Project Wildfire

CSPB 4502 – Group 3

Team

- Natalie Dreher
- Ronald Durham
- Grant Fairbairn

Questions Sought to Answer

- Have wildfires become more or less frequent over time?
  - There was no significant correlation between the number of fires and years between 1992 and 2015.
- In this dataset, what is the most common cause of fires?
  - The "Debris Burning" category was the most common cause in this dataset (by a long shot). But that category is quite vague.
- Which cause of fire most commonly leads to the largest fire size?
  - From 1992 to 2015, the data shows that lightning strikes were the identified cause of the largest wildfires.

Data Preparation Work

- We utilized the data set "1.88 Million US Wildfires" for wildfires occurring in the United States between 1992 and 2015.
    - Short, Karen C. 2017. Spatial wildfire occurrence data for the United States, 1992-2015 [FPAFOD20170508]. 4th Edition. Fort Collins, CO: Forest Service Research Data Archive. https://doi.org/10.2737/RDS-2013-0009.4.
- The dataset was massive, with 1.88 million rows of tabulated data, requiring that it be broken into four CSV segments before importing.
- We needed to do some data cleaning as some of the rows were incomplete. We utilized the attributes that were useful in answering our questions and removed excess or ones not applicable to this project.

Tools Used

CODING:

- Jupyter Notebook
- Python

DATA PROCESSING AND REGRESSION:

- Pandas
- Numpy
- Sklearn
- Statsmodels

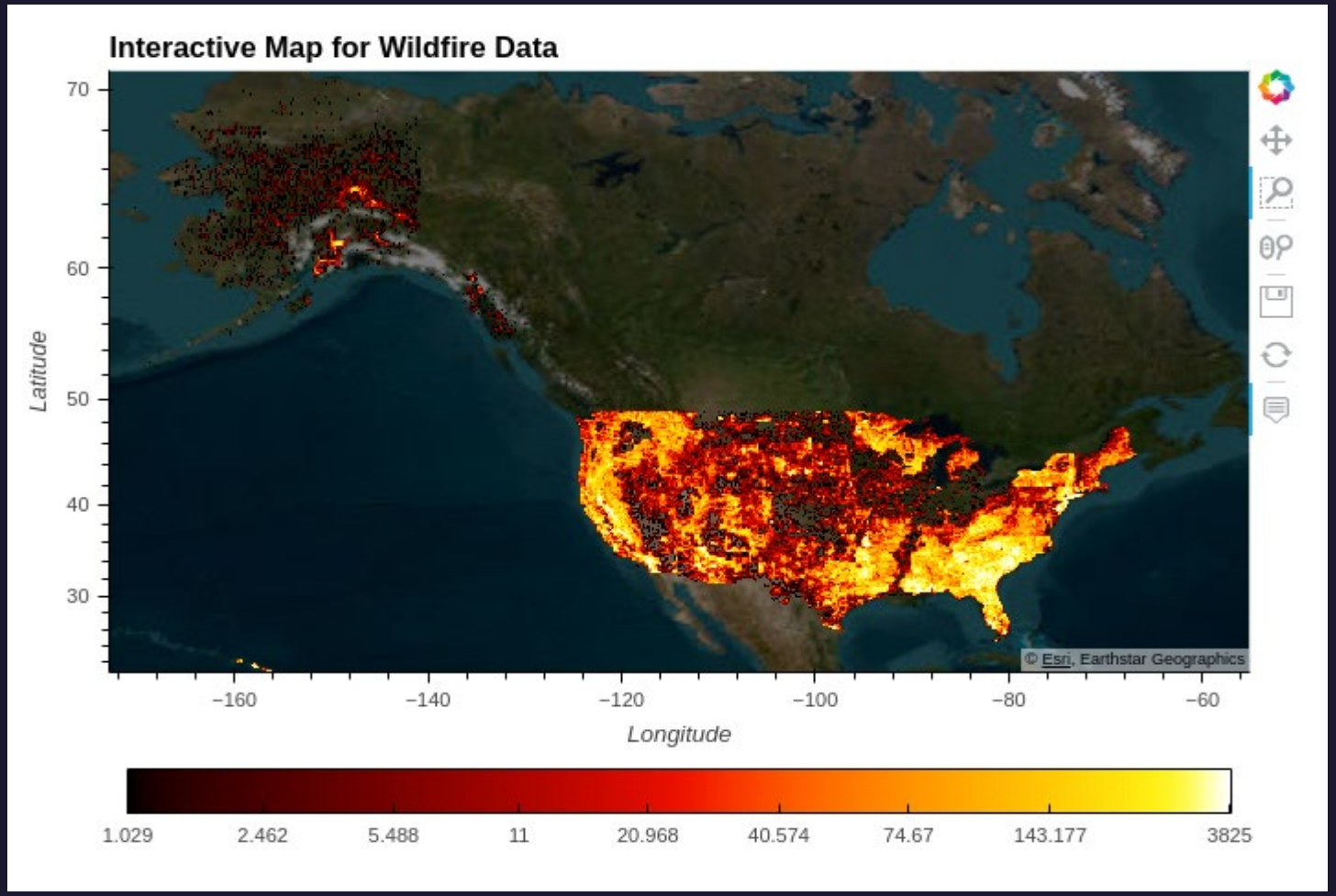
VISUALIZATION TOOLS:

