CSC110 Project Proposal: Impact of climate change on wildfires in America

Yuzhi Tang, Zeyang Ni, Junru Lin, Jasmine Zhuang

Friday, November 6, 2020

Problem Description and Research Question

Wildfire and climate change are two hot topics nowadays. The frequent occurrence of the catastrophic wildfire, one of the most serious ecological disasters, has aroused people's concerns around the world. Meanwhile, climate change has also caused a series of ecological and environmental problems. Then we may think: does climate change have anything to do with wildfire?

Climate change has manifested itself in many ways. In recent years, there are observable changes in temperature, precipitation, soil moisture, relative humidity, wind speed, etc. at various geological levels. High temperature contributes to a better state and larger amount of the fuel. More precipitation means more plants that can become fuels in a wildfire, and less possibility for wildfire spreading. As for soil moisture, relative humidity and wind speed, these factors can also affect the occurrence and propagation of wildfires.

Since climate change influences wildfires, we would like to explore **how much climate change has impacted the frequency and intensity of wildfires in California and Texas**. Among a number of aspects of climate change, we choose the change of precipitation and temperature to explore how the fire size and the occurrence vary from 1994 to 2013. California and Texas are two states in America that are shown to have the highest frequency of wildfire occurrence. Therefore, the two states are chosen as typical research objects.

In our project, precipitation and temperature will be independent variables, while the occurrences and area of wildfires will be dependent variables. We choose the number of wildfires per year to represent the frequency of the occurrence of wildfires, which is the most intuitive reference factor for it.

The majority of the damage caused by climate change is invisible, while wildfires caused significant socioeconomic damage every year. Thus, we would like to show one of the implicit hazards of climate change in terms of wildfires. By exploring the quantitative relationship between the two, we want to demonstrate that climate change has significantly increased the severity of wildfires, thus calling for people to pay more attention to this global issue, climate change and adopt to a low-carbon lifestyle.

Dataset Description

- **Dataset 1:** Monthly Precipitation and Temperature Data for California U.S. and Texas, U.S. from Jan, 1989 to Oct, 2020

Dataset description: This dataset includes monthly precipitation and temperature data made by many weather stations in California and Texas. Some relevant attributes include the mean, mean max, and mean min temperature in a given month, and total precipitation in a given month.
Source: National Centers for Environmental Information.
Website: https://www.ncdc.noaa.gov/cdo-web/search.
Data format: csv.
Sample data:

Station: **ASHEVILLE AIRPORT, NC US GHCND:USW00003812**

Date	Temperature (F)												Precipitation (Inches)								
Elem ->	TAVG	TMAX	TMIN	HTDD	CLDD	EMXT		EMNT		DXP0	DX32	DT32	DT00	PRCP	EMXP		SNOW	EMSD		DP01	DP05	DP10
Month	Mean	Mean Max.	Mean Min	Heating Degree Days	Cooling Degree Days	Highest	High Date	Lowest	Low Date	Number of Days				Total	Greatest Observed		Snow, Sleet			Number of Days		
										Max >= 90	Max <= 32	Min <= 32	Min <= 0		Amount	Date	Total Fall	Max Depth	Max Date	>=.10	>=.50	>=1.0
Apr	57	69	44	144	9	84	28	27	10		0	2	0	2.50	1.07	30	0.0	0.0	30	7	2	1

- **Dataset 2:** Geo-referenced Wildfire Records in the U.S. from 1992 to 2015

Dataset description: This dataset includes 1.88 million geographic-referenced records of wildfires that occurred in the U.S. from 1992 to 2015. The wildfire records were acquired from the reporting systems of federal, state, and local fire organizations. Basic error-checking was performed and redundant records were identified and removed, to the degree possible.
Source: Kaggle.
Website: https://www.kaggle.com/rtatman/188-million-us-wildfires.
Data format: SQLite database (extension: .sqlite).
Sample relevant data:

STATE	FIRE_SIZE	STAT_CAUSE_CODE	FIRE_YEAR
CA	0.1	9	2005
CA	0.25	1	2004
CA	0.1	5	2004
CA	0.1	1	2004

Computational Plan

- **Data Preparation:**

Before we implement any operations on our datasets, we would need to transfer all the datasets to csv file where a new python library sqlite3 will be used to extract the data from sqlite file and store it in csv file. We would first use sqlite3.connect to create a Connection object that represents the database, then we will create a cursor and and call its execute() method to perform SQL commands.

