.Climate Risk Viewer – Project Report

Author: Huu Dang Nguyen, Ammar Al-baadani
Date: 11/27/2023
Class: ITSS 4V95

Table of Contents:

Overview

Problem Statement

The mission of the National Forest Service is to identify and address vulnerabilities, threats, and risks associated with wildfires and climate change. This includes a focus on key resources and values within the National Forest System, such as water and watersheds, biodiversity and species at risk, forest carbon, and mature and old growth characteristics. The mission extends to evaluating and adjusting current management practices to enhance the effectiveness of adaptation and mitigation actions in response to these identified challenges.

Purpose

The project aims to improve existing methods and assessments on Climate Risk Viewer for National Forest System managers, enabling them to areas with the most urgent need for climate change adaptation. This approach facilitates informed decision-making, providing crucial support for the development of strategies that promote climate resilience and sustainable development.

Climate Risk Viewer introduction

Climate Risk Viewer, a tool containing over 70 map layers. It breaks down the overlap of various resource values with climate exposure, vulnerability, and management. This Climate Risk Viewer by National Forest System serves a wide range of audiences from policymakers, researchers to communities, non-profit organizations, and individuals that seeking to understand and respond to the challenges posed by a changing climate

Focusing on Firesheds and Watersheds, our team believes they are helpful to see the full picture of environmental issues – connecting the dots between wildfire risks and water resources. It is key to crafting strategies for adapting and standing strong in the face of climate change.

Data Overview and Data Pre-processing

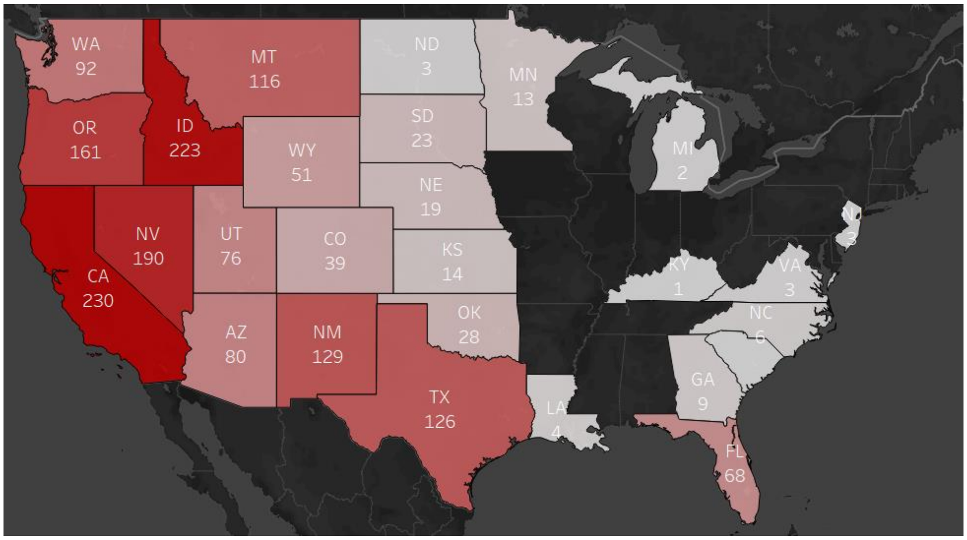
Data overview

- Dataset Scope: from 1992 to 2015, encompassing 1.88 million wildfires in the United States.
- Data Source: acquired from federal, state, and local fire organizations' reporting systems
- Storage Format: stored as an SQLite database file
- Criteria: records required discovery date, fire size, and a point location as precise as the Public Land Survey System section (1-square mile grid).
- Data Transformation: aligned with the National Wildfire Coordinating Group standards for consistency.
- Quality: some error checks performed, and redundant records were removed.
- Geographical Representation: 1.88 million geo-referenced wildfire records, covering an extensive 140 million acres (about the area of Texas) burned over the 24-year period.
- Accessibility: the dataset is available at https://www.fs.usda.gov/rds/archive/Catalog/RDS-2013-0009.4/.

