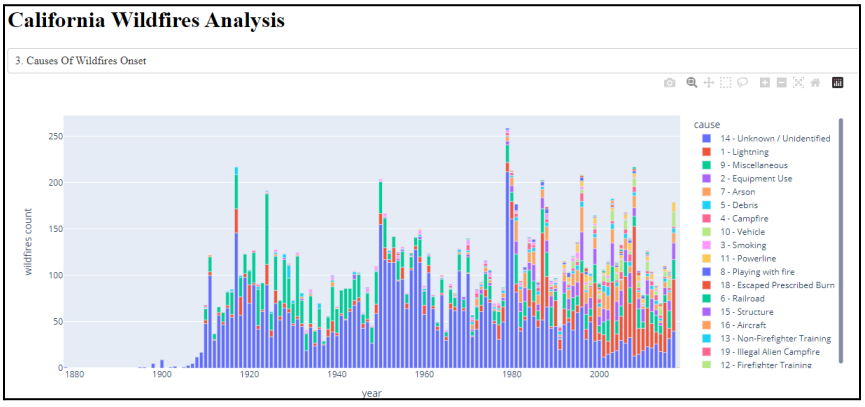
Expressiveness	Geographical data is natural to express on a geographical map for exploration tasks. Bubble map is appropriate for showing the magnitude of a quantitative attribute at a certain geographical location. The animation / slider by Year attribute helps to see changes over the time.
Effectiveness	Use spatial geometry for placing marks (circles) to the appropriate location. The size and color of the circles indicate the magnitude of the quantitative attribute (area burned). The color scheme is distinguishable for variance of color deficiencies.
Findings	The fire intensity (area burned) is increasing over time.
Challenges	Geofencing doesn't work perfectly for this dataset, so the manual tweaking needed to be done to get the right geo coordinates.
Weakness and lessons learned	It is hard to point to an exact geographical pattern, using bubble maps, fires of different sizes occur all over the state . The fire perimeters would be more representative for this task.

Task 2. What are the wildfire trends in terms of their count and area burned?



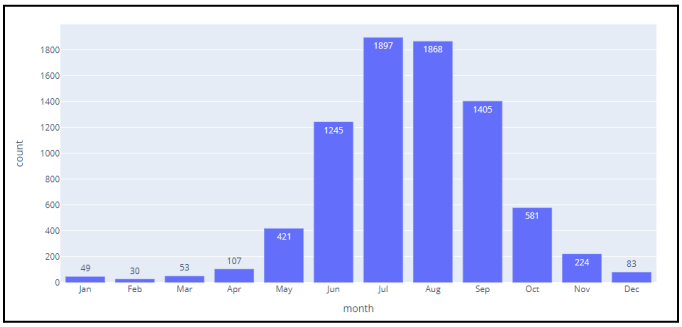
Expressiveness	Trends are very well seen on the line charts. Trend lines help even more if the obvious change over time can't be identified.
Effectiveness	Blue color was chosen for the count line, red color was chosen for area burned line. The lines are separable even on the same chart. Dashed lines of the same colors were chosen for trend lines to indicate belonging to a particular chart, but, at the same time, be separable.
Findings	The increasing trends are clearly visible for both fire counts and area burned. Also, there is some correlation/causation between two charts.
Challenges	The implementation of the trend lines required using an external library, which took unexpected, additional effort.
Weakness and lessons learned	One might think that four lines in one chart is too busy. But it saves space and the filtering feature resolves this problem.

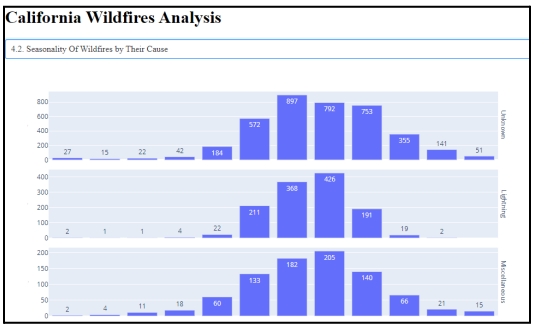
Task 3. What are the main causes of wildfires?



