CSC110 Project: The Effects of Climate Change and Ozone Concentration on Torontonian Residents' Respiratory Health

Katherine Luo, Kevin Xia, & Jennifer Cao

January 14, 2022

Problem Description and Research Question

Because of climate change, the temperature of Toronto has been steadily increasing over the last few decades (NASA Global Climate Change). This is due to the increasing population of vehicles, which burns hydrocarbon as fuel and produces CO_2 and water. The high temperatures and pressures inside the combustion engine also causes the reaction between molecular nitrogen and molecular oxygen to combine and produce nitrogen oxide (Climate Central).

Also, volatile organic compounds (VOC) are produced from the emission of unreacted gasolines from a combustion engine as well as from trees. VOC act as a intermediate that takes nitrogen oxide and the oxygen atoms in air to produce nitrogen dioxide and the ozone (Climate Central). Both are pollutants that can produce smog, damage crops, and cause difficulty breathing especially for those with asthma and other respiratory problems. More specifically, the ozone is produced when nitrogen dioxide is exposed to sunlight heat. With an increase in temperature, more ozone is produced, which can become a significant threat towards Torontonian's respiratory health (EPA).

The Province of Ontario and the federal government have 2 air quality standards related to the concentration of ozone: The Ontario 1-hour Ambient Air Quality Criterion (OAAQC) of 80 ppb and the Canadian 8-hour Ambient Air Quality Standard (CAAQS) of 63 ppb (Ontario Ministry of the Environment and Climate Change). They are put in place as an attempt to keep the ozone concentration in control. However, the CAAQS have been exceeded quite a few times in 2020 alone (Ministry of the Environment, Conservation and Parks).

Our research question is: **How would an increase in temperature due to climate change cause a further increase in the concentration of ozone at ground-level, and how would that affect an Torontonian resident's respiratory health?**

Dataset Description

The dataset for the concentration of pollutants come from NAPS, the National Air Pollution Surveillance Program run by Environment and Climate Change Canada. It contains data at different times of the day for different atmospheric gases in different cities in Canada in 2018 (Environment and Climate Change Canada). The datasets were modified by the UofT course CHM135H1 lab 1: "air quality analysis" to include other computed variables as well. This resulted in two datasets based in Toronto. We combined these two datasets and removed any missing values before using. We only used the temperature (in Celsius), the concentration of nitrogen dioxide (in ppb), and the concentration of ozone (in ppb). This is a CSV file we named "toronto_atmospheres.csv".

The dataset for average temperatures in Toronto comes from Environment and Climate Change Canada on the website, toronto.weatherstats.ca (Environment and Climate Change Canada). We only used the date (from January 1st, 1940 to December 11th, 2020) and average daily temperature (in Celsius). This is the 1st and 4th column from the original dataset (we modified it to only include the required two columns sine there were so many variables). This is a CSV file we named "weatherstats_toronto_daily.csv".

Computational Overview

In the `data_wrangling.py` module, we created two dataclasses: `TorontoAtmosphere` and `TorontoTemperatureDaily`. These represent the columns/variables from our datasets. We stored the values of each row of the datasets in a list of their corresponding dataclass type using the functions `read_csv_data1` and `read_csv_data2`. We also created a third dataclass, `TorontoTemperatureYearly`, which stored the converted average daily temperatures to average yearly temperatures. We converted this using the `daily_to_yearly` function.

In the `data_analysis.py` module, we performed linear regression fives times: (1) year vs average temperate, (2) average temperature vs nitrogen dioxide concentration, (3) average temperature vs ozone concentration, (4) year vs predicted nitrogen dioxide concentration, and (5) year vs predicted ozone concentration. We calculated the slope, the y-intercept, the correlation, and the coefficient of determination (R^2) for all of these relationships. This was done with the `simple_linear_regression` function. The slope and the y-intercept create a regression line (i.e., line of best fit) for each relationship. The correlation tells us how strong the relationship between these variables are (i.e., how strong the regression line is) and whether it is a positive or negative association (MathIsFun.com). The coefficient of determination tells us how much variability in the relationship is explained by this regression line (Bloomenthal). Note that these relationships are all based on Toronto temperature and emission.