Data preprocessing

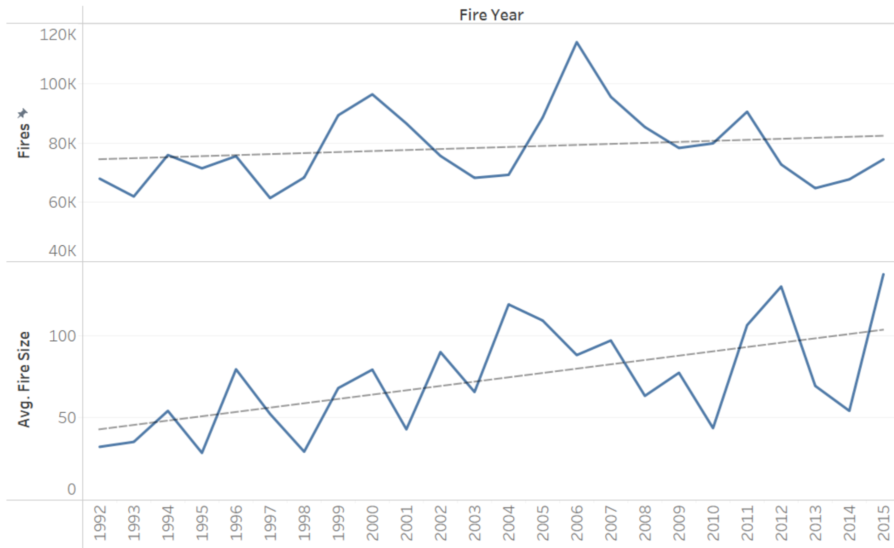
The dataset went through preprocessing within the SQLite database and Python, aligning with National Wildfire Coordinating Group standards, performing error checks, removing redundancies, and date reformatted. These steps have resulted in a clean, reliable dataset and ready to use.

Exploratory Data Analysis- EDA

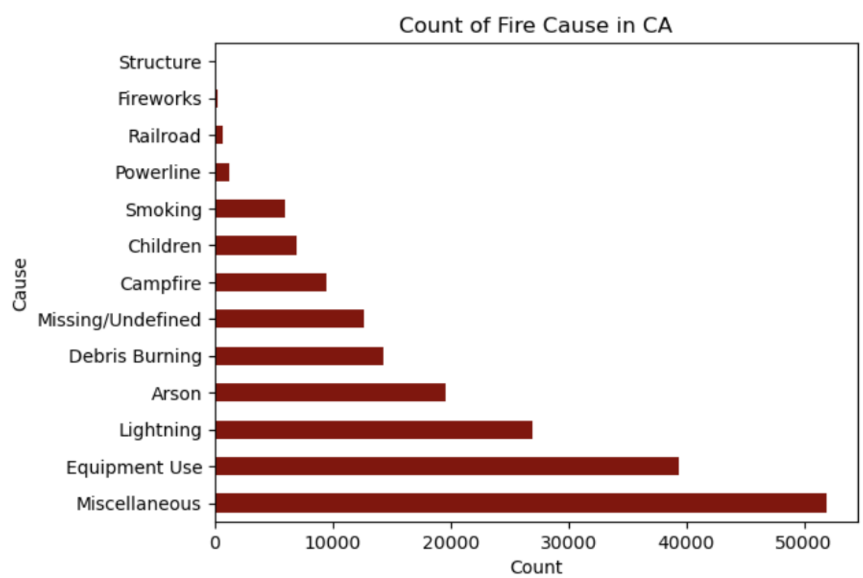


Wildfires happen a lot on the West Coast because of the hot, dry summers and mild and wet winters. The vegetation makes it easier for fires to start and spread. The strong, dry winds, such as the Santa Ana and Diablo winds, play a big role by carrying embers and making fires spread faster. Human belongings like building homes in fire-prone areas and sometimes accidents like power line issues also make wildfires more common. And now, with climate change bringing hotter temperatures and longer droughts, wildfires are even more likely and worse on the West Coast.

