Part 4 - Final Report
Raman SV
11-Dec-23

Introduction

Bozeman, Montana, is a city in the Gallatin Valley, and hence, it is part of the aptly named Gallatin County. It is a picturesque place with amazing mountain views, surrounded by forests and is home to the largest educational institution in the state - Montana State University. Bozeman is famous among outdoor enthusiasts for activities such as hiking, skiing, and fly fishing. It has all the charm of a quaint western city combined with the neo-classical aspects of an American College Town.

Smaller towns still face large issues, smoke from the nearby lush forests is a major cause of concern. Especially for towns and cities unused to vehicular and industrial pollution, acute smoke can wreak havoc. Denizens of such places are often ill-equipped, in terms of ventilation or even air purifiers. Further, given that, Bozeman houses MSU, a significant amount of its population comprises of ephemeral student residents, for whom, such devices are a costly, if not unaffordable, investment.

Further, one can transform an economy of a college region by retaining the skilled students who become a valuable addition to the workforce. Look at Seattle for example, the prime university here is the renowned University of Washington. Although it was initially funded heavily by Boeing, the presence of a skilled workforce brought in (or led to the growth) of tech behemoths such as Microsoft, Amazon and Google. The key to retaining students would be a great experience over their college periods. This would in-turn attract other companies to the area and improve the overall standards of living.

A major part of Bozeman's economy revolves around education and its secondary industries such as hospitality and transportation. The figure below illustrates the makeup of Bozeman's industry –

