California Wildfires Analysis

DSE241 Sergey Gurvich

Motivation

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.

Main Reasons:

- 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.

The given dataset visualizations should show that our concerns are real and data driven. The action is required as soon as possible.

Sources:
https://emlab.ucsb.edu/sites/default/files/documents/wildfire-brief.pdf
https://earth.org/what-causes-california-wildfires/

Problems

1. Explore wildfires geography over time. Does the activity increase over time? What regions are the most affected?
2. What are the wildfire trends in terms of their count and area burned?
3. What are the main causes of wildfires?
4. What is the seasonality of wildfires? Can we confirm using this dataset visualization, that the seasonality changes to fall and winter months?
5. What is the trend of wildfires' duration over time? Is the average duration of wildfires increasing or decreasing?

The Dataset: Historical California Wildfire Data

Source: https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcQxmUVirXtPwk6P3S4224ilFQDCm62Fiihfntb8AB4HiDHRpUs&s

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

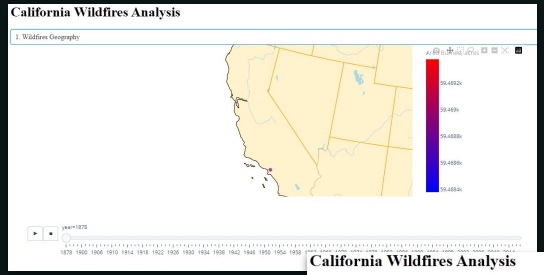
- YEAR
- STATE
- AGENCY
- UNIT_ID
- FIRE_NAME
- INC_NUM (incident number)
- ALARM_DATE
- CONT_DATE (containment date)
- CAUSE
- COMMENTS
- REPORT_AC (acres burned)
- GIS_ACRES (acres burned: GIS data)
- C_METHOD
- OBJECTIVE

