STA 141B Final Project

Bird Densities and Fire Ecology

Jonathan Casas-Ramirez, Adam Hetherwick

Instructor: Peter Kramlinger

12/15/2023

Figure 1: Overhead View of California Wildfire (CalFire)

Introduction

According to the AR6 Synthesis Report: Climate Change 2023, a report conducted by the Intergovernmental Panel on Climate Change (IPCC), "human activities have caused global warming and increased greenhouse gas emissions, which has lead to an increase in global surface temperature reaching 1.1°C above 1850-1900 in 2011-2020" (IPCC, 2023, p. 2). Due to increasing temperatures as a product of global warming, wildfire frequency has drastically increased since the turn of the century. It is generally understood that wildfires impose difficult situations on wild animals in the form of displacement, habitat destruction, and nutrition depletion, though the true extent to which such wildfires affect wild animal populations remains fairly unstudied. This is likely because measuring animal density and biodiversity in given locations is not easy to accomplish, and takes time and resources. Understanding such information will help biologists and professionals in related fields gauge the magnitude of wildfires and their accompanying harm.

This evident displacement could be seen as the reason bird densities may have decreased throughout counties in California as a product of fire ecology growth. Nearly all bird species migrate based on geographic locations and scenic cues, therefore the wellbeing of their preferred migration locations is increasingly important. Examining and visualizing data representations of wildfire frequency in relation to bird densities will help us understand this unfortunate dynamic.

California's wildfire data is made available to us via the CalFire official website (fire.ca.gov). This information was presented in a format that required some data handling and modification techniques, which my partner and I were prepared to handle. We chose to utilize the undocumented API to retrieve such data because the request response was in a convenient format for analysis (more information on Data Acquisition and Processing section).

Obtaining bird species densities relative to given locations is unfortunately not as convenient in comparison to the California wildfire data. Measuring and quantifying such bird populations is difficult and time consuming. The process entails having trained observers stationed out in the field for days or weeks at a time to record bird sightings. Thankfully for us, birding is a common hobby among naturalists.

Our bird population density data was sourced from the publicly available eBird app (The Cornell Lab of Ornithology), which is a platform for naturalists to record their own bird sightings and observations. This given takeoff of utilizing data that is sourced from the public poses the issue of credibility, though the analysis was continued with the assumption that each county and each species had an equal chance of being misidentified. The eBird database was webscraped using numerous Python libraries, and the species observations for each county were gathered.

Our eBird database being part of a common hobby for naturalists poses downsides as well, for example we have had a drastic bloom in data in recent years as the hobby increased. To accommodate for this, we omitted years that were visibly affected by such bloom, and instead utilized a California State Geoportal database via a documented API, which obtained data from California's Department of Fish and Wildlife. This database includes general avian biodiversity data per county, which is resistant to birding's recent popularity explosion.

These three datasets provide us with the necessary information we need to assess our question of how California wildfire frequency and magnitude affect migratory bird populations. We hypothesize that as wildfire size, duration, and intensity increase, bird populations will decrease as a product of unstable habitat conditions and lower food or nutrient availability. We will assess such hypotheses by accessing and handling our data to retrieve the important information, comparing and contrasting the important predictors, and visualizing our findings via interactive maps, plots, and tables.

Data Acquisition and Processing

To assess our claim that bird population densities will decrease as fire frequencies increase, we need to access credible data sources that will assist us in research. We accessed California's natural wildfire data via fire.ca.gov, maintained by The Department of Forestry and Fire Protection (CalFire). Initial availability of this data was in the form of a webscrapable table with fire name, county, start date (date only, no hour timestamp), total acres, and percent containment. This information could have been sufficient for analysis, but we figured that fire duration may have an affect on bird populations as well. To obtain this information, it is possible to click on each fire's name, and retrieve the start date with hour timestamp, extinguished date with hour timestamp, latitude and longitude of fire, and other information not needed for our analysis. This entailed utilizing the undocumented API for all fires from 2016-2023 (8 separate web pages), and then individually utilizing the API for each fire to gather the remaining information. The function created to retrieve such information initiated many API requests, thus resulting in a function that took some time to run (23 minutes for my potato computer).