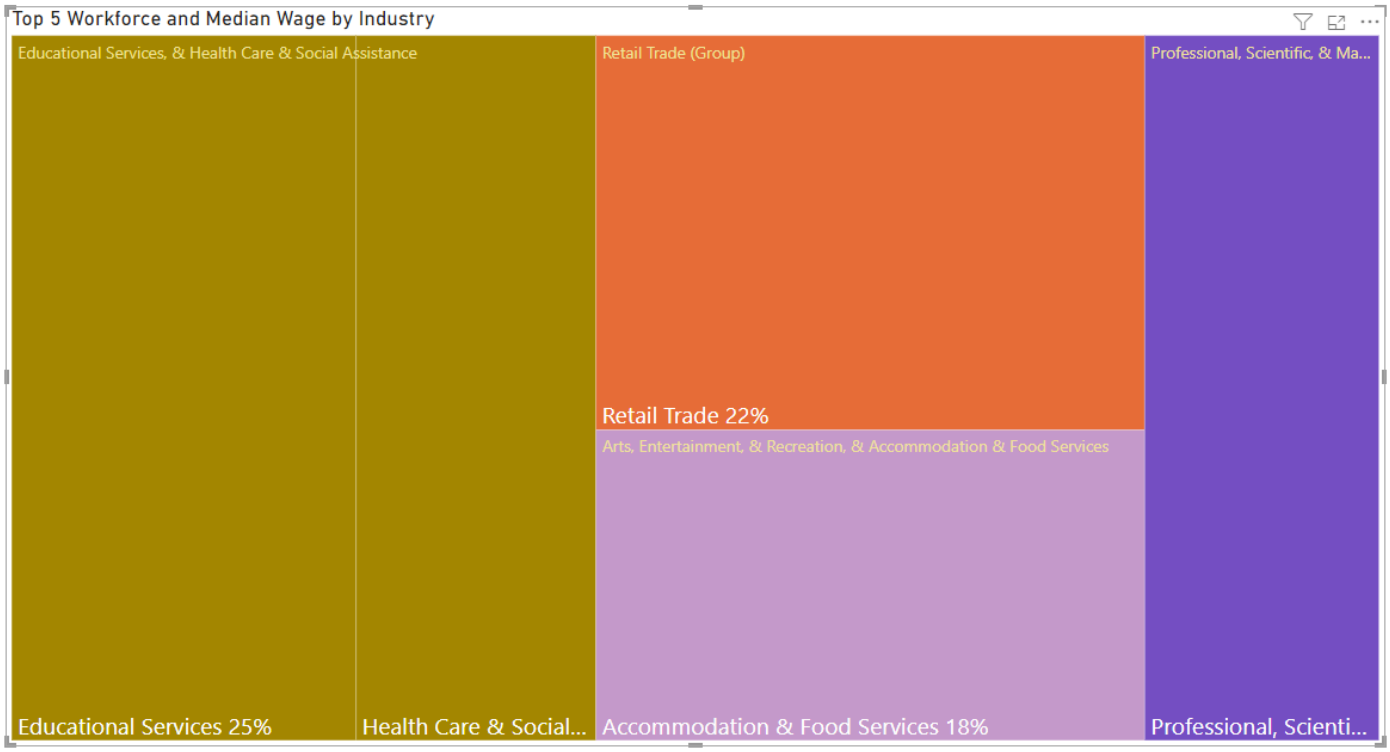

Figure 1 – Industry in Bozeman – top 5 by Workforce and Median Wage

Thus, it is important to understand the impact of wildfire smoke, on Bozeman as a whole as well as its student population.

Background and Related Work

In a large study in China, university students exposed to worse air had lower levels of happiness and higher levels of depression, compared with those exposed to less air pollution [1]. This study is reference in the paper – "The Impacts of Air Pollution on Mental Health: Evidence from the Chinese University Students". Further, the study found that male students had 1.6 times greater odds of increased mental health problems. This is an interesting stat and it gains importance in the context of MSU, as evidenced by the enrolment rate by gender, in the below figure (fig. 2). There are more males than females at the university and smoke pollution on account of wildfires can have a larger impact. Often times, when people move to a new place and feel down, they tend to look back on all the good times and nostalgia at the places they are from. This yearning, could lead to more students leaving a university area after graduation. This study is a major driving force behind this analysis – is there an impact of wildfire smoke on Bozeman with regards to education and employment and what does the future hold.