These are the formulas we used:

Variable	**Source**	**Formula**
slope	Assignment 1 of this course	$\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_1 - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$
y-intercept	Assignment 1 of this course	$\bar{y} - \text{slope} \cdot \bar{x}$
correlation	MathIsFun.com	$\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_1 - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$
R^2	Assignment 1 of this course	$1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

where x_i and y_i are the i^{th} x-coordinate and y-coordinate point respectively for each relationship, n is the number of x-coordinates/y-coordinates, $\bar{x}$ and $\bar{y}$ are the average of x-coordinates and the average of y-coordinates respectively, and $\hat{y}$ is the predicted y_i value for the i^{th} y-coordinate based on the regression line.

In the `plots.py` module, we graphed these five relationships using the `plotly` library. There were a total of seven scatter plots added using the `add_trace` and `Scatter` methods in plotly: one for each of the five relationships, one where we combined temperature vs the nitrogen dioxide and ozone concentrations, and another where we combined the year vs the nitrogen dioxide and ozone concentrations. This "combining" feature of graphing was done using a "visibility" option in plotly. The corresponding regression line(s) were also on these plots using title features in plotly. On the top-right of each graph, the slope, y-intercept, correlation, and coefficient of determination (R^2) of the regression line(s) were displayed rounded to 3 decimal places. We created a drop-down menu that allows a user to switch between these seven scatter plots using the Custom Buttons feature in plotly to create buttons in the format of a dictionary. We added these buttons using the `update_layout` method in plotly (Plotly). This makes it possible to "hide" the other scatter plots from the user and only show the wanted ones based on the drop-down option. Displaying these plots was done with the `display_plots` function. Creating the buttons for the drop-down menu was done with the `create_buttons` functions. These added features of "visibility", special title and axes labels, and Custom Buttons for a drop-down menu are the greater extents of plotly we used.

Changes Since the Proposal

Originally, we planned on using the dataset for how temperature will change over the years from ClimateData.ca. However, the dataset did not go as far back as we had hoped and we had to find a new source. We also planned on scrapping this data using the `scrapy` library. However, the TA marking our proposal recommended that we just download the datasets we needed directly to save time.

We had also planned on graphing only two different relationships before: time vs temperature and year vs ozone concentration. However, we decided to extend our research to other relationships as well (see "Computational Overview" for these relationships). As a result, we did not need to create subplots with the `make_subplots` function in the `plotly` library as we had originally planned. Instead, creating a drop-down menu using the Custom Buttons feature was better for displaying so many relationships in an organized-fashion, without cluttering the screen.

In terms of computations, we decided to include the coefficient of determination (R^2) along with correlation. Before, we thought to exclude it since we've already done it in Assignment 1 of this course. However, we believe that it is still useful for our discussion later in addition to correlation, so we included it.

Instructions

We used two datasets. The datasets are kept at the same level as `main.py` and the other modules, not in subfolders.

The first dataset is from the website *here*. Once on the website, choose the first option: "Climate Daily/Forecast/Sun" and then press "Download". Note that this dataset is updated regularly. In our analysis, we only included dates from "1940-01-01" to "2020-12-11" and the two variables "date" and "avg_temperature" (i.e., the first and fourth column). Our modified dataset can be downloaded through https://send.utoronto.ca/ using the following codes (make sure the file name is "weatherstats_toronto_daily.csv"):

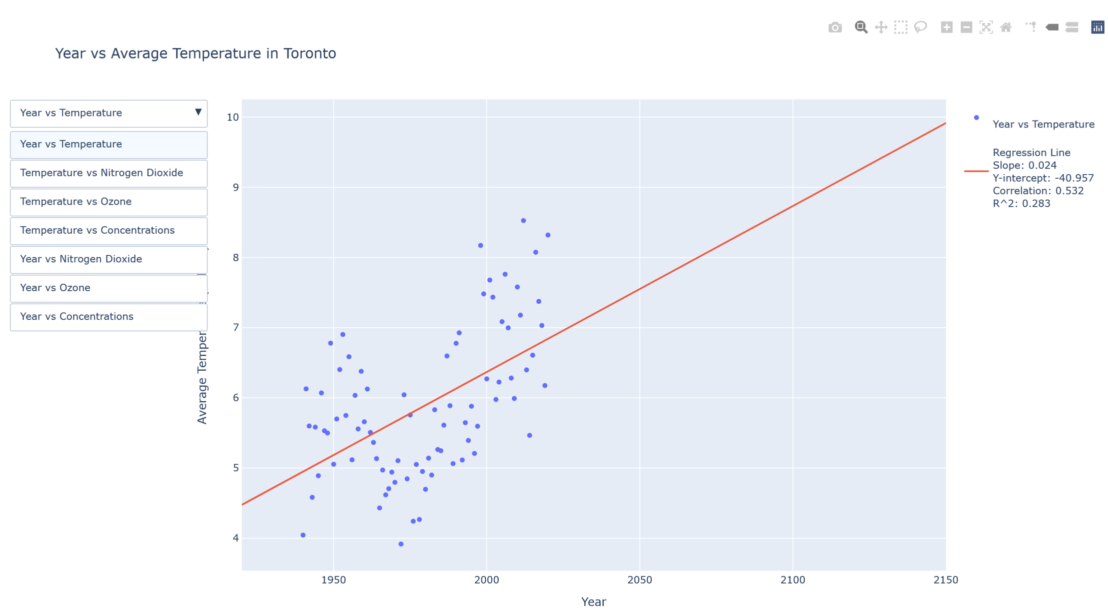
Claim ID: 8q8uJ6ZfksEBAEds   Claim Passcode: nZ4fFREeJ92Q3Gc9