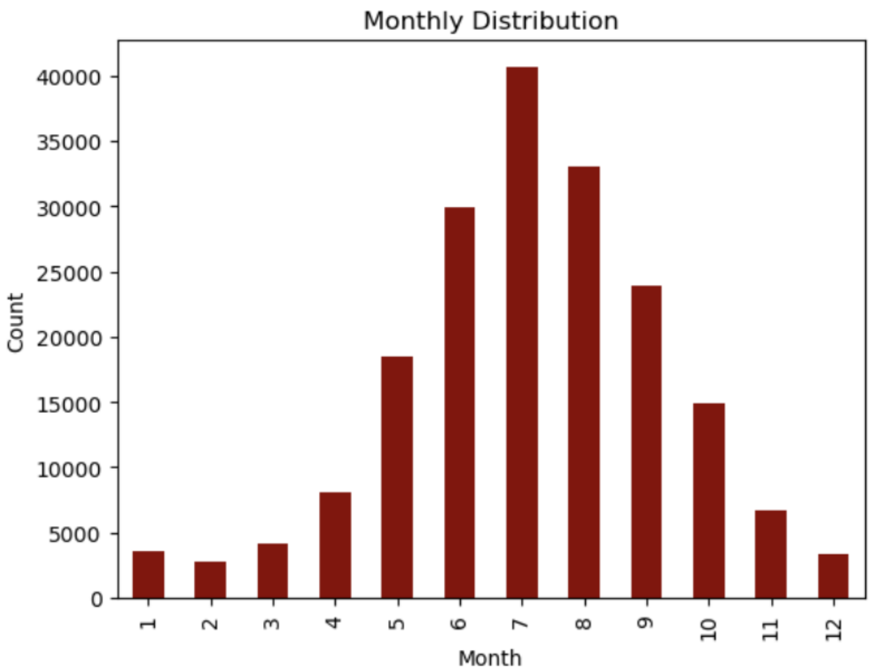
Idaho is a special case where it is supposed to have fewer wildfires due to better location and weather temperature. However, a large percentage of Idaho's forests are at dangerously high risk of severe fire because of dense and overcrowded conditions with dead and dying trees.



- Number of fires are slightly increasing over the years
- The average size of fires is increasing significantly higher each year
- Larger forest fires in California are happening more because of climate change. Hotter temperatures and longer dry spells make it easier for fires to start and spread.



- There are big numbers of fire causes that National Forest Service are unable to define in the State of California
- Equipment Use and Lightning also caused many wildfires in California.
- Arson is another alarming situation that California needs to mitigate.



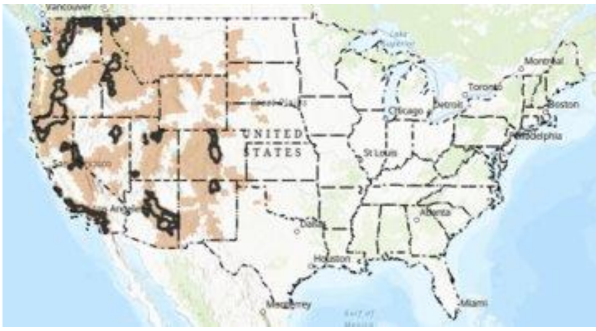
- During summer months, California experiences dry and hot weather conditions, with the lack of rainfall and high temperatures
- During summer and early fall, Southern California with Santa Ana winds and Northern California with Diablo winds, known for strong and dry winds, can rapidly spread wildfires and make it harder to stop the fires
- Also, activities such as outdoor picnics or campfires tend to increase during summer months. Negligent behaviors like unattended campfires or discarded cigarette butts can lead to fires

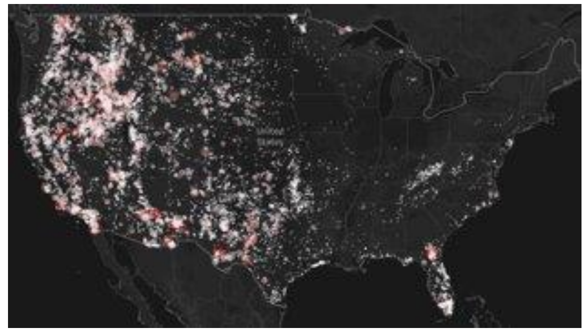
Managing wildfires becomes way more challenging during the dry season. In California, the summer and early falls months tend to be the period of the highest fire risk, which is reflected in the distribution of fires across May to September.

