Zach Price
12/11/2023
DATA 512

Final Report

Introduction

Climate change has rapidly increased over the past 50+ years throughout the world, bringing a variety of environmental alterations along with it. With the rising temperatures brought on by climate change, forest fires have become an increasingly popular area of focus. As global temperatures rise, forest fires are becoming more and more prevalent.

The city of Kingman, Arizona is no exception to this trend. Located in the arid environment of northwest Arizona, Kingman has experienced a great increase in the number of nearby forest fires over the past few decades. It's apparent that the increase in forest fires has introduced large damages to the forests themselves, but it's not apparent how this increase in forest fires has affected the people of Kingman, Arizona.

The goal of this project is to determine the impact that forest fires within the surrounding area of Kingman, Arizona over the past half century have had on the people of the city. This is a broad objective and nearly impossible to investigate with one project; thus, we will narrow our goal of understanding the impact of forest fires. Specifically, we will determine the impact that forest fires that occurred within 1,250 miles of Kingman, Arizona between 2010 through 2020 have had on the four year graduation rate of high schools throughout Mohave County, Arizona; the county that Kingman is located in.

This project attempts to answer an unresolved research question about the potential affects increased forest fires have on four year high school graduation rates in Mohave County, Arizona. Our initial thoughts are that increased smoke from a higher amount of forest fires would negatively impact performance in the classroom due to the inability of students to spend time outside as they would in their normal routines. Obtaining a high school diploma is vital for the success of many young Americans; thus, we hope to do our best in mitigating any factors that might hurt students' chances of graduating from high school.

This research is interesting and useful from both scientific and practical perspectives. Through the scientific lens, it's intriguing to quantify the potential impact that increased smoke has had on education. This helps us understand correlations and leading indicators for changes in four year graduation from high school. From a practical perspective, this helps the city of Kingman,

Arizona better prepare for the impacts of future forest fires. For instance, assume we find a negative correlation between smoke impact and four year graduation rate. The mayor of Kingman, Arizona could use this information to proactively increase schooling efforts and so forth to combat falling graduation rates as forest fires rise.

From this project, we hope to learn whether there is a correlation between our smoke impact estimate metric and four year high school graduation rates in Mohave County, Arizona. Furthermore, we hope to learn if we can use this relationship to predict four year high school graduation rates with our smoke impact estimate. If this turns out to be possible, we can provide a helpful indicator to Kingman, Arizona for understanding changes in four year high school graduation rates, allowing them to better prepare for future impacts of forest fires on their education system.

Background & Related Work

There does not exist an abundance of research on the effects of wildfires on education; however, there is some work that can help guide our analysis on this project. Wen and Burke's paper, *Lower test scores from wildfire smoke exposure*, provides an insightful look into the effects that increased smoke from wildfires has had on test scores. The study investigates United States standardized test scores from 2009-2016 for 11,700 school districts, in relation to satellite-derived estimates of smoke exposure. The researchers found that smoke exposure during the school year reduces test scores by roughly 0.15% of a standard deviation. Furthermore, the work estimates that smoke exposure in 2016 reduced discounted future earnings by approximately $111 per student. Lastly, the study found that these negative effects were most apparent in disadvantaged school districts, younger students, and racial/ethnical minority groups. Overall, this research agrees with our hypothesis that increased smoke from forest fires has a negative effect on four year high school graduation rates in Mohave County, Arizona.

Hamideh et al. provide another interesting viewpoint about wildfire impact on education through their paper *Wildfire impacts on education and healthcare: Paradise, California, after the Camp Fire*. This study gives a qualitative analysis on the effects on education and healthcare in Paradise, California following the 2018 Camp Fire. Through interviews with teachers, counselors, and administrators, the researchers cultivate a qualitative understanding of the 2018 Camp Fire's effect on the local educational institutions. Overall, the study found that many of the students experienced trauma and stress from the evacuation caused by the 2018 Camp Fire,

making it difficult for them to refocus on their educational responsibilities. This supports our hypothesis that increased smoke from forest fires has a negative effect on four year high school graduation rates in Mohave County, Arizona; if students are focused on the forest fires in their surrounding area, they might be less likely to graduate in four years from this lack of focus on their education.

