Part 2 - Extension Plan

Raagul Nagendran

Motivation/Problem Statement

There has been an alarming rate of increase in the frequency and severity of wildfires in the U.S. This has made us all realize that there is a pressing need to understand their broader impacts on community health, especially in urban areas where people may face health risks due to dense populations and limited mitigation options. Wildfire smoke, which releases harmful pollutants like carbon monoxide (CO), particulate matter, and smoke compounds, degrades air quality, leading to respiratory and cardiovascular health issues in humans.

The motivation behind this analysis is that by finding out more about these patterns between wildfires and their resultant health conditions, we could make informed decisions for emergencies, health aid allocation, and public health announcements during wildfire events. By finding patterns between mortality data and carbon monoxide exposure trends with wildfire data, the analysis seeks to provide a more nuanced view of how air quality degradation due to wildfires translates into measurable health outcomes. This could be valuable for local authorities, who could use the findings to implement more targeted interventions during wildfire episodes, and healthcare providers, who might anticipate and manage unusual increases in wildfire-related health issues.

Ultimately, this analysis hopes to highlight the human cost of wildfire pollution. This analysis could also be scaled to other cities and regions, offering a robust framework for analyzing the impact of wildfires across different regions. This analysis also aims to improve community awareness and provide local stakeholders with the evidence needed to advocate for improved air quality regulations and emergency response measures.

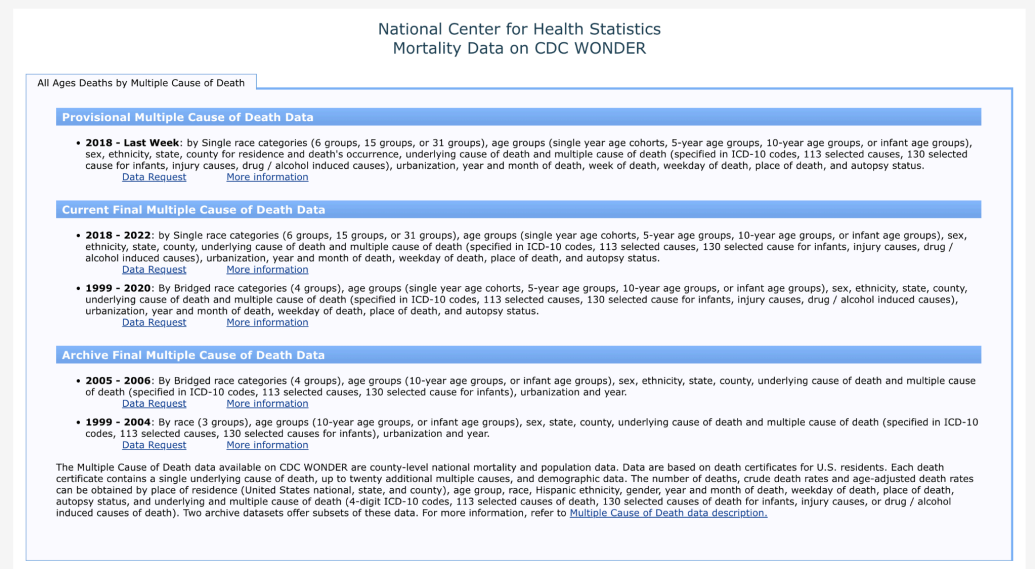
Impact Focus

For this extension, we will focus on the healthcare section in Arlington, Texas. We will specifically focus on the possible issues that could arise due to the smoke and particulate pollutants released in wildfires. This will involve deaths caused due to respiratory diseases in and around Arlington, including the whole county of Tarrant (where the city is located). Another analysis we will be doing will involve studying the relationship between people being exposed to carbon monoxide (calls made to the Texas Poison Center Network). Understanding these impacts is critical, as it highlights how wildfire smoke might cause severe health issues, which puts additional strain on healthcare systems and emergency services.

Data/Model to be used

We will perform this analysis by using two datasets: CDC Wonder Mortality Data and carbon monoxide exposure data from the Texas Poison Center Network (TPCN). Specifically, we want to analyze and quantify how wildfire events impact mortality rates and carbon monoxide-related health incidents, providing insights into the level of the health burden on Arlington city/Tarrant county residents.

