JOURNAL OF CHEMICAL EDUCATION

Editors' Choice

pubs.acs.org/jchemeduc

Article

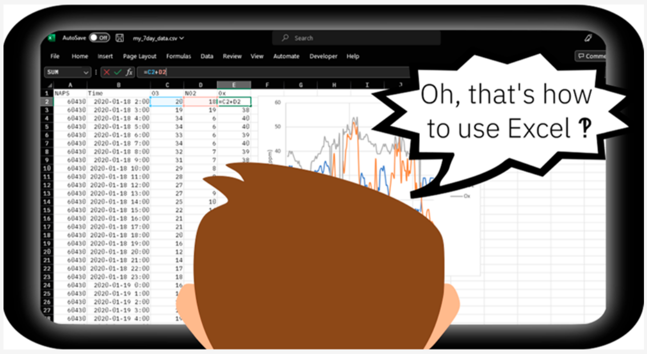
How's the Air Out There? Using a National Air Quality Database to Introduce First Year Students to the Fundamentals of Data Analysis

David Ross Hall and Jessica C. D'eon*

Cite This: *J. Chem. Educ.* 2023, 100, 3410−3418

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Chemistry is increasingly data centric and the undergraduate curriculum needs to adjust to keep up. To address this, we created the *Air Quality Activity*, a new first-year undergraduate activity where students use Microsoft Excel to analyze a unique subset of atmospheric ozone (O_3) and nitrogen dioxide (NO_2) measurements from the Canadian National Air Pollution Surveillance (NAPS) program. Through this activity students develop their numeracy, graphicacy, and proficiency with Excel. Moreover, students are equipped with a foundational approach to data analysis they can leverage throughout their studies. To make this activity possible, we developed an open-source webbook detailing pertinent Excel operations for first-year students, and an interactive web-app for the generation, distribution, and exploration of NAPS data. Students were excited by the analysis of real-world chemical phenomena in comparison to traditional first-year lab exercises and appreciated their acquired Excel skills. The *Air Quality Activity* is readily adaptable for both virtual and in-person implementation, entirely open-source, and readily deployable at any institution wishing to teach data analysis in a chemistry context.

KEYWORDS: *Environmental Chemistry, Interdisciplinary, First-Year Undergraduate/General, Computer Based Learning, Inquiry-Based, Web-Based Learning, Chemometrics, Laboratory Computing/Interfacing*

■ INTRODUCTION

We live in the age of data; from climate models[1] to machine-learning driven organic synthesis,[2] every domain of chemistry is becoming increasingly data-driven. The omnipresence of data has placed data-science, a union of computer science and statistics, at the forefront of every field of research.[3] The popularization of data-science often focuses on exotic statistical analysis (i.e., machine learning), detracting from the fundamental skills of numeracy and graphicacy which refer to the ability to understand, reason, and apply numerical[4] and graphical concepts.[5−7] Time constraints in the undergraduate laboratory mean that students are often left with single replicates that do not allow for an analysis of larger data sets. In addition to logistical restrictions on lab work, students often encounter significant hurdles with the introduction of new software. When data analysis software (i.e., Microsoft Excel) is presented, it is often as a means to solve a specific problem, rather than a tool that students can leverage across their studies.[8] The repercussions are that software meant to help students becomes a source of anxiety and frustration as their focus shifts from understanding subject matter to wrestling with unfamiliar software.[9]

This paper describes an exercise implemented in an introductory general chemistry class, which we call the *Air Quality Activity*, that incorporates a simple introduction to atmospheric chemistry and real atmospheric measurements to explicitly introduce essential data-science and Microsoft Excel skills. This activity originated as a virtual lab in summer 2020, as a means to integrate real data into the undergraduate experience, as opposed to many virtual laboratories that used simulated data.[10] This was accomplished using data generated by Environment and Climate Change Canada's National Air Pollution Surveillance (NAPS) Program which has been recording hourly measurements of ambient ground-level concentrations of several atmospheric pollutants from stations across Canada since 1969.[11] The monitored pollutants include ozone (O_3) and nitrogen dioxide (NO_2), which students used in their analysis. These compounds were chosen because of their unique interdependence, and the impact of their chemistry on the daily lives of students through the issuance of air quality advisories, which in Toronto are governed by the

