Disparities in Climate-Induced Health Outcomes in the Greater Toronto Area.*

Sarah Mansoor

07 July 2022

The growing impact of climate change in Canada demonstrates a clear and current threat to human health. These threats are becoming revealed in Canada as the country continues to face the consequences of warming almost at twice the global rate. The focus for this paper is on tempreture increases, lower air quality and more pollution and a decline in health for Toronto. Once the data for this research was collected, the data cleaning process begun which entailed of using packages in R like tidyverse, janitor, dplyr, …. Duplicate and irrelevant observations were removed, as well as unwanted outliers. The process was complete once the data was correct, consistent and usable.

Table of contents

*Code and data are available at: https://github.com/sarahmansoorr/SUDS2022

Introduction

As climate change worsens, the danger it presents to human health worsens as well. The health impacts due to climate change are already being demonstrated in parts of Canada. Canada is warming at almost twice the global rate and if this continues, emissions will exceed 2 degrees Celsius worldwide.

The Paris Agreement within the United Nations Framework Convention on Climate Change includes 175 countries that have agreed to limit the average global temperature rise to below 2 degress Celcius. Canada had signed to the Paris Agreement as well and committed to cutting emission 30% below the 2005 levels by 2030. However, Canada's emission increased in 2016 compared to 1990. As a response to this increase, the Government of Canada developed a national climate change plan called *The Pan-Canadian Framework on Clean Growth and Climate Change*. This framework notes that human activities are driving severe changes to the Earth's climate which poses substantial risks to human health.

Urgent action is needed and the science shows that it will make a tangible difference. The health impacts from climate change can be prevented by improving efforts both in societies and health systems in Canada. Changes in measures can prevent health impacts and build health systems and facilities that are more resilient to climate change.

Data

Data Collection

Temperature data was collected from the Government of Canada's website on Weather, Climate and Hazard's page containing past weather and climate data for the station "Toronto City". This data was collected from January 2002 to May 2022. The original data contains daily data on maximum temperature, minimum temperature, mean temperature, heat degree days, cool degree days, total rain (mm), total snow (cm), total precipitation (mm), snow on ground (cm), direction of maximum gust, and speed of maximum gust (km/h). For this research, I will focus on only maximum temperature, minimum temperature, and mean temperature.

Pollution data was collected from the Government of Canada's National Air Pollution Surveillance (NAPS) Program website. This website contains pollution data from 2002 to 2020. Each year contains a file with continuous data and integrated data. I will be using continuous data for each year. The continuous data contains annual summaries and hourly data. I will be looking at the hourly data for CO (Carbon Monoxide), NO (Nitric Oxide), NO2 (Nitrogen Dioxide) , O3 (Ozone), and SO2 (Sulfur Dioxide). This data includes the city (which is Toronto in our case), the levels of each pollutant for each hour of every day. I reduced the data to the mean of pollution levels by pollutant of each day.

Variables of Interest

Temperature variables of interest:

- City
- Maximum Temperature
- Minimum Temperature
- Mean Temperature
- Day
- Month
- Year

Pollution variables of interest:

- City
- Pollutant
- Mean pollution for pollutant each day
- Day
- Month
- Year

Data Processing

Temperature data

The data for temperature was processed using R studio and R. I started off by renaming the column names for each temperature data set from 2002 to 2022. Once this was done I combined the 2002 to 2022 temperature data together. The data set initially contained 31 variables. From these I selected the city, date, year, month, day, maximum temperature, minimum temperature, and mean temperature using the dplyr package. I then cleaned the names of the data set using the janitor package. I added a column for the names of the month by mutating the data set and changed the city rows to represent only Toronto data. From here, I dropped any NA values from the maximum, minimum, and mean temperature columns. Lastly, I saved the cleaned data into the repository as a csv file.

Pollution Data

The pollution data was also processed using R studio and R. I first loaded the different data sets for each pollutant for each year. Each year had 5 data sets for each pollutant (NO, NO2, CO, O3, SO2). I started off by also renaming the column names for the data sets from 2002 to 2020. Once this was done I combined the pollutant data sets for each year into one data frame for the year. The data initially contained 27 columns. The data sets initially contained pollution levels for each hour of every day in the respective year. I replaced these 24 columns (for each hour of the day) with one column that contained the mean pollution level for that day and pollutant. I filtered for Toronto from the list of cities in the data set using the dplyr package and removed the duplicate rows where the pollutant and date repeated. The missing values in this data set were labeled as "-999", so I replaced these values with NA and then with 0 for consistency. This process was repeated for all years from 2002 to 2020. Once this was completed, the data was combined for all the years into one data frame and the following columns were selected: city, pollutant, mean pollution, day, month, year. This cleaned data was then saved into the repository as a csv file.