The last work we will discuss is not a research paper like the prior two pieces, rather it is the state of Washington's "Summary Wildfire Smoke Guidance for Closing Schools" by their Department of Health. Unfortunately, a document such as this could not be found publicly available for Arizona, but this report serves the same purpose. Overall, this shows that schools have to close once certain levels of smoke from wildfires are surpassed. While Arizona likely does not have the exact same guidelines as Washington, they still have to close schools when smoke levels become too unbearable for students and potentially harmful to their health. This further supports our hypothesis that increased smoke from forest fires has a negative effect on four year high school graduation rates in Mohave County, Arizona; if students are forced to miss school due to closures from high smoke levels, they might be less likely to graduate in four years from a lack of time spent in the classroom.

Data on forest fires throughout the United States was taken from the United States Geological Survey. This data was used to gather information on every forest fire that occurred within 1,250 miles of Kingman, Arizona from 1963-2020. Relevant information on the forest fires includes a year field with the year of the fire, an id field with the id number of the fire, a name field with the name of the fire, a size field with the number of acres burned by the fire, a type field with the type of the fire, a close_lat field with the closest point's latitude to Kingman, Arizona for the fire, a close_lon field with the closest point's longitude to Kingman, Arizona for the fire, and a distance field with the distance in miles of the closest perimeter point of the fire to Kingman, Arizona. In addition, we pulled data from the United States Environmental Protection Agency's sensors in Mohave County, Arizona to acquire daily measurements of the air quality index near Kingman, Arizona from 1996 through October 2023.

In order to link the wildfire data to education in Kingman, Arizona, we had to find another reliable data source that tracks a meaningful education metric within the area. Thus, we used data from the Arizona Department of Education to gain information on four year high school graduation rates for Mohave County, Arizona from 2010 through 2020. This data can be found

here under Accountability & Research Data → Graduation Rate, Dropout Rate, and Enrollment Reports → Graduation Rates → Four Year Graduation Rates.

The Arizona Department of Education states:

> The Arizona Department of Education (ADE) is committed to maintaining transparency of government and providing actionable information to parents, educators, and the community. To that end, the ADE provides public datasets and reports that are available and may be immediately downloaded and utilized. These are aggregated files that do not include student level data. The Data Governance Team collects and maintains the frequently used ADE datasets on this page.

The Arizona Department of Education website does not display much beyond this message in the sense of licensing and terms of use, unless one is making a specific data request. Thus, we interpret the phrase "may be immediately downloaded and utilized" as appropriate for approving our use of the data in this project.

The data itself is not too complex but provides the information necessary for completing this project. The data is annual and comes in a separate XLS file for each year. Thus, there are eleven datasets in total with the same schema that need to be combined into a single dataset for analysis. Each dataset includes eleven columns: Cohort Year for the year of graduation, Graduation Rate Type to indicate four year graduation, LEA Entity ID for the unique code of each school, LEA Name for the name of each school, School Entity ID for another unique code of each school, School Name for another name of each school, County for the county the school resides in, Subgroup for the demographic of students, Number Graduated for the number of students graduating, Number in Cohort for the number of students that began in each cohort, and Percent Graduated for the graduation rate.

Methodology

In this section, we will walk through the steps taken throughout our analysis on the effect of wildfires on four year graduation rates of high schools in Mohave County, Arizona, detailing both our analytical and ethical considerations.

The first step following the acquisition of the forest fire data was to find a metric that estimates the level of smoke produced by the forest fires. We detailed in the prior section the acquisition of the AQI data from the United States EPA. We wanted to go beyond these AQI measurements for

estimating air quality because we felt using the AQI data was somewhat unethical due to its potential misleading flaws. First, the EPA only had one reliable sensor in Mohave County from 1996-2020; this meant some years involved more data than others when averaging. Second, the sensors only captured a few types of particle pollutants, many left out $PM_{2.5}$, which is vital to consider when measuring air quality. Lastly, the data only went back to 1996; we wanted our smoke estimate to extend earlier in time, as to fully utilize the wildfire data.

For these reasons, we decided to create our own metric for air quality, labeling it the smoke impact estimate. The smoke impact estimate formula is simple. For each individual fire, we divided the number of acres it burned by the distance its closest perimeter point was from Kingman, Arizona. This means fires that burned more acres and were closer to Kingman, Arizona have a higher smoke impact estimate. Then, we grouped by year and summed the smoke impact estimate for all fires in each year and divided this number by 184. We divided by 184 for two reasons. First, this is the number of days in the fire season. Second, this made our smoke impact estimate similar in scale to our AQI measurements. When building our estimate, we wanted to include a factor to account for prescribed burns because we assumed these produce less smoke. However, we did not fully understand the difference in smoke production between wildfires and prescribed burns; thus, we felt it was unethical to create a factor for prescribed burns without fully understanding their effects. Once these steps were complete, we were left with wildfire data describing smoke estimates for each year from 1963-2020.