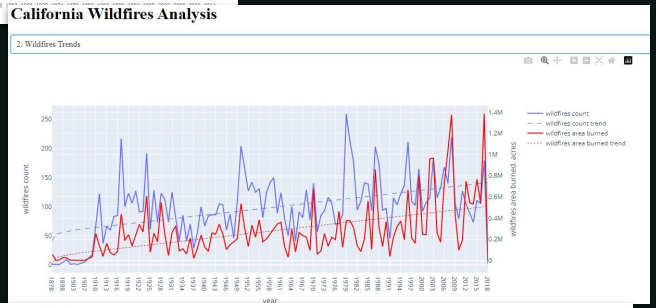
YEAR	UNIT_ID	FIRE_NAME	ALARM_DATE	CONT_DATE	CAUSE	REPORT_AC	GIS_ACRES	C_METHOD	OBJECTIVE
2008	Shasta-Trinity National Forest	SHU LIGHTNING-KIRKMAN	6/20/2008	<Null>	1 - Lightning	2054	2053.58	2 - GPS Air	Suppression (Wildfire)
2008	Sequoia National Forest	MOSES	10/15/2008	10/31/2008	1 - Lightning	105	231.5684	2 - GPS Air	Resource Benefit (WFU)
2008	Tehama - Glenn CAL FIRE	MILL	6/21/2008	6/29/2008	1 - Lightning	13580	13511.76	8 - Unknown	Suppression (Wildfire)
2008	Plumas National Forest	FOUR MILE	6/21/2008	<Null>	1 - Lightning	<Null>	788.7921	8 - Unknown	Suppression (Wildfire)
2008	San Joaquin River National Wildlife Refuge	RIVER	6/24/2008	6/25/2008	9 - Miscellaneous	580	573.1215	8 - Unknown	<Null>
2008	Lava Beds National Monument	JACK	8/17/2008	8/31/2008	1 - Lightning	5397	6067.5	7 - Mixed Collection Methods	Suppression (Wildfire)
2008	Sequoia - Kings Canyon NP	HIDDEN	9/10/2008	10/8/2008	1 - Lightning	3685	3685.978	7 - Mixed Collection Methods	Suppression (Wildfire)
2008	Sequoia - Kings Canyon NP	TEHIPITE	7/19/2008	12/1/2008	1 - Lightning	4140	11648.09	3 - Infrared	Resource Benefit (WFU)
2008	CA Desert District - BLM	MONTGOMERY	4/7/2008	4/7/2008	4 - Campfire	290	289.9371	1 - GPS Ground	Suppression (Wildfire)
2009	CRA	DEER	7/6/2009	7/21/2009	14 - Unknown / Unidentified	<Null>	330.1603	8 - Unknown	Suppression (Wildfire)
2009	Vandenberg Air Force Base	HIGHWAY	9/30/2009	10/3/2009	11 - Powerline	562.127013	562.13	8 - Unknown	Suppression (Wildfire)
2009	Angeles National Forest	TUJUNGA	7/5/2009	7/9/2009	1 - Lightning	180	210.5635	7 - Mixed Collection Methods	Suppression (Wildfire)
2009	Angeles National Forest	OSITO	7/13/2009	7/13/2009	2 - Equipment Use	300	332.0352	2 - GPS Air	Suppression (Wildfire)
2009	Angeles National Forest	MORRIS	8/25/2009	11/30/2009	7 - Arson	2163	2236.504	7 - Mixed Collection Methods	Suppression (Wildfire)
2009	Angeles National Forest	STATION	8/26/2009	9/22/2009	7 - Arson	160371	160833.1	7 - Mixed Collection Methods	Suppression (Wildfire)
2009	San Bernardino National Forest	MCKINLEY	6/3/2009	6/3/2009	1 - Lightning	150	166.3999	6 - Hand Drawn	Suppression (Wildfire)
2009	San Bernardino National Forest	ELM	7/7/2009	7/7/2009	9 - Miscellaneous	350	348.3582	6 - Hand Drawn	Suppression (Wildfire)
2009	San Bernardino National Forest	COTTONWOOD	8/27/2009	8/31/2009	9 - Miscellaneous	2409	2410.618	7 - Mixed Collection Methods	Suppression (Wildfire)
2009	San Bernardino National Forest	SHEEP	10/3/2009	10/24/2009	9 - Miscellaneous	7128	7237.873	7 - Mixed Collection Methods	Suppression (Wildfire)
2009	San Bernardino CAL FIRE	CRAFTON	9/23/2009	9/27/2009	14 - Unknown / Unidentified	347	348.1416	1 - GPS Ground	Suppression (Wildfire)
2009	San Bernardino CAL FIRE	FORT	2/5/2009	2/7/2009	18 - Escaped Prescribed Burn	945	943.448	1 - GPS Ground	Suppression (Wildfire)
2009	San Bernardino CAL FIRE	OAK GLEN 3	8/30/2009	9/10/2009	7 - Arson	1159	983.2415	2 - GPS Air	Suppression (Wildfire)
2009	San Bernardino CAL FIRE	PENDLETON	8/31/2009	9/3/2009	7 - Arson	860	838.701	2 - GPS Air	Suppression (Wildfire)
2009	Monterey - San Benito CAL FIRE	BRYSON	8/26/2009	8/28/2009	15 - Structure	<Null>	2256.828	7 - Mixed Collection Methods	Suppression (Wildfire)
2009	Monterey - San Benito CAL FIRE	GLORIA	8/28/2009	9/8/2009	9 - Miscellaneous	<Null>	6435.952	2 - GPS Air	Suppression (Wildfire)
2009	CA Desert District - BLM	LAMB	7/11/2009	7/11/2009	10 - Vehicle	288	288.1363	7 - Mixed Collection Methods	Suppression (Wildfire)
2009	CA Desert District - BLM	MIDWAY	4/17/2009	<Null>	2 - Equipment Use	112	111.6097	1 - GPS Ground	Suppression (Wildfire)
2009	San Mateo - Santa Cruz CAL FIRE	LOCKHEED	8/12/2009	8/23/2009	14 - Unknown / Unidentified	7817	7783.064	1 - GPS Ground	Suppression (Wildfire)
2009	San Mateo - Santa Cruz CAL FIRE	LOMA	10/25/2009	10/27/2009	14 - Unknown / Unidentified	650	669.3967	1 - GPS Ground	Suppression (Wildfire)
2009	Hoopa Valley Tribe	MILL CREEK 4	10/7/2009	11/18/2009	9 - Miscellaneous	3689	2831.125	8 - Unknown	Suppression (Wildfire)
2009	Humboldt - Del Norte CAL FIRE	PRATT	9/11/2009	9/11/2009	2 - Equipment Use	140	141.965	2 - GPS Air	Suppression (Wildfire)

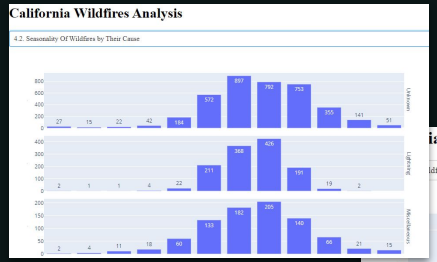
Dataset Source: https://cvw.cac.cornell.edu/PyDataSci1/wildfires