This function involved using numerous coding concepts and python libraries to manipulate the data and obtain what we want. The API request result was thankfully readable in a nice json format, though the data itself required some handling. The date for example was in a non hashable format, therefore we converted it using the pandas library. Some of the starting date and extinguishing dates were either in the form of *Na* or *None*, or contained a negative value in the hour timestamp. To accommodate for this we set up *if* statements and string slice indexing to see if the dates were in acceptable format or not. The times were converted from milliseconds to hours, minutes, seconds via another function. After accessing all 1,919 individual fire APIs from

2016-2023 and 23 minutes later, we retrieved our initial California Fire dataframe (More data handling techniques were done later as needed).

To contrast our fire dataframe, we webscraped data from eBird. The initial site to scrape was in the form of California counties, with an outlink at each county to examine the species observations for each date. This meant for all 58 counties in California, we had to web scrape and find the *href* identifier for each county then parse the html from the outlink. To start, we split the initial html and found all the county's *hrefs*, then for each county, opened their individual observation page. We used the *re* package to parse and split the html by certain identifiers to obtain the url needed for the request. After making the request, the result had too much information for Jupyter Notebook to print. We had to omit unnecessary html and locate the *div* with the *id='place-species-observed-results'*. We continued to split the html into smaller chunks to gather the observation date, observation number, and county. This dataframe was assembled with 19,064 total observations, one for each entry for each county.

[7]:

	County	Species_Count	Date
0	Los Angeles	1	12-15-2023
1	Los Angeles	1	12-14-2023
2	Los Angeles	4	12-14-2023
3	Los Angeles	3	12-14-2023
4	Los Angeles	10	12-14-2023
...	...	...	...
19060	Calaveras	1	11-22-2018
19061	Calaveras	1	11-22-2018
19062	Calaveras	1	10-11-2018
19063	Calaveras	2	05-25-2018
19064	Calaveras	1	09-10-2016

Figure 2: Initial eBird species observation webscraped data frame

To access the California State Geoportal database, we interacted with the documented API that was available to us. Using this was by far the easiest method of the three and took little to no data handling and manipulation. The result from the API request was put into a *json* format which was then put into a pandas data frame.

To continue with our data manipulation process, we calculated the total duration of each fire in hours, by converting the start time and extinguish time in our fire dataframe to the appropriate format, then found the difference. This raised some errors because not all fires had a fire extinguish timestamp, and some didn't have a fire start timestamp. Using if statements and error handling techniques like the try and except block, we input the total burn duration into our fire dataframe.

Another issue was that our fire dataframe had some fires that stretched into multiple counties, or sometimes even into Arizona or Oregon. To accommodate for this, we split each fire that stretched into multiple counties or states into individual rows. We did this with nested for loops and *.loc* indexing to implement correct data into our dataframe.

[18]:

	County	Year	Total_Burn_Time	AcresBurned	Log_Burn_Time	Log_Acres_Burned
0	Alameda	2016	14	211	2.639057	5.351858
1	Alameda	2017	15879	537	9.672753	6.285998
2	Alameda	2018	31946	1128	10.371802	7.028201
3	Alameda	2019	130	533	4.867534	6.278521
4	Alameda	2020	1199	396985	7.089243	12.891654
...	...	...	...	...	...	...
381	Yuba	2019	217	268	5.379897	5.590987
382	Yuba	2020	142	2467	4.955827	7.810758
383	Yuba	2021	315	1353	5.752573	7.210080
384	Yuba	2022	376	223	5.929589	5.407172
385	Yuba	2023	54	187	3.988984	5.231109

Figure 3: Final fire dataframe per county and year

We then calculated the aggregated total burn time and total acres burned for each year for each county into a new dataframe. This was pivotal for our analysis because it is the basis in which we could merge it with our eBird and California Geospatial biodiversity data.

After omitting the irrelevant counties, we noticed a large variance in total acres burned on a year to year basis. Some counties would have over 1,000,000 acres burned then the next year be less than 1,000. Plotting such data is very difficult and hard to gauge. To accommodate for this, we decided to utilize a log transformation, to decrease variance and center the data.