1. Center for Disease Control and Prevention (CDC) – Wonder


National Center for Health Statistics
Mortality Data on CDC WONDER

All Ages Deaths by Multiple Cause of Death

Provisional Multiple Cause of Death Data

- **2018 - Last Week**: by Single race categories (6 groups, 15 groups, or 31 groups), age groups (single year age cohorts, 5-year age groups, 10-year age groups, or infant age groups), sex, ethnicity, state, county for residence and death's occurrence, underlying cause of death and multiple cause of death (specified in ICD-10 codes, 113 selected causes, 130 selected cause for infants, injury causes, drug / alcohol induced causes), urbanization, year and month of death, week of death, weekday of death, place of death, and autopsy status.
  Data Request More information

Current Final Multiple Cause of Death Data

- **2018 - 2022**: by Single race categories (6 groups, 15 groups, or 31 groups), age groups (single year age cohorts, 5-year age groups, 10-year age groups, or infant age groups), sex, ethnicity, state, county, underlying cause of death and multiple cause of death (specified in ICD-10 codes, 113 selected causes, 130 selected cause for infants, injury causes, drug / alcohol induced causes), urbanization, year and month of death, weekday of death, place of death, and autopsy status.
  Data Request More information
- **1999 - 2020**: By Bridged race categories (4 groups), age groups (single year age cohorts, 5-year age groups, 10-year age groups, or infant age groups), sex, ethnicity, state, county, underlying cause of death and multiple cause of death (specified in ICD-10 codes, 113 selected causes, 130 selected cause for infants, injury causes, drug / alcohol induced causes), urbanization, year and month of death, weekday of death, place of death, and autopsy status.
  Data Request More information

Archive Final Multiple Cause of Death Data

- **2005 - 2006**: By Bridged race categories (4 groups), age groups (10-year age groups, or infant age groups), sex, ethnicity, state, county, underlying cause of death and multiple cause of death (specified in ICD-10 codes, 113 selected causes, 130 selected cause for infants), urbanization and year.
  Data Request More information
- **1999 - 2004**: By race (3 groups), age groups (10-year age groups, or infant age groups), sex, state, county, underlying cause of death and multiple cause of death (specified in ICD-10 codes, 113 selected causes, 130 selected causes for infants), urbanization and year.
  Data Request More information

The Multiple Cause of Death data available on CDC WONDER are county-level national mortality and population data. Data are based on death certificates for U.S. residents. Each death certificate contains a single underlying cause of death, up to twenty additional multiple causes, and demographic data. The number of deaths, crude death rates and age-adjusted death rates can be obtained by place of residence (United States national, state, and county), age group, race, Hispanic ethnicity, gender, year and month of death, weekday of death, place of death, autopsy status, and underlying and multiple cause of death (4-digit ICD-10 codes, 113 selected causes of death, 130 selected causes of death for infants, injury causes, or drug / alcohol induced causes of death). Two archive datasets offer subsets of these data. For more information, refer to Multiple Cause of Death data description.

The Multiple Cause of Death data available on WONDER are county-level national mortality and population data spanning the years 1999-2022. Data are based on death certificates for U.S. residents. Each death certificate contains a single underlying cause of death, up to twenty additional multiple causes, and demographic data. The number of deaths, crude death rates, age-adjusted death rates, and 95% confidence intervals for death rates can be obtained by cause of death. We will make use of the following fields for now (might include more during analysis) -

1. Year
2. County - Tarrant County (the county Arlington is in)
3. Gender - to see if there is any difference between the gender demographics
4. Ten-year age groups - to see if there is any difference between the ten-year age groups demographics
5. Deaths (due to diseases of the respiratory system) - count of the deaths caused due to issues with respiratory system. These consider “underlying cause of death” only,

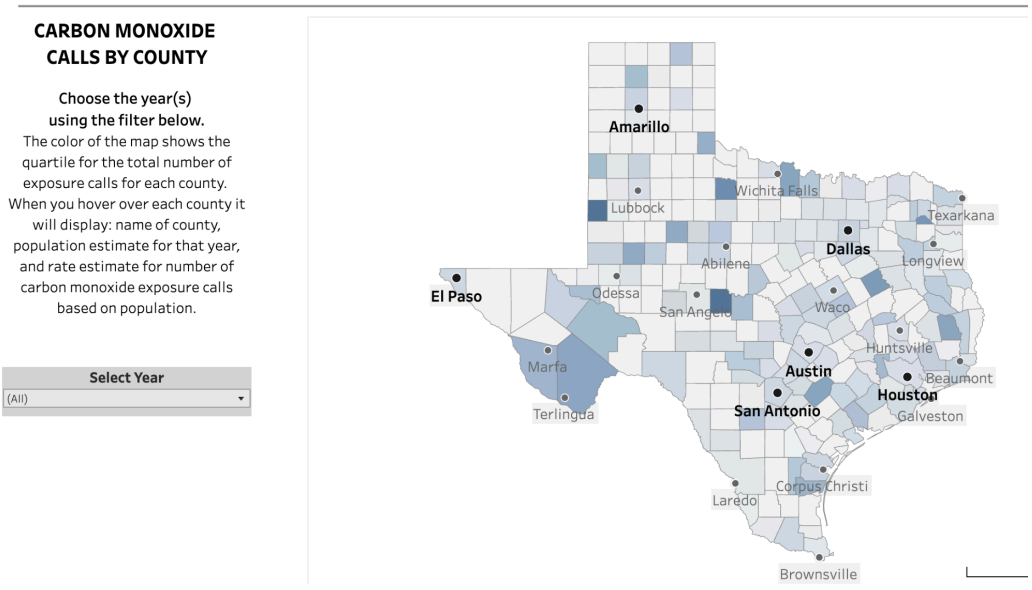
and will look into the “multiple cause of death” field during analysis to see if there’s any significant difference.