Sample

The sample includes data from Toronto of daily temperature maximum, minimum and means from 2002 to 2022 and daily pollution levels for NO, NO2, CO, O3, SO2 from 2002 to 2020.

The Data

Table 1: Extracting rows from the Temperature data from 2002 - 2022

Year	Month	Day	Max Temp	Min Temp	Mean Temp
2002	11	14	14.3	4.4	9.4
2003	7	6	30.1	19.9	25.0
2004	6	22	23.9	15.2	19.6
2006	1	16	-3.0	-12.4	-7.7
2007	6	1	26.6	18.6	22.6
2008	3	23	1.5	-5.6	-2.1
2011	4	8	9.3	4.5	6.9
2014	1	8	-7.7	-16.2	-12.0
2016	10	26	6.2	0.9	3.6
2019	12	12	1.9	-7.6	-2.8

Table 1 shows the rows of the Temperature data from 2002 to 2022. Variable "Year" indicates the year, variable "Month" indicates the month, and variable "Day" indicates the day. Variable "Max Temp" represents the maximum temperature on the day with the following month and year. Variable "Min Temp" represents the minimum temperature on the day with the following month and year. Variable "Mean Temp" represents the mean temperature on the day with the following month and year.

Table 2: Extracting rows from the Pollution data from 2002 - 2020

City	Pollutant	Mean Pollution	Day	Month	Year
Toronto	CO	0.77	10	4	2002
Toronto	NO2	18.79	27	9	2002
Toronto	O3	9.54	15	4	2002
Toronto	SO2	3.92	9	2	2002
Toronto	SO2	5.46	6	12	2002
Toronto	CO	0.00	24	6	2003
Toronto	NO	33.71	6	11	2003
Toronto	NO2	16.21	2	9	2003
Toronto	SO2	4.92	11	11	2003
Toronto	O3	15.17	8	9	2004
Toronto	NO	2.04	4	6	2005
Toronto	SO2	1.00	1	3	2005
Toronto	NO2	8.04	20	5	2007
Toronto	CO	0.30	14	2	2008
Toronto	SO2	0.12	23	1	2014

Table 2 shows the rows of the Pollution data from 2002 to 2020. Variable "City" indicates the city from which the data was collected. Variable "Pollutant" indicates the pollutant and includes 5 types: CO, NO, NO2, O3, and SO2. Variable "Mean Pollution" represents the mean pollution levels of the following pollutant on the following day of the month and year. Variable "Day" indicates the day, variable "Month" indicates the month, and variable "Year" indicates the year. Variable "Max Temp" represents the maximum temperature on the day with the following month and year.

Visualizing Temperature from 2002-2022

Visualizing Temperature by Month and Year

I'm interested in finding the trends of the maximum temperatures from 2002 to 2022 by the month and year.

Figure 1 shows an increasing pattern for the maximum temperature from year 2002 to 2022 from December to March. Temperatures in March continue to be increasingly higher than any of the other months in this plot. It is noticeable that the temperatures for all months are the highest in 2012. According to NASA, the average temperatures in 2012 was warmer than the mid-20th century baseline and is considered to be an outlying of seasonal extremes that are warmer. Even the lowest maximum temperature in January is higher than other years in 2012. The plot shows that the average maximum temperature from December to March has been between -19.2 degrees Celsius in February of 2015 and 24.3 degrees Celsius in March of 2012.