We also decided to group our fire data by year to help us gauge the year to year changes and compare that with bird population changes. We did the same with the bird populations, and created a basis to merge our dataframes. Doing so resulted in a 192 row dataframe with each row signifying a unique year and county, with both raw counts and log transformations of acres burned, total duration burned, and total bird species observed.

[37]:

	County	Year	Species_Count	Log_Species_Count	Total_Burn_Time	AcresBurned	Log_Burn_Time	Log_Acres_Burned
0	Alameda	2016	2	0.693147	14	211	2.639057	5.351858
1	Alameda	2017	4	1.386294	15879	537	9.672753	6.285998
2	Alameda	2018	5	1.609438	31946	1128	10.371802	7.028201
3	Alameda	2019	5	1.609438	130	533	4.867534	6.278521
4	Alpine	2016	3	1.098612	288	0	5.662960	-inf
...	...	...	...	...	...	...	...	...
187	Yolo	2017	6	1.791759	4487	1700	8.408940	7.438384
188	Yolo	2018	7	1.945910	13787	90836	9.531481	11.416811
189	Yolo	2019	6	1.791759	208	2253	5.337538	7.720018
190	Yuba	2016	2	0.693147	5	389	1.609438	5.963579
191	Yuba	2019	2	0.693147	217	268	5.379897	5.590987

Figure 4: Final eBird and CalFire merged dataframes

This gave us ample opportunity to plot and visualize the data to investigate our hypothesis. The following section dives into such predictions and shows our findings.

Data Visualization

Using the data frames created, we could begin to visualize the behavior of our data over time. Doing so would be the first step in helping us answer if California wildfires affect bird populations. The total log transformed fire burn time by county over each year will show us our trend over time.

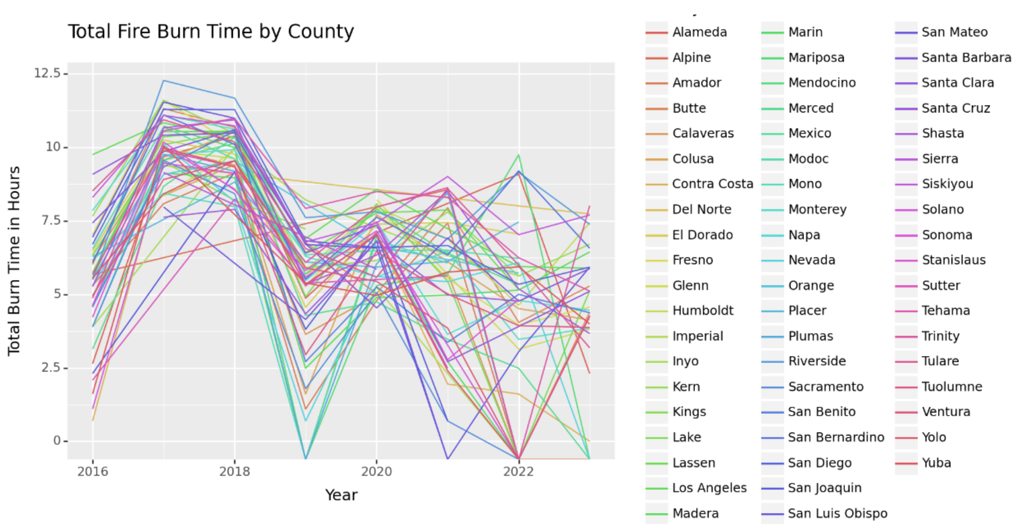


Figure 5: Total fire burn time by county

This plot leads us to believe that there is an apparent trend in total fire burn time by county. We can see that the fire burn times were very high in 2017 and 2018, then decreased in 2019, increased in 2020, then decreased until present. The decrease of burn time that we see from our plot can be a result of a mixture of different factors. One possible explanation for the decrease of burn time is that CalFire has increased response and extinguish times. Another factor that may decrease the amount of burn time is the recent effort to implement the use of beneficial fires called prescribed burn (Wildfire and Forest Resilience Task Force). Prescribed burns help remove overgrown vegetation and promote forest health. Whether or not this information has a correlation to the bird densities in this area remains to be determined. We can also visualize the total acres burned for each county over each year.

