Elliott Sanger
DATA 512
Final Writeup: https://github.com/Elliott-1/Elliott-512-Final

Introduction:

Wildfires are a significant environmental and societal challenge, necessitating a nuanced understanding of their impacts and management. While wildfires are a natural component of many ecosystems, playing a role in forest regeneration and ecological balance, their increasing frequency and severity pose grave threats to public health, property, and wildlife. Attempting to eliminate wildfires entirely can lead to unintended consequences, such as the accumulation of burnable biomass, which can fuel uncontrollable, catastrophic fires. This paradox highlights the importance of developing strategies to mitigate wildfire risks while allowing natural processes to occur in a controlled manner.

This project focuses on analyzing wildfire trends and their associated costs, by looking at their impact on one city in particular: **Sioux Falls, South Dakota**. By examining the historical trends of wildfire acreage and smoke over the past six decades, we aim to understand the growing challenges posed by wildfires and their implications for local communities. Preliminary findings suggest that the total acreage burned within a 650-mile radius of Sioux Falls has likely increased from 1961 to 2020. Additionally, a smoke metric constructed for this study reveals a similar upward trend in smoke levels over the same period, indicating heightened risks to public health and quality of life. Notably, certain years emerge as outliers, with significantly higher levels of wildfire activity, pointing to periods where existing containment measures failed or proved inadequate.

The second part of this analysis examines the financial ramifications of wildfires, particularly the impact of increased smoke and fire activity on government spending for wildfire suppression. Public concern over taxation and government expenditure makes this aspect particularly relevant to stakeholders. By correlating the smoke metric with annual wildfire suppression costs, this study reveals that higher smoke levels often precede increased spending. Such findings underscore the importance of proactive wildfire management to stabilize budgets and prevent unexpected financial burdens associated with severe fire years.

This project's broader goal is to offer actionable insights to policymakers and stakeholders. Demonstrating that wildfire acreage and smoke levels have historically increased - and that these factors drive up suppression costs - provides a compelling case for prioritizing wildfire management reforms. Preventative measures such as prescribed burns, forest thinning, and debris removal could help mitigate the risks of uncontrollable wildfires[1] while offering a more predictable framework for budgeting and resource allocation. A preventative strategy not only curbs the environmental and societal impacts of wildfires but also optimizes fiscal management.

South Dakota's current approach, as evidenced in the state's 2019 budget report, acknowledges the unpredictability of wildfire-related expenditures. They state, "due to the unpredictable nature of these items, the State does not budget estimated expenditures for fighting wildfires"[2] This reactive stance may contribute to inefficiencies and heightened financial volatility. Based on the trends observed in Sioux Falls and surrounding areas, it appears that a more strategic, preventative policy could yield significant benefits. By reducing the fuel load in forests and implementing controlled burn programs, the state can transition to a more sustainable and cost-effective wildfire management system. In doing so, South Dakota could not only protect its citizens and natural resources but also lay the groundwork for a stable and efficient budgetary process.

Background/Related work:

The study of wildfires, their environmental impacts, and associated costs has garnered significant attention in recent years due to their increasing frequency and severity. For this project, existing research and datasets provided a foundational basis for understanding wildfire trends, smoke production, and financial implications. The primary datasets for this analysis were sourced from reputable government agencies, ensuring reliability and legal compliance for research purposes. The wildfire data used to identify fires within 650 miles of Sioux Falls, South Dakota, was obtained from the US Geological Survey (USGS)[Data Source 1], while the US Environmental Protection Agency (EPA) Air Quality System API provided additional data to validate these wildfire findings[Data Source 2]. I also used CPI data[Data Source 3] in order to adjust the budgets for inflation over the years.

Expanding on this analysis required investigating financial data related to wildfire suppression efforts in South Dakota. However, South Dakota's small population and largely rural landscape pose challenges in accessing granular financial data at the city or county level. Fortunately, Sioux Falls, as the state's largest city by population[3], serves as a reasonable proxy for estimating the tax burden associated with wildfire management. Approximately 22% of the state's sales tax revenue originates from Sioux Falls, the largest share by any single city[4]. This proportion supports the assumption that a significant portion of wildfire suppression costs would be borne by this urban center. For historical financial trends, state budget records provided the most comprehensive data available. These budgets, publicly accessible and documented in the data sources section, were analyzed for wildfire suppression expenses and related metrics[Data Source 4].