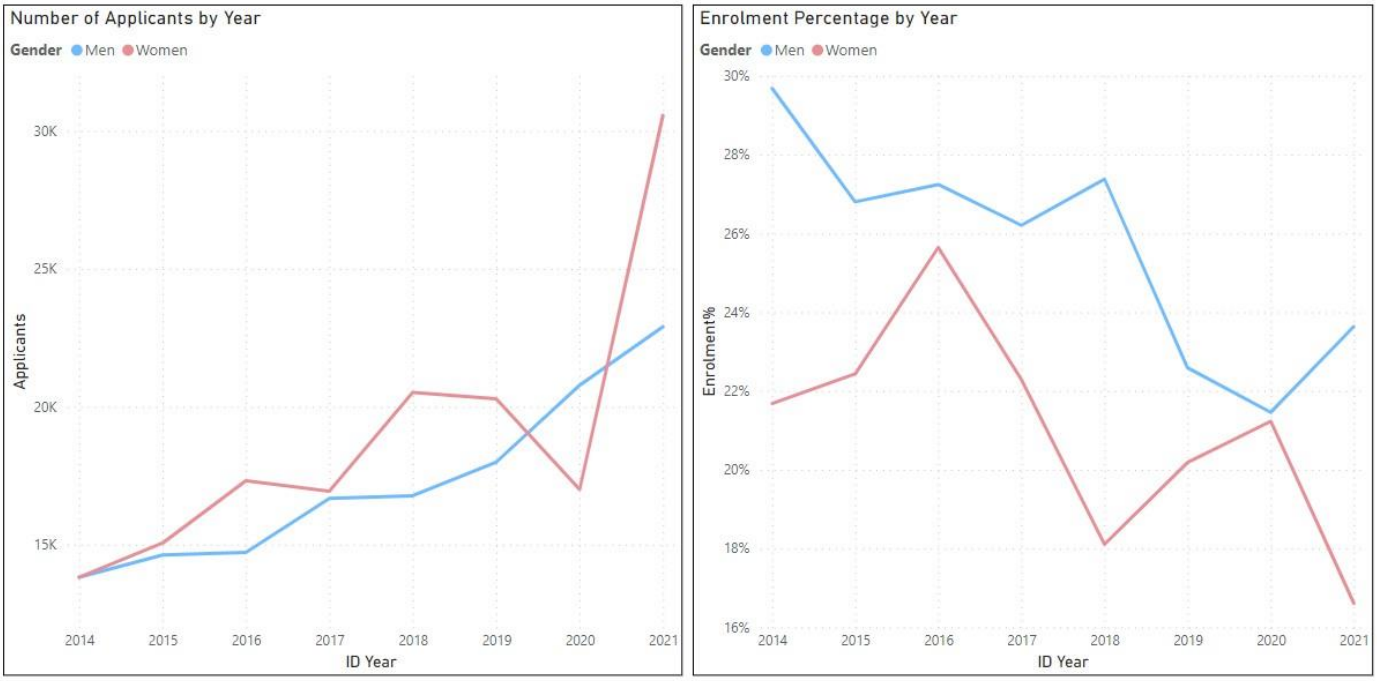

Figure 2 – MSU applicants and enrolment by year

Methodology

To analyze the impact of wildfire smoke on the city of Bozeman, wildfire data was collected from the US Geological Survey site. The distance of the fires from the city was calculated based on this data and below (fig. 3) is a figure that illustrates the number of fires since 1963 and their distances from Bozeman. The dataset also contains a description of the various fires. The primary reason for using this data is to generate an estimate for smoke based on the parameters tracked and compare this to the Air Quality Index (AQI) data to account for the accuracy of the model. AQI data is not widely available (especially pre 1985) and creating a model that predicts fires based on the source and not the impact would help predict fires around various regions, even those without nearby AQI tracking stations. The code for processing the data and generating the model is made open source and is readily available on GitHub for others to improve upon the model. This is particularly helpful for those who are closer to the ground and are looking for a way to analyze smoke impact in their region, and this model already takes care of getting data and processing it for the previous recorded fires as of 2023.

Apart from being open-source and free to use, this analysis is aimed at the city council members and as such, goes for a visual representation using simple graphs and doesn't include heavy data science terms. An important part of human centered data science is to always keep the audience in mind. Even the best models won't get picked up if they are inapproachable. Data scientists are story tellers, and the story is effective only if the author knows their audience.

