# *Forest Fires in the United States*

**Gabrielle Beinars**
**DSC680-T301 Applied Data Science (2221-1)**
**https://github.com/gbeinars**

## Which Domain?

The publication of this data was in support of the national Fire Program Analysis (FPA) system where the records were originally collected from systems of federal, state and local level.[1]  The FPA is meant to replace older systems including National Fire Management Analysis System (NFMAS), FirePro and FireBase.  The FPA supports the wildland fire program and is considered important for the Forest Service and the Department of the Interior.[2]

## Which Data?

The dataset that will be analyzed for this project can be found at https://www.kaggle.com/rtatman/188-million-us-wildfire.[3]  This dataset includes forest fires in the US from 1992 to 2015.  There are 1.88 million records, with 130 columns, where each observation is a different fire, and total of all fire sizes span 140 million acres.  The primary features to be analyzed are the discovery date, final fire size, and location (precise within 1 square mile).  The data has already undergone basic error checks and duplicate records were removed. [3]

[1] https://www.fs.usda.gov/rds/archive/Catalog/RDS-2013-0009.5  (Data Publication)

[2] https://www.blm.gov/policy/ib-2009-040

[3] https://www.kaggle.com/rtatman/188-million-us-wildfires (Dataset)

[4] https://prepperswill.com/understanding-forest-fire-safety-preparation-and-survival/

[5] https://www.nwcg.gov/glossary/a-z (Glossary of Wildland Fire)

[6] https://williamkoehrsen.medium.com/data-wrangling-with-python-and-sqlite-900d21bc5a53

[7] https://www.fs.usda.gov/science-technology/fire/information

[8] https://www.nifc.gov/fire-information/nfn

https://www.usgs.gov/products/data-and-tools/real-time-data/wildfire

[9] https://dailydatablog.wordpress.com/2017/10/01/sqlite-tutorial-bike-share-company-data-analysis/#:~:text=SQLite%20is%20a%20Python%20package%20which%20allows%20the,Hubway%20is%20a%20bike-share%20company%20based%20in%20Boston. (Forest Fire Information)

[10] https://earthdata.nasa.gov/learn/articles/feature-articles/wildfire-articles

**Research Questions? Benefits? Why analyze these data?**

This data is stored in SQLite database and will be analyzed using exploratory data analysis, with a focus on the questions below. This data can serve a variety of purposes. For example, if the cause of fires could be predicted, then preventative measures can be put in place. This could save thousands of lives, the forests, and the wildlife. Predictions of fire size can also allow for prior action to be taken.

- Are wildfires increasing over the years?
- Do certain geographic areas have more fires than others?
- Are there any features that show correlation with fires?
- Is it possible to predict the cause of a fire based on certain features?
- What is the most common cause of forest fires?

Exploratory data analysis will involve removing empty values and any required data cleaning. Relationships between features will be explored and various visualizations will be created.

**What Method?**

During exploratory data analysis, I plan to create various visuals, narrow down the data selection to important features. After exploratory analysis and data cleaning is complete, I plan to create a model that can predict cause or a model that can predict fire location.

**Potential Issues?**

I anticipate the data cleaning may be challenging as there is a lot of data, and many containing codes that I will need to explore further. The project may fall off schedule if the data cleaning has any major obstacles due to the amount of data. SQLite is also not a frequent language for myself, but a great opportunity to improve my skills.

**Concluding Remarks**

The ability to predict what locations are more prone to fires, the most common causes of fires, or the size of a fire, can provide endless benefits. By understanding more about forest fires, preventative measures can be taken and possibly save lives, homes, forests and wildlife. Fires can be violently destructive and difficult to manage. Fires are known to destroy millions of acres of forest every year and can spread and grow quickly. Some causes of forest fires are controlled, accidental, intentional or natural. Three types of forest fires are surface, ground and crown.. Surface only impacts a small area of the forest floor and are easiest to stop. Ground fire, also referred to as underground fire, burns where there is dead and dry vegetation, spread slow, but difficult to put out because of its unpredictable pattern. Crown fires are serious, destructive and deadly.[4] This data will provide insight into forest fires with the purpose of developing a better understanding, and determine whether any factors of forest fires can be predicted, thus allowing for preventative measures to be taken.