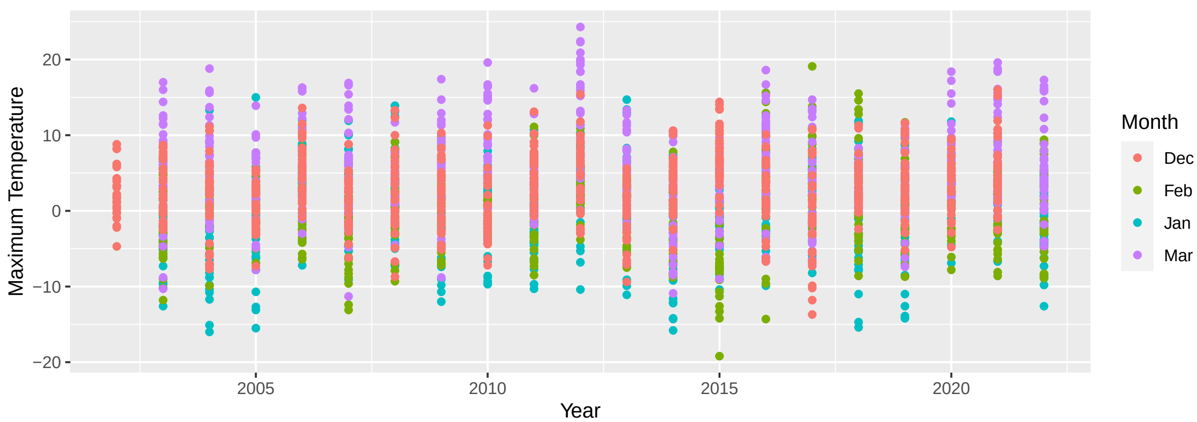

Figure 1: Temperature 2002 to 2022 for December to March

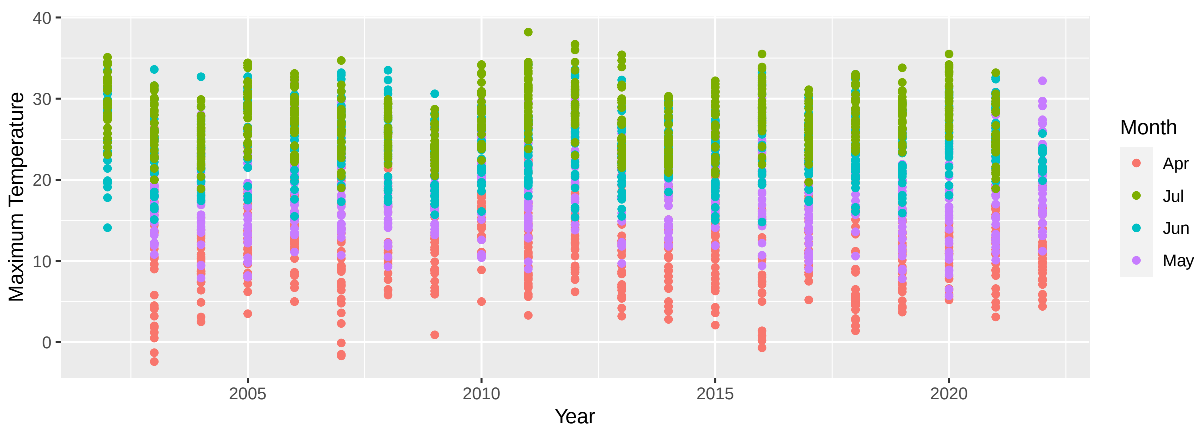

Figure 2: Temperature 2002 to 2022 for April to July

Figure 2 shows an increasing pattern for the maximum temperature from year 2002 to 2022 from April to July. Temperatures in July are consistently higher throughout the years than any of the other months in the plot and temperatures in April are consistently lower throughout the years than any of the other months. There are some years where June has the highest temperature. The plot shows that the average maximum temperature from April to May has been between -2.4 degrees Celsius in April of 2003 and 38.2 degrees Celsius in July of 2011.