The second dataset is a combination of datasets from the website *here*. It was acquired by the UofT course CHM135H1 lab 1: "air quality analysis" and modified by them to include other variables computed based on the ones from that website. Although the website contains large datasets, we only focused on Toronto. The course had two separate datasets, which we combined into one and removed any missing values before doing the data wrangling. Our modified dataset can be found using the same codes as listed above (make sure the name of the file is "toronto_atmospheres.csv").

After running the `main.py` file, a webpage should pop-up displaying a scatter plot of year vs average temperature in Toronto. On the top-left, there is a drop-down menu, which allows a user to navigate to different scatter plots. On the top-right, there are the values for the slope, y-intercept, correlation, and coefficient of determination (R^2) for the regression line of the particular relationship being shown. It is also possible to press on legend options in this top-right section to hide the line of best fit or the scatter plot points in the graph.

The drop-down menu consists of seven options for the seven possible scatter plots to view discussed in the "Computational Overview": (1) year vs average temperate, (2) average temperature vs nitrogen dioxide concentration, (3) average temperature vs ozone concentration, (4) average temperature vs both those concentrations, (5) year vs predicted nitrogen dioxide concentration, (6) year vs predicted ozone concentration, and (7) year vs both predicted concentrations.

Below is a screenshot of what should pop-up as well as the seven possible drop-down options:



Discussion

The results of the simple linear regression data analysis for each of the five relationships discussed in "Computational Overview" are in the table below. Recall that x_i is the i^{th} x-coordinate and $\hat{y}_i$ is the predicted y_i value for the i^{th} y-coordinate based on the line of best fit.

Relationship	**Slope**	**Y-intercept**	**Correlation**	R^2	**Regression Line**
Year vs Average Temperature	0.024	-40.957	0.532	0.283	$\hat{y}_i = 0.024x_i - 40.957$
Average Temperature vs Nitrogen Dioxide Concentration	-0.345	13.169	-0.612	0.374	$\hat{y}_i = -0.345x_i + 13.169$
Average Temperature vs Ozone Concentration	0.477	24.566	0.548	0.301	$\hat{y}_i = 0.477x_i + 24.566$
Year vs Predicted Nitrogen Dioxide Concentration	-0.008	27.311	-1.0	1.0	$\hat{y}_i = -0.008x_i + 27.311$
Year vs Predicted Ozone Concentration	0.011	5.039	1.0	1.0	$\hat{y}_i = 0.011x_i + 5.039$

A positive slope and positive correlation mean that the association between the relationships are positive and a negative slope and negative correlation mean that the association between the relationships are negative. The magnitude of correlation ranges from 0 to 1, inclusive, with 0 meaning no relationship between the x and y variables and 1 meaning a perfect linear relationship (MathIsFun.com). Note that the last two relationships in the table show a perfect linear relationship because it is just the predicted concentrations for each based on the regression lines. Generally, the first three relationships have a good correlation, meaning the line of best fit for each show moderate to good strength in explaining the association between the x and y variables (MathIsFun.com).