The state budget includes two crucial data points utilized in this analysis. First, the general fund allocation for wildfire suppression and related expenses was examined. These allocations, while consistently present, are subject to template changes across years, requiring careful navigation through the documents. Second, the budget's statistical section provides data on wildfire acreage suppressed, which has been recorded for the past decade and intermittently prior. The records analyzed include the 2011, 2017, and 2023 budgets, with detailed information only available from 2001 onward and acreage statistics tracked since 2002.

A critical aspect of this project is comparing the wildfire acreage data with the estimated smoke metric developed during the analysis. While smoke and acreage are not perfectly correlated -

smokier fires may indicate greater severity and require additional suppression resources - the comparison provides valuable insights. This comparison and the methodology for constructing the smoke metric will be discussed in detail in the methodology section.

Building on these datasets, this project investigates the correlation between wildfire activity (measured through acreage and smoke) and the financial costs associated with wildfire suppression. The overarching research question asks: *To what extent is the tax money spent on wildfires correlated with wildfire acreage and smoke, and what is the forecasted estimate for wildfire suppression costs over the next 20 years (2021–2046)?*

While this analysis does not involve an independent audit of the provided data, the sources - compiled by state officials and audited by internal and external groups - are presumed accurate for the purposes of this research. This assumption allows for the development of models and forecasts based on reliable historical trends. Finally, since all data was sourced directly from government agencies, it is freely available for research use under public domain provisions. This accessibility ensures that the methods and findings of this project remain transparent and reproducible, further supporting their validity and utility in informing future wildfire management strategies.

<u>Methodology:</u>

The methodology of this project was designed to investigate the relationship between wildfire activity, smoke production, and the financial cost of wildfire suppression, with a focus on interpretability and simplicity. The approach began with the wildfire dataset, filtered to include only wildfires within a 650-mile radius of Sioux Falls, South Dakota. This range was selected to capture fires likely to impact the city’s air quality while excluding fires too distant to be relevant. To focus on the peak wildfire season, only fires occurring between May 1st and October 31st were considered.

Development of the Smoke Metric

To estimate the impact of wildfire smoke on Sioux Falls, a custom smoke metric was developed. The metric was defined as:

$$Smoke\ Metric\ =\ \frac{Area\ of\ Fire\ (in\ square\ miles)}{(Distance\ from\ Sioux\ Falls\ (in\ miles))^2}\times 5000$$

This formula reflects two key assumptions: larger fires produce more smoke, and smoke dissipates with distance from the fire source. The distance was squared to emphasize its substantial impact on smoke dispersion, and a scaling factor of 5000 was applied to ensure the resulting values were within a magnitude comparable to normal EPA Air Quality Index (AQI) levels, avoiding small decimals. While the units of the metric are arbitrary, the goal was to convey that larger, closer fires contribute more significantly to smoke levels in Sioux Falls.

To validate the smoke metric, it was compared with EPA AQI data for the fire season. Annual AQI values were calculated for October 31st or the closest available date, and the smoke metric was summed for fires occurring between May 1st and October 31st. Correlations between the computed smoke metric and the AQI showed that the metric reasonably approximated actual smoke levels and reflected the size and severity of nearby wildfires.

Predictive Modeling

Using historical smoke metric data from 1961 to 2020, a linear regression model was trained to predict smoke levels for the next 25 years (2021-2046). This predictive model was then combined with financial and acreage data from South Dakota's annual financial reports. These reports provided two key metrics: wildfire suppression budgets and wildfire acreage. Since the state reports operate on a fiscal year (July 1-June 30), a ±1 year shift was applied to align the data with the calendar-year-based smoke metric. The shift with the highest correlation between smoke metric and wildfire acreage was selected to synchronize the datasets.