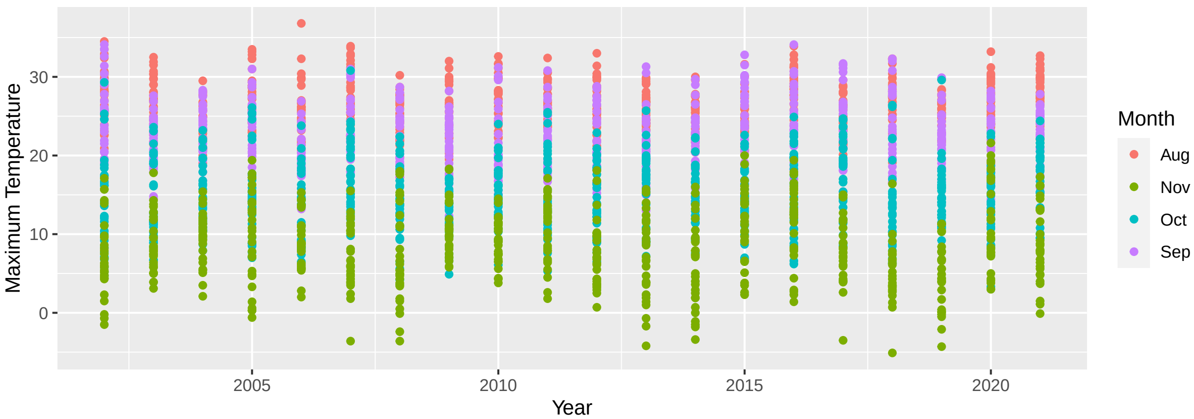

Figure 3: Temperature 2002 to 2022 for August to November

Figure 3 shows an increasing pattern for the maximum temperature from year 2002 to 2022 from August to November. Temperatures in both August and September are the highest throughout the years compared to November and October. Temperatures in November are consistently the lowest throughout the years compared to the other months. There are a few high temperatures in October throughout the years as well and this is typically seen when the August and September have some of the highest temperatures compared to other years. The plot shows that the average maximum temperature from April to May has been between -5.1 degrees Celsius in November of 2018 and 36.8 degrees Celsius in August of 2006.

Visualizing Pollution from 2002-2020

Visualizing Pollution by Year

I'm interested in finding the trends of the mean pollution levels from 2002 to 2020 by year. The units of NO, NO2, O3, and SO2 are the same (ppb) whereas the units for CO are different (ppm). For this reason, these plots have been made with regards to those units.

Figure 4 shows a decreasing pattern for the mean pollution from year 2002 to 2020 for NO by year. We can notice that O3 and SO2 levels remain mostly the same throughout the years. The highest levels of pollution are as follows: 150.3 ppb for NO in 2002, 65 ppb for NO2 in 2005, 65.95833 ppb for O3 in 2003, and 22.79167 ppb for SO2 in 2002.