The next step in our process was to transform the education data into yearly aggregated graduation rates for high schools in Mohave County, Arizona. We chose four year graduation rates for understanding educational impact for two reasons. First, the data comes from a reliable source, the Arizona Department of Education. Second, we felt that four year graduation is a fair indicator of a young student's success in high school. We considered using standardized test scores, like Wen and Burke's in *Lower test scores from wildfire smoke exposure*, but we felt test scores do not fully encompass educational success as much as graduation rates. Many students do not care about their standardized test scores because they understand they have retake options for some tests and many of the tests do not directly affect their educational metrics. Thus, we wanted to avoid this potential bias by using graduation rates. In order to use the data for our purposes, we had to filter the County column to Mohave. Furthermore, we needed to aggregate over each year, school, and demographic subgroup to get total graduation rate for each year over

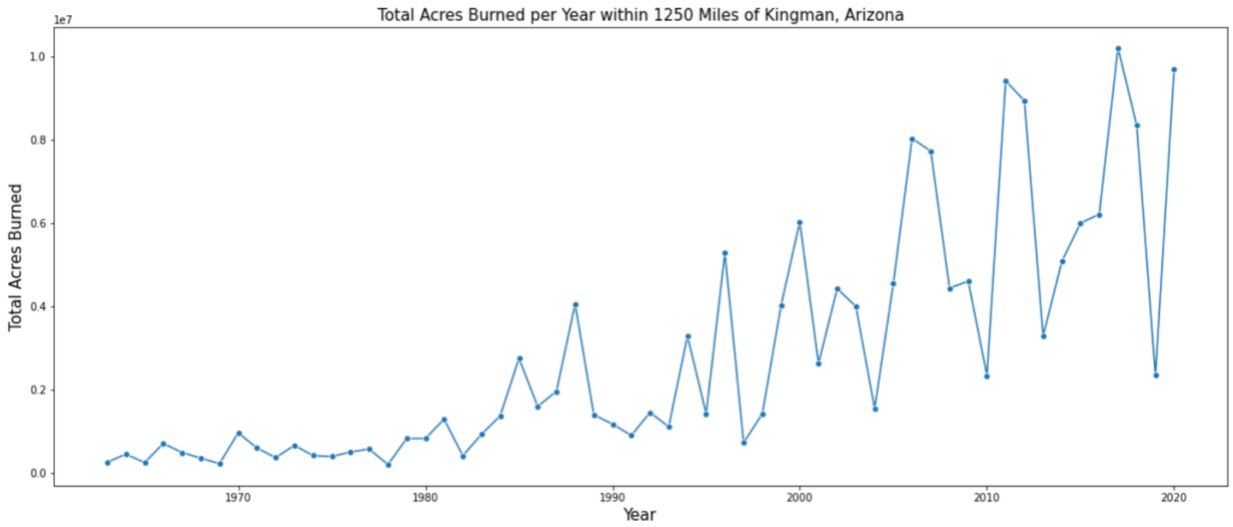
all schools in Mohave County. While we wanted to use graduation rate at the city level of Kingman, we had to settle for using county level data. Kingman makes up a large portion of Mohave County's population; thus, they are generally representative of each other.

Now that we had annual smoke impact estimates from 1963-2020 and annual four year graduation rates for high schools in Mohave County, Arizona from 2010 through 2020, we were ready to investigate potential impacts that forest fires had on four year graduation rates in Kingman, Arizona. Our goal was to build a model that could predict four year graduation rate using our smoke impact estimate. This would allow us to understand if a relationship exists between the two and provide future predictions on four year graduation rates so the city council of Kingman, Arizona could better prepare for the effects of upcoming fires on their educational institutions.