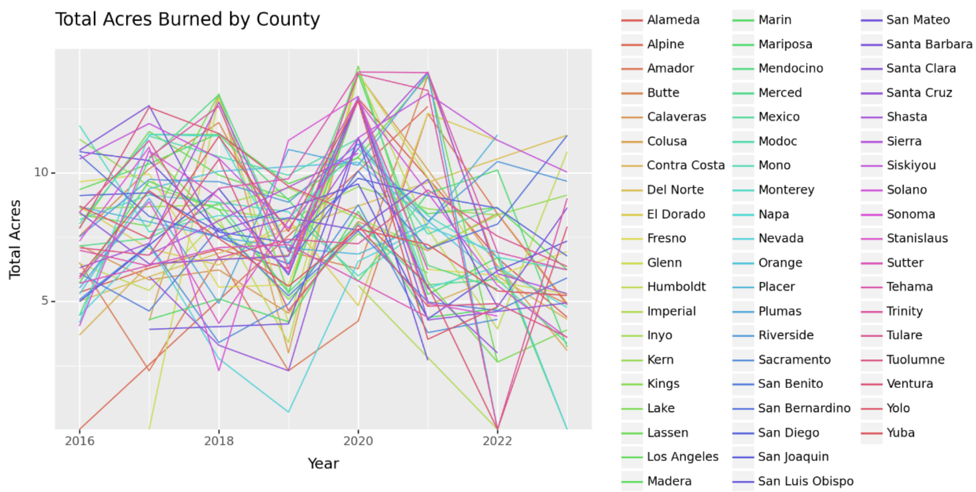
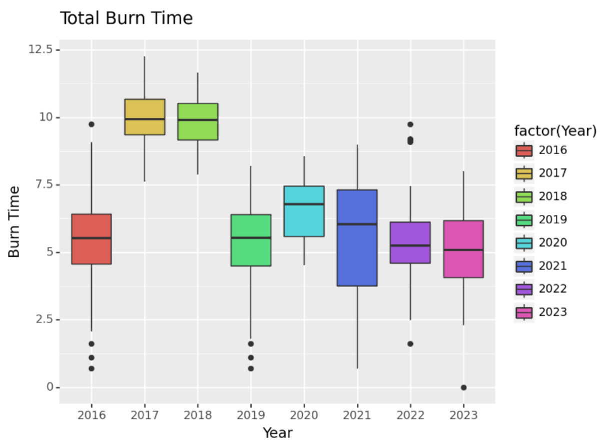


Figure 6: Total Acres Burned by County

This plot however leads us to believe there is not much of a trend in total acres burned each year for each county. This could be due to California being so long North and South that the southern areas received more precipitation than compared to the hotter central valley areas or northern areas. Grouping our results by year may give us an easier to interpret visual.



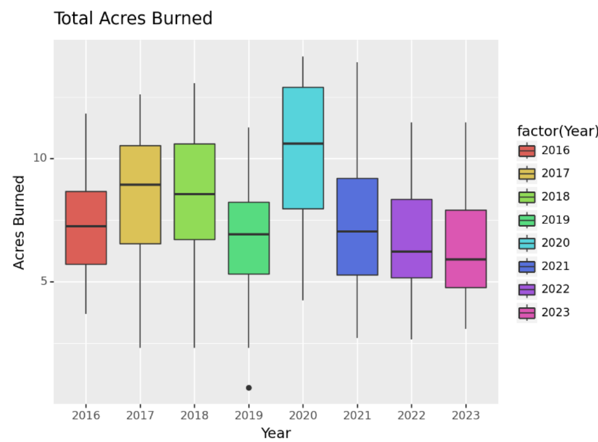


Figure 7: Total burn time by year for all counties

Figure 8: Total acres burned by year for all counties

These plots indicate that grouping by year is much more optimal for analysis, because it is much easier to visualize and understand the year to year changes for California's fire behavior. We notice that both burn time and acres burned decrease after 2020. Next we can understand the behavior of the bird species observations per year for each county from the eBird database.

