Activity: Seeking Patterns in Ground-Level Ozone Pollution and Temperature through an Exploratory Data Analysis

Arlene M. Fiore, Olivia E. Clifton, and Michael J. Passow

Goals:
> Gain insights into the seasonal and year-to-year variations in ground-level ozone pollution and temperature
> Enhance skills in analyzing data sets and constructing graphs that describe these relationships

Introduction
This activity is created as part of the “Air Quality-Pollution-Vegetation Interactions” workshop presented by Arlene Fiore and Olivia Clifton for the Earth2Class Workshops for Teachers at the Lamont-Doherty Earth Observatory of Columbia University.

Before you begin these activities, you may choose to view the slideshows explaining the science concepts involved. View the Introductory Slideshow and the Workshop Presentation (also accessible through this Fiore Atmospheric Chemistry Group link.).

We will use ozone and temperature measurements collected at a location in Pennsylvania from 1987 – 2014 to create charts and analyze them to identify patterns and trends. These data have been used in scientific research, so the activities will help you understand how investigators examine data and present results. You will also build skills necessary to create various types of graphs and charts using Excel spreadsheets. Finally, based on the data patterns and the scientific information provided in the Earth2Class Workshop resources, answer questions and present your conclusions.

About the Data Sets

Source: The data are publically available from the U.S. EPA Clean Air Status and Trends (CASTNET) network of rural ground-level monitoring sites (http://www.epa.gov/castnet). For more information about these data, see Appendix 1.

Ozone values are monthly averages of maximum daily average 8-hour (MDA8) \( \text{O}_3 \) values, calculated from hourly measurements. MDA8 is the metric used by the U.S. EPA to determine compliance with the National Ambient Air Quality Standards (NAAQS).

Note: a value of “-999.99” indicates that there were insufficient measurements to calculate a representative monthly average and thus that month of data is “missing.” Using the values of -999.99 will create misleading patterns in our analysis, so they should be removed from the data sets prior to creating the graphs.

Temperature values are monthly averages recorded in Kelvin. Most users are more familiar with the Celsius (centigrade) and Fahrenheit scales. We have provided the measurements in all three systems in the accompanying spreadsheets.

More details about the measurements and data are provided in Appendix 1.
To Begin in Excel

We provide the data in several formats, so you can select the level of challenge you wish. **Open the accompanying “Ozone-Temperature-Data_Sets” Excel spreadsheet, then save it as a new file by renaming it “Your_Name_data_set.”** This will avoid over-writing the data so you can start again with the original spreadsheet if you wish.

**The instructions below will use the format labelled “Cleaned”_data.** This provides only ozone and temperature in Fahrenheit, and has had “missing” data removed.

For those who want a greater challenge, Appendix 2 provides instructions for using the data presented in the “Original_measurements” and “Processed_data.” With the data in the “Original_measurements” tab, you are able to work with actual data collected at the field station. The “Processed_data” tab provides temperatures already converted to the °C and °F values, and has removed missing data (e.g., instead of having the “-999.99” value, there is just an empty cell).

**Activity 1) Seasonal variability in ozone and temperature**

**Preparing your spreadsheet:**

1. At the bottom of the spreadsheet, open “Sheet 2” by clicking on it. Then re-name this sheet “Seasonal_Patterns.” **SAVE YOUR WORK OFTEN!**

**Part 1.** First, we will create graphs showing ozone patterns over a year

2. Go back to Sheet 1 and copy Cells 2A – M (months) onto Row 1 in sheet 2.
3. Choose one year from the early 1990s that has a full 12 months of data (no missing cells). Copy Cells A – M for that year “onto Row 2 of your Seasonal Patterns” sheet.
   - Put the year in Cell 1A (next to January)
   - Change Cell 2A from the year to “Ozone.”
4. **Follow the steps of your Excel program** to make a “Line with markers” chart to show the pattern of ozone values (y-axis) vs. month (x-axis).
   - Right-click on the chart title to label it “Ozone – Your Year”
   - Place the chart about 10 rows below the table.

**It might help helpful to copy the tables and charts you make to a new Word document in order to stay organized.**

Q1. Describe what this pattern shows about ozone during this year. In which season is it highest? Lowest?