Received: April 17, 2023
Revised: June 29, 2023
Published: August 11, 2023

Canadian Ambient Air Quality Standards (CAAQS)[12] and the Ontario Ambient Air Quality Criteria (OAAQC).[13]

DEVELOPMENT OF LEARNING RESOURCES

This *Air Quality Activity* was developed for the one-term, first-year, general chemistry course *CHM135 Chemistry: Physical Principles* at the University of Toronto. CHM135 is offered three times per year (fall, winter, and summer) with enrolments that vary from 100 to 2000 students. The class meets for three lecture hours each week over 12-weeks and includes five biweekly laboratories. As of winter 2023, the *Air Quality Activity* has been completed by over 7000 students.

This activity was originally designed as a fully virtual activity and was executed as part of a fully online version of the class wherein the activity was presented to students as a self-guided exercise (summer 2020) then as an exercise done, at least partly, over zoom with support from a teaching assistant (fall 2020 to winter 2022). In summer 2022, with the return of in-person laboratories, a slightly abridged version of the online-only *Air Quality Activity* was introduced as a self-guided activity completed after the first lab period. In all iterations of the activity, lab office hours and a discussion board were available to the students for help and guidance.

Execution of this activity, at the scale imposed by our class sizes, required the development of three learning resources: a brief introduction to atmospheric chemistry, a centralized resource for pertinent Excel operations, and a method to introduce experimental data in formats suitable for students. These resources are discussed below, and copies can be found in the Supporting Information.

Introduction to Atmospheric Chemistry

The cornerstone of this introduction is a 6 min video that walks students through the role of nitrogen oxides (NO_x), volatile organic compounds (VOCs), and ozone in urban air quality.[14] To summarize the video, NO_2 is a secondary pollutant produced from the reaction of VOCs and nitric oxide (NO, see the unbalanced reaction 1) which are primary pollutants emitted during the operation of internal combustion engines.[15]

$$\mathrm{VOC} \rightarrow \mathrm{RO_2^{\bullet}} + \mathrm{NO} \rightarrow \mathrm{RO^{\bullet}} + \mathrm{NO_2} \quad (1)$$

During the day, NO_2 is photolyzed by wavelengths shorter than 424 nm to produce a free oxygen atom (reaction 2) that quickly combines with molecular oxygen (O_2) to generate O_3 (reaction 3), another secondary pollutant. A major loss mechanism for O_3 is via a reaction with NO that regenerates O_2 and NO_2 (reaction 4).

$$\mathrm{NO_2} \xrightarrow{h\nu} \mathrm{NO} + \mathrm{O} \quad (2)$$

$$\mathrm{O} + \mathrm{O_2} \rightarrow \mathrm{O_3} \quad (3)$$

$$\mathrm{O_3} + \mathrm{NO} \rightarrow \mathrm{O_2} + \mathrm{NO_2} \quad (4)$$

Reactions 2−4 form a null cycle and demonstrate the intimate connection between the atmospheric concentrations of NO_2 and O_3. In an urban atmosphere with regular inputs of primary pollutants, O_3 is typically the dominant pollutant observed during the daytime when photon fluxes drive reaction 2, whereas NO_2 becomes the dominant pollutant observed overnight when reaction 2 ceases and reaction 4 dominates. This interdependent diurnal cycle is so intimate that the sum of their concentrations is given the term "odd oxygen" or O_x (eq 5) and is frequently used in discussions of pollutant loadings in the atmosphere.[16]

$$[\mathrm{O_3}] + [\mathrm{NO_2}] = [\mathrm{O}_x] \quad (5)$$

The set of reactions described above is a simplified model of the possible reactions taking place in the atmosphere, and readers interested in a more robust understanding of the chemistry taking place are encouraged to read Chapter 6 *Chemistry of the Troposphere*, in particular 6.8 *Relative Roles of VOC and NO_X in Ozone Formation*, in the book *Atmospheric Chemistry and Physics* by Seinfeld and Pandis[17] or a similar atmospheric chemistry resource. In addition to exploring the relationships described here, the activity also pushed students to explore the gap between simple theoretical models (i.e., reactions 2−4) and the complex chemical systems they attempt to explain, a long-standing goal of science pedagogy.[18]