Expressiveness	Stacked bar charts not only show the major causes, but also are good for trends analysis.
Effectiveness	Categorical colormap makes stacked bars be separable. The bar's length allows one to see the magnitude.
Findings	Most of the wildfires' causes are unidentified. The biggest identified cause is lightning.
Challenges	It is hard to find a color map with so many categories that would be appropriate for all color deficiencies and would be distinguishable at the same time.
Weakness and lessons learned	Probably, it would be better to reduce the number of categories (causes) to eliminate the non-significant ones.

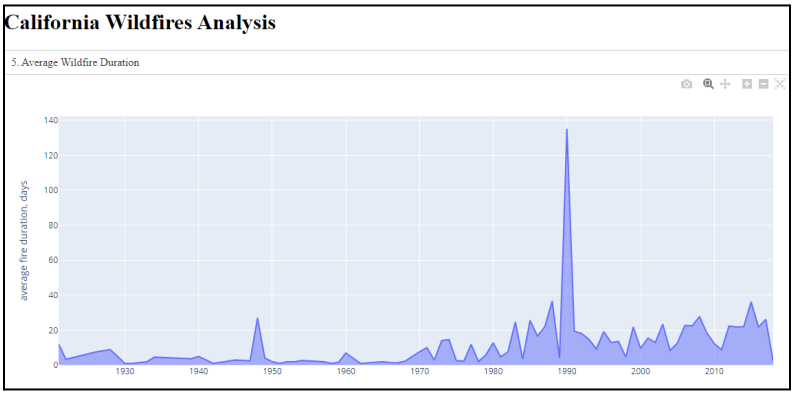
Task 4. What is the seasonality of wildfires? Can we confirm using this dataset visualization, that the seasonality changes to fall and winter months?





Expressiveness	Bar charts of quantitative attribute (Count) over one ordered attribute (Month) allow to see when the most fires occur.
Effectiveness	Line marks express value attribute with aligned vertical position, separate key attribute with horizontal position. Partition by key Cause using facets can help to separate categories.
Findings	It is clear that May-November is the main season of the wildfires onset overall and also if comparing by causes.
Challenges	The animation doesn't confirm the hypothesis of the shifting of seasonality for the given dataset. Probably, more data is needed for the recent years.
Weakness and lessons learned	Faceted views were not very informative. The Overall chart better represents the seasonality.

Task 5. What is the trend of wildfires' duration over time?



Expressiveness	The area chart helps to see the increased filled area which indicates the trend.
Effectiveness	The height of the line shows the magnitude. The filled area under the line helps to comprehend the change in magnitude over the time.
Findings	The duration of wildfires grows over time (with some fluctuations).
Challenges	The initial idea was to show if the fire departments are managing better over the time and reduce the duration of the fires. However, there are too many dependent variables such as warming climate, fuel density. etc . One can't judge the performance only with this dataset.
Weakness and lessons learned	Showing pure fire duration might be misleading. There can be some normalization done, but other data attributes are required to do this.

Conclusion

The visualizations point to increased trends in counts, geographical spread, area burned and duration of California wildfires. According to Buechi, the increasing trends are also seen in damages to structures and to people. Also, these results are consistent with wildfire activity in the western U.S since the mid-1980s. Although wildfires can be beneficial to the ecosystem, better control and management is needed to avoid the destructive megafires that occurred recently all over the West Coast states.

Bibliography

1. Dataset Source: Historical California Wildfire Data. https://cvw.cac.cornell.edu/PyDataSci1/wildfires
2. Olivia Lai, What Causes California Wildfires? (2022) https://earth.org/what-causes-california-wildfires/
3. Hanna Buechi, Long-term trends in wildfire damages in California. (2021) https://emlab.ucsb.edu/sites/default/files/documents/wildfire-brief.pdf
4. Michael Goss, Climate change is increasing the likelihood of extreme autumn wildfire conditions across California. (2020) https://iopscience.iop.org/article/10.1088/1748-9326/ab83a7