Zachary Bowyer

Data 512

11/13/2023

Extension Plan for Wildfire smoke impact on Leavenworth, Kansas

Motivation/problem statement:

Leavenworth, Kansas, is a small city [1] with a population of 37,086 as of 2022. While the first part of this project has shown that the air quality of Leavenworth has historically remained at healthy levels, the extension of this work plans on supplementing previous data with demographic data and research results to come up with a policy suggestion regarding the impact wildfire smoke has on the city's healthcare sector. Specifically, this work will address the relationship between exposure to wildfire smoke, the health of elderly people, health insurance, and persons with long term health issues. The final result of this work will be an outlined report that include findings, conclusions, and a final recommendation to the city of Leavenworth on whether or not this issue is currently important enough to warrant action.

The issue of how wildfire smoke impacts healthcare sector is specifically important to the community of Leavenworth, Kansas, because, according to [4], the healthcare industry makes up 16.4% of civilian employment for Leavenworth, which is slightly above the national average. Additionally, data from the US Census Bureau shows that the old age dependency ratio [5] has increased steadily from 2010 to 2021 [3]. This is relevant because older people are more likely to have health conditions [6] that may be exacerbated by wildfire smoke [7]. Overall, with these facts in mind, the author believes that city officials would be interested in knowing whether or not the future impact of wildfire smoke is an issue they should be concerned with due to the relative increase in the older population. To give a potential example of an issue arising from smoke, if the smoke trend continues to increase, then the healthcare sector may have to increase its numbers of employees to keep up with the number of elderly people. However, if the proportion of elderly people is too high, then the city may have to hire externally.

This analysis is potentially interesting to the author because it will help improve the author's understanding of the relationships between wildfire smoke and health, provide the author with an opportunity to work with US Census Bureau data [3], and be a good exercise for the author in trying to create a policy recommendation based on data and statistical techniques. With that being said, in this work, the author hopes to learn how to use the US Census Bureau API and create a good policy recommendation.

Impact focus:

The specific focus area of this study will be on healthcare. As stated in the problem statement section of this document, the analysis will be focused on how the increasing elderly population, who are more likely to have health issues that can exacerbate the negative effects of wildfire smoke, could impact the healthcare industry in Leavenworth, Kansas.

Data or model to be used:

In terms of the data being used for this work, the two datasets [8, 9] from the first part of this project will be reused for smoke and air quality analysis. The first dataset comes from the USGS [8], which, in short, contains wildfire data (polygons and metadata) that occurred from the 1800s to 2021. This wildfire dataset is planned to be used to estimate and project the smoke impact on Leavenworth using historical fire data and statistical techniques such as linear regression modeling and autoregressive integrating moving average (ARIMA) modeling. The reason the estimated smoke impact metric is important is to be able to show policymakers whether or not Leavenworth is at risk in the future of having a lower averaged air quality. In terms of permissible use, this dataset is "freely redistributable with proper metadata and source attribution" and has one requirement stated to be "The U.S. Geological Survey requests to be acknowledged as originator of these data in future products or derivative research." [8]

The second dataset comes from [9], which is an API designed "to obtain row-level data from the EPA's Air Quality System database" [9]. This dataset contains historical and current air quality index (AQI) measurements across the United States. To be specific, air quality index is a measurement that ranges from 0-500, and has five main measurements, which are ground-level ozone, particle pollution, carbon monoxide, sulfur dioxide, and nitrogen dioxide [10]. The reason this dataset will be leveraged is to provide an estimate for the historical AQI levels of Leavenworth, Kansas, to help determine if any trends exist between wildfire smoke, air quality, and public health in Leavenworth specifically. The terms of service for this dataset can be found on the dataset website [16].

The last data source this analysis plans on incorporating is from the US Census Bureau [3]. Using an API, important temporal information about the city of Leavenworth, such as demographic, health, and employment information, can be extracted and analyzed. For this work, the plan is to use such data in conjunction with smoke estimates and AQI estimates to derive new estimates that measure both the current and historical burden on the healthcare industry in Leavenworth. While the author is not currently aware of any specific healthcare-based estimates, research will be done as a part of this report to find appropriate metrics. For example, the initial plan is to research models that estimate health conditions based on age, research smoke impacts based on health conditions, and come up with healthcare impact metrics from said models. In terms of usability, the US Census Bureau offers its API as a service, which is offered conditionally subject to "acceptance of the terms and conditions contained" in its terms of service document [11].

Unknowns and dependencies:

This extension plan would not be complete without highlighting the various flaws and assumptions that may ultimately nullify the outcomes of this work. Below, five different issues with this work will be discussed.

One major flaw with this analysis, apparent in the first part of the project, is that the smoke estimate metric used is inherently flawed due to multiple reasons. First, the metric only considers wildfires that occurred within 1250 miles of Leavenworth, Kansas; it is common knowledge that wildfires can travel thousands of miles. Second, the smoke estimate assumes that the USGS wildfire polygons, dates, and other associated data are correct. Third, the smoke estimate uses a formula that incorporates distance from Leavenworth, which is assumed to be the average distance from all polygon points of the fire to Leavenworth's coordinates. Fourth, the smoke estimate incorporates two other estimates: the average

amount of wood per square feet in a forest and the volume of smoke produced by burning wood. These estimates are likely inaccurate due to the fact that the first estimate does not incorporate what type of trees were being burned and the second estimate does not incorporate how dense the forests were. Fifth, smoke dispersion is very complicated, and in state-of-the-art methodology, it is often estimated with an atmospheric transport model such as HYSPLIT [12]. In our case, since we do not have the resources to incorporate such expensive methods, we must assume smoke dispersion is inversely proportional to distance, which is a gross oversimplification. To summarize, the estimated smoke metric is likely very inaccurate due to the reasons listed above.