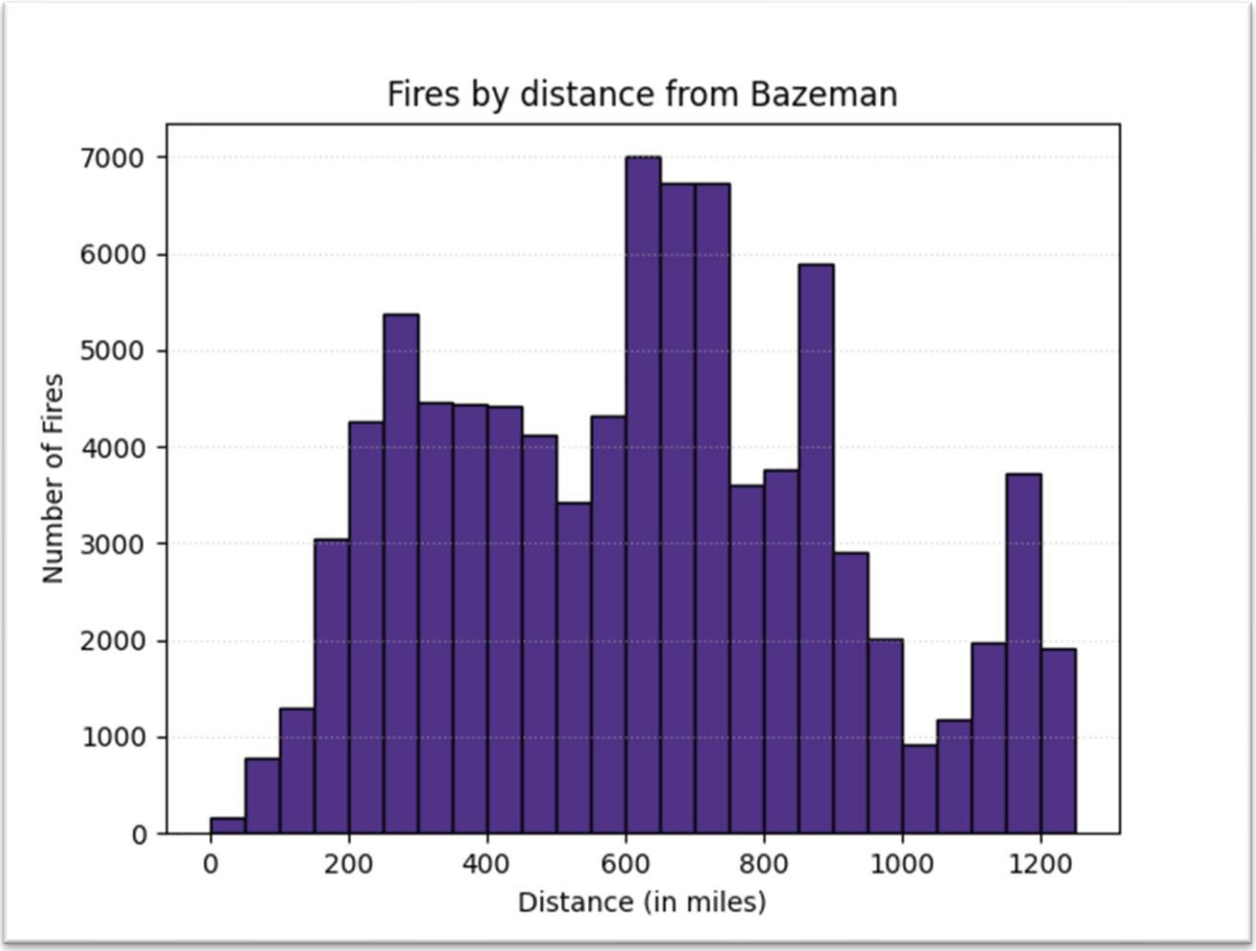

Figure 3: Fires by distance from Bozeman, MT

The model itself is a simply weight score for each fire based on these parameters, the weight ascribed to each parameter is represented in parenthesis –

1. Fire Type (5%) – this field comprises these values, with their weights in brackets: Wildfire (5) , Unknown - Likely Wildfire(2), Prescribed Fire(3), Likely Wildfire (4), Unknown - Likely Prescribed Fire(1).
2. Hectares affected (15%) – the hectares affected are sorted into quintiles with the most affected getting a weight of 5 and lowest a weight of 1.
3. Area affected by fire (15%) – similar to the above, this covers the area affected by the fire - sorted into quintiles with the most affected getting a weight of 5 and lowest a weight of 1.
4. Proximity of fire to Bozeman(35%) – the distance of the fire from Bozeman, the weights are assigned as: within 100 miles:5, 100-250mi:4, 250-500mi:3, 500-750mi:2, >750mi:1
5. Circle factor of the fire (30%) – this covers the circleness of the fire - sorted into quintiles with the most affected getting a weight of 5 and lowest a weight of 1.

For the AQI, data was obtained from the EPA Air Quality System (AQS) site. The API processing details are provided in the code file and can be used by anyone provided they get a code from the EPA AQS site.

Findings

Based on the above model, the fires were assigned a score and aggregated over a year and their comparison with the AQI scores are presented below in figure 4, as mentioned above, the AQI scores start only from the mid 1980s.