- Matplotlib
- Pyplot
- Seaborn
- Plotly
- Colorcet
- Holoviews

Techniques Applied

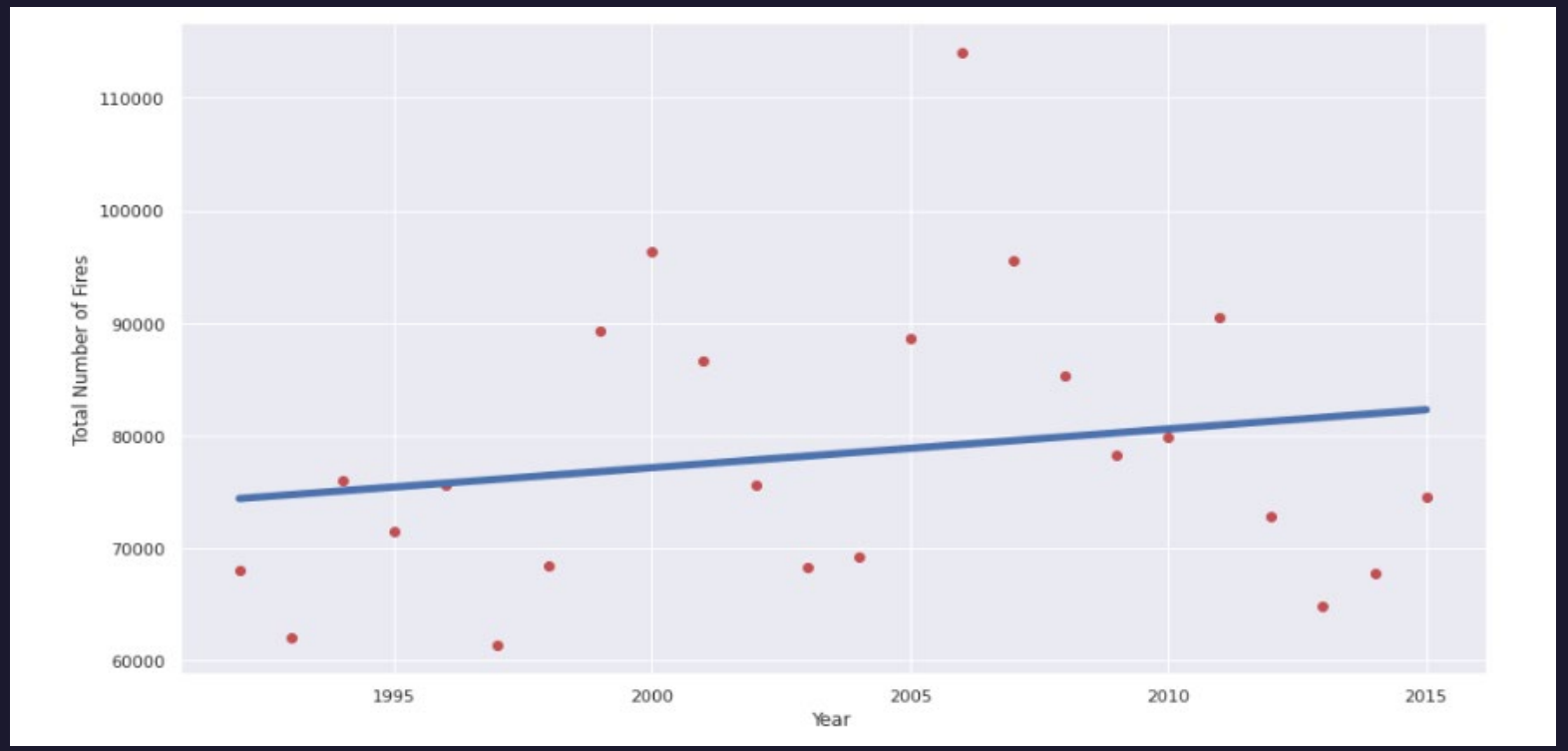
- We utilized visualization tools (Matplotlib and Seaborn) to analyze the data and look for interesting patterns to analyze, and we created a "heatmap" showing the prevalence of fires from the dataset.
- Once we had an idea of what we were looking for, we set out to look for significance in correlations.
- For example, we used the linear regression tools from the Python statsmodel library to analyze whether wildfires were becoming more or less common over time.