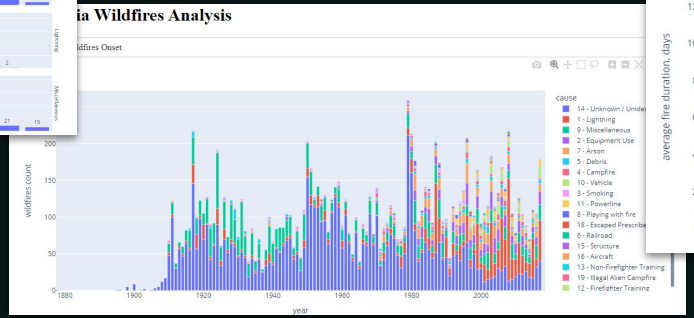
Solution

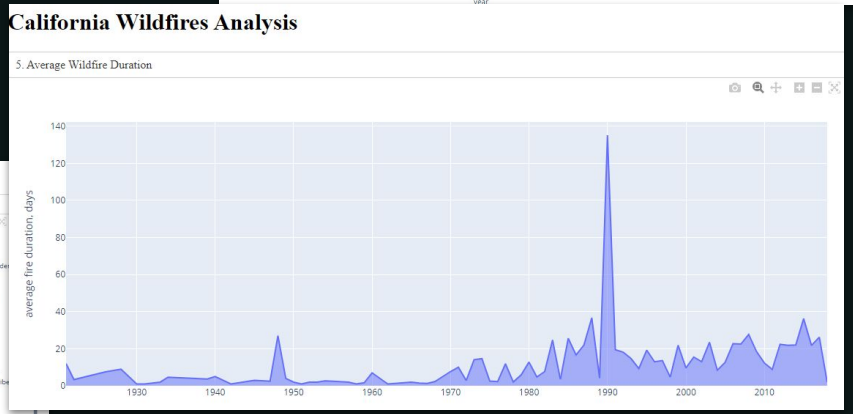
- ★ Programming Language: Python
- ★ IDE: Jupyterlab
- ★ Data Wrangling: Pandas, Scikit-Learn
- ★ Geofencing: Radar API
- ★ Visualization: Plotly, Plotly Express
- ★ Web Application / Dashboard: Plotly Dash