The coefficient of determination (R^2) ranges from 0 to 1, inclusive, with values close to 0 indicating very little variability is explained by the regression line and values close to 1 indicating most of the variability is explained by the regression line (Bloomenthal). Note again that the last two relationships have a R^2 of 1.0 because they are the predicted concentrations based on the regression lines themselves. For the first three relationships, it seems that little variability is explained by the regression lines. As a result, this is one of the limitations of these regression lines and this study. A possible reason for such a low R^2 value may be because ozone concentration also depends on other factors besides temperature. One of the biggest factors would be car emission.

As observed in the year vs average temperature graph and the positive slope and correlation in the table, the mean temperature of Toronto is showing an increase trend. We know that this is due to climate change (NASA Global Climate Change).

From the average temperature vs ozone concentration graph and the positive slope and correlation in the table, we can see how the ozone concentration increases as the average temperature increases. Furthermore, from the temperature vs NO_2 concentration graph and the negative slope and correlation in the table, we can see that as temperature increases, NO_2 concentration decreases. Thus, there is an inverse relationship between the concentration of NO_2 and the concentration of ozone (O_3). This inverse relationship is due to the conversion of the 2 chemical species via the following endothermic reaction:

$$NO_2 + photon \longrightarrow NO + O$$
$$O + O_2 \longrightarrow O_3$$

As we can see, sunlight exposure will provide proton for the first reaction and produce more O atoms. The O atoms will then combine with the diatomic oxygen in air to form ozone (as equation 2 shows). Sunlight exposure is positively correlated with temperature so as temperature increases, the ozone concentration will also increase (Shapley).

With this knowledge as well as the fact that the temperature is showing an increasing trend as time proceeds (in years), we combined the year vs average temperature graph and average temperature vs NO_2/ozone concentration graph to create the year vs NO_2/ozone concentration relationships we seen in the last two rows of the above table. The decrease trend of the year vs NO_2 concentration is evidence backing up the increase trend of year vs ozone concentration. However, how does this help with determining the effects of respiratory health?

Studies have shown that ozone concentration is significantly associated with the number of individuals suffering respiratory diseases and deaths. This is because inhalation of ozone can cause lung malfunctions, which can result in respiratory diseases such as chronic lower respiratory diseases (COPD), asthma, and many more (Jerrett). We will take COPD and elderly asthma as our example for the purpose of this project.

As the paper "The short-term effects of air pollution on hospital admissions in four Australian cities" suggested, as the ozone concentration increase by 1 ppb, the related risk of COPD admissions and elderly asthma admissions to hospital increase by 1.0014 (Simpson). Australia and Canada have similar healthcare qualities (both are first-world country with public free healthcare, etc.) so the related risks should be similar.

Based on the statistics provided by "Chronic obstructive pulmonary disease (COPD), 35 years and over" and "Asthma, by age group" at statcan.gc.ca, in 2018, there were 329,300 COPD patients in Ontario aged 35 and over, and 189,800 asthma patients in Ontario aged 65 and over (Statistics Canada).

From our dataset, we calculated the average of ozone concentration in 2018 to be 28.252 ppb. Therefore, using the regression line of the year vs ozone concentration graph, we can predict the average ozone concentration of 2150 to be 29.29 ppb, which is approximately 1 ppb greater than 2018's average ozone concentration. Using the related risk of 1.0014, we would expect approximately 549 more patients with COPD and approximately 266 more patients with elder asthma in 2150.

These might seem like small numbers, but these estimates are limited. They both underestimate and assume climate change will be a positive linear association and that ozone emission will be constant. In reality, temperature is likely to increase faster as time goes on (Mulkern), and ozone emission is unlikely to stay constant as the number of Toronto residents increase, as it has been for the past 15 years (as the population increase there are more people driving) (Data Commons). Therefore, there may be much more COPD and elderly asthma patients than what we predicted.

Furthermore, there are some additional limitations. As stated in the previous paragraph, because we are assuming climate change will be linearly increasing and ozone emission will be constant, we are greatly underestimating the number of patients. If we had more precise functions for how temperature will increase as times goes on and how does ozone emission increase with population size, we can attain a more accurate prediction for the additional number of COPD and elderly asthma patients.

Also, our predictions are only based on extrapolations of the graph. If we could include more factors that affects climate change and ozone emission (e.g. trees actually produce a compound that is converted to ozone), we can increase the accuracy of our prediction.

In addition, we couldn't find any datasets on the medical progress for COPD and asthma. If there were medical precautions that prevents individuals from contracting such diseases before 2150 (which is very likely), then we would be expecting much less COPD and asthma patients.