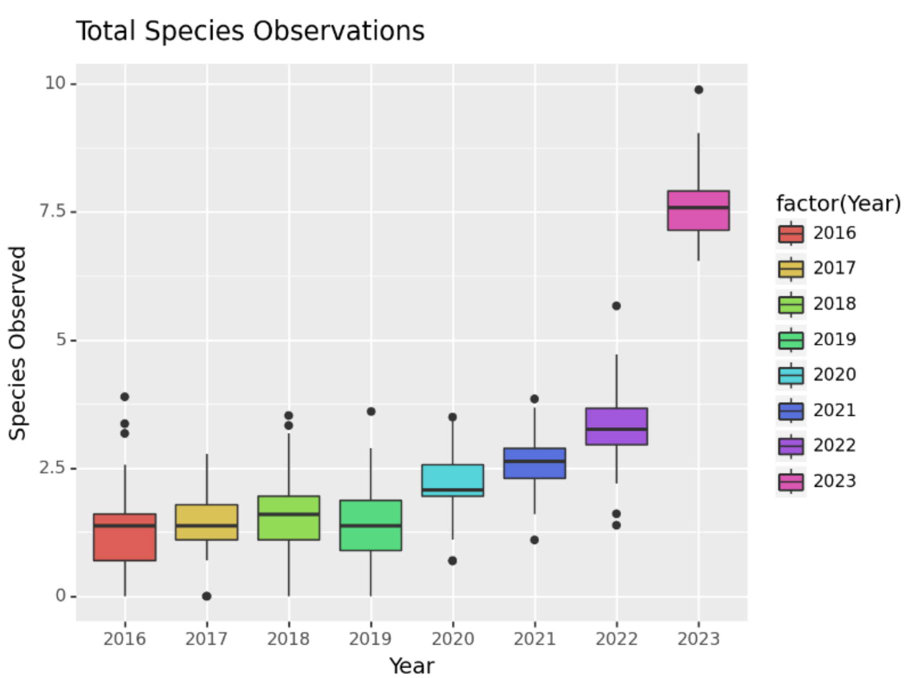


Figure 9: eBird species observations for each year grouped by county

Here we see the predicted bloom in 2023 observations, as represented through the increasing popularity of birding as a hobby. Because of this clear bloom in 2023, and the upward trend after the year 2020, my partner and I took this into account and decided to compare the fire acres and fire duration with the species biodiversity dataset from California State Geoportal (accessed via documented API).

[125]:

	County	SpBioRnkEco	Year	Total_Burn_Time	AcresBurned	Log_Burn_Time	Log_Acres_Burned
0	Siskiyou	145	2021	8096	478345	8.999125	13.078088
2	Modoc	1318	2021	662	2035	6.495266	7.618251
3	Humboldt	265	2021	967	2462	6.874198	7.808729

Figure 10: Ecological Biodiversity rank per county in 2021

The Cal State Geoportal from 2021 compared with the total fire burn time and acres burned data leads us to believe that there is not much difference on bird species prevalence. This is seen by the species biodiversity rank for Siskiyou being lower (indicating more biodiversity) than Modoc despite having much more burn hours and acres. The visualization of the burns geographically can help us understand why some counties might be more prone to wildfires than others.