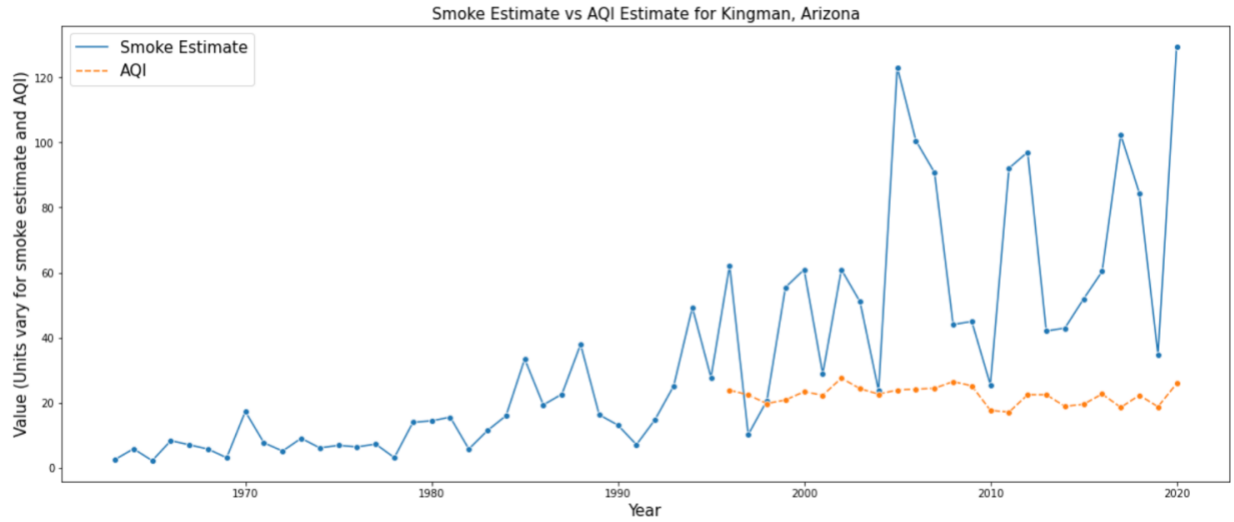
The model we used did not leverage any form of an existing model for understanding forest fire impact on education in Arizona. We decided to use a simple linear regression with smoke impact estimate as the predictor, a constant, and four year graduation rate as the response variable. We chose this model for a few reasons. First, linear regression is a very explainable model, allowing us to fully examine the relationship between smoke impact estimate and four year graduation rate. Second, we wanted to keep the model simple, rather than providing misleading estimates that would better fit the data. We felt it was unethical to fit a high degree polynomial model to the data because it could easily provide misleading results that overfit the data.

Findings

Let's start by taking a look at some of the findings from the forest fire data. This will help us gain a better understanding of the magnitude of the forest fire impact over the past couple decades. Below we investigate a variety of visualizations representing the forest fire data.



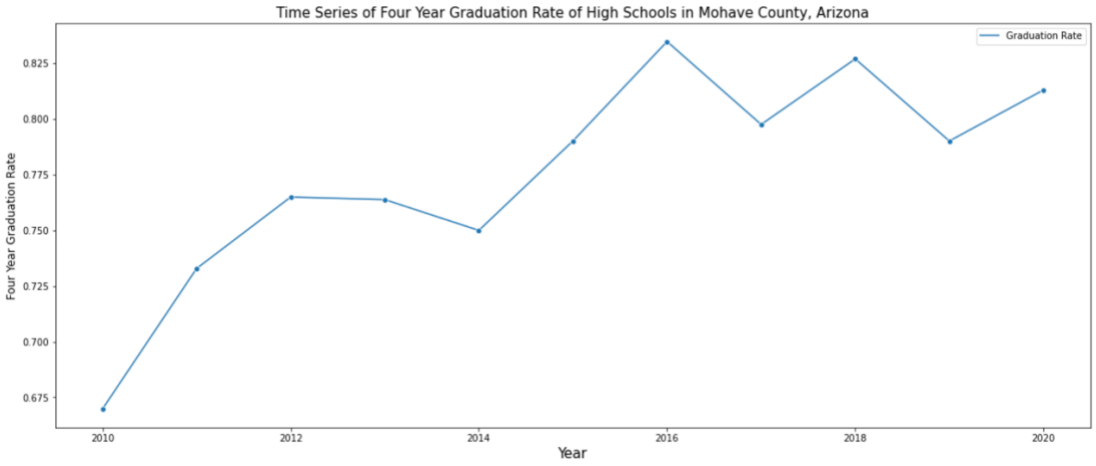
The figure above is a time series that shows the total acres burned by fires within 1250 miles of Kingman, Arizona from 1963 through 2020. The points represent individual total acres burned counts for each year, and the lines simply connect the points in the shortest possible path sequentially. This helps show the change in total acres burned over time. The viewer can read the visual as a trend over time of total acres burned by fires within 1250 miles of Kingman, Arizona from 1963 through 2020. Each point represents the number of total acres burned by fires for its respective year. Overall, we see an upward trend over time of total acres burned by fires from 1963 through 2020.



