ANALYSIS OF AIR QUALITY AND ASTHMA

Proposal

By

Mary Robinette

Florida Institute of Technology

CIS 5898 – Projects in Computer Information Systems

for

Dr. Bernard Parenteau

September 12, 2021

Introduction

Air quality affects everyone. Some individuals and segments of the population are more sensitive to poor air quality than others. According to the U.S. Environmental Protection Agency (EPA), air pollution can trigger asthma attacks and can make asthma symptoms worse.

Asthma is a chronic, non-communicable respiratory disease characterized by wheezing, coughing, shortness of breath, and chest tightness. It can be life threating.

Two key pollutants affecting asthma are ozone and particle pollution (EPA). Ozone pollution is found in smog, with ozone levels typically higher on hot summer days. Particle pollution is found in dust, smoke, and haze, with levels independent of season.

Indications of air pollution can sometimes, but not always, be observed. In the U.S., the EPA's tool for communicating air quality to the public is the U.S. Air Quality Index (AQI). The AQI covers five major pollutants—ozone, particle pollution, carbon monoxide, nitrogen dioxide, and sulfur dioxide. These pollutants, regulated by the Clean Air Act, require monitoring by state, local, and tribal air pollution control agencies (Air Now).

AQI values run from 0 to 500, with higher values associated with greater levels of air pollution. The AQI is divided into six categories, corresponding to levels of concern: good, moderate, unhealthy for sensitive groups, unhealthy, very unhealthy, and hazardous.

The EPA's Air Quality System (AQS) is the repository for air pollution data. The data can be used to assess air quality. I will use the data to analyze and visualize air quality data in the U.S. and visualize the relationship between air quality and asthma.

Algorithms/Project Solution

The first part of my project is based on Roger Peng's data analysis: Changes in Fine Particle Air Pollution in the U.S. (Peng, 2020). This study described changes in outdoor air pollution in the U.S. between 1999 and 2012, focused on one pollutant—PM 2.5. PM 2.5 refers to particulate matter that is 2.5 microns or less in diameter. The author used the R programming language to analyze and visualize the EPA raw text file datasets of PM 2.5 for 1999 and 2012. The study hypothesis was that the levels of PM 2.5 in the U.S. decreased from 1999 (the first year of required PM 2.5 monitoring) to 2012 (the most current dataset at the time of the study publication), ostensibly because of passage of the Clean Air Act.

I do not have a copy of the original datasets used by Dr. Peng.

According to the EPA's About AQS data:

> Historical data can change at any time. Many quality assurance review processes are made on an entire year's worth of data, so it might not be until the middle of this year until the final review and changes have been made to last year's data by a submitter. Also, historical monitoring or calculation methods may be found to be problematic and require that older data be changed. Finally, there is no "versioning" or freezing of data in the Data Mart, so if other people may need the data exactly as it was retrieved to verify or continue an analysis, the user must preserve a copy.

The EPA presently provides pre-generated downloadable CSV files of pollutants at the site https://aqs.epa.gov/aqsweb/airdata/download_files.html. Sample (raw) data is accessed via the AQS API https://aqs.epa.gov/aqsweb/documents/data_api.html#meta and requires

registration. As the data has likely changed, I cannot expect my retrievals to correspond exactly with Dr. Peng's. Thus, I will use the EPA's downloadable CSV datasets for PM 2.5 from 1999 and 2012 rather than raw files. I will upload the datasets to the project GitHub repository, preserving them in their retrieved form. I'll use Python to analyze the datasets to see how my results compare to Dr. Peng's.

After the initial comparison review, I will repeat the analysis for changes in PM 2.5 between 1990 and 2020. Basic questions and visualizations include (1) How does the level of PM 2.5 in the U.S. compare between 1999, 2012, and 2020? (2) Which states have the highest and lowest levels of PM 2.5?

Next, I will download and analyze the ozone datasets years 1990, 2012, and 2020. The datasets will be uploaded to the project GitHub repository in their retrieved form. Basic questions and visualizations include (1) How does the level of ozone in the U.S. compare between 1999, 2012, and 2020? (2) Which states have the highest and lowest levels of ozone?