Two separate linear regression models were developed for forecasting. The first model used historical smoke metrics to predict future wildfire suppression budgets. The second model relied on wildfire acreage data, fitting a regression to historical and predicted values to forecast associated costs. Both models were smoothed using Locally Weighted Scatterplot Smoothing (LOWESS)[5], a non-parametric algorithm that captures underlying trends without imposing overly simplistic linear assumptions. This was necessary as raw linear regression often predicted implausible values, such as budgets decreasing to zero or negative amounts. LOWESS provided a more realistic forecast by accommodating non-linear trends.

Ethical and Practical Considerations

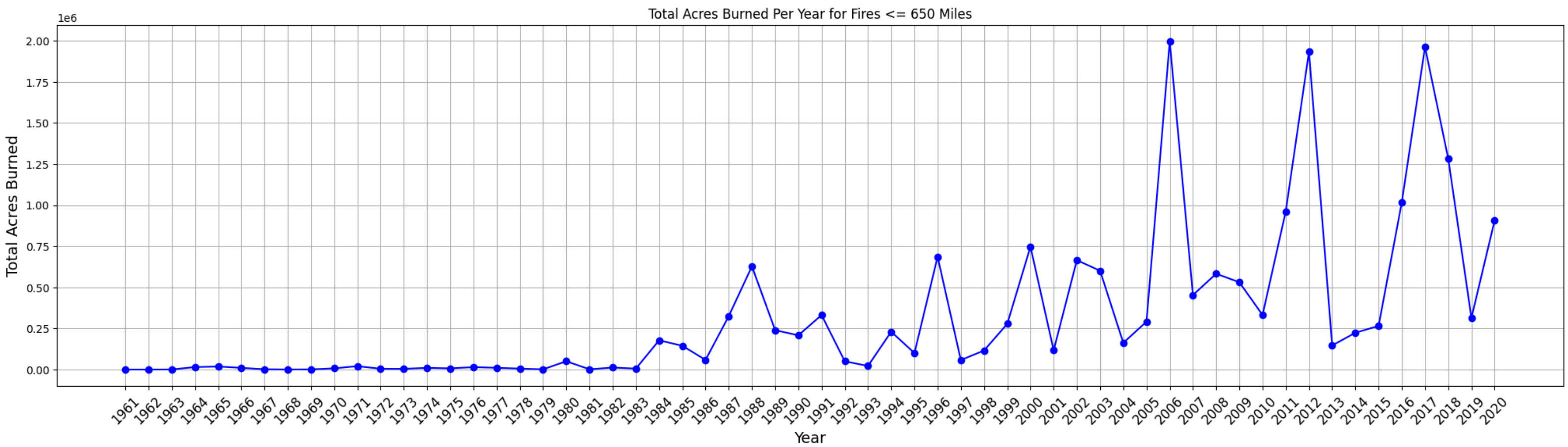
The analysis was intentionally designed with simplicity and interpretability as guiding principles. Given limited familiarity with wildfire dynamics and the complexities of financial data, overly complex models risked being both ineffective and opaque. By employing straightforward metrics, such as the smoke formula, and widely understood analytical tools, such as linear regression, the results were kept accessible to non-expert audiences, including policymakers and stakeholders.

This approach prioritizes transparency and encourages informed decision-making. The smoke metric provides a clear visual representation of how wildfire proximity and size impact Sioux Falls, while the regression models highlight the financial consequences of wildfire trends. By focusing on the need for proactive wildfire management, the methodology underscores the importance of reducing wildfires and their associated financial burdens through preventative measures such as prescribed burns, forest management, and early intervention strategies.

In summary, the methodology combined simplicity, transparency, and accessibility to produce actionable insights while adhering to ethical research practices. By using publicly available, government-sourced data, the analysis ensured compliance with legal standards and reproducibility, further strengthening its value as a tool for addressing wildfire management challenges.