5. Go down to Row 6 and repeat the steps above for another year sometime between 2005 and 2014. Place the charts next to each other.

Q2. Compare and contrast the chart patterns. What might explain differences?
Q3. What new questions can you create to investigate based on these patterns? What would be your next steps in order to address/answer these questions?

Part 2. Now we will look at temperature patterns.

6. First, we will temporarily copy the Months in Row 1 again into the cells in Row 3.
7. Return to the Data Set (Sheet 1) and copy the temperature data (Cells N – Y) for the same year in Row 4 of your “Seasonal Patterns” sheet.
   Label Cell 4A “Temp.”
8. Follow the steps of your Excel program to make a Line with Markers chart that shows the pattern of temperature values (y-axis) vs. month (x-axis).
   Drag this chart and place it beneath the Ozone chart.
   Add the year to the title
9. Repeat these steps for your other selected year.

Q4. Compare and contrast the temperature patterns for these years.

Part 3. We will now compare the ozone and temperature patterns

10. Delete Rows 3 and 8 (the second row of months) to place the temperature values directly below the ozone value.
11. Highlight all the values in Rows 1 – 3 and make another ‘Line with Markers” chart. Click-and-drag it below the other charts for this year.

12. When you first look at it, both lines appear to have the same y-axis. You can create y-axes on both sides To do this:
    > place the arrow on top of one of the temperature markers
    > right-click on one of the temperature values
    > select “Format Data Series” and “Secondary Axis,” then “Close.”
13. Repeat these steps for your other selected year.

Q5. Describe the patterns you see when you compare the ozone and temperature graphs.

Q6. Based on what we learned about how ozone is produced in the atmosphere, discuss one reason for the seasonal timing of the surface ozone maximum at this site in Pennsylvania.

Q7. What new questions might you investigate based on these patterns? What would be your next steps?
Part 4. Making “Scatter Charts” and statistical comparisons

So far, much of what we have done may be familiar to you. Now, we will use Excel features that you may not have used before, but which scientists often use to gain additional insights from data sets. We will construct “scatter charts” to illustrate the patterns that emerge when we pair the ozone value for a given time period with the temperature value for the same time period. We can use Excel tools to indicate if there is a statistical relationship between ozone and temperature.

Before you begin, you have to make sure that temperature is on the x-axis and ozone is on the y-axis. In Excel, the top row of data is by default the x-axis.

> One simple way to do this is to place the arrow over the “3” in Row 3 and right-click “Cut.”
> Move the arrow up to the “2” and right-click “Insert cut cells.”
> Do the same for the other year’s data.

14. Highlight the values in Cells 2B – M and 3B – M.
15. Use your Excel program to select “Scatter with only Markers.”
   Click-and-drag the chart to below the others for that year.
   Re-name the title to “Ozone vs Temperature YEAR”.
16. Repeat these steps for your other year’s data.

Q8. Compare the scatter charts with the line charts. What additional information about the patterns (e.g., the patterns from the figures that you created during part 1 and 2) can you find using this type of chart?

You will now use some of the advanced tools in Excel to learn more about these patterns. We will add ‘best-fit’ lines constructed by statistical calculations with these data.

17. Place the arrow over one of the points in the Scatter Chart and right-click.
   Select “Add Trendline” and at the bottom, also on “Display Equation on chart” and “Display R-squared value on chart.”

In your Algebra class, you may have learned that a line can be represented by the equation $y = mx + b$. If you look closely, you will see that the equations have this form.

The equation indicates your “m” (this is your slope) and your “b” (your y-intercept) values. Your y-intercept is the value of ozone when temperature is zero.

The slope of the line indicates how ozone changes with the change in temperature. There are two parts of the slope that you should pay attention to. One is whether the slope negative or positive. If the slope is positive, ozone increases with an increase in temperature, and if the slope is negative, ozone decreases with an increase in temperature. The second is the magnitude of the slope – is the slope a large or small number compared to the temperature value? The magnitude of the slope tells you about how sensitive ozone is to temperature.
For example, if the magnitude of the slope is very small as compared to the value of temperature, then ozone will not change a lot with a change in temperature.