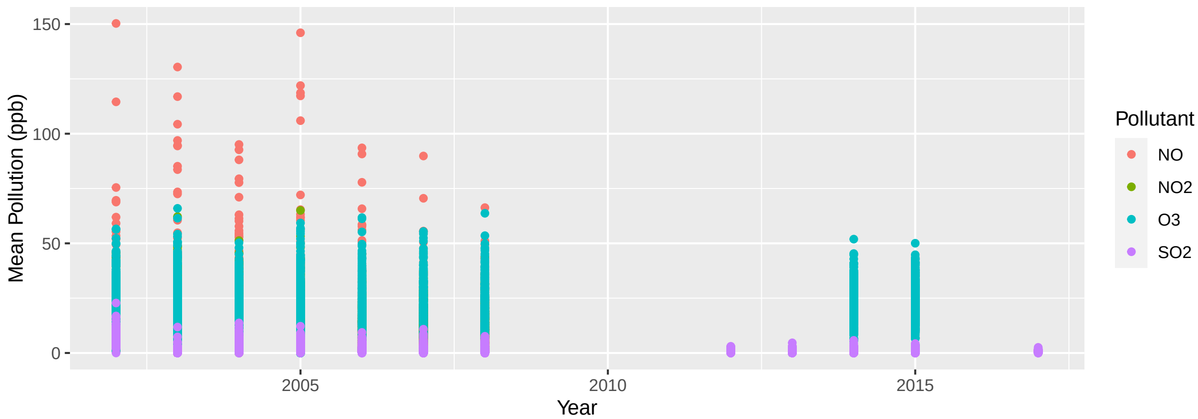

Figure 4: Mean Pollution 2002 to 2020 for NO, NO2, O3, SO2 by Year

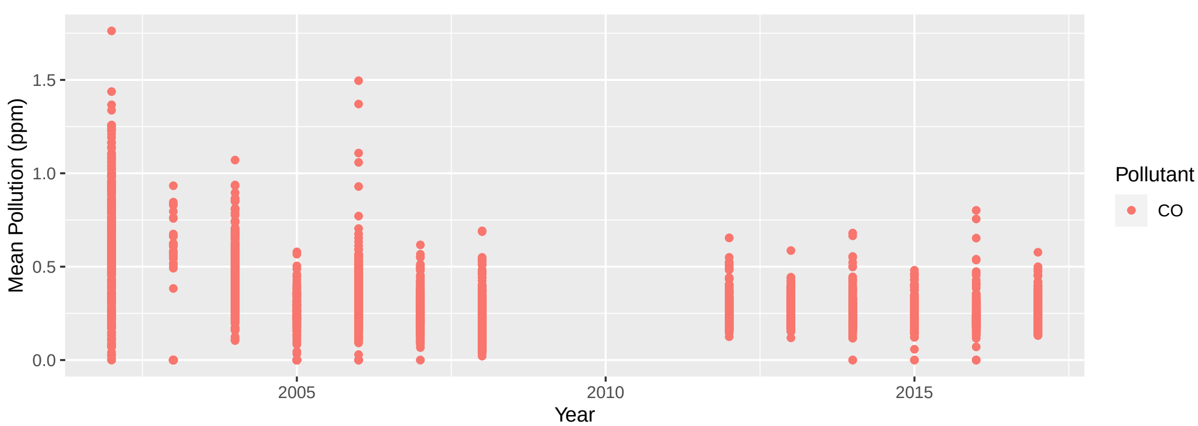

Figure 5: Mean Pollution 2002 to 2020 for CO by Year

Figure 5 shows a decreasing pattern for the mean pollution from year 2002 to 2020 for CO by year. The highest level of pollution for CO is 1.7625 ppm in 2002. We can notice that there is a drop in the pollution levels from 2003 to 2005 but it rises again in 2006. Looking back at figure 3, this was the same year some of the highest temperatures were recorded from August to November.

Visualizing Pollution by Month

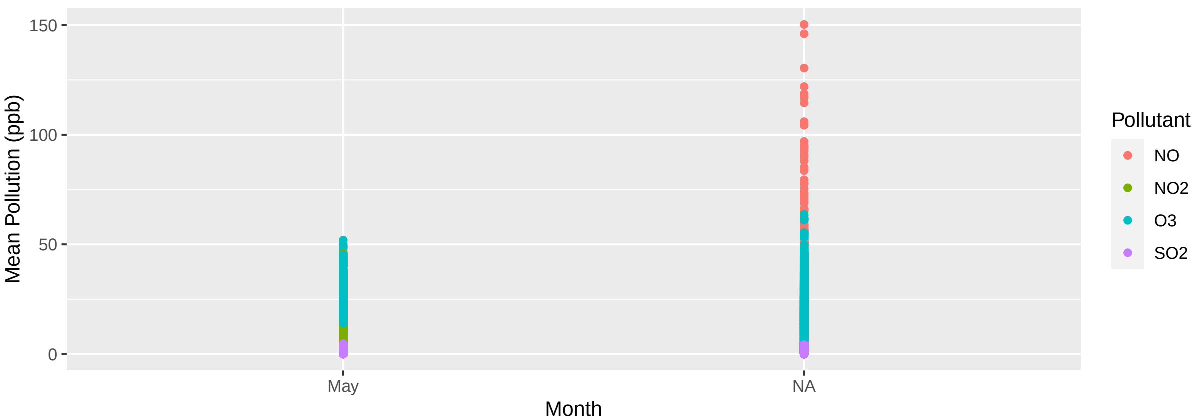

Figure 6: Mean Pollution 2002 to 2020 for NO, NO2, O3, SO2 by Month

Figure 6 shows an ... pattern for the mean pollution from year 2002 to 2020 for NO, NO2, O3, and SO2 by month.