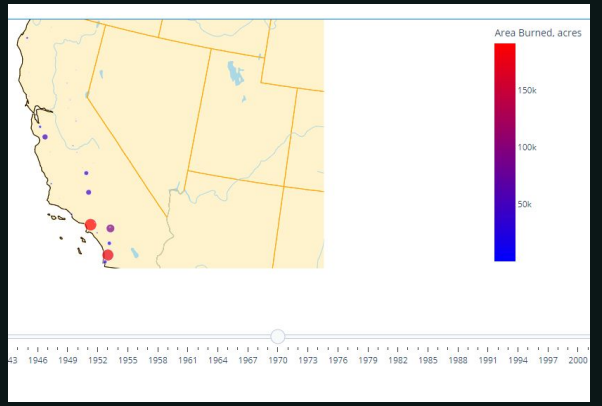


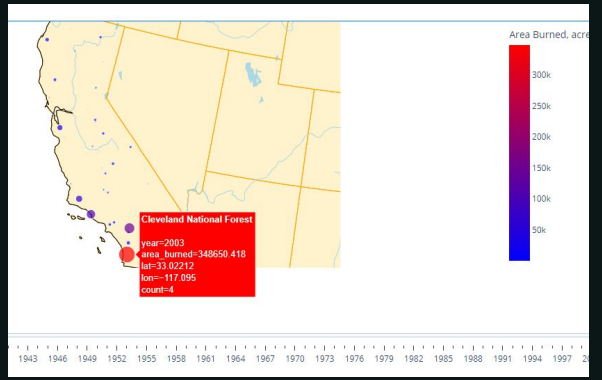




Task 1. Wildfires Geography

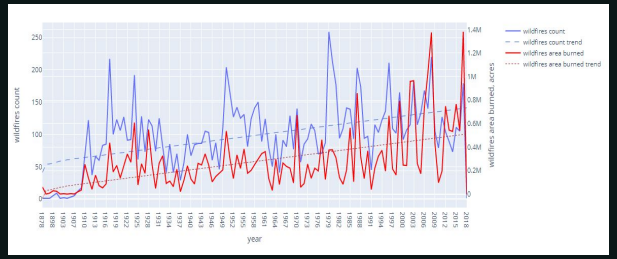
Data Wrangling:	Get geographical coordinates of fire locations by using geofencing API. Grouping to get counts and area burned by years.
Idiom:	Bubble map
Interaction:	Zoom, pan, hover to see details, animation over years
What (Data) :	Two quantitative value attributes Total Fires, Area Burned One ordered attribute: Year Geographic coordinates
Why (Tasks) :	Explore the most affected geographical areas in California. Compare them over the time.
How (Encode) :	Space: use given geometry for area mark boundaries. Color: continuous color map for area burned attribute Circle size: area burned attribute
Findings:	The fire intensity is increasing over the time. It is hard to point exact geographical pattern, using bubble map, the fires of different sizes occur all over the state. The fire perimeters would be more representative for this task.

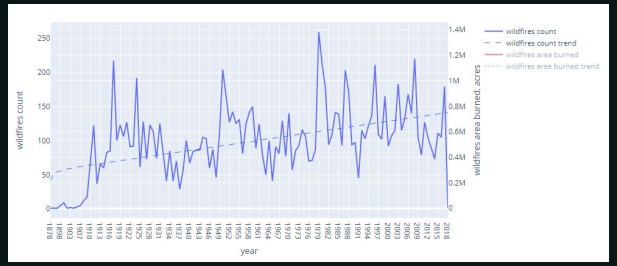


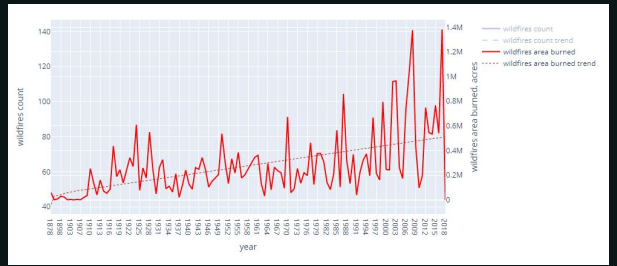


Task 2. Wildfires Trends

Data Wrangling:	Grouping to get counts and area burned by years. Calculating trendlines for both charts
Idiom:	Dual Axis Line Charts
Interaction:	Filter, zoom, hover to see details
What (Data) :	Two quantitative value attributes Total Fires, Area Burned. One ordered attribute: Year
Why (Tasks) :	Show trends of wildfire counts and area burned over the years.
How (Encode) :	Dot chart with connection between dots. Line plot for fitted quadratic. Colors: red and blue for separability of attributes
Findings:	The increasing trends are clearly visible for both fire counts and area burned.

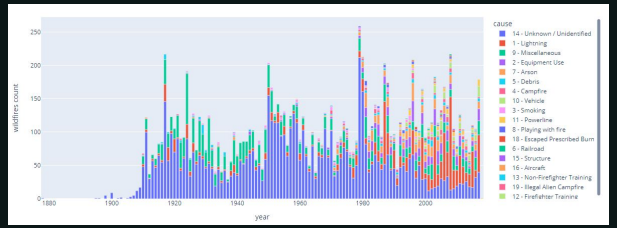


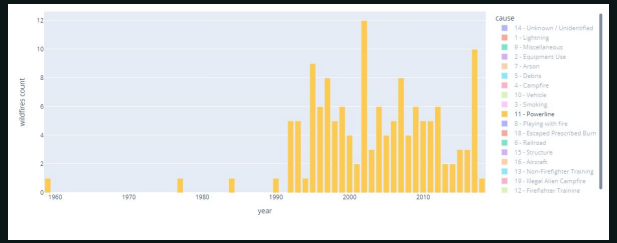


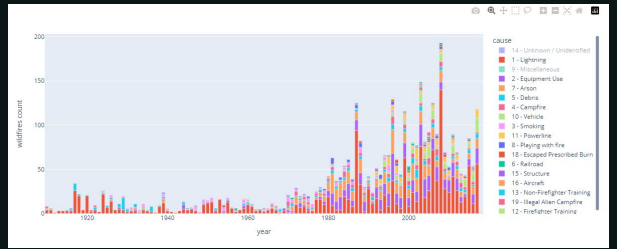


Task 3. Wildfires Causes

Data Wrangling:	Grouping to get counts by categories and years. Ordering by causes.
Idiom:	Stacked Bar Chart
Interaction:	Filter categories, zoom, hover to see details
What (Data) :	One quantitative value attribute Wildfires Count. One categorical attribute: Cause One ordered attribute: Year
Why (Tasks) :	Explore the main causes of wildfires
How (Encode) :	Bar glyph with length-coded subcomponents of value attribute for each category of secondary key attribute. Separate bars by category of primary key attribute.
Findings:	The main cause falls under ‘Unknown’ category. The next biggest cause is lightning. ‘Power Lines’ appear mostly in the recent decades.