Excel4Chem Webbook

The first iteration of the *Air Quality Activity* provided students with an *Excel Tip Sheet* containing instructions on how to perform the specific Excel operations required; a copy of the *Tip Sheet* is available in the Supporting Information. This resource has since been expanded into a webbook called *Excel for General Chemistry* or *Excel4Chem* for short,[19] that covers all Excel operations required to complete all of the CHM135 laboratories. The webbook format offers many advantages; it is readily disseminated via a simple URL, and it includes embedded short-animated GIFs of screen recordings. We believe this is an optimal approach for teaching software as students do not need to watch a lengthy video, while the GIFs retain the ease of visualizing mouse-click actions that is lost in written instructions or static images.

The choice was made to tailor the beginning of the webbook to the *Air Quality Activity* so students in CHM135 feel supported to complete the activity independently. Section 1 of the webbook provides all of the necessary background on air quality described above, as well as a downloadable sample data set which anyone outside the CHM135 course can use to complete the activity. Afterward the webbook introduces more general concepts such as advanced functions (e.g., = COUNT(), = IF()) required in subsequent CHM135 laboratories. The webbook is open-source, and available to anyone interested in learning basic data analysis skills in Excel.[19]

Exploring Air Quality Data Shiny App

To streamline the dissemination of data sets and facilitate students' ability to explore a variety of geographical locations and seasons, we created a web-app,[20] written in R and built using the Shiny package for interactive web applications (Figure 1).[21−23] The web-app consists of an interactive map showing the location and local population of every NAPS surveillance station across Canada. Students can select both the station of interest and the time span during that year using simple graphical widgets (e.g., button and drop-down options) and an interactive time-series plot of $[O_3]$, $[NO_2]$, and $[O_x]$, and a corresponding scatter plot of $[O_3]$ vs $[NO_2]$ are automatically generated and displayed. These visualizations are similar to the ones students produce, allowing them to readily interpret the results and incorporate their observations into the report discussion questions. Additional details about the web-app can be found in the Supporting Information.