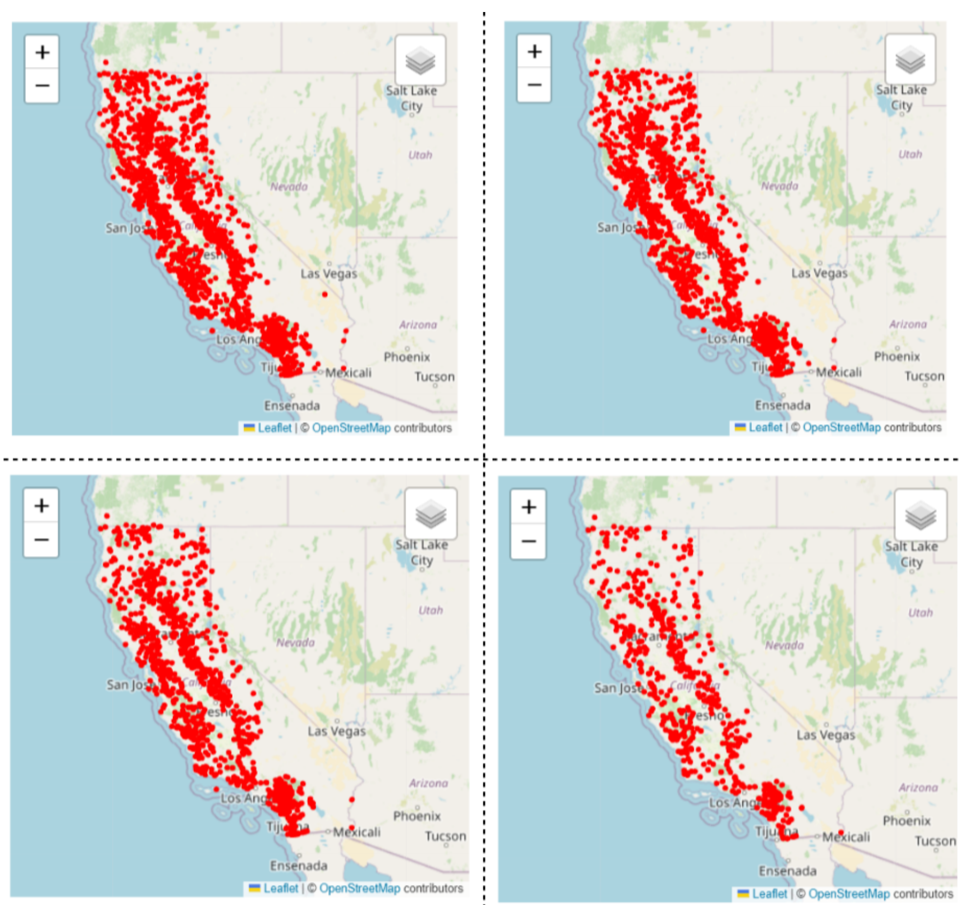


Figure 11: Geographical interpretation of all fires by year
(top left 2023-2016, top right 2021-2016, bottom left 2019-2016, bottom right 2017-2016)

With this information, we can visualize the locations of each fire and how some areas may be more vulnerable to larger scale fires. Proximity to fire stations and dense urban areas plays a factor into such variables. These larger fires will likely drive down bird species biodiversity. Our final visualization plot can assist us in viewing the species observations over time in comparison to fire sizes.