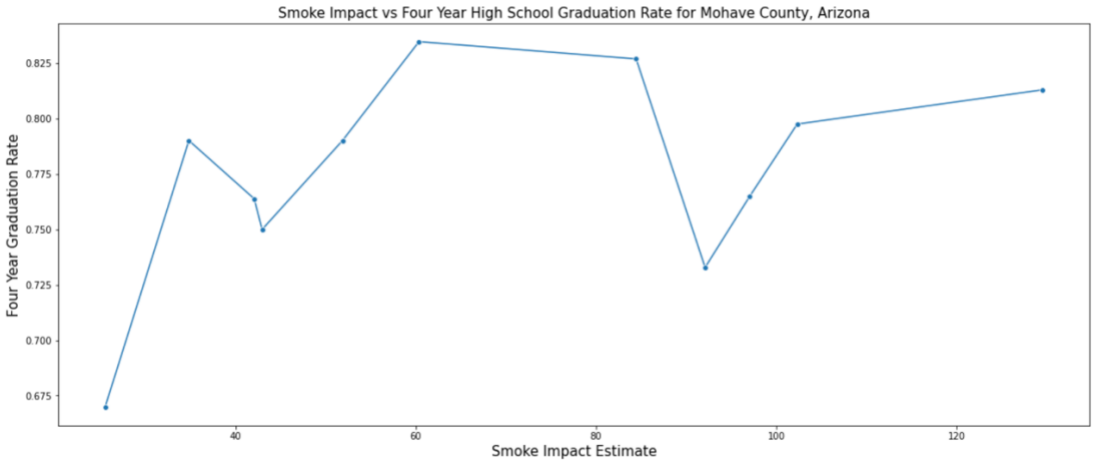
The visualization above is a time series showing our smoke estimate and the US EPA's air quality index estimate over time for Kingman, Arizona. The viewer can read the figure in a similar manner as we did for the previous visualization. The figure represents the trend over time in our smoke estimate compared to the US EPA's AQI estimate. Each point is an individual

measurement for its respective metric, with lines connecting points sequentially to provide an idea of trend. We can see that our smoke estimate generally increases over time from 1963 through 2020. However, the AQI estimate stays generally flat from 1996 through 2020. It's important to note that we're looking at trends rather than magnitudes in this graph because the units vary between the two metrics.

Next, we take a look at the education data and how it relates to the smoke impact estimate. Below we investigate a variety of visualizations representing the four year graduation rate and smoke impact estimate data.



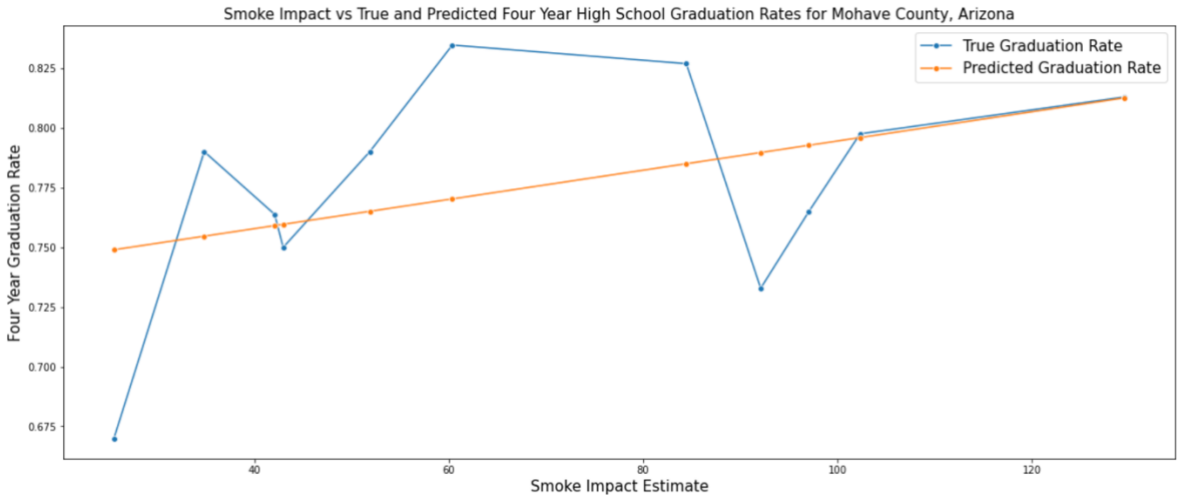
The figure above is a time series that shows the four year graduation rate of high schools in Mohave County, Arizona from 2010 through 2020. The data is annual; thus, there is a point for each of the eleven years in the data. Overall, we can see a clear upward trend in four year high school graduation rates in Mohave County, Arizona over the last elven years.