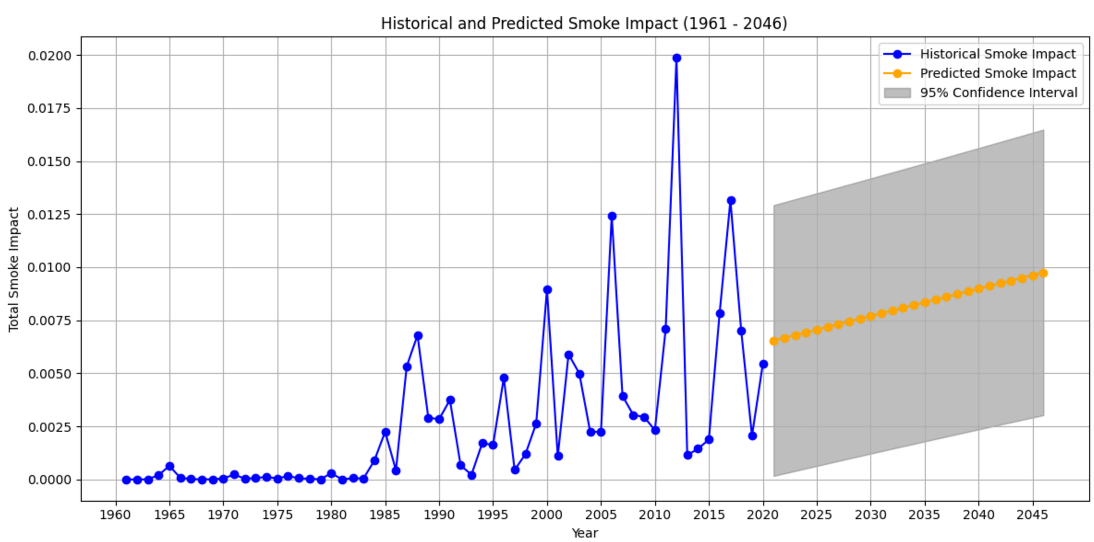
Findings:

The total wildfire acreage burned within 650 miles of Sioux Falls has shown a clear upward trend over time.



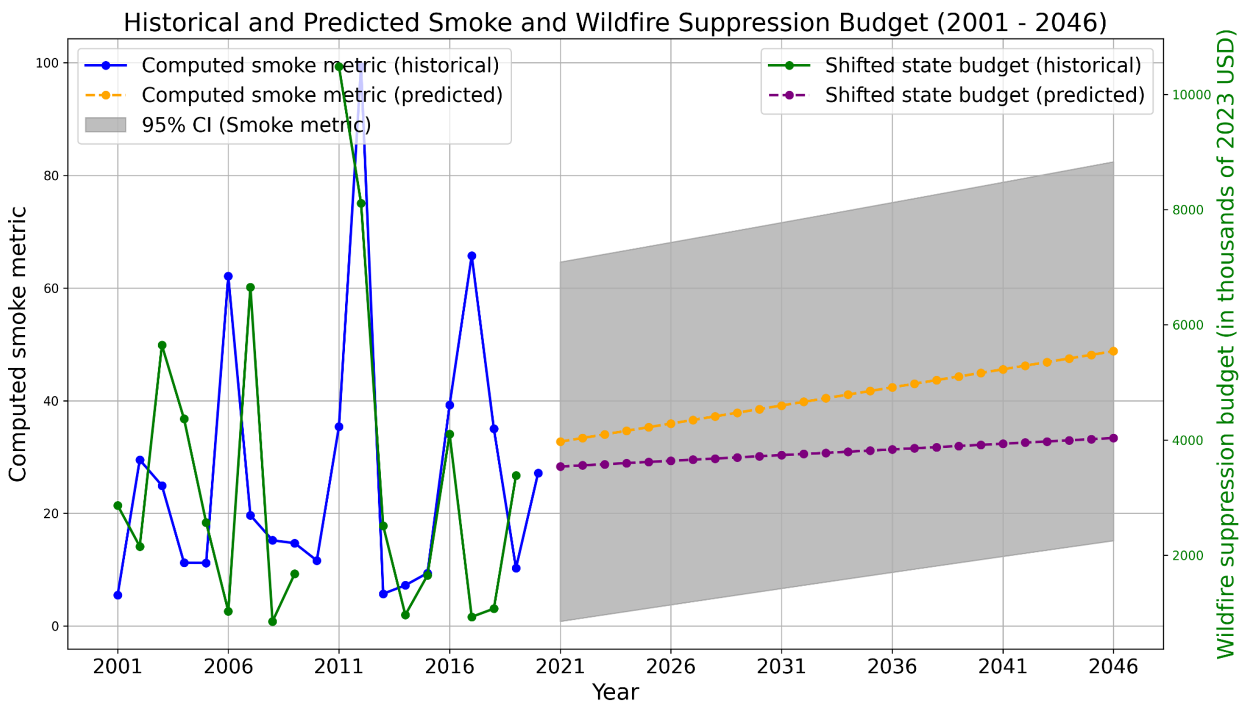
Since 2005, there have been several outlier years with exceptionally high burned acreage, contributing significantly to the overall increase. These years likely represent periods of intense wildfire activity, which may result from a combination of environmental factors and resource limitations in wildfire suppression.