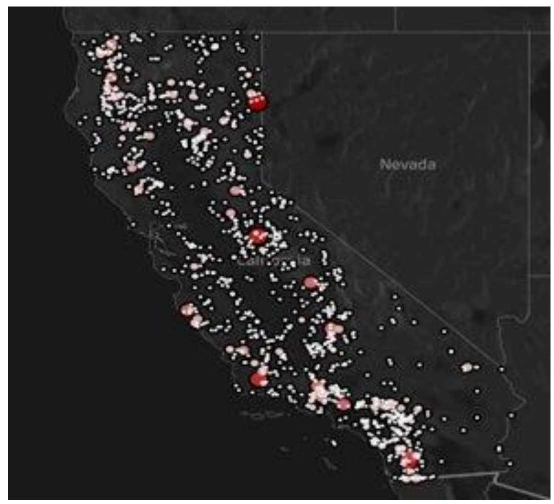
Wildfire – Improvements

Geographic Precision

Having geographic precision in the Climate Risk Viewer by adding the location of wildfires, representing their size as bubble shapes, and varying the intensity of the red color to indicate the severity of fires would significantly improve the viewer's functionality. This addition would offer users a more detailed visualization of ongoing wildfire events within Firesheds and wildfire crisis strategy landscapes. The interactive display of real-time wildfire data provides a comprehensive view of spatial distribution to make more informed decisions, allocate resources strategically, and respond effectively to the climate risks in the critical regions.

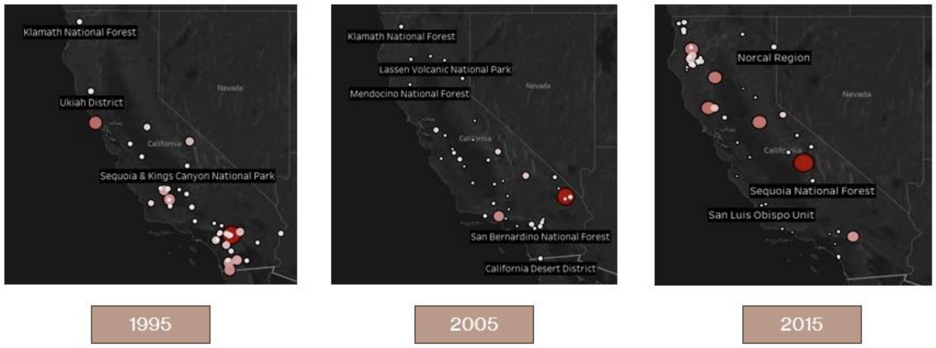






Temporal Analysis

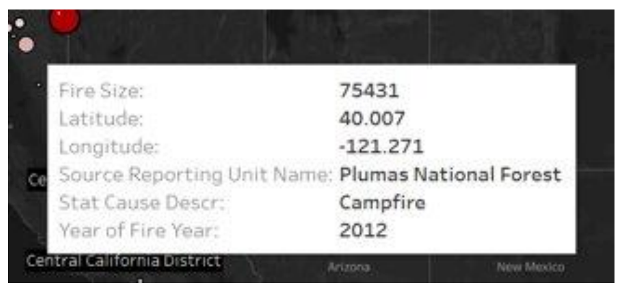
Integrating a year-based filter into the Climate Risk Viewer is very important for a better understanding of climate risks, which is wildfire risk in this case. This feature allows National Forest System managers to track how wildfires change over different years, helping them make more insightful decisions. It is like zooming in on specific timeframes, making the Risk Climate Viewer more useful and practical. It should look like this:



This upgrade means the Climate Risk Viewer becomes a more flexible and powerful tool, giving managers the insights that they need to manage climate risks in the National Forest.

Interactive information card

Including interactive information cards in the Climate Risk Viewer's dashboard is a big plus. These cards give users quick, real-time insights in an easy-to-understand way. It is like having dynamic snapshots that provide key info. With this feature, users can quickly get important data, helping them make smart decisions and better grasp the risks of changing climate.



Wildfire – Machine Learning Model

Objective

Integrating a machine learning model to predict the cause of fire into the Climate Risk Viewer serves as an early warning system and enhances precision in risk assessment by identifying specific factors contributing to fires, enabling more targeted and effective climate risk management.

I am building this model only for California for 2 reasons. First is because the data is so big and it would take so long to train and test the models; testing on California is to get the ideal, we can always apply same idea for rest of the states. Secondly, choosing California for model development offers diverse climate conditions, varied landscapes, and abundant historical wildfire data, providing a foundation for a comprehensive and adaptable predictive model.