Then, all relevant attributes will be extracted from the datasets. For Dataset 1, these include: the state where the weather station is in, the date when the monthly temperature/precipitation data was recorded, the average temperature of the month, and the total precipitation in the month. For Dataset 2, these include: the state where the wildfire occurred, the size of fire, the statistical cause of the fire, and the year when the fire occurred. The raw data extracted will be transformed into reasonable Python data types (e.g. string, float, datetime) to allow for easy computation.

For the extracted data of Dataset 1, the monthly temperature and precipitation data will be transformed into a yearly format. Average temperature in a year of a state will be calculated by taking the average of all monthly temperature data recorded in that state, in that year. Average precipitation in a year of a state will be calculated by taking the average of all monthly precipitation data recorded in that state, in that year. Then, we would have the yearly temperature and precipitation data for California and Texas, stored using a custom Python dataclass.

For the extracted data of Dataset 2, all wildfires that occurred outside Texas and California, or ones that are not caused by natural means (e.g. man-made wildfires) will be filtered out. The wildfire data will also be transformed into a yearly format. Average wildfire size in a year of a state will be calculated by taking the average of the size of all wildfires in that state, in that year. Total wildfire occurrence in a year of a state will be calculated by summing up all wildfire occurrences in that state, in that year. Then, we would have the yearly wildfire size and occurrence data for California and Texas, stored using a custom Python dataclass.

- **Computational Models/Algorithms:**

To evaluate the impact of climate change on the frequency and intensity of wildfires, we would conduct several linear regressions for both California and Texas.

  - Linear regression between year(independent variable) and average year temperature(response variable)
  - Linear regression between year(independent variable) and average year precipitation(response variable)

By doing these two regressions, we would be able to predict the temperature change in the following decades and precipitation changes in the following decades.

- Linear regression between average year temperature(independent variable) and wildfire frequency(response variable)
- Linear regression between average year temperature(independent variable) and wildfire size(response variable)
- Linear regression between average year precipitation(independent variable) and wildfire frequency(response variable)
- Linear regression between average year precipitation(independent variable) and wildfire size(response variable)

By doing these four regressions, we would find the relationship between climate and wildfire, since we'd already predicated the future climate data, we could then use this model to predict the future wildfire size and frequency.

We would take the following two procedures to conduct linear regression.

I. Measure the strength of linear association:

$$R^2 = \frac{\Sigma_{i=1}^n (y_i - \bar{y}^2) - \Sigma_{i=1}^n (y_i - \bar{y})^2}{\Sigma_{i=1}^n (y_i - \bar{y})^2}$$

The coefficient of determination R^2 is the proportion of the variability in y which is explained by our fitted regression line.

- *R^2 close to 1 indicates that most of the variability in y is explained by the regression model*
- *R^2 close to 0 indicates that very little of the variability in y is explained by the regression model.*

II. Calculate the fitted linear equation:

$$\beta_0 = \frac{\Sigma_{i=1}^n y(\Sigma_{i=1}^n x^2) - (\Sigma_{i=1}^n x)(\Sigma_{i=1}^n xy)}{n(\Sigma_{i=1}^n x^2) - (\Sigma_{i=1}^n x)^2}$$
$$\beta_1 = \frac{n(\Sigma_{i=1}^n xy) - (\Sigma_{i=1}^n x)(\Sigma_{i=1}^n y)}{n(\Sigma_{i=1}^n x^2) - (\Sigma_{i=1}^n x)^2}$$

β_0*: intercept parameter*
β_1*: slope parameter*

The estimated equation should be:

$$y = \beta_0 + \beta_1 x$$

- **Results Reporting:**

We will use plotly to display the change of frequency of fire occurrence, fire size, temperature and precipitation to see if there exists any trend of them.
plot_a: how the frequency of fire occurrence, temperature and precipitation change with time.
plot_b: how fire size, temperature and precipitation change with time.
plot_c: the relation between the frequency of fire occurrence and temperature.
plot_d: the relation between the frequency of fire occurrence and precipitation.
plot_e: the relation between fire size and temperature.
plot_f: the relation between fire size and precipitation.

References

National Oceanic and Atmospheric Administration. (2020, November). *National centers for environmental information.* https://www.ncdc.noaa.gov/cdo-web/search

Plotly. (2020, November).*Plotly python open source graphing library fundamentals.* https://plotly.com/python/plotly-fundamentals/

Sqlite3. (2020, November). *DB-API 2.0 interface for SQLite databases.* https://docs.python.org/3/library/sqlite3.html#module-sqlite3

Tatman, R. (2020, November). *1.88 million US wildfires: 24 years of geo-referenced wildfire records.* https://www.kaggle.com/rtat
million-us-wildfires