Similarly, the computed smoke metric, which incorporates wildfire acreage and the proximity of fires to Sioux Falls, also indicates a rising trend in smoke levels.



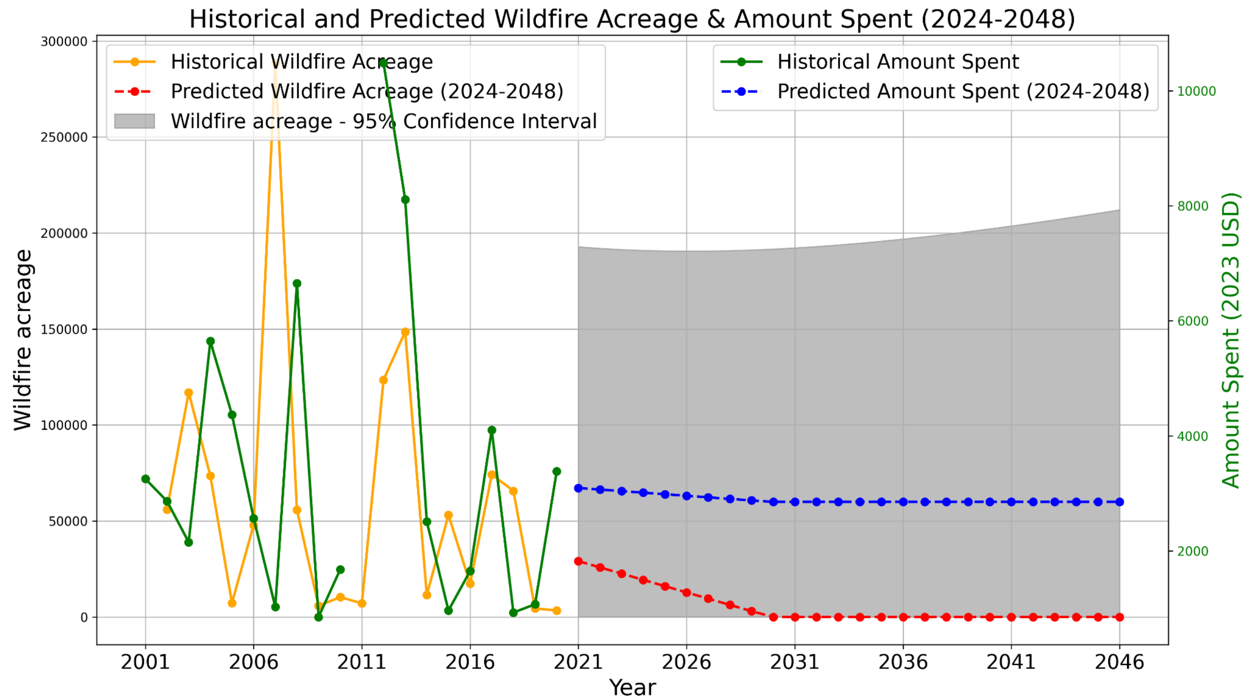
When projected into the future using a linear regression model, the results suggest that smoke levels will continue to increase. The linear model yielded a coefficient of 0.64 per year, suggesting a gradual rise in smoke levels. However, the 95% confidence interval for this prediction was broad, indicating high variability and limited statistical significance.

When examining the relationship between the computed smoke metric and the annual wildfire suppression budget, the findings showed that spending has remained mostly flat over time, with only a slight increase.