Preparing data for modeling

- Labeling all fire causes from 1 to 13 in the dataset through the process of data labeling
- Dividing the dataset into training (75%) and testing (25%) sets to facilitate effective model training and evaluation
- I use all fire data, not only wildfire, and most of the features to build the model

```
#drop unnecessary column
del df['STATE']

# converting needed features to numeric data types
label_en = preprocessing.LabelEncoder()
df['STAT_CAUSE_DESCR'] = label_en.fit_transform(df['STAT_CAUSE_DESCR'])
```

```
#drop unneeded columns for machine learning
df = df.drop('DATE',axis=1)
df = df.dropna()
```

```
X = df.drop(['STAT_CAUSE_DESCR'], axis = 1).values
y = df['STAT_CAUSE_DESCR'].values

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25, random_state = 42)
```

Modeling

- I tested random forest algorithm first

```
# random forest algorithm

# Train the RandomForestClassifier
rf = RandomForestClassifier(n_estimators=50)
rf.fit(X_train, y_train)

# Make predictions on the test set
y_pred = rf.predict(X_test)

# Evaluate the accuracy of the model
accuracy = rf.score(X_test, y_test)
print(f"Accuracy: {accuracy}")
```

```
Accuracy: 0.4800160378154807
```

- Then, I tested Decision Tree algorithm

```python
# Use DecisionTreeClassifier
dt = DecisionTreeClassifier()
dt.fit(X_train, y_train)

# Make predictions on the test set
y_pred = dt.predict(X_test)

# Evaluate the accuracy of the model
dt_accuracy = dt.score(X_test, y_test)
print(f"decision tree accuracy: {dt_accuracy}")
```

```
decision tree accuracy: 0.40079767029627755
```

- Since both algorithms did not yield very good accuracy, I did regroup them to see whether it would give better results.

```python
def labeling(category):
    cause = 0
    act_of_god = ['Lightning']
    human_caused = ['Structure', 'Fireworks', 'Powerline', 'Railroad', 'Smoking', 'Children',
                    'Campfire', 'Equipment Use', 'Debris Burning']
    malicious = ['Arson']
    other = ['Missing/Undefined', 'Miscellaneous']

    if category in act_of_god:
        cause = 1
    elif category in human_caused:
        cause = 2
    elif category in malicious:
        cause = 3
    else:
        cause = 4
    return cause


df['LABEL'] = df_original['STAT_CAUSE_DESCR'].apply(lambda x: labeling(x))
```

- Then I ran the two algorithms, here are the results

```
random forest accuracy: 0.6092048619903773
```

```
decision tree accuracy: 0.5330885456233646
```

Clearly, the Random Forest algorithm gives better accuracy for predicting fire cause in California. I also tested the algorithm in all states, it gave a better accuracy score. So, California is a little bit more challenging to predict than the US in general.

To make the fire cause prediction model better for California, we can try some adjustments in the features used for predictions, testing out different ways the model learns, adjusting some setting to make it work better, and making sure it does not favor one type of fire cause over another. Also, check how well it does use different sets of data and keep refining the model based on what works best. Consider adding more info that might help and keep testing things to get better results over time.

Applying this machine learning model to predict the cause of fires in California helps early intervention strategies, aids in targeted resource allocation, and contributes to effective wildfire prevention and management.

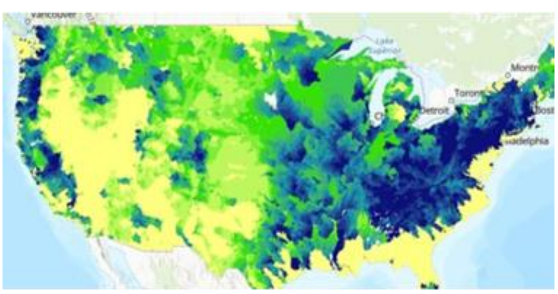
Watershed – Overview and Assessment Layers

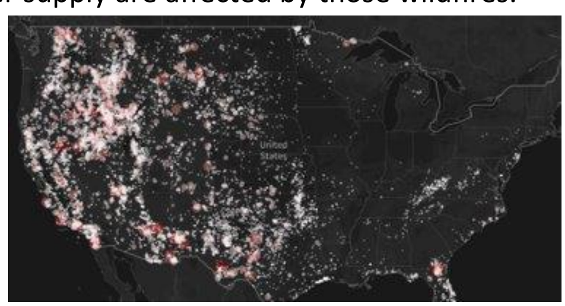
Watersheds Overview

Watersheds are one of the largest resources of water and in the United States approximately 60 million people (about twice the population of Texas) rely on forest water to meet their demands. The forests provide more than half of the United States' water supply. In the West, forests provide 65% of the water supply. Therefore, any negative impacts on the forests lands means water supply would be affected too.

