

# Capstone Project Proposal

Name: Matthew Caldwell

## Objective

The goal is to create a tool (dashboard most likely) which will help Californians determine what areas are most prone to fires. This can help them to determine what areas are safest to build new developments, purchase real estate, or take preemptive action to defend their homes from fires. It will also provide a guideline which could allow state authorities to better optimize resource allocation, as well as determine what areas of wilderness may require additional fire observation posts.

## Impact

This will allow private users to determine what areas may be safest, and preferably, it could also be used by state authorities to help determine which areas require more resources to help prevent fires. On a microlevel, individuals could use it to inform themselves on how ready they need to be in the event of a fire in the area. On a macrolevel, it could allow for state fire departments to predict, or at least in some capacity, anticipate where a fire may be, and divert additional engines and air assets before the outbreak of a fire. This could massively help containment, and prevent fires growing out of control.

## Dataset(s)

[List your data sources with links to them. If you have already uploaded them to your capstone repository on GitHub, please mention the location. In addition, briefly discuss the datatypes and the reliability of the data.]

<https://www.fire.ca.gov/incidents> - This database contains a csv and geo json fill containing all fires in California in recent years. This will be the basis of the map, as it will allow us to observe temporal trends, and diagnose which areas have seen repeat fires. This is regulated and published by the state, so it is likely highly reliable.

<https://www.fire.ca.gov/what-we-do/fire-resource-assessment-program/gis-mapping-and-data-analytics> - This database contains the rest of the data needed. Including critical resources such as land type, vegetation coverage, and past fire incident and their causes. Similarly, it contains key tangential data, such as administrative boundaries and cities. This is regulated and published by the state, so it is likely highly reliable.

<https://data.ca.gov/> - This is a massive repository of open source data, including most important, utility lines and maps, which have historically caused many fires. This is state regulated, despite being open source, and is likely well vetted and reliable.

## Approach

[Talk about how you plan on approaching this capstone through several steps. List the steps below.]

1. Preprocess data - Due to the mass amounts of data, the first goal will be to process it into a state where it can work together. Things as simple as ID columns will need to be checked and verified in order to ensure the data can work between databases. The solution may be the creation of a single massive database, but more intense work with the data will need to be done to determine viability.
2. Experiment with causes - Before a final map/dashboard can be published, the causes of fires needs to be tested. Standard correlation tests should suffice, and this will hopefully allow for some guidance when it comes to weighing factors in terms of importance.
3. Compile - Once the causes and weights are established, the data will be used to calculate a layer within ArcGIS that will show what areas are more or less prone to fires. Scale will be tested here, to determine whether county, or census tract level, will be the most efficient way to visually represent the data.
4. Data verification - Due to a long history of fires, California has many causes that can be crossed referenced with the data to ensure accuracy. This will be a time when new discoveries could be noted, or where errors could be caught. If significant areas are caught, then step 2 will be repeated to find a more accurate model.
5. Visualization/UI creation - In order to allow users a somewhat personalized experience, a GIS dashboard will be created. Using this, users will be able to click around information and see greater specifics. It will also allow for more visualization to be shown outside the initial map. Secondary maps, charts, and other visualization tools will be displayed here.
6. Beta testing - The database will be sent to some other users to ensure that it remains intuitive to non-data audiences, while also being technically advanced enough to be used by more specialized consumers.
7. Publish

## Timeline

[Edit the following example timeline:

This is a rough timeline for this project:

- (3 Weeks) Data Processing
- (3 Weeks) Feature Importance/Experiment with causes
- (1 Weeks) Compile
- (3 Weeks) Data verification
- (2 Weeks) Visualization/UI creation

- (2 Week) Writing up the report/Beta testing
- (1 Week) Poster and Final Presentation

]

## Possible Issues

[List some of the prospective challenges and issues and discuss how you envision overcome them]

Due to the large amounts of data, there may be challenges trying to link different data sources to each other. Some work arounds may include using informal IDs like the names of counties, and then from there a formal numbered ID system will be implemented. GIS tools also can be difficult to work with, so there exists a number of potentially unforeseen challenges originating from the rigidity of GIS tools, and potential limitations with their tools. GWU has many licenses for advanced GIS tools, so there will likely be alternative methods to solve these problems, but it will take time to implement them.