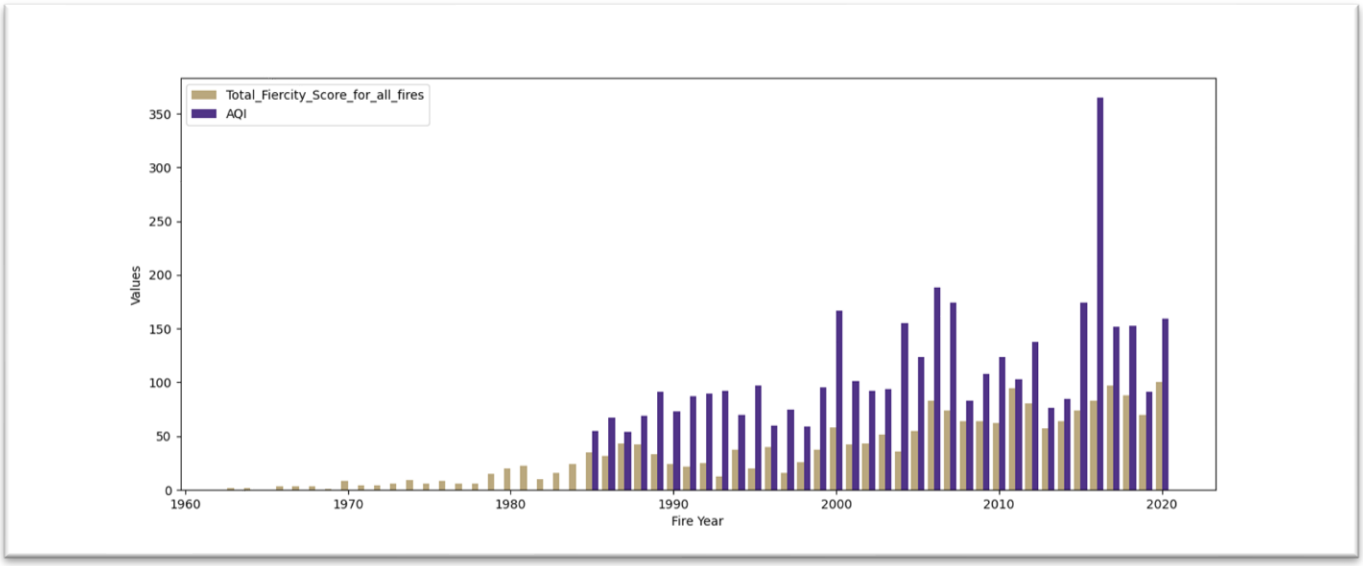

Figure 4: Comparison between the calculated "fiercity score" with the measure AQI

The comparison above is between the calculated fiercity score based on various input features and the overall AQI index over the period of 1963-2023 within 1250 miles of Bozeman, MT. The X-axis shows the years, and the Y-axis represents 2 data points for each year as bars: the predicted score and the AQI index for that year.

Further analysis on the employment composition of Bozeman and the MSU enrolment are available on a power BI file named – Bozeman Analysis.

Implications

One of the findings is the contrast that is evident from the visuals (fig. 5) from the enrolment statistics for MSU compared to the employment in sectors that could end up hiring students. For the purposes of this visual, the left hand side comprises median earnings for the Accommodation and Food services, art, and educational services industry. The right hand side is an assumption that graduating students who stay in the area would be in services such as Finance & Insurance, Information, and, Professional, Scientific and Technological services.

The enrolment vs entering the industry is staggered, as students would typically graduate in 3-4 years and going by this logic, the spike in enrolment around 2016 corresponds to a spike in 2019 for the RHS visual. From figure 4 we see that 2016 had the worst AQI in a long time and 2019 was the lowest in that time range.