Western Region Focus

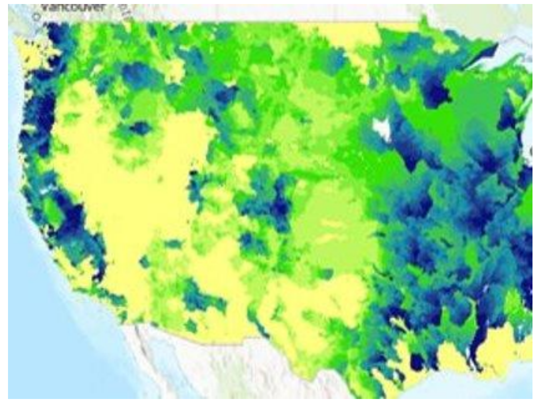
Our analysis focused on the areas where fires and watersheds relation are impacted by one another. Dry and hot weather conditions, with the lack of rainfall and high temperatures to name a few, are some of the challenges forests faces. These conditions serve as the starting point of climate crisis and disasters. The western side of the US happened to fit the conditions while the rest of the US didn't seem to be as impacted by them. A significant number of fires occurred in the western side of the US. The other issue is that a large number of watersheds that are essential for drinking water and water supply are affected by those wildfires.





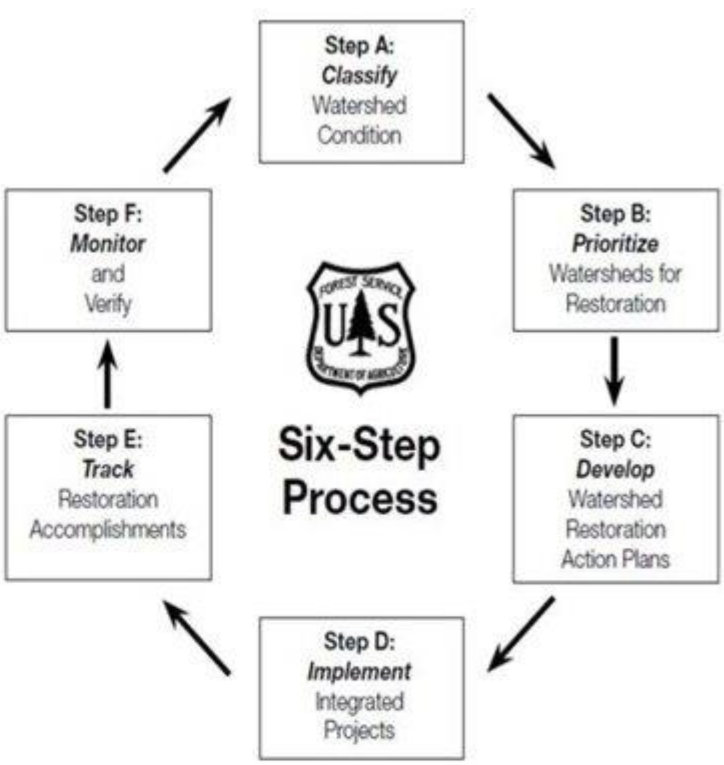
Watershed Layers

The Climate Risk Viewer offers several types of layers to choose from. The layers help in highlighting important information and present the viewer with a better understanding of where things stand currently with the watershed. The layers cover some topics such as watershed importance and watershed conditions. Using these layers, actions or plans are created to fix the specified goals of the analysis.



NFS Restoration Model

The model that the National Forest Services use consists of 6 steps to help them track their progress and success in restoring the forest. Our focus is the first 2 steps which are the assessment and prioritization efforts. The first step is assessing the level of importance of the watersheds and their importance to the overall benefit of the forest as well as the communities around them. Next step is to use that information to focus the NFS capital, resources and efforts into the most important areas.



Watershed – Improvements

Current Viewer Issues & Solution

While the CRV offers some good insights into a situation it isn't really enough to be used in the restoration model since there would still be a large margin of error in the analysis. The proposed solution is to add and create two other layers. Originally this layer was supposed to be only one, but I decided to split so that there is more freedom in controlling the factors as well as reducing the noise associated with the other factors.