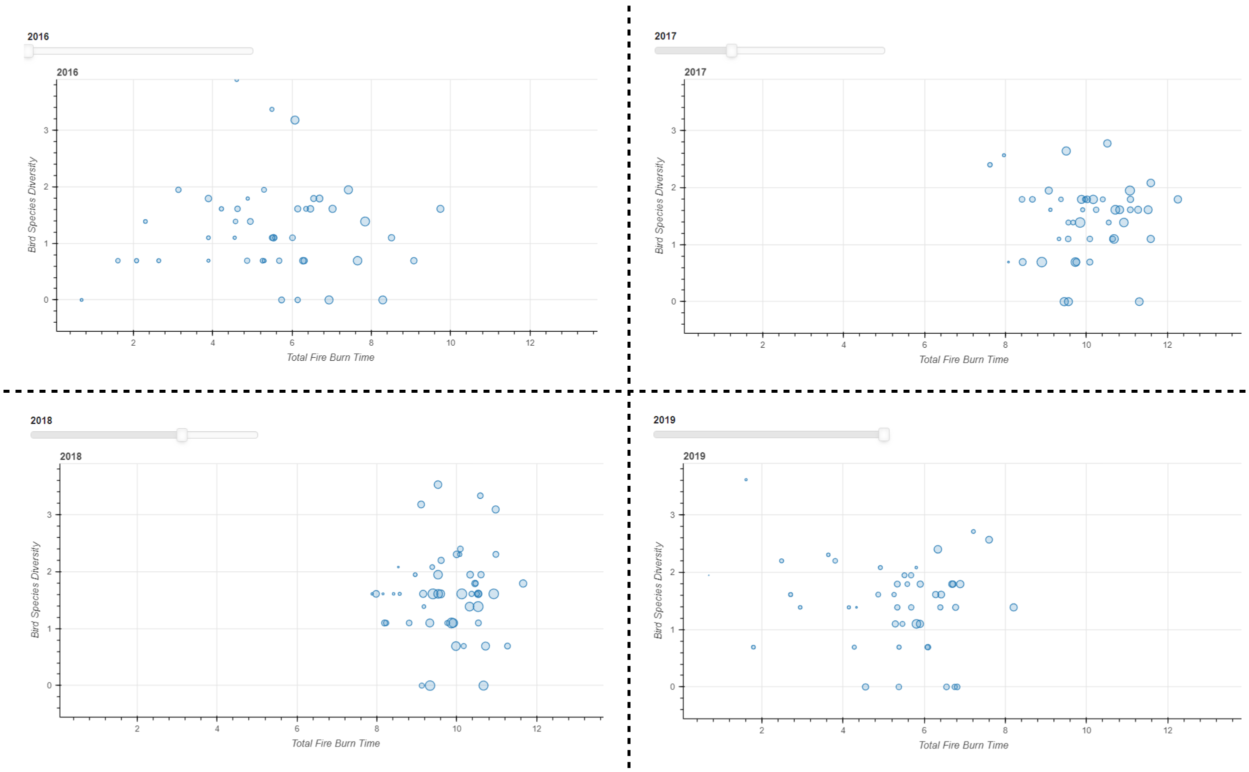


Figure 12: Bird species diversity (y-axis) by total fire burn time (x-axis)
(2016 top left, 2017 top right, 2018 bottom left, 2019 bottom right)

These slider plots indicate that there is no apparent change in bird density in comparison to fire burn time. This also supports our findings in Figure 10. This could be due to the fact that the birds that are native to California are used to migrating through/around burned areas and have evolved to accommodate for that. This is contradictory to our initial hypothesis. Our visualizations of the data however proved to be incredibly helpful and helped us determine our conclusion.

Conclusions

In this analysis we investigated how California fire ecology affects bird populations. We examined how total fire acres and total hour duration of all fires per county relates to the bird populations. We conducted our analysis with data from CalFire, eBird, and California State Geoportal. We utilized numerous data handling and visualization techniques learned this quarter to apply our analysis, such as and not limited to: API interaction (both undocumented and documented), web scraping, html parsing, error handling, data frame merging, aggregation, geographical and slider based visualizations, using plotnine, and others.

We hypothesized that as both fire duration and acres burned increased, general bird species sightings would decrease. Our findings supported contradictory conclusions however, and we found that total acres burned and total duration seemed to have no effect on bird populations. We chose to refine our study years to 2016 to 2019, which were not impacted by eBird's recent data bloom because of more and more people enjoying the birding hobby. The California State Geoportal biodiversity data for 2021 also supported this claim, as seen through the biodiversity not changing in comparison to the total acres burned and total fire duration.

Overall, this project was extremely helpful for our learning and was a great opportunity for Jonathan and I to apply our knowledge learned from STA141B this quarter. Although our data visualization did not align with our initial hypothesis, we both still learned a lot and were glad to apply our learnings to a subject we are both passionate about. We appreciate Professor Kramlinger and the teaching assistants for their efforts teaching us such tools and techniques this quarter.

Works Cited

1. Intergovernmental Panel on Climate Change. "AR6 Synthesis Report: Climate Change 2023." IPCC, United Nations, 2023, https://www.ipcc.ch/report/sixth-assessment-report-cycle/.
2. Overhead View of California Wildfire (CalFire). (n.d.). NASA Earth Observatory https://earthobservatory.nasa.gov/images/144225/camp-fire-rages-in-california
3. Wildfire and Forest Resilience Task Force. California's Strategic Plan for Expanding the Use of Beneficial Fire. March 2022, https://34c031f8-c9fd-4018-8c5a-4159cdff6b0d-cdn-endpoint.azureedge.net/-/media/calfire-website/what-we-do/natural-resource-management/prescribed-fire/californias-strategic

-plan-for-expanding-the-use-of-beneficial-fire-march-16_2022.pdf?rev=f2694844e0d64c9f81e26380af608631. Accessed 12/1/2023

Code Appendix

The code used for this assignment is available via GitHub.

https://github.com/AJHetherwickstats/STA141B_Project