https://doi.org/10.1021/acs.jchemed.3c00333
J. Chem. Educ. 2023, 100, 3410−3418

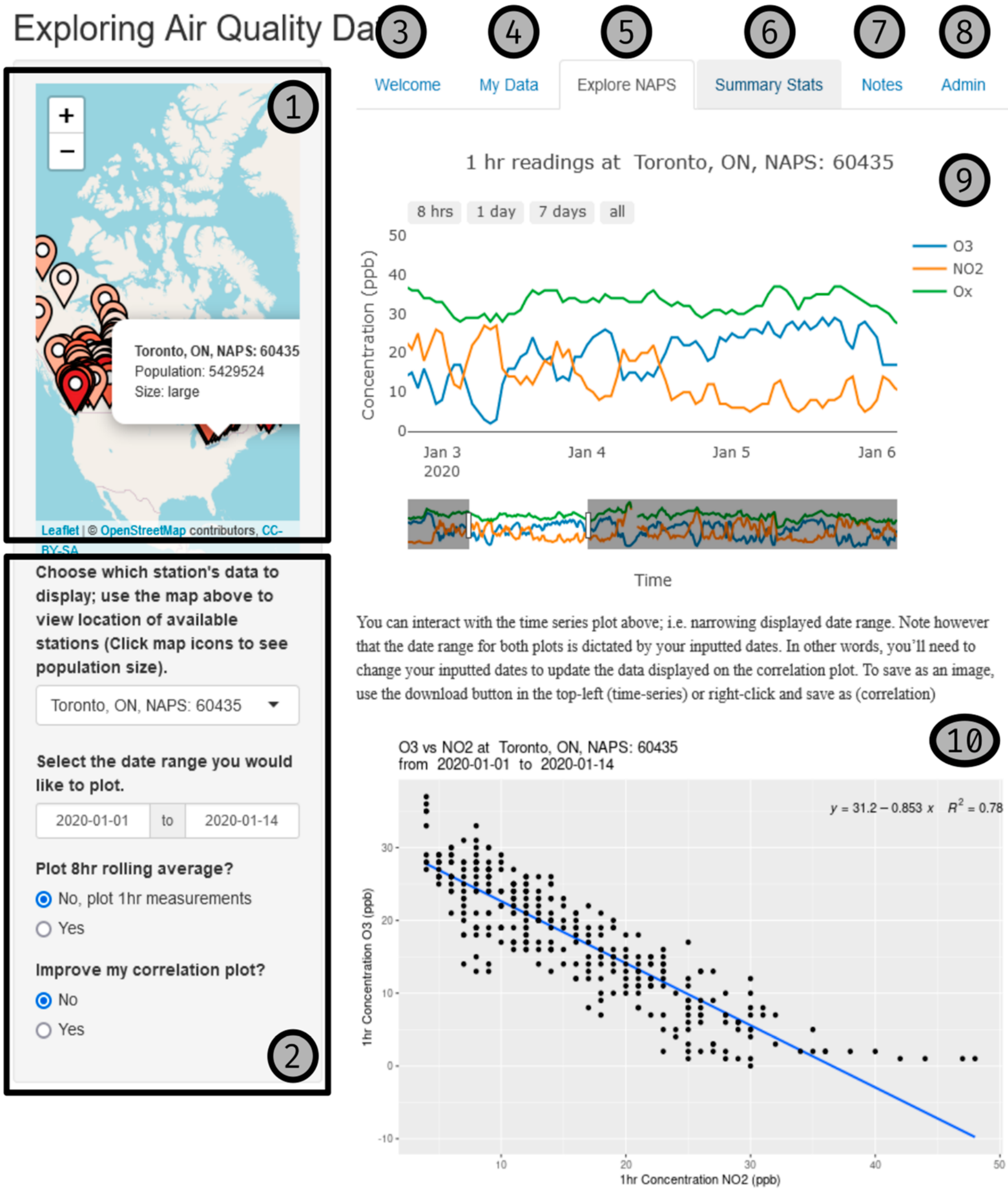


Figure 1. Screenshot of the *Air Quality Activity* web-app interface. (1) Interactive map with all NAPS stations and their local population; (2) Inputs to select station, date range, and plotting options; (3) Splash page tab, contains instructions on using the web-app; (4) "My Data" tab where students download their data set; (5) Tab with plots based on selected inputs; (6) Tab with summary stats based on selected data; (7) Additional notes and sources; (8) Password protected Admin tab for marking; (9) Interactive time-series plot; (10) Scatter plot with equation and *R*-squared.

■ DATA ANALYSIS WORKFLOW AND LEARNING GOALS

A motivating factor for the creation of the *Air Quality Activity* was to equip students with an adaptable, yet consistent, approach to data analysis. The data analysis workflow in this activity is built on the work of Hadley and Grolemund,[24] which is often recognized by experienced researchers as one they have implicitly learned. We hoped that by explicitly introducing it to students at the outset of their undergraduate careers, they will have a better understanding of how they should approach data analysis problems and possess a framework into which they can integrate the new skills they learn as their studies progress. Briefly, the steps of this data-analysis workflow are

1. *Importing* raw data (typically in a .csv format) into software for further analysis, in our case Microsoft Excel.
2. *Cleaning* data, which is the transformation of raw data into data that is ready for further analysis.
3. *Transforming* data using mathematical operations.
4. *Visualizing* data via plots and tables.
5. *Modeling* mathematical relationships between variables.
6. *Communicating* and exploring results.

This data analysis workflow is built into the *Excel4Chem* webbook and used explicitly with the CHM135 students to provide them with a clear framework for why they are performing the relevant Excel operations. Building on this framework there are specific learning goals we would like students to achieve upon completion of this activity:

1. Students should establish a baseline proficiency with Microsoft Excel, which includes:
   a. Installing a working copy of Excel.

https://doi.org/10.1021/acs.jchemed.3c00333
J. Chem. Educ. 2023, 100, 3410−3418

b. Performing simple data manipulation and mathematical operations.

c. Creating plots to visualize numerical data.

2. Students should be able to interpret and discuss their data visualizations in the context of the lab subject matter.
3. Students should possess a qualitative understanding of correlation and simple statistics and be able to interpret results in the chemical context of the lab.
4. Students should be equipped with an adaptable data-analysis workflow they can use to approach future data-analysis problems.