The figure above shows four year graduation rate plotted as a function of smoke impact estimate. We connected the points to highlight potential trends in the relationship between smoke impact estimate and four year graduation rate. Overall, this figure gives a good understanding of how four year graduation rates react to changes in smoke impact estimate. Ideally, these points would form a perfectly straight line because that would increase the usefulness of our linear regression model. The points do form a somewhat straight line but show patterns that will negatively affect the strength of the linear regression. Furthermore, four year graduation rate and smoke impact estimate clearly share a positive relationship with one another; they have a correlation of 0.43. This is surprising and goes against our hypothesis that increased smoke levels would negatively impact four year graduation rates.

	coef	std err	t	P>|t|
const	0.7334	0.032	22.571	0.000
smoke_impact	0.0006	0.000	1.438	0.184

The figure above shows the summary of the linear regression model. We see that the constant in the model is fit to 0.7334 and is a significant predictor in the model. We see that smoke impact estimate is fit to a coefficient of 0.0006, resulting in an insignificant p-value of 0.184 at the 0.05 significance level. This means that our model has not determined that smoke impact estimate is a significant predictor for four year high school graduation rate in Mohave County, Arizona.



The figure above displays the true four year high school graduation rates and predicted graduation rates in Mohave County as a function of the smoke impact estimate. The best fit line

provides a fairly good prediction of the true graduation rate; we see that many of the true and predicted graduation rates are fairly similar. However, the value of these predictions is mainly coming from the constant that is significant in the model, rather than the information provided by the smoke impact estimate.

Discussion & Implications

Our findings from the model show that our smoke impact estimate is not a significant predictor of four year graduation rate of high schools in Mohave County, Arizona. This implies that the increased amount of smoke from the higher number of annual forest fires has no effect on the four year graduation rate of high schools in Kingman, Arizona. While this is not the most exciting result, it does provide some useful information and potential actions for the city council to partake.

Smoke impact estimate and four year graduation rate were actually positively correlated. While this does not imply that more forest fires leads to higher graduation rates, it is surprising and completely goes against our hypothesis that increased smoke is affecting students' ability to succeed in school.

Furthermore, the city council should better investigate wildfire effects on education in Kingman, Arizona. To do this, they should utilize deeper metrics on forest fires and education. The city should try using a different metric to estimate smoke impact, potentially trying the satellite-derived smoke level estimates as Wen and Burke did in *Lower test scores from wildfire smoke exposure*. Additionally, the city should try using test scores, like Wen and Burke did in *Lower test scores from wildfire smoke exposure*, to estimate educational success. Changing the predictor or response in the model might lead to more significant results.

Lastly, the city of Kingman, Arizona should attempt to use a more complex model. This involves increasing the number of predictors and changing the type of model. Increasing the number of predictors can help to control for confounding variables and provide a more informative model. Changing the model type can increase the predictive power of the model over something such as linear regression. All of these solutions for improving the investigation should be done within the next few weeks in order to be prepared for future forest fires. While we did not find a negative impact of forest fires on education, other researchers have before. Thus, we need to fully

determine if a negative impact exists before the start of the next fire season so the city of Kingman, Arizona can fully prepare for maintaining the success of their students.