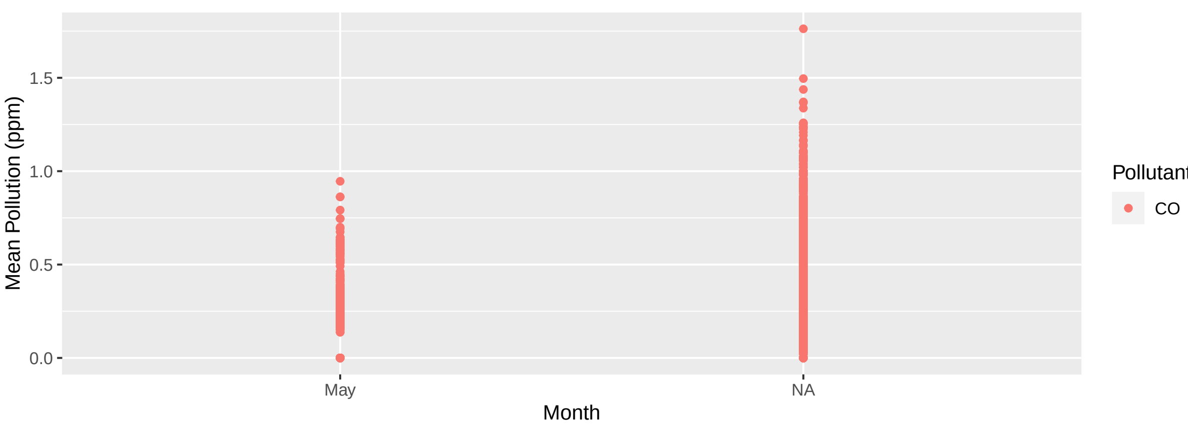

Figure 7: Mean Pollution 2002 to 2020 for CO by Month

Figure 7 shows an ... pattern for the mean pollution from year 2002 to 2020 for CO by month.

Visualizing Pollution and Temperature from 2002-2020

Visualizing Mean Pollution by Temperature, Pollutant and Month

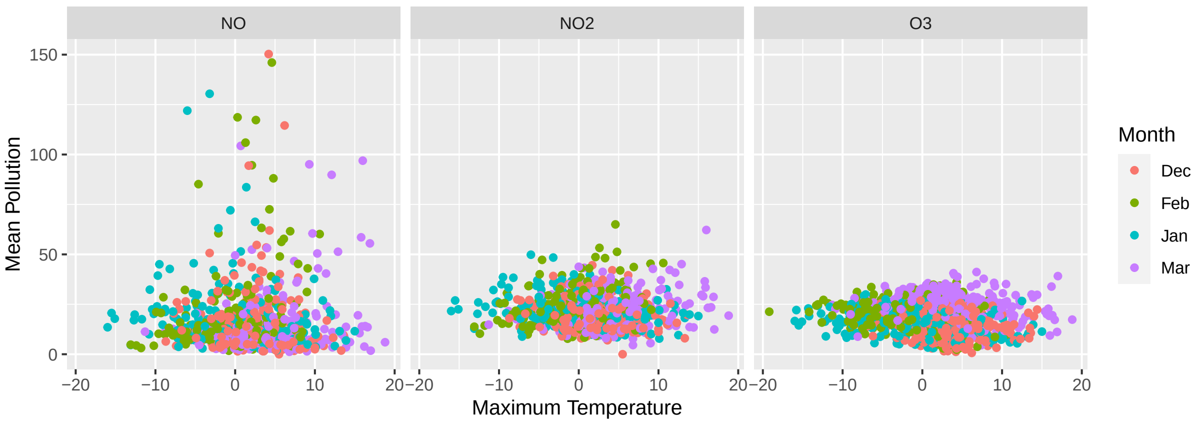

Figure 8: Mean Pollution 2002 to 2020 for NO, NO2, O3 by Month, Maximum Temperature, and Pollutant for December to March

In the above plot…

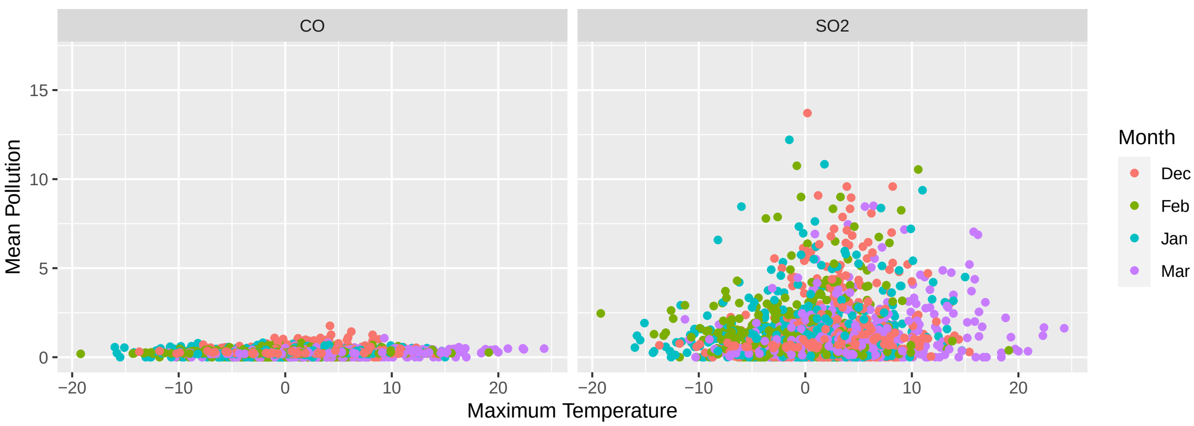

Figure 9: Mean Pollution 2002 to 2020 for CO and SO2 by Month, Maximum Temperature, and Pollutant for December to March