The linear model fitted to this data yielded a y-intercept of 2532.49 (thousands of dollars) and a slope of 30.78 per year, suggesting a base budget of approximately $2.5 million and an annual increase of about $30,000. For example, at around the year 2021, this base budget would equate to around $3.5 million.

Interestingly, the historical data from the South Dakota financial reports presents a contrasting narrative. Since 2001, wildfire acreage has reportedly decreased, and the annual budget for wildfire suppression has adjusted accordingly.



When these trends are smoothed and projected into the future using a linear model, the predictions suggest a significant reduction in wildfire acreage, approaching near zero by 2046. However, the wildfire suppression budget is predicted to remain steady at approximately $3 million annually.

Overall, these findings highlight a complex and evolving relationship between wildfire activity, smoke levels, and financial suppression costs. While the computed smoke metric and historical wildfire acreage trends suggest increases in wildfire impact, the state's financial reports indicate a more optimistic view of decreasing wildfire acreage. These conflicting perspectives underscore the importance of refining predictive models and aligning financial resources with actual wildfire trends to improve long-term planning and mitigation efforts.

Discussions/Implications:

The findings from this study provide critical insights into the historical and projected trends in wildfires, smoke impact, and the financial burden of wildfire suppression. They highlight a potential shift in the state's approach to wildfire management. In the early 2000s, there appeared to be a reactive spending pattern, where years with high smoke levels and acreage were met with spikes in wildfire suppression budgets. This reactive strategy likely led to inconsistent wildfire management, with periods of low activity followed by years of devastatingly high wildfire impact.

In recent years, however, the observed decorrelation between smoke levels and suppression budgets, along with a reduction in both acreage and smoke, suggests a transition toward a more preventative approach. Evidence for this shift can be seen in the 2023 budget, which includes funding for infrastructure projects like the “construction of a maintenance shop for the Wildland Fire Suppression Division”[6]. This investment reflects a forward-thinking approach, prioritizing wildfire preparedness and suppression capabilities over reactive measures.

If discussing these findings with city leadership, such as the city council, city manager, mayor, or residents, the focus should be on the progress made in wildfire suppression strategies. Acknowledging these successes is crucial to sustaining momentum and community support. However, it’s equally important to highlight the historical pattern of cyclical wildfire activity - periods of low wildfire acreage often precede years of extreme fire activity due to biofuel accumulation. This underscores the need for continued and consistent investment in preventative measures.

To maintain control over wildfires and their associated costs, policymakers should focus on proactive wildfire management. Strategies might include controlled burns, biofuel removal, investments in firefighting infrastructure, and public education campaigns to reduce human-caused wildfires. By emphasizing the connection between effective prevention and reduced financial burden, it’s possible to frame these efforts as both an environmental and economic imperative.

The urgency to act should be stressed. With the historical pattern of fluctuating wildfire activity, the window for implementing effective preventative measures may be short. A proactive, sustained wildfire management plan should be put in place as soon as possible to prevent a recurrence of devastating wildfire years. Waiting too long could result in another uncontrolled wildfire season with severe financial and environmental consequences.

Finally, human-centered data science principles informed the decision-making throughout this project. Simplicity and transparency were prioritized to ensure that findings were accessible and actionable for non-expert audiences. This approach aligns with ethical considerations by promoting informed decision-making among stakeholders and empowering communities to address wildfire challenges collaboratively.

Limitations:

The limitations of this project are numerous and have a significant impact on the accuracy and interpretability of the findings. One major limitation is the use of state-level data from South Dakota as a proxy for the wildfire impact on Sioux Falls. While this is reasonable given the arbitrary units of the smoke metric, it may not fully capture the localized impact on the city. Adjusting the budget data to reflect Sioux Falls' specific contribution could refine the analysis but would not change the overall trends. Additionally, the smoke metric is limited to wildfires within a 650-mile radius of Sioux Falls. Fires beyond this range could still affect air quality, and significant wildfires in neighboring states might influence the smoke exposure in South Dakota without being reflected in the state budget.