Another major flaw present in this work is that the AQI estimates being used to analyze the historical air quality of Leavenworth also are operating under many assumptions. For example, since wildfire smoke is the main concern, the AQI estimate used excludes pollutants such as Carbon Monoxide, Lead, Nitrogen Dioxide, Ozone, and Sulfur Dioxide [13]. Additionally, no air quality monitoring stations directly exist in Leavenworth, which means AQI estimates must consider multiple air quality systems in the surrounding counties, which leads to some inaccuracies in the AQI estimate.

One major assumption that this work will make is that data from the US Census Bureau is correct. Luckily, the data explored from the US Census Bureau thus far has stated uncertainties, but it is currently unknown how those uncertainties were formed. For the most part, US Census Bureau data are formed via polling, which means that the uncertainties between datasets may differ as well.

From there, the main unknowns of this work are mainly related to the development of quantitative healthcare impacts from smoke. For example, a question that must be answered in this work is "By what basis can healthcare impact be quantified". Attempting to answer this question may lead down two main paths. First, one can likely leverage existing research papers to try to understand and quantify how influxes of patients can overwhelm support services, how smoke can cause specific health defects, how susceptible elderly people are to wildfire smoke, etc. Another option is to come up with custom healthcare impact metrics and not worry about relative scale, but rather overall trends.

Finally, the biggest unknown is whether or not any proposed policies based on the analysis would have a positive effect. While policymaking by itself is quite difficult, this project adds another layer of difficulty by attempting to suggest a policy using many assumption-based estimates as trustworthy information. It seems that providing concrete proof to whether or not policy recommendations would have the desired outcome may be impossible, assuming the results of the analysis are trustworthy. Overall, for the sake of the work, it is likely that by the time suggestive measures are proposed, metrics will be assumed to be correct.

Timeline to completion:

Below will list each significant task that will need to be completed in order to have enough information ready to compile a final, extensive report. Specifically, each task will be sectioned in its own paragraph that includes a task description, time estimate, and justification for the time estimate.

The first major task that will need to be completed is to form census datasets for the city of Leavenworth, Kansas. These datasets will be formed by using the census API [14], and will likely contain historical demographic, employment, and health data. The estimated time cost for this task will be

roughly twelve hours due to the nature of learning a new API, making multiple requests to a free service over the internet, download time, and formatting data into a better format for the general analysis.

Next, once demographic data is collected, the difficult challenge of estimating healthcare impacts on both the general population and the elderly population will be next. In short, this task will require extensive reading of the current body of healthcare research to determine both if any significant impact could affect Leavenworth from smoke and what those effects may be. The estimated time cost for this task may be greater than 30 hours, as this task will require a lot of research and interfacing between the AQI, wildfire, and census datasets.

After that task is completed, the next task will follow a similar pattern to previous work, which involves creating a predictive model that will both predict current healthcare impacts and project the future healthcare impacts for Leavenworth going into the year 2050. Assuming the previous task effectively combines datasets, this task should take no longer than four hours, as this work will likely continue to employ both a linear regression and autoregressive moving integrated average model for predictions and projections respectively.

The next major task is to go through all of the work completed from previous tasks and summarize findings. This task will include developing multiple visuals, validating models, and finding out what the overarching story of the analysis is. It is likely the task will take roughly five hours, as developing good visuals often requires some extra programming and data manipulation.

Finally, given the summary of completed work and overarching story, the most important task is to come up with policy recommendations for the city of Leavenworth. This task will require careful consideration of Leavenworth's history, economy, and background as well as knowledge of how other cities have dealt with similar issues over time. This stage of the work is likely to take ten or more hours due to the research that must be involved.

After the main work has been completed, to comply with project grading criteria, a set of presentation slides will be developed that highlight and showcases the most important aspects of the project. Specifically, there is a requirement for the presentation to consist of eleven slides in the "PechaKucha" style presentation [15]. This portion will likely take two or three hours at most, as creating slides at this point will not require any additional technical work or research.

Lastly, to again comply with project grading criteria, a final report will be written that includes methods used, results, and information from previous documents. Specifically, the report must include introduction, background, methodology, finding, discussion, limitation, conclusion, reference, and data source sections. It is likely that this task will take twelve hours at a minimum, as writing a good report takes time.

References:

[1] https://nces.ed.gov/programs/edge/docs/locale_classifications.pdf

[2] https://lvcountyed.org/site-selectors/leading-employers/

[3] https://data.census.gov/

[4] https://statisticalatlas.com/place/Kansas/Leavenworth/Industries

[5] https://data.oecd.org/pop/old-age-dependency-ratio.htm#:~:text=age%20dependency%20ratio-,The%20old%2Dage%20to%20working%2Dage%20demographic%20ratio%20is%20defined,rates%2C%20fertility%20rates%20and%20migration.

[6] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5732407/

[7] https://www.epa.gov/wildfire-smoke-course/why-wildfire-smoke-health-concern

[8] https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81

[9] https://aqs.epa.gov/aqsweb/documents/data_api.html

[10] https://www.airnow.gov/aqi/aqi-basics/

[11] https://www.census.gov/data/developers/about/terms-of-service.html

[12] https://www.arl.noaa.gov/research/atmospheric-transport-and-dispersion/

[13] https://www.epa.gov/air-trends

[14] https://www.census.gov/data/developers/data-sets.html

[15] https://en.wikipedia.org/wiki/PechaKucha

[16] https://aqs.epa.gov/aqsweb/documents/data_api.html#terms