Finally, I will download the annual AQI by county for years 1990, 2012, and 2020. The datasets will be uploaded to the project GitHub repository in their retrieved form. Basic questions and visualizations include (1) Which states have reports of hazardous air quality index (AQI)? (2) Which states have reports of very unhealthy AQI? (3) Which states have reports of unhealthy AQI? (4) Which states have reports of unhealthy for sensitive individuals AQI? (4) What does a map of AQI look like?

In the second part of my project, I will analyze the relationship between air quality and asthma. My primary dataset for asthma comes from the Centers for Disease Control and Prevention. Additional datasets for states having high reported incidence of asthma will be

located. Basic questions include (1) What does the reported incidence of asthma look like by state? (2) Which states have the highest and lowest reported incidence of asthma? (3) What are ways to visualize the relationship between air quality and asthma?

Implementation

Technologies and resources used to implement the project:

- The project will be implemented on an HP OMEN laptop PC with a 64-bit operating system and x64-based processor running Microsoft Windows 10 Home Edition, Version 20H2.
- Anaconda open-source Individual Edition (www.anaconda.com) for Windows was downloaded and installed. The Anaconda installation contained Python 3.8.8, the programming language that will be used along with open-source Python libraries NumPy, Pandas, and Matplotlib for data analysis and visualization. NumPy (numpy.org) enables numerical computing with Python. Pandas (pandas.pydata.org) is a data analysis and manipulation tool. Matplotlib (www.matplotlib.org) is used to create publication quality visualizations. The Anaconda installation also contained Jupyter Notebooks 6.3.0, which is an interactive computing notebook environment. Jupyter Notebooks will be used to run Python code and to create readable documents.
- Tableau for Students software (www.tableau.com) was downloaded and installed with the intent to use it for data visualization of summarized or otherwise simplified air quality or asthma datasets, time permitting. Tableau is commercial software; the

Tableau desktop version installed is Tableau 2021.3. The student version provides a free one-year academic license to the Tableau professional version.

- Microsoft Office 365 will be used for creating Microsoft Word and Excel documents used in the project.
- GitHub (www.github.com) will be used as a repository for documenting and posting project results.
- The data sources for this project include air quality datasets (multiple years, TBD) from the United States Environmental Protection Agency (EPA) Air Quality System (AQS) database. More specifically, daily summary data for ozone and particulate matter (PM) 2.5 will be analyzed. Separate datasets for each of these components are available by year at https://aqs.epa.gov/aqsweb/airdata/download_files.html#Daily. Annual summary data for AQI by County will also be analyzed. This dataset is available at https://aqs.epa.gov/aqsweb/airdata/download_files.html#Annual.
- The second part of the project looks at an asthma dataset from the Centers for Disease Control and Prevention (CDC). This dataset is available at https://chronicdata.cdc.gov/Chronic-Disease-Indicators/U-S-Chronic-Disease-Indicators-Asthma/us8e-ubyj. Additional asthma datasets from state public health departments (TBD) will be examined for states found to have the highest air quality indices, reported as AQI.

Schedule

TASK	WEEK	START	END
Phase 1 - Research and Proposal Writing			
Select project topic	1	8/23/21	8/29/21
Finalize project/scope	2	8/30/21	9/5/21
DELIVERABLE - project proposal	3	9/6/21	9/12/21
Phase 2 - Design and Implementation			
Update proposal (as required)	4	9/13/21	9/16/21
Set up GitHub Repository for Project	4	9/13/21	9/14/21
Find/download air quality (AQ) datasets	4	9/13/21	9/19/21
Clean AQ datasets	5	9/20/21	9/26/21
DELIVERABLE - First Progress Report	5	9/25/21	9/26/21
Update GitHub Repository	5	9/26/21	9/26/21
Exploratory analysis (EA) of AQ datasets	6	9/27/21	10/3/21
Compare results of AQ EA with case study	6	9/27/21	10/3/21
Create AQ visualizations	7	10/4/21	10/10/21
DELIVERABLE - Progress Report 2	7	10/9/21	10/10/21
Update GitHub Repository	7	10/10/21	10/10/21
Find/download asthma datasets	8	10/11/21	10/17/21
Clean asthma datasets	9	10/18/21	10/24/21
DELIVERABLE - Post-Midterm Progress Report	9	10/23/21	10/24/21