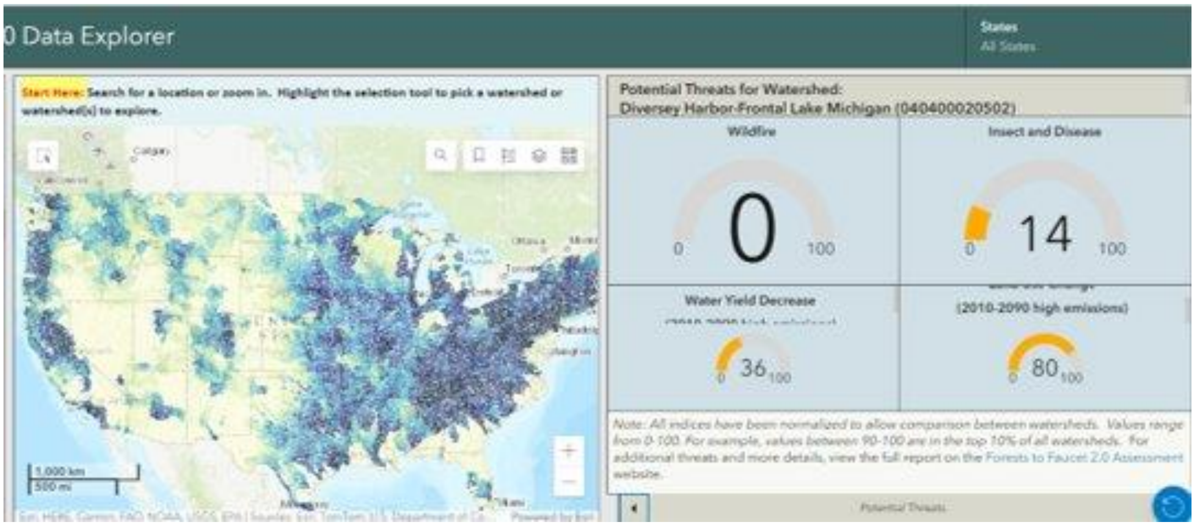
The first layers would cover the threat to surface drinking water which is calculated based on a selection of some potential threat factors referred to as *(PT), index of* importance to surface drinking water model (*IMP*), and Ability to produce clean water index model (*APCW*). The second layer would be over Wildfire threat which measures the threat level of the fires to the surrounding water sources. The benefits of these layers are that they offer an enhanced experience in identifying unstable areas and areas that are more susceptible to fail faster visual analysis.

Threats to surface drinking water (THREAT) model	$THREAT=(IMP \times APCW \times PT)/10{,}000$	*IMP* is the index of importance to surface drinking water model *APCW* is the ability to produce clean water index model *PT* is percentage of a watershed that is threatened from a contamination to surface water. *PT* is defined differently for each threat: climate change that decreases water yield (PT_{YIELD}); land use change (PT_{LUC}); insects and disease (PT_{ID}); or wildfire potential (PT_{WFP}).

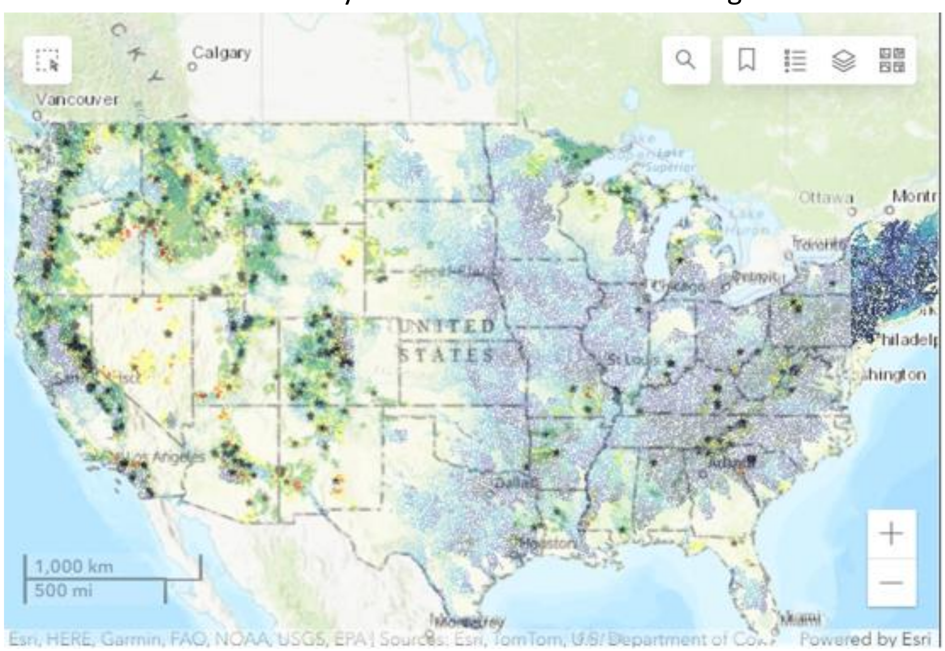
Features to Implement

Dashboard View

Creating a dashboard for the National Forest Service helps us keep an eye on things in real-time. It lets us quickly spot issues and assess risks related to watersheds, wildfires, and climate change. With this dashboard, we can respond faster simplify decision-making, and create plans to better manage resources sustainably.