The \( R^2 \) value provides an estimate of the ability of the x-variable to predict the y-variable. The \( R^2 \) value would tell you whether monthly changes in temperature a good predictor of monthly changes in ozone?

Q9. What additional understanding do you gain from utilizing the slopes, equations, and \( R^2 \) values?

**Summary of Seasonal Variability**

Write a brief summary of what you have learned about how ground-level ozone varies by season over the eastern United States, including: importance for making measurements, observational methods, patterns and their possible causes. Also provide two questions that you could investigate in further research.

**Activity 2) Inter-annual variability in ozone and temperature**

Next, we will consider how ozone and temperature compare from year to year (scientists refer to year-to-year differences as “inter-annual variability”). We will look at ozone during July, when ozone levels are usually highest.

Prepare your spreadsheet by clicking on “Sheet 3” and re-naming it “Inter-annual_patterns.”

1. Return to your Data Set (sheet 1) and select the Year values in Cells A2 – 30. Paste these in Column A on your new spreadsheet.
2. Now copy the July ozone column (Cells H2 – 30) in Column B on that spreadsheet.
3. Label the top of Column A (Cell A1) “Year” and Cell B1 “Ozone.”

Q10. During which years in the period from 1988 to 2014 do the highest and lowest ozone values occur? What is the range (e.g., what is the lowest value of ozone and how does it compare with the highest value of ozone)? In which years do the highest and lowest values of ozone occur? What does looking at the range of ozone values during 1998 to 2014 tell you about “inter-annual variability” of ozone during July in Pennsylvania?

2. Highlight the July ozone values and make a line chart. Place it below the table.

Q11. What causes the gap in the line? Does it greatly change your ability to see the overall pattern?
Q12. What additional information do you now have about long-term (i.e., over the 26 years) changes in ozone?

3. Now we will add the July temperature values for this period.
   > Copy the July temperature column (Cells U2-30) in Column C.
   > Label Cell C1 “Temp.”

Q10. When were the highest and lowest values? What is the range of temperatures over the 1988-2014?

4. Highlight the temperature values and create a line chart. Place it next to the ozone chart below the table.

Q11. What might you conclude about ozone and temperature patterns by looking only at these charts?

   Let’s add some of the advanced Excel tools you used before to learn more about these patterns.

5. Highlight both the year, ozone, and temperature values to make a “Scatter Chart with Markers”. You may be able to place it to the right of the table above the other charts.
   > Right-click on an ozone marker, select “Format Data Series” and “Secondary axis.” Note that the left-hand axis is for Temp and the right-hand axis is for ozone.

Q12. What are the most obvious patterns (e.g., in terms of the relationship between temperature and ozone) as illustrated by this chart?

6. Now click on temp and ozone data points and select “Add Trendline.”

Q13. What additional information do you now have about the relationship between year-to-year changes in average July ground-level ozone and temperature? What do the slope of the best-fit line tells us?

Q14. Are year-to-year changes in July average temperatures as good of a predictor of year-to-year changes in July average ozone as we saw for monthly temperature versus monthly ozone in Q5? What might explain similarities and differences?

**Activity 3. Comparing “Before-and-After” Emission Controls**

Since the late 1990s, new laws have phased in controls on emission of ozone precursors (chemicals that create ozone) in the eastern United States. These affect power plants and motor vehicles. Let’s explore the impact that these emission changes may have had on the observed relationship between ozone and temperature. [An extra:
To see the decline in a key precursor, nitrogen dioxide (NO$_2$), check out the animation on this NASA website: [http://airquality.gsfc.nasa.gov/].

Open a new spreadsheet and re-name it “Emission_controls.”

1. As you did for your “Seasonal_patterns” spreadsheet, copy the temperature and ozone data for 1988 and 2013 onto the “Emissions_control” spreadsheet. Make line graphs showing the patterns for each year.

Q15. How similar are the temperatures? How similar are the ozone levels? Comment on the apparent impacts of controls on emission of ozone precursors implemented between these time periods.