Heatmap Visualization



Knowledge Gained

- We did not find a significant trend in the growth or reduction of fires over time. The R-squared value of "Fires" against "Years" was low at 0.036:



OLS Regression Results

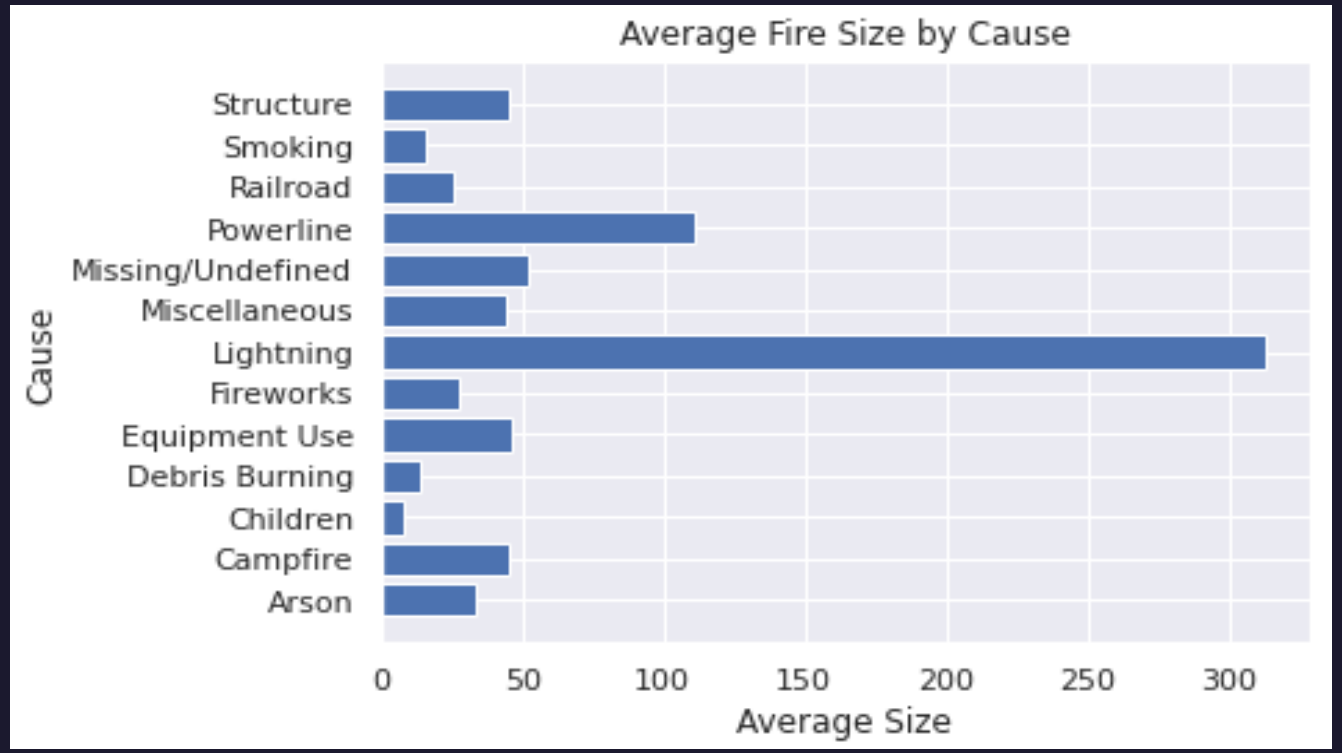
Dep. Variable:	Fires	**R-squared:**	0.036
Model:	OLS	**Adj. R-squared:**	-0.008
Method:	Least Squares	**F-statistic:**	0.8265
Date:	Tue, 29 Nov 2022	**Prob (F-statistic):**	0.373
Time:	02:55:59	**Log-Likelihood:**	-260.00
No. Observations:	24	**AIC:**	524.0
Df Residuals:	22	**BIC:**	526.4
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-6.095e+05	7.57e+05	-0.806	0.429	-2.18e+06	9.6e+05
Year	343.3443	377.678	0.909	0.373	-439.911	1126.600