Therefore, some further exploration we could do is to explore how climate change would impact the ozone concentration in other cities and in other countries, not just Toronto. We could study how an increase in ozone concentration would impact the respiratory health of the residents elsewhere as well. In particular, we would like to explore this impact on the resident's respiratory health in developing countries, where there is less sufficient medical resource and less methods to reduce emission. What will the number of additional patients be? What would be the additional deaths due to COPD and elderly asthma be, taking in account the less sufficient medical resource? In addition, we could research what the most effective methods are for reducing ozone emissions and slowing down climate change as a whole so that there are less COPD and asthma patients.

References

Bloomenthal, Andrew. "How the Coefficient of Determination Works." *Investopedia*, Investopedia, 14 July 2020, www.investopedia.com/terms/c/coefficient-of-determination.asp.

Climate Central. "Explainer: How Ground-Level Ozone Is Formed." *Climate Central*, Climate Central, 31 July 2019, www.climatecentral.org/gallery/graphics/explainer-how-ground-level-ozone-is-formed.

CSC110 Department. *CSC110 Fall 2020 Assignment 1: Data and Functions*, CSC110 Department, 2020, www.teach.cs.toronto.edu/∼csc110y/fall/assignments/a1/handout/.

Data Commons. "Toronto." *Data Commons*, Data Commons, 2017, https://datacommons.org/place/wikidataId/Q172

Environment and Climate Change Canada. "Data Download for Toronto." *Weather Dashboard for Toronto*, Environment and Climate Change Canada, 2020, https://toronto.weatherstats.ca/download.html.

Environment and Climate Change Canada. "HourlyData-DonneesHoraires." *Environment and Climate Change Canada Data*, 21 Nov. 2018, data.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/Data-Donnees/2018/ContinuousData-DonneesContinu/HourlyData-DonneesHoraires/?lang=en.

EPA, Environmental Protection Agency. "Health Effects of Ozone in Patients with Asthma and Other Chronic Respiratory Disease." *EPA*, Environmental Protection Agency, 23 June 2020, www.epa.gov/ozone-pollution-and-your-patients-health/health-effects-ozone-patients-asthma-and-other-chronic.

Jerrett, Michael, et al. "Long-Term Ozone Exposure and Mortality: NEJM." *New England Journal of Medicine*, 25 June 2009, www.nejm.org/doi/full/10.1056/NEJMoa0803894.

MathIsFun.com. "Correlation." *Math Is Fun Advanced*, MathIsFun.com, 2018, www.mathsisfun.com/data/correlation.html.

Ministry of the Environment, Conservation and Parks. "Canadian Ambient Air Quality Standards." *Ontario*, Ministry of the Environment, Conservation and Parks, 5 June 2019, www.ontario.ca/document/air-quality-ontario-2017-report/canadian-ambient-air-quality-standards.

Mulkern, Anne. "Today's Climate Change Proves Much Faster Than Changes in Past 65 Million Years." *Scientific American.* Scientific American, 2 Aug. 2013, https://www.scientificamerican.com/article/todays-climate-change-proves-much-faster-than-changes-in-past-65-million-years/

NASA Global Climate Change. "The Effects of Climate Change" *Facts*, NASA Global Climate Change, 2020, https://climate.nasa.gov/effects/

Ontario Ministry of the Environment and Climate Change. "Ontario's Ambient Air Quality Criteria - Sorted by Contaminant Name." *Ontario*, Ontario Ministry of the Environment and Climate Change, 30 Apr. 2019, www.ontario.ca/page/ontarios-ambient-air-quality-criteria-sorted-contaminant-name.

Plotly. "Plotly Python Open Source Graphing Library." *Plotly Graphing Libraries*, Plotly, 2020, https://plotly.com/python/.

Shapley, Patricia. Ground Level Ozone Formation, *University of Illinois*, 2012, butane.chem.uiuc.edu/pshapley/genchem2/a12/2.html.

Simpson, Rod, et al. "The short-term effects of air pollution on hospital admissions in four Australian cities." Australian and New Zealand journal of public health vol. 29,3 (2005): 213-21.

Statistics Canada. "Chronic Obstructive Pulmonary Disease (COPD), 35 Years and Over." *Statistics Canada*, Statistics Canada, 14 Dec. 2020, www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310009619.