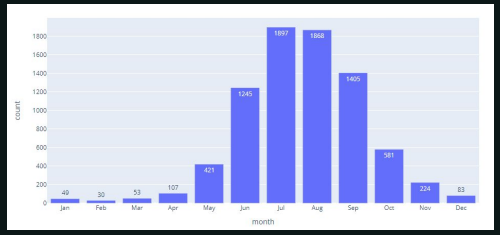


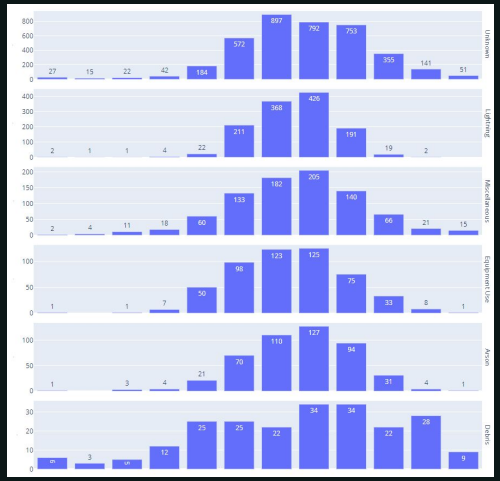


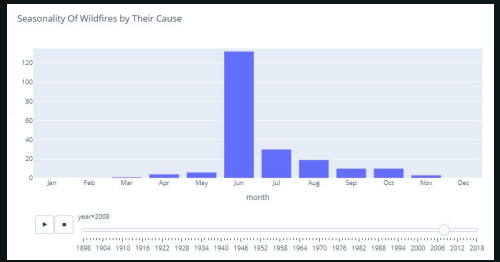


Task 4. Wildfires Seasonality

Data Wrangling:	Grouping by months, causes and years to get counts Join with another dataset which had all years and months to avoid ‘jumping’ while playing the animation.
Idiom:	Bar charts, Faceted Bar Charts
Interaction:	Zoom, hover to see details, animation over years
What (Data) :	One quantitative value attribute Wildfires Count One categorical attribute: Cause Two ordered attributes: Month, Year
Why (Tasks) :	Explore, discover, compare seasonality of wildfires onsets, hypothesis confirmation
How (Encode) :	Line marks express value attribute with aligned vertical position, separate key attribute with horizontal position. Partition by key Cause.
Findings:	It is clear that May-November is the main season of the wildfires onset overall and also if comparing by the causes. The animation DOESN’T confirm the hypothesis of the shifting of seasonality for the given dataset.

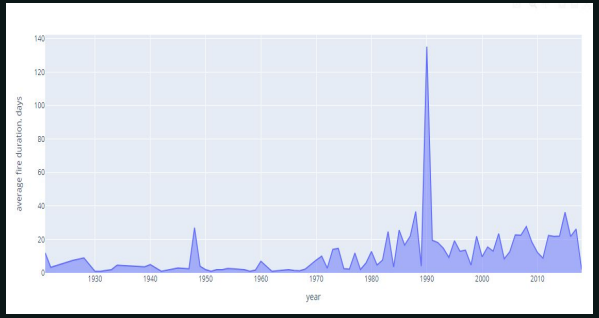






Task 5. Wildfires Duration

Data Wrangling:	Calculate duration of each fire Group by year to get average durations
Idiom:	Area Chart
Interaction:	Zoom, hover to see details
What (Data) :	One quantitative value attribute: Wildfires Duration in Days One ordered attribute: Year
Why (Tasks) :	Show Trend
How (Encode) :	Dot chart with connection between dots, color filled area below the dots.
Findings:	The duration of wildfires grows over time.



Conclusions

1. The annual number of wildfires and total area burned has clear increasing trend over the years all over the state
2. Major identified cause of the wildfires is natural (lightenings)
3. As expected, the wildfires occur mainly between May and November. However, the visualization didn’t reveal any obvious shifting/extending from summer to fall as one of the articles states.
4. The duration of fires over years also has positive trend.


Thank you.