Another limitation is the likely inclusion of prescribed burns in the wildfire acreage data. Prescribed burns, which are controlled fires used for vegetation management, could skew the results, particularly in years when they represent a large portion of reported acreage. Furthermore, the models used to analyze the data may extrapolate negative trends in wildfire acreage or smoke to zero or negative values, which is unrealistic. This issue is somewhat mitigated by the wide confidence intervals and the challenges faced in fitting simple linear models to the data.

The methodology also includes a one-year shift to align wildfire data with South Dakota's fiscal year budget. This approach oversimplifies the relationship between smoke exposure and spending, as the budget often reflects multiple factors, including preventative measures and reclamation projects from prior years. This implies a direct causal relationship between smoke and spending that is unlikely to be entirely accurate. Compounding this issue, the smoke metric itself is overly simplistic, relying only on wildfire size and distance from Sioux Falls. It lacks important variables such as weather patterns, vegetation type, or historical fire data, which would significantly improve its accuracy and predictive value.

Inflation adjustments in the analysis are also simplistic, relying solely on CPI data without accounting for sector-specific cost changes or other economic variables that might influence wildfire suppression budgets. Additionally, the wildfire suppression budget may not exclusively cover wildfire-related expenses, as it could include funding for other issues like invasive species management. While most years indicate a strong focus on wildfires, the precise breakdown of spending remains uncertain.

Lastly, the use of the LOWESS smoothing algorithm to model trends presents challenges. While effective for approximating nonlinear trends, LOWESS is sensitive to outliers and does not provide a formal statistical model. This could result in exaggerating local noise or obscuring broader trends, making it less reliable for long-term projections. These limitations underscore the inherent challenges of working with real-world data and the need for more nuanced approaches to ensure the robustness and accuracy of analyses.

Conclusion:

The primary goal of this study was to understand how wildfire activity and its associated smoke impact Sioux Falls, South Dakota, and to explore how these trends correlate with the state’s wildfire suppression budget. Using a novel smoke metric based on wildfire size and distance, alongside state budget and wildfire acreage data, I hypothesized that there would be a correlation between increased wildfire activity, smoke impact, and suppression spending. I also sought to examine how these variables might project into the future, highlighting areas for proactive policy intervention.

The findings generally aligned with my hypotheses, indicating that while wildfire acreage has fluctuated, periods of high smoke and fire activity often corresponded with increased suppression spending, particularly in earlier years. In recent years, there appears to be a shift toward a more preventative approach, with investments in wildfire management infrastructure. However, this study also revealed significant periods where spending and smoke levels were decorrelated, suggesting either a lag in budgetary response or the influence of non-wildfire factors in suppression expenditures. Projections using linear models and smoothing techniques suggested that wildfire acreage could decrease in the future, but the budget for suppression is expected to remain steady.

This study underscores the importance of human-centered data science principles in addressing complex environmental and societal challenges. Key methodological choices, such as focusing on interpretability and simplicity, were motivated by the need to create insights accessible to a diverse audience, including policymakers and the general public. By designing a straightforward smoke metric and employing linear regression, the analysis prioritized clarity over complexity, ensuring that the findings could effectively inform decision-making. However, this approach also highlighted the ethical responsibility to acknowledge limitations and ensure transparency about the assumptions and constraints inherent in the analysis. Moving forward, this framework could serve as a basis for expanding the analysis with more robust data and techniques, while maintaining the core principle of accessibility and relevance to human concerns.

References:

1. “SD Wildland Fire: Operations.” *SD Wildland Fire | OPERATIONS*, wildlandfire.sd.gov/operations/prescribedfire.aspx.
2. *South Dakota Comprehensive Annual Financial Report 2019 page 29*, South Dakota Bureau of Finance and Management, 31 June 2019, bfm.sd.gov/ACFR/SD_CAFR_2019.PDF.
3. “U.S. Census Bureau Quickfacts: Sioux Falls City, South Dakota.” *QuickFacts Sioux Falls City, South Dakota*, United States Census Bureau, www.census.gov/quickfacts/fact/table/siouxfallscitysouthdakota/PST045223.
4. *SOUTH DAKOTA SALES and USE TAX REPORT Page 8*, South Dakota Department of Revenue, 1 Sept. 2024, dor.sd.gov/media/wslnk1xx/0924countybycitypercentoftotal.pdf.
5. “Statsmodels.Nonparametric.Smoothers_lowess.Lowess¶.” *Statsmodels.Nonparametric.Smoothers_lowess.Lowess - Statsmodels 0.15.0 (+522)*, statsmodel, www.statsmodels.org/dev/generated/statsmodels.nonparametric.smoothers_lowess.lowess.html.
6. *South Dakota Comprehensive Annual Financial Report 2023 page 28*, South Dakota Bureau of Finance and Management, 31 June 2019, https://bfm.sd.gov/ACFR/SD_ACFR_2023.PDF#view=fit.