Now open another spreadsheet and rename it Time_period-comparisons

   > Copy the July “Year,” and “Ozone” values for 1988 – 2001 in Columns A-B.
   > Copy the “Ozone” values for 2002 – 2014 in Columns C.
   > Make line graphs showing the ozone values for these time periods, and add Trendlines.

Q16. Compare the patterns of the graphs and slopes of the two charts. What can you conclude about the impact of reducing ozone precursor emissions on the relationship between ozone and temperature?

**Enrichment: Changing seasonal cycles**

QE1. Do you notice a shift in the ground-level ozone from a summertime maximum to a broader spring-summer maximum?

To make this shift clearer, average over all the years falling within each period you selected.

   > create an empty row below the lines containing individual years monthly averages
   > go to the Jan cell in the empty row and type ‘average(’. Then use your cursor to highlight the cells you want to average over. Then type ‘)’.
   > Do this for each month to create 12 multi-year monthly averages.
   > Now plot the months versus these multi-year monthly averages.

The shift in the seasonal cycle from a summertime maximum to a broader spring-summer maximum is largely due to reductions in the emissions of gases (specifically nitrogen oxides) that lead to ozone production in the lower atmosphere. These emissions reductions began in the late 1990s and have continued as more stringent air pollution control legislation has been implemented. The seasonal cycle with the broader spring-summer maximum (i.e., the one that is occurring during the later years) is commonly observed at measurement sites in regions that do not have a lot of air pollution. Therefore, the shifting seasonal cycle is another piece of evidence that air over the eastern United States is becoming cleaner.
CONCLUSIONS

Write a 3 – 5 paragraph ‘science-style’ report about the procedures and important concepts learned in these activities.
Appendix 1 Data Set Description

Source. The data from the U.S. EPA Clean Air Status and Trends (CASTNET) network of rural ground-level monitoring sites is publically available at http://www.epa.gov/castnet. For an overview of measurements collected, see http://www3.epa.gov/castnet/docs/CASTNET_Factsheet_2013.pdf. In this activity, we are using measurements recorded at a U.S. EPA Clean Air Status and Trends site in Pennsylvania (Penn State at 40.720902°N, 77.931759°W and 364m elevation; http://www3.epa.gov/castnet/site_pages/PSU106.html) that has measured both ozone and temperature fairly continuously from 1987-2014. Note that as of 2011 this network is part of the EPA’s formal ozone monitoring network (http://www.epa.gov/castnet/ozone).

Measurements. The meteorological measurements including temperature that we use here are generally taken at 2m above the surface, though many were discontinued in 2010. In 2011, all ozone instruments were upgraded to be compliant with requirements for sites used as part of EPA’s compliance monitoring network. These instruments measure the absorption by ozone in the ultraviolet. For more information, check out the CASTNet Annual report: http://www3.epa.gov/castnet/docs/CASTNET/AR2013-main.htm#chapter1top.

Data processing. For ozone, we downloaded hourly data and calculated the maximum daily average 8-hour (MDA8) metric that the EPA uses to assess compliance with its National Ambient Air Quality Standard (http://www3.epa.gov/ttn/naaqs/criteria.html) for ground-level ozone. Specifically, we calculate 24 8-hour averages, one that begins at each hour of the day. We then take the maximum of these 24 values as the MDA8 for that day. We then take the average across all MDA8 values within each month and enter those into the spreadsheet. At each step along the way, we screen the data, requiring at least 75% of the data within a day or within a month to be valid. For months where we do not have sufficient high-quality measurements, we enter a -999.99 into the spreadsheet. For temperature, we downloaded hourly data and find the maximum value each day and then average across all days within each month, applying similar data screening criteria.

Units. The MDA8 ozone data are in parts per billion (ppb), which is $10^9$ times the mole fraction (also called mixing ratio) of ozone in air.

The temperature data were originally measured in units of Kelvin. If preferred, you can convert K to the Celsius scale by subtracting 273.15 from each value. To convert to Fahrenheit, we first subtract 273.15 from each value, then multiply by 9/5 and add 32.
Appendix B Instructions Using the More Complex Data Sets

> Open the PSU106_monthly_ozone_temp_data_1987-2014.xlsx data set and save it as a new Excel spreadsheet by renaming the file. This avoids over-writing the data as you can always start over from the original spreadsheet.

