Zach Price
11/17/2023
DATA 512

Extension Plan

Motivation

The goal of this project is to determine the impact that forest fires throughout the United States over the past half century have had on our society. 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.

We showed in Part One – Common Analysis that the number of forest fires and acres burned by forest fires within 1,250 miles of Kingman, Arizona generally increased from 1963 through 2020. Using these metrics, along with proximity to Kingman, Arizona, we created a smoke impact metric; this also has been generally increasing from 1963 through 2020. Now, we wish to understand how this increasing smoke impact estimate relates to education throughout Mohave County, Arizona.

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.

Impact Focus

The focus area of this assignment is education. Specifically, we are investigating four year high school graduation rates in Mohave County, Arizona from 2010 through 2020. These include both public, charter, and private schools. The general thought is that increased forest fires over the past decade leads to increased smoke. The increased smoke leads to missing school, being unfocused in the classroom due to external circumstances, and not being able to go about life normally for an adolescent. All of these factors caused by increased smoke levels could negatively impact four year graduation rates from high school. Our goal is to investigate if this trend actually exists, model the situation, and provide informative advice to the city of Kingman, Arizona on their educational system in response to rising forest fires.

Data

We will be using the same data from Part One – Common Analysis for gathering forest fire information and creating smoke impact estimates. In addition, we will be using 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.

In order to use the data for our purposes, we will have to filter the County column to Mohave. Furthermore, we will need to aggregate over each year, school, and demographic subgroup to get total graduation rate for each year over all schools in Mohave County. If desired, we could also investigate four year high school graduation rates by demographic subgroups to understand further trends.

Model

This project will not be using any form of an existing model for understanding forest fire impact on education in Arizona. The model will consist of a simple linear regression using our smoke impact metric as an estimator for predicting four year high school graduation rates for Mohave County. We hope to find that the smoke impact estimate is a significant factor in predicting four year graduation rates in Mohave County, Arizona. We expect the two to share a negative correlation, if any relationship exists.

Unknowns and Dependencies

There are some factors outside of our control that might impact our ability to fully answer the questions we are posing about forest fires and education in Kingman, Arizona. First, we are relying on the quality of data gathered by the Arizona Department of Education. While this is a trusted source, organizations can still make mistakes in their data and misrepresent statistics accidentally. Second, the data is not as granular as we would like. The data is county-wide, rather than city-wide; data for schools just within Kingman, Arizona is not accessible, causing us to use county data on schools. This is 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. Third, COVID-19 introduces 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. Lastly, 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 goes back to 2010, and we could not find any Mohave County education data that predates what we currently have. Our analysis will still cover an eleven year span, but we wished for more range.

Timeline to Completion

Data Collection and Cleaning

This step involves downloading the data, combining the eleven datasets, and aggregating to get yearly graduation rates for Mohave County, Arizona. The deadline for this step is November 19th, 2023.

Model Building

This step involves creating a simple linear regression model with our smoke estimate as a predictor for four year high school graduation rate. The deadline for this step is November 22nd, 2023.

Model Analysis

This step involves determining the significance of our predictor and the quality of fit. Following this, more predictors might need to be added to the model for improvement. The deadline for this step is November 24th, 2023.

Result Visualization

This step involves creating helpful visualizations for understanding the impact our smoke estimate has on graduation rates and providing a general picture of the larger forest fire problem. The deadline for this step is November 28th, 2023.

Course Project – Part Three

This step involves creating a presentation for the completed project. The deadline for this step is November 30th, 2023.

Process Documentation

This step involves documenting the entire process, from accessing the data all the way to finalizing the model. It includes documentation throughout code notebooks and within the readme file of our repository. The deadline for this step is December 6th, 2023.

Course Project – Part Four

This step involves finalizing a repository with all code, documentation, data, and a final report for the entirety of this project. The deadline for this step is December 12th, 2023.