Data sources:

1. Welty, J.L., and Jeffries, M.I., 2021, Combined wildland fire datasets for the United States and certain territories, 1800s-Present: U.S. Geological Survey data release, https://doi.org/10.5066/P9ZXGFY3.
2. *Air Quality System (AQS) API*, United States Environmental Protection Agency, aqs.epa.gov/aqsweb/documents/data_api.html.
3. *Consumer Price Index, 1913-: Federal Reserve Bank of Minneapolis*, Federal Reserve Bank of Minneapolis, www.minneapolisfed.org/about-us/monetary-policy/inflation-calculator/consumer-price-index-1913-.
4. Budget information:

Year	Amount (in thousands)	Link	Page according to PDF	Wildfire acreage
2001	1890	https://bfm.sd.gov/ACFR/SD_CAFR_2001.PDF	60	

2002	1692	https://bfm.sd.gov/ACFR/SD_CAFR_2002.PDF	75	55976
2003	1300	https://bfm.sd.gov/ACFR/SD_CAFR_2003.PDF	31	116933
2004	3500	https://bfm.sd.gov/ACFR/SD_CAFR_2004.PDF	31	73585
2005	2800	https://bfm.sd.gov/ACFR/SD_CAFR_2005.PDF	32	7309
2006	1700	https://bfm.sd.gov/ACFR/SD_CAFR_2006.PDF	33	47860
2007	700	https://bfm.sd.gov/ACFR/SD_CAFR_2007.PDF	33	288616
2008	4700	https://bfm.sd.gov/ACFR/SD_CAFR_2008.PDF	35	55731
2009	600	https://bfm.sd.gov/ACFR/SD_CAFR_2009.PDF	37	5803
2010	1200	https://bfm.sd.gov/ACFR/SD_CAFR_2010.PDF	33	10335
2011				7100
2012	7900	https://bfm.sd.gov/ACFR/SD_CAFR_2012.PDF	25	123484
2013	6200	https://bfm.sd.gov/ACFR/SD_CAFR_2013.PDF	27	148604
2014	1950	https://bfm.sd.gov/ACFR/SD_CAFR_2014.PDF	26	11421
2015	750	https://bfm.sd.gov/ACFR/SD_CAFR_2015.PDF	26	53145
2016	1300	https://bfm.sd.gov/ACFR/SD_CAFR_2016.PDF	24	17359
2017	3300	https://bfm.sd.gov/ACFR/SD_CAFR_2017.PDF	26	74228
2018	766	https://bfm.sd.gov/ACFR/SD_CAFR_2018.PDF	14	65659
2019	900	https://bfm.sd.gov/ACFR/SD_CAFR_2019.PDF	30	4413
2020	2874	https://bfm.sd.gov/ACFR/SD_CAFR_2020.PDF	104	3404
2021	4501	https://bfm.sd.gov/ACFR/SD_ACFR_2021.PDF #view=fit	104	39771
2022	8441	https://bfm.sd.gov/ACFR/SD_ACFR_2022.PDF #view=fit	103	17059
2023	5500	https://bfm.sd.gov/ACFR/SD_ACFR_2023.PDF #view=fit	102	5655

Wildfire acreage sourced from the following two PDFs (both contain up to 10 year historical data starting in 2002):

https://bfm.sd.gov/ACFR/SD_CAFR_2017.PDF page 193 https://bfm.sd.gov/ACFR/SD_ACFR_2023.PDF#view=fit page 203 of the PDF