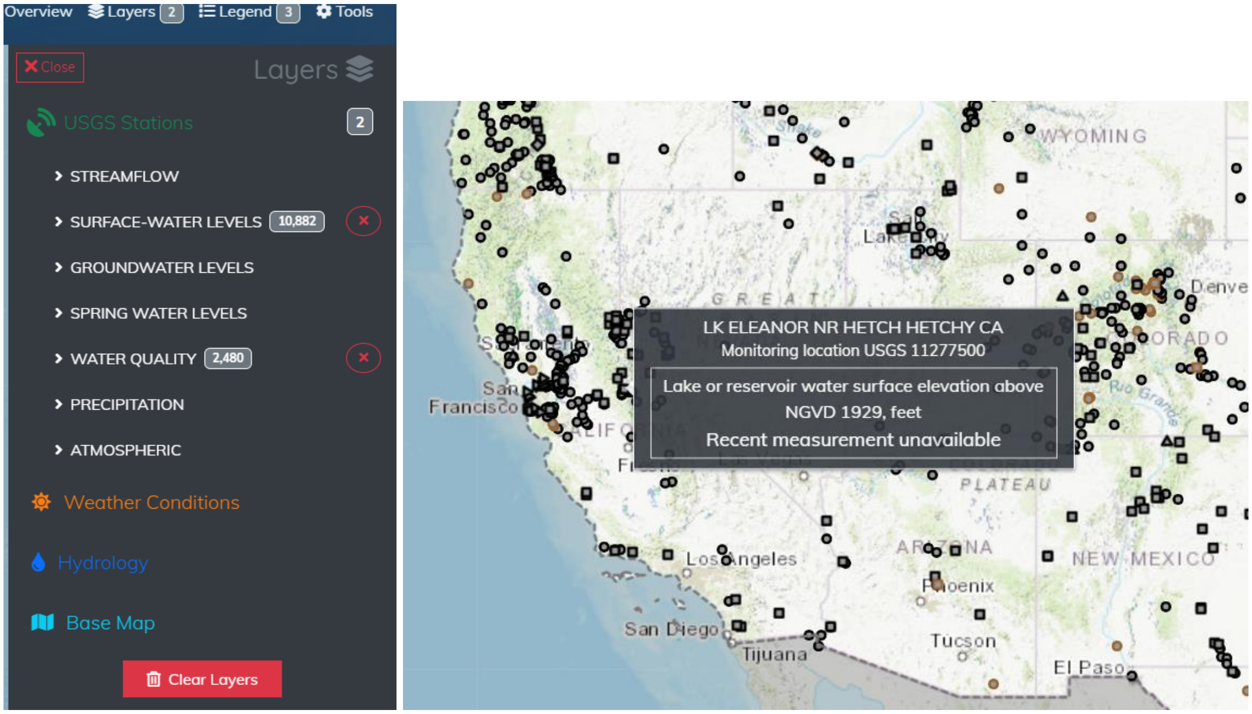


Opacity adjuster
With the Opacity Adjuster, users can layer different maps together and tweak their transparency levels. This feature enables a seamless blend of various map elements, offering a clear and customizable view. By adjusting opacity, users gain the flexibility to highlight specific details from different layers, enhancing the overall map visualization experience and reducing the time needed to switch different layers to reach a conclusive insight.



Filters and Info Cards
Filters are arguably the most important feature to be implemented. Not only do they make it easier to extract meaningful insights without being overwhelmed by irrelevant details, but they also enhance the ability to detect subtle patterns crucial for early identification of potential

issues. In addition to that an interactive information card provides additional insights when the cursor hovers over the selected area.



Conclusion

The proposed enhancements to the Climate Risk Viewer aim to enable a forward-thinking approach. These improvements address current challenges and emphasize scalability (can be added and applied to other layers), customization freedom (filtering and info cards enable a focused examination of diverse environmental factors), multi-tasking capabilities (multiple layers used in parallel without obstructing one another), and predictive possibilities (helps set improved risk mitigation plans in case of disasters happening).