Omnibus:	5.718	**Durbin-Watson:**	0.925
Prob(Omnibus):	0.057	**Jarque-Bera (JB):**	3.859
Skew:	0.938	**Prob(JB):**	0.145
Kurtosis:	3.585	**Cond. No.**	5.80e+05

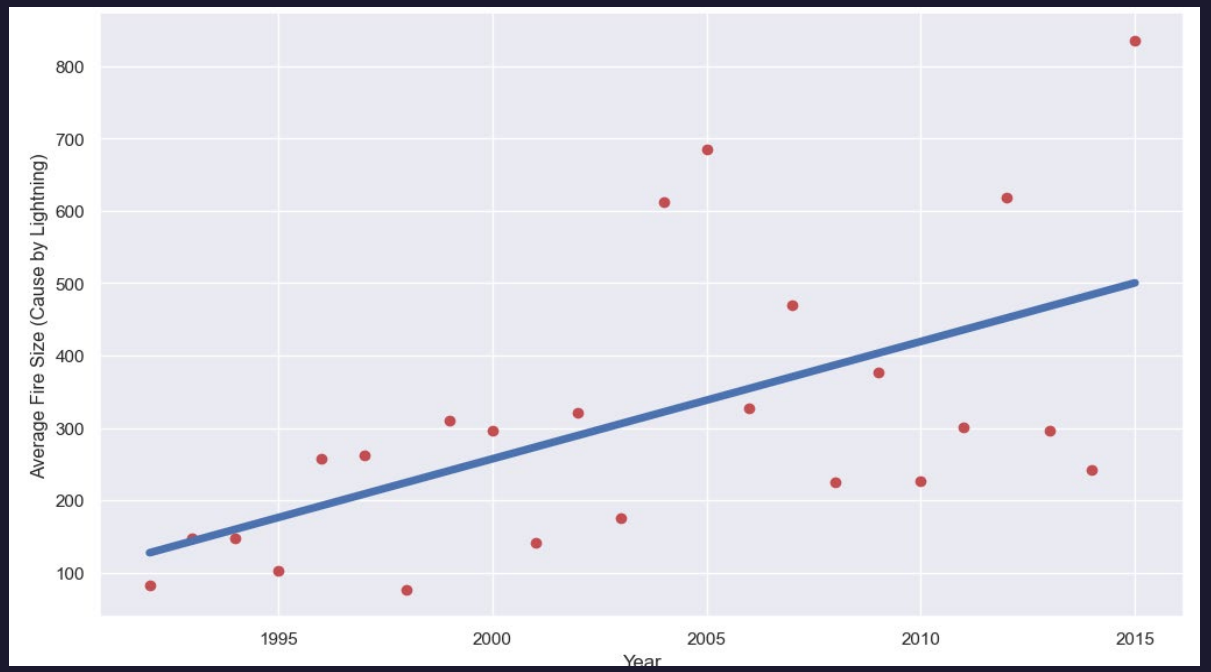
Knowledge Gained

- But we did find an interesting result in comparing fire size by cause. Lightning strikes accounted for the largest average fire (by a wide margin):



Knowledge Gained

- That led us to analyze the average fire size caused by lightning over time, revealing a positive correlation:



OLS Regression Results

Dep. Variable:	Size	R-squared:	0.334
Model:	OLS	Adj. R-squared:	0.303
Method:	Least Squares	F-statistic:	11.02
Date:	Tue, 06 Dec 2022	Prob (F-statistic):	0.00312
Time:	13:18:52	Log-Likelihood:	-155.65
No. Observations:	24	AIC:	315.3
Df Residuals:	22	BIC:	317.6
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-3.217e+04	9786.229	-3.287	0.003	-5.25e+04	-1.19e+04
Year	16.2129	4.885	3.319	0.003	6.083	26.343

Omnibus:	2.779	Durbin-Watson:	1.975
Prob(Omnibus):	0.249	Jarque-Bera (JB):	2.090
Skew:	0.717	Prob(JB):	0.352
Kurtosis:	2.821	Cond. No.	5.80e+05

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 5.8e+05. This might indicate that there are strong multicollinearity or other numerical problems.

Applications

- The largest area of application from studying wildfire data is the potential for learning about prevention and containment.
- Wildfires are a necessary part of the global ecosystem, and lightning strikes are natural causes.
- While lightning strikes cannot be prevented, early detection of strikes in forested areas prone to the largest fires (e.g., the western portion of the United States) could reduce the burn area.

Thank You

https://github.com/rodu4835/CSPB4502_ProjectWildfire