In the above plot…

In the above plot…

In the above plot…

In the above plot…

In the above plot…

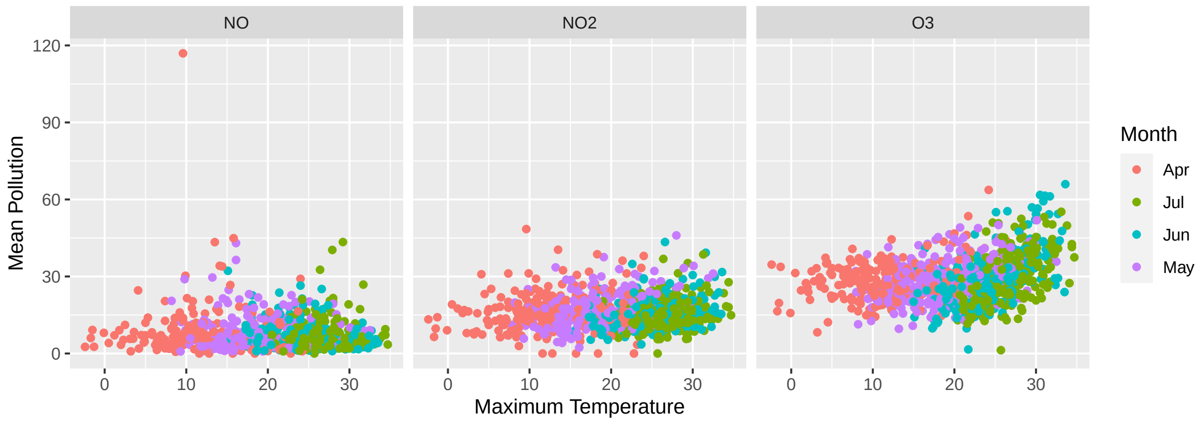

Figure 10: Mean Pollution 2002 to 2020 for NO, NO2, O3 by Month, Maximum Temperature, and Pollutant for April to July

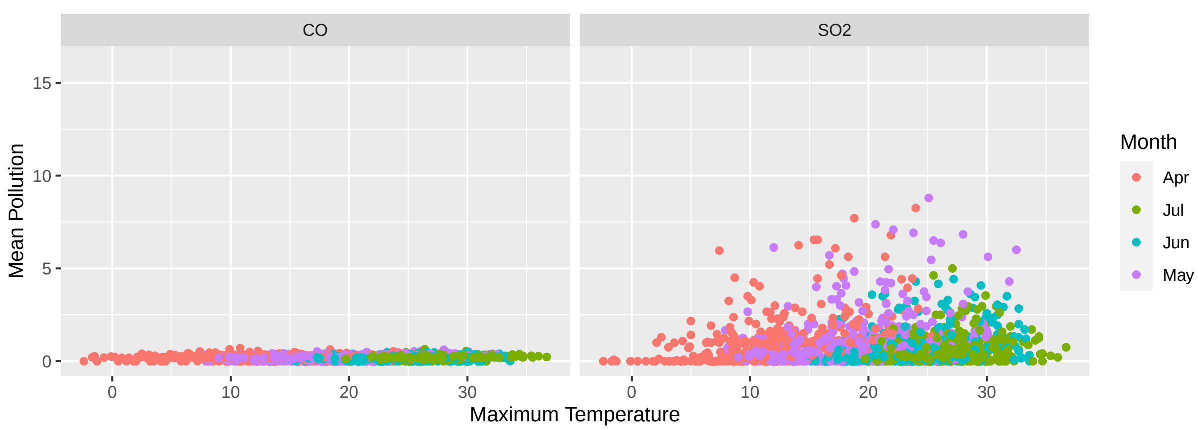

Figure 11: Mean Pollution 2002 to 2020 for CO and SO2 by Month, Maximum Temperature, and Pollutant for April to July