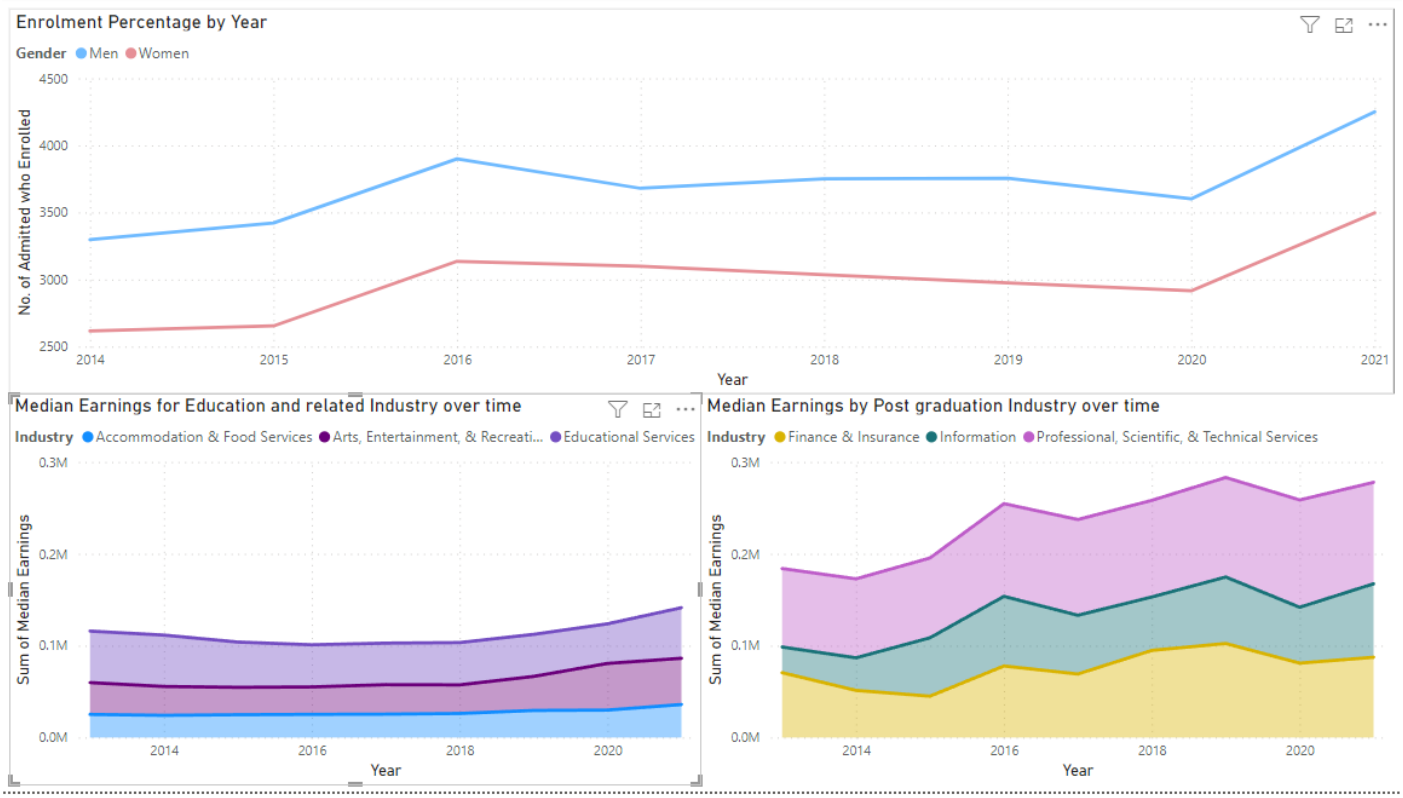

Figure 5. Enrolment percentage and Median earnings for various industries in Bozeman

Limitations

The above is evidence that is correlation at best, but we cannot say with certainty that this was a cause. We would need more data, especially around MSU grad statistics to confirm where the students go after graduation. But this is very granular data and it might need a certain level of obfuscation and anonymity as we are still dealing with young adults. Further, the analysis is around the impact of wildfire and this is a university project that takes a look at available data to come up with inferences and visual representations. This is not a replacement for the research reports that are objective and have performed site visits and confirmed observations on the ground.

Further, we are going through the worst phases of climate change so far and the incidences of wildfires might get higher in certain regions, this is not something the model takes into account. The below 2 graphs from figure 6 show the impact of wildfires over the last 35-40 years. It seems that the west coast is the more affected region and looking at Idaho, it is not a stretch to assume Montana could be next and the city needs to start working on mitigation measures.