Information from the above website will be used following their [Data Use Restrictions and Policies](#).

2. Texas Poison Center Network (TPCN) – Carbon Monoxide exposure data

Texas Poison Center Network (TPCN) was created in 1993 with the goal of providing 24-hour access to free immediate medical advice for Texans. Anyone can access these services by calling a 24-hour, toll-free hotline, any day of the year. These centers help provide the public, hospitals, emergency responders, and health care providers with information and treatment for poisonous substances, hazardous substances, and similar issues. We collect the aggregated TPCN data from their [public dashboard](#) for Carbon Monoxide exposures. This is a dashboard that shows the total number of calls to the TPCN about exposures to carbon monoxide in Texas by year, county, and based on demographic information. We will be manually entering the information that we require from the dashboard, as the download option only allows exporting the visualizations and dashboards as a whole. We will make use of the following fields for now (might include more during analysis) -

1. Year
2. County - Tarrant County (the county Arlington is in)
3. Rate estimate for the number of CO exposure calls - this will be an estimate based on the total population of the county



Information from the above website will be used following their Data Use Restrictions and Policies.

Unknowns and Dependencies

At this stage of the project, here is a non-exhaustive list of unknowns and dependencies that I foresee -

- **Effects of COVID-19** - Almost all datasets and existing dashboards mention years with heavy COVID-19 cases as exceptions, which implies a significant skew in those years. Care must be taken to mention such anomalies in our final reports and visualizations, to not lead the readers astray.
- **Data inconsistencies -** Both datasets may have reporting delays or inconsistencies in geographic and temporal coverage, which could complicate the analysis.
- **Confounding factors -** There may be confounding environmental factors, such as regional industrial pollution or weather patterns, that influence air quality and health outcomes independently of wildfire smoke, potentially skewing the results.
- **Matching the different datasets -** A dependency for this project is the ability to accurately match wildfire events near Arlington city / Tarrant county to the actual AQI fluctuations and related health impacts, as localized meteorological data on wind patterns and pollution dispersion are necessary for accurate modeling (which we will not be performing due to the additional complexity it adds to our short project).
- **Correlation does not mean causation -** While performing the analysis, even if we get a good level of correlation between the different factors we are considering, it would not mean that wildfires would be directly causing the carbon monoxide poisonings or be the cause of the deaths due to cardiovascular diseases. It is very difficult to isolate the specific impact of wildfire smoke from other environmental factors. Furthermore, long-term versus short-term impacts may have different patterns in the data. The complexity of such analysis is beyond the scope of this project's goals.

Timeline to completion

1. **Collect data** - Complete by Nov 12th
    a. Get the CDC Wonder data and save them for processing
    b. Get the carbon monoxide data from the public dashboard and save it for processing

2. **Build a model** - Complete by Nov 16th
    a. Clean the datasets and suitably format them by standardizing dates and time intervals
    b. Identify the model to use and build the model
3. **Visualize the results** - Complete by Nov 23rd
    a. Analyze the results from the model
    b. Perform correlation and related analysis to find out data trends
4. **Presentation slides** - Complete by Nov 26th
    a. Prepare charts for the correlation, prepare the slides for the presentation
5. **Documentation** - Complete by Dec 4th
    a. Write final report, finish project repository, organize code files

References

1. Centers for Disease Control and Prevention. (n.d.). *Underlying Cause of Death, 1999-2020*. WONDER Online Database. Retrieved from https://wonder.cdc.gov/mcd.html
2. Texas Department of State Health Services. (n.d.). *Carbon Monoxide Related Poison Calls*. Health Data. Retrieved from https://healthdata.dshs.texas.gov/dashboard/environmental-health/carbon-monoxide-related-poison-calls
3. Tarrant County Public Health. (n.d.). *Chronic Disease Data and Information*. Retrieved from https://www.tarrantcountytx.gov/en/public-health/public-health-administration/health-data-and-information/chronic-disease.html?linklocation=Public%20Health%20Data%20and%20Information&linkname=Chronic%20Disease%3C/a
4. North Carolina State University. (2021, December). *Explainer: How Wildfires Start and Spread*. Retrieved from https://cnr.ncsu.edu/news/2021/12/explainer-how-wildfires-start-and-spread/