Throughout this entire project, human centered data science principles informed our decision making. First, when choosing a metric for education, we wanted a statistic that fully encompassed a student's academic success. We felt that standardized test scores were a small part of student success and that graduation rate better captured the full picture of student success. Second, we kept our model simple because we wanted it to be explainable to the Kingman, Arizona city council. While we felt that predictions were important, we considered inference from the model to be our main goal. A linear regression gives a solid understanding of the relationship between smoke impact estimate and graduation rate that can be easily explained to council members. This allows them to act without getting caught up in the technicalities of the model. Third, when building the smoke impact estimate metric, we did our best to involve factors that would most impact humans. This included the acres burned by the fires, the distance from Kingman, Arizona, and the number of days in the fire season.

Limitations

As with any large project, there are various limitations we encountered with this investigation. First, the education data is not as granular as we would have liked. The data is county-wide, rather than city-wide; data for schools just within Kingman, Arizona was not accessible, causing us to use county data on schools. This was not too much of an issue because the city is relatively representative of the county it resides in; however, we still would have liked data that only focuses on the city of Kingman, Arizona. In addition, COVID-19 introduced some interesting external circumstances to our investigation. Generally, students did worse academically during COVID-19, causing it to be hard to attribute potential drops in 2020 graduation rates to impacts of COVID-19 or increased forest fires. Furthermore, the data does not span as many years as we would have liked. Ideally, the graduation rate data would reach back to 1963, when the forest fire data begins. This would allow for more extensive investigation into trends over time. Unfortunately, this data only goes back to 2010, and we could not find any Mohave County education data that predates what we currently have. We would have liked education data reaching back to 1963, the start of the wildfire data. Finally, the education data is inherently difficult to align with forest fires. Peak fire season occurs over the summer when schools are

often not in session; this might explain the lack of a clear relationship between smoke impact estimate and graduation rate.

Another major limitation we faced was accounting for prescribed burns. We considered using prescribed burns when building our smoke impact estimate; however, we were not fully sure how to do so. We understand that prescribed burns likely produce less smoke than wildfires, but we're not sure how much less. On the other hand, we didn't want to fully remove prescribed burns from our data because they still produce some smoke. Thus, we decided to leave the prescribed burns in the data and treat them the same as a normal forest fire.

The final major limitation we encountered through this investigation had to do with the AQI measurements pulled from the United States Environmental Protection Agency. Ideally, we would have used these as predictors, instead of creating the smoke impact estimate, for four year high school graduation rates in Mohave County, Arizona; however, we felt there were some issues with the AQI measurements. Overall, we believed using the AQI data was somewhat unethical due to its potential misleading flaws. First, the EPA only had one reliable sensor in Mohave County from 1996-2020; this meant some years involved more data than others when averaging. Second, the sensors only captured a few types of particle pollutants, many left out $PM_{2.5}$, which is a vital consideration in measuring air quality. Lastly, the data only went back to 1996; we wanted our smoke estimate to extend earlier in time as to fully utilize the wildfire data.

Conclusion

Overall, this investigation attempts to understand the potential effects that rising forest fires over the past couple decades has had on education in Kingman, Arizona. Specifically, we attempted to answer whether increased smoke impact estimates from a higher number of wildfires from 2010 through 2020 affected four year high school graduation rates in Mohave County, Arizona. We hypothesized that poorer air quality from a higher number of fires would disrupt students' daily lives, leading to worse academic performance. In the end, we found that the smoke impact estimate had no significant effect on four year graduation rates of high schools in Kingman, Arizona. We recommend that the city council of Kingman, Arizona further investigate this topic through different metrics on smoke and education. Throughout this entire process, we have displayed the importance of human centered data science. Ultimately, this project focuses on the effects that fires are having on humans. We ensured that humans were a large part of our decision

making at every step, including building the smoke estimate, choosing metrics, selecting the model, and so forth.

References

Hamideh, S., Sen, P., & Fischer, E. (2021, October 9). *Wildfire impacts on education and Healthcare: Paradise, California, after the Camp Fire - natural hazards*. SpringerLink. https://link.springer.com/article/10.1007/s11069-021-05057-1

Washington State Department of Health. (2023, June 1). Summary Wildfire Smoke Guidance for Closing Schools. Washington.

Wen, J., & Burke, M. (2022, September 29). *Lower test scores from wildfire smoke exposure*. Nature News. https://www.nature.com/articles/s41893-022-00956-y

Data Sources

USGS Wildfire Data - https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81

US EPA AQI Data - https://aqs.epa.gov/aqsweb/documents/data_api.html

Arizona Department of Education Data - https://www.azed.gov/accountability-research/data