The first goal enumerated here requires students to complete steps 1−4 of the data analysis workflow outlined above. Learning goals 2 and 3 are an application of steps 4 and 5 of the workflow, respectively, and Goal 4 is a synthesis and digestion of the entire data analysis workflow. The success of the *Air Quality Activity* in achieving learning goals 1 to 3 was assessed using students' performance on their graded reports. Learning goal 4 was assessed qualitatively using feedback from student surveys as well as the success students had completing subsequent lab reports using the Excel skills they developed in the *Air Quality Activity*.

■ ACTIVITY OVERVIEW

All iterations of the *Air Quality Activity* start with students receiving a data set that consists of 7 days of hourly O_3 and NO_2 measurements collected in Toronto in the winter (January or February), as the data collected in the winter tends to be more consistent with the introductory material. Data sets are provided to students as a comma-separated values file (.csv), with at least one experimental error (concentration of −999) within the measurement period.

To perform the necessary Excel operations, the .csv file must be saved as an .xlsx file to an appropriate directory. To support all learners, this first step is accompanied by a discussion of file management titled "Getting Setup for Success" in the *Excel4Chem* webbook. This section explicitly outlines how to use folders to organize files and how to use cloud platforms (e.g., OneDrive, Google Drive, Dropbox, etc.) to backup files and access them from multiple devices. This may seem redundant for a generation of students who have grown up using computers, however that is not our experience. Many of the students in our classes were relying on keyword searches of their Google Drives to find files (as opposed to any folder or file structure within the drive) and were not reliably backing up files when they were not working on an online version of the software. Our student body is incredibly (and proudly) diverse and so by establishing these practices across the entire class we ensure all students begin on an equal footing regarding file management and data analysis.

Once students have their data set saved as an .xlsx file, they are guided through the data cleaning component of the activity, which is the transformation of raw data into data that is ready for further analysis. Specific data cleaning exercises include formatting cells to properly display the date and time, and the use of the "find and replace" function to remove the "−999" missing data values. This is a crucial step to emphasize with students early in their University careers, as many believe any data manipulation or transformation is an academic offense.

With their data set now ready for further analysis the students calculate the concentration of O_x (reaction 5) by adding the concentration of O_3 and NO_2 using a relative reference and copying the formula down the column, and then produce a time-series plot of O_3, NO_2, and O_x over the time span of their data set (Figure 2A). Students also produce a correlation plot between O_3 and NO_2, that includes a linear trendline with its associated line-of-best fit equation and R^2 value displayed, to explore the relationship between O_3 and NO_2.

In the virtual iteration of the activity students received a second data set collected in the summer (July or August) in Toronto and performed the operations listed above on both data sets so they could explore seasonal differences in the relationship between O_3 and NO_2. The virtual activity also included several additional mathematical operations. Students calculated a mean and standard deviation for O_3, NO_2, and O_x in both data sets. They also calculated a rolling 8-h mean, also known as a moving average, of the concentration of O_3 using the = AVERAGE() function in Excel with a relative cell reference and then used the = MAX() function to determine if either the CAAQS or OAAQC air quality standards were exceeded in their data set.[12,13] With the return to in-person laboratories the concepts and skills from this cut content were reapportioned within the subsequent lab activities (i.e., calculating means of student's experimental triplicates).

Copies of the data analysis instructions and report sheet questions for both the in-person and virtual-only iterations of the *Air Quality Activity* can be found in the Supporting Information.

■ PEDAGOGICAL OUTCOMES

In the report submitted by the students, each of the data analysis steps outlined in the previous section is accompanied by discussion questions meant to assess the student's ability to interpret and discuss the values and plots produced in relation to the chemistry context of the activity (learning goal 2). Examples of visualizations and responses to these report prompts are provided here as evidence of students learning. Although it is not possible to provide a comprehensive picture of all student responses, overall grades were high on these reports with the course average for the Fall 2020 term (final enrolment of 1722 students) being 83.4% (A−), indicating the majority of students were successful in these tasks.