> Also go ahead and rename the first sheet as “Data_Set.” **Save your work (often!)**

> To remove -999.99: Simply delete the -999.99 entries and leave the cell blank.

> To convert temperature from K to C to F: Copy the years and months header column and row to a new place in your spreadsheet. Place your cursor into the cell for the entry of the first month and year in your new blank table. Type ‘=’ to get into equation mode. Then highlight the temperature entry in K in the table provided, which should cause the cell number to appear in the ‘fx’ line at the top, and then type ‘- 273.15’. Then copy this formula to every cell in your new table and it should automatically fill in, converting the data provided in Kelvin to Celsius. To convert to Fahrenheit, we multiply Celsius by 9/5 and add 32. Note that if you are copying cells containing formulas (i.e., they are calculated from values in other cells) to another worksheet, you may see ‘#REF!’ when you paste. If this occurs, right click and choose the option ‘paste values only’.
Appendix 3  Selected Pertinent NGSS DCIs

ESS2.A: Earth’s Materials and Systems
- All Earth processes are the result of energy flowing and matter cycling within and among the planet’s systems. This energy is derived from the sun and Earth’s hot interior. The energy that flows and matter that cycles produce chemical and physical changes in Earth’s materials and living organisms. (MS-ESS2-1)
- The planet’s systems interact over scales that range from microscopic to global in size, and they operate over fractions of a second to billions of years. These interactions have shaped Earth’s history and will determine its future. (MS-ESS2-2)

ESS2.D: Weather and Climate
- Weather and climate are influenced by interactions involving sunlight, the ocean, the atmosphere, ice, landforms, and living things. These interactions vary with latitude, altitude, and local and regional geography, all of which can affect oceanic and atmospheric flow patterns. (MS-ESS2-6)

ESS2.A: Earth Materials and Systems
- Earth’s systems, being dynamic and interacting, cause feedback effects that can increase or decrease the original changes. (HS-ESS2-1),(HS-ESS2-2)
- The geological record shows that changes to global and regional climate can be caused by interactions among changes in the sun’s energy output or Earth’s orbit, tectonic events, ocean circulation, volcanic activity, glaciers, vegetation, and human activities. These changes can occur on a variety of time scales from sudden (e.g., volcanic ash clouds) to intermediate (ice ages) to very long-term tectonic cycles. (HS-ESS2-4)

ESS2.D: Weather and Climate
- The foundation for Earth’s global climate systems is the electromagnetic radiation from the sun, as well as its reflection, absorption, storage, and redistribution among the atmosphere, ocean, and land systems, and this energy’s re-radiation into space. (HS-ESS2-2)(HS-ESS2-4)
- Gradual atmospheric changes were due to plants and other organisms that captured carbon dioxide and released oxygen. (HS-ESS2-6),(HS-ESS2-7)

ESS2.E Biogeology
- The many dynamic and delicate feedbacks between the biosphere and other Earth systems cause a continual co-evolution of Earth’s surface and the life that exists on it. (HS-ESS2-7)

ESS3.B: Natural Hazards
- Natural hazards and other geologic events have shaped the course of human history; [they] have significantly altered the sizes of human populations and have driven human migrations. (HS-ESS3-1)

ESS3.C: Human Impacts on Earth Systems
- The sustainability of human societies and the biodiversity that supports them requires responsible management of natural resources. (HS-ESS3-3)
- Scientists and engineers can make major contributions by developing technologies that produce less pollution and waste and that preclude ecosystem degradation. (HS-ESS3-4)

ESS3.D: Global Climate Change
• Though the magnitudes of human impacts are greater than they have ever been, so too are human abilities to model, predict, and manage current and future impacts. (HS-ESS3-5)

• Through computer simulations and other studies, important discoveries are still being made about how the ocean, the atmosphere, and the biosphere interact and are modified in response to human activities. (HS-ESS3-6)

**ETS1.B: Developing Possible Solutions**

• When evaluating solutions, it is important to take into account a range of constraints, including cost, safety, reliability, and aesthetics, and to consider social, cultural, and environmental impacts. *(secondary to HS-ESS3-2),(secondary HS-ESS3-4)*