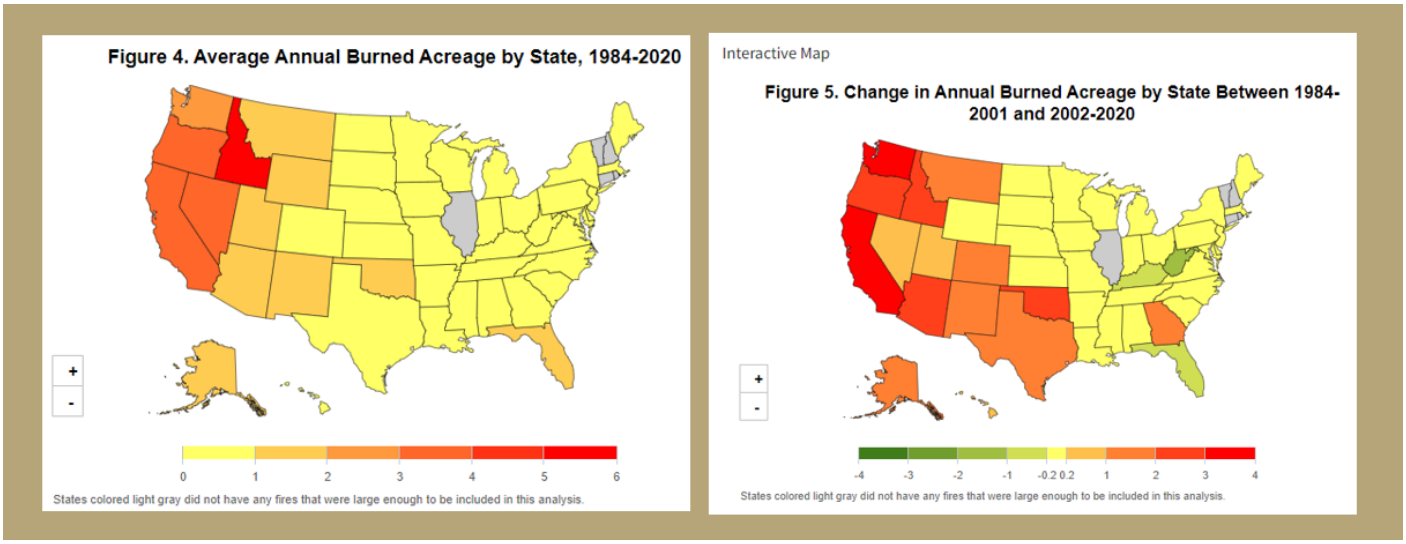

Figure 6. Spreading fires

Conclusion

The biggest takeaway of this activity was the chance to collaborate with various classmates to get their impressions on the project, their approaches as well as open-ended discussion around cause and effect. These meetings expound the importance of collaborative data science, there is a lot to learn when we socialize not just our results but our pathways and thought processes. One overarching theme that came through this research was the worsening AQI over multiple areas. Further, this research helped me understand the importance of preparing results and deriving methodology aimed at open and collaborative data science.

On a personal level, I have always been interested in the societal impacts of climate change and the discourse is wildly opinionated on the west coast social media sites I was able to look at. There is a larger impact of climate change on our lives and while fires definitely set the heart racing, the droughts that cause the fires are probably a bigger cause for concern. Drying rivers, over populated desert cities and unbridled commercial agriculture are major factors to keep an eye out for over the next few years.

I hope to continue analyzing these topics as I gain access to more data and I hope to be able to deliver a version that concentrates on statistical and data analysis techniques to add context to these visuals.

References

[1] The Impacts of Air Pollution on Mental Health: Evidence from the Chinese University Students, Daqing Zu et al. Int J Environ Res Public Health. 2020 Sep; 17(18): 6734. Published online 2020 Sep 16. doi: 10.3390/ijerph17186734

[2] https://www.ehs.washington.edu/environmental/wildfire-smoke Wildfire Smoke, UW's smoke safety requirements

Data Sources

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

Combined wildland fire datasets for the United States and certain territories, 1800s-Present (combined wildland fire polygons). This is the data for the wildfires that was processed to obtain the dataset that was the source for the smoke modelling.

[2] https://aqs.epa.gov/aqsweb/documents/data_api.html - this is the dataset that contains the AQI figures used to validate the smoke scores.

[3] DataUSA – readable link - This is a listing of the past 10 years of data for Bozeman that contains details on the workforce numbers, the industry, median earnings etc. Most of this data is sourced from governmental agencies

[4] U.S. Department of Housing and Urban Development - This covers the statistics used for housing and demographic related research including property value, rent vs owned statistics and social needs

[5] National Center for Education Statistics -This covers the statistics used for enrolled students, their demographics, completion rate etc.

[6] United States Department of Transportation - This covers the statistics used for analyzing freight movement, trade growth and traffic conditions