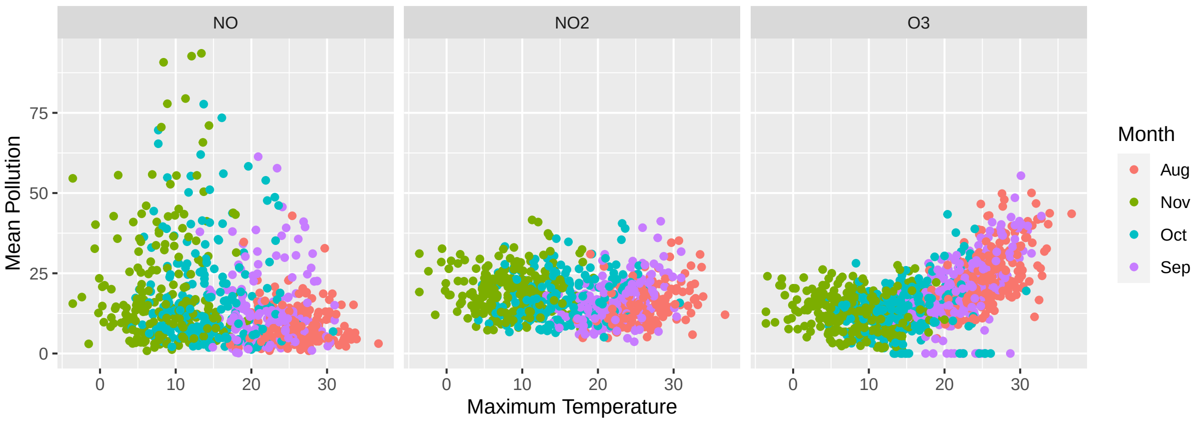

Figure 12: Mean Pollution 2002 to 2020 for NO, NO2, O3 by Month, Maximum Temperature, and Pollutant for August to November

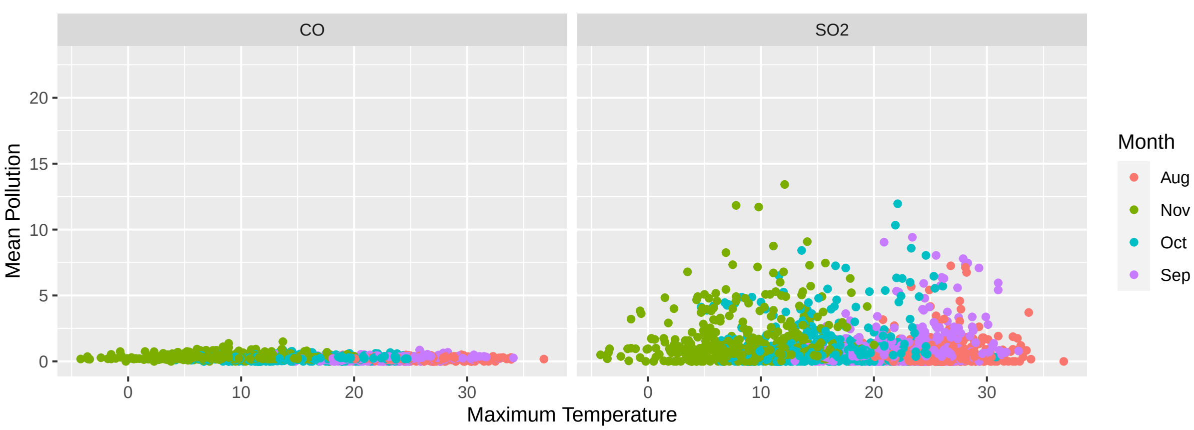

Figure 13: Mean Pollution 2002 to 2020 for CO and SO2 by Month, Maximum Temperature, and Pollutant for August to November

Limitations

In this limitations section, I will focus on the data used and look into why it is not easy to take the implications mentioned above directly.

Missing Pollution Data from 2021-2022

Results and Discussion

Temperature

Pollution

Health

Temperature and Pollution

Temperature and Health

Pollution and Health

Temperature, Pollution and Health

Next Steps

Following on from the previous section, we can make good suggestions on ways to improve the data so that the findings will be less ambiguous and more ethical to use with confidence. First off, we can start by collecting pollution data from 2021 to 2022. It is possible that this data is available but not published yet on the website. In doing this, the data scientist who gets a hold of the data will have the liberty to conduct their analysis with the full data sets for temperature and pollution for 20 years. This will not only ease the job of the data scientist but will also improve the reproducibility of their workflow and finding.