Update GitHub Repository	9	10/24/21	10/24/21
EA of asthma datasets	10	10/25/21	10/31/21
Create asthma visualizations	11	11/1/21	11/7/21
DELIVERABLE - Second Half Progress Report 2	11	11/6/21	11/7/21
Update GitHub Repository	11	11/7/21	11/7/21
Plan for/standardize joining of datasets	12	11/8/21	11/14/21
Join AQ and asthma datasets	13	11/15/21	11/17/21
EA of combined datasets	13	11/18/21	11/21/21
DELIVERABLE - Last Interim Progress Report	13	11/21/21	11/21/21
Update GitHub Repository	13	11/21/21	11/21/21
Create AQ and asthma visualizations	14	11/22/21	11/28/21
DELIVERABLE - Last Interim Progress Report	14	11/28/21	11/28/21
Update GitHub Repository	14	11/28/21	11/28/21
Phase 3 - Final Report/Presentation			
Create Final Project	15	11/29/21	12/5/21
Add Lessons Learned Section	15	11/29/21	12/5/21
Add Appendices (as required)	15	11/29/21	12/5/21
Finalize Project Report and Materials	16	12/6/21	12/10/21
Update GitHub Repository	16	12/10/21	12/10/21
Submit DELIVERABLE - Final Project Report & Materials	16	12/10/21	12/10/21

References

AirNow. (n.d). *Using the Air Quality Index.* AirNow. Retrieved September 10, 2021, from https://www.airnow.gov/aqi/aqi-basics/using-air-quality-index/

Centers for Disease Control and Prevention. *U.S. Chronic Disease Indicators: Asthma.* [Data set]. https://chronicdata.cdc.gov/Chronic-Disease-Indicators/U-S-Chronic-Disease-Indicators-Asthma/us8e-ubyj

Dominici, F., Peng, R. D., Barr, C. D., & Bell, M. L. (2010). Protecting human health from air pollution: shifting from a single-pollutant to a multipollutant approach. *Epidemiology* , *21*(2), 187–194. https://doi.org/10.1097/EDE.0b013e3181cc86e8

Environmental Protection Agency. *Annual AQI by County*. [Data set]. EPA. https://aqs.epa.gov/aqsweb/airdata/download_files.html#Annual

Environmental Protection Agency. (n.d). *Asthma and Outdoor Air Pollution, EPA-452-F-04-002.* Environmental Protection Agency. Retrieved September 10, 2021, from https://www.airnow.gov/sites/default/files/2018-03/asthma-flyer.pdf

Environmental Protection Agency. *Daily Ozone*. [Data set]. EPA. https://aqs.epa.gov/aqsweb/airdata/download_files.html#Daily

Environmental Protection Agency. *Daily PM2.5*. [Data set]. EPA. https://aqs.epa.gov/aqsweb/airdata/download_files.html#Daily

Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019*. Lancet, 396(10258):1204-22.* https://doi.org/10.1016/S0140-6736(20)30925-9

Guarnieri, M., & Balmes, J. R. (2014). Outdoor air pollution and asthma. *Lancet*, *383*(9928), 1581–1592. https://doi.org/10.1016/S0140-6736(14)60617-6

Keet, C. A., Keller, J. P., & Peng, R. D. (2018) Long-Term Coarse Particulate Matter Exposure Is Associated with Asthma among Children in Medicaid. *Am J Respir Crit Care Med, 197(6), 737-746.* https://doi.org/10.1164/rccm.201706-1267OC

Peng, R. D. (2020). Data Analysis Case Study: Changes in Fine Particle Air Pollution in the U.S. In *R programming for data science*. Leanpub. https://leanpub.com/rprogramming

World Health Organization. (n.d.). *Asthma Factsheet*. World Health Organization. Retrieved September 11, 2021, from https://www.who.int/news-room/fact-sheets/detail/asthma.