The first major output of the activity is a time series that plots the concentrations of O_3, NO_2, and O_x over time for the 7 days of the data set (Figure 2A). The first learning goal of this activity can be assessed by visual inspection of the time series plot as the presence of the plot itself speaks to the student's ability to visualize data in Excel; the properly formatted x-axis speaks to their ability to adjust the formatting of those cells into machine-readable date-time formats; the presence of O_x in the time series demonstrates the student was able to complete a mathematical calculation in Excel using a relative reference; and the fact that there are no "−999" concentration values displayed confirms the students removed these, likely using the find and replace function.

After creating the time-series plot in Figure 2A, students were asked the following question: *Looking at your [time series] data set, do O_3 and NO_2 show the diurnal trend (meaning changes that happen over 1 day) that was discussed in the introductory video? Explain your answer*. Visual inspection of plots and graphs is often undervalued by students and this exercise was

https://doi.org/10.1021/acs.jchemed.3c00333
J. Chem. Educ. 2023, 100, 3410−3418

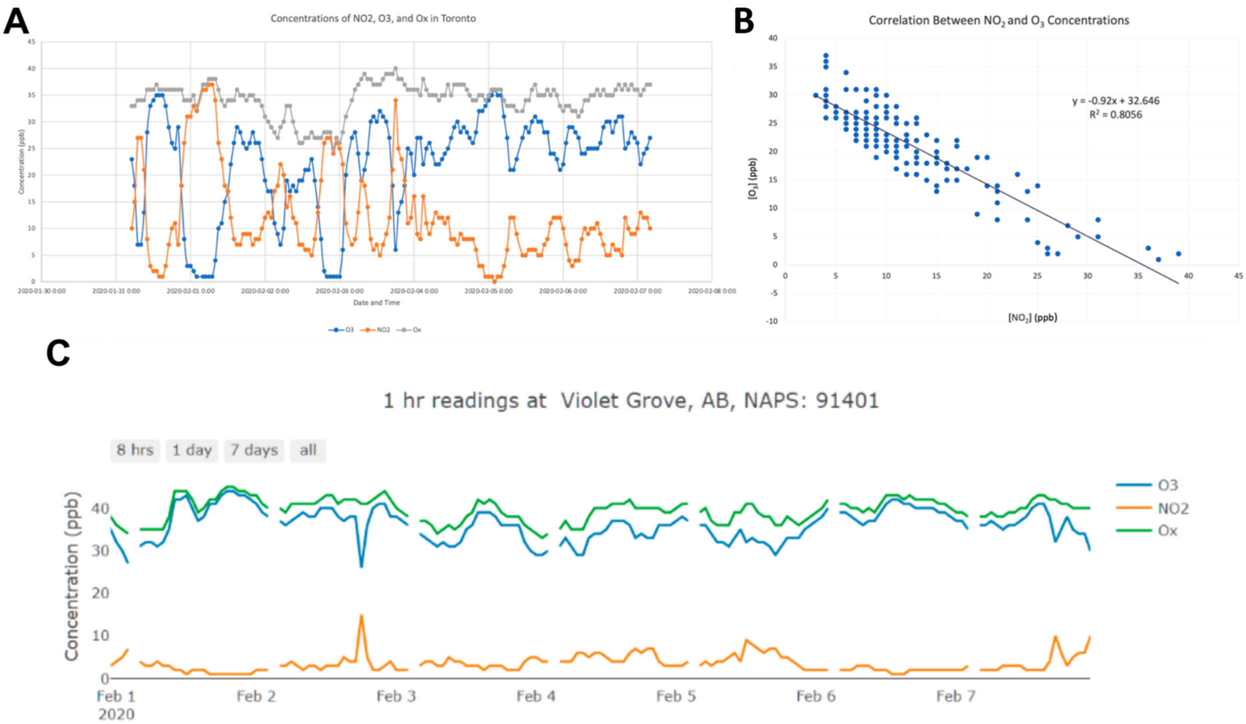


Figure 2. Overview of plots produced during the *Air Quality Activity*. (A) Student produced plots of hourly airborne pollutant concentrations in the City of Toronto during winter 2020. (B) Student produced correlation plot of hourly O_3 and NO_2 in the Greater Toronto Area during winter 2020; note A and B were produced by different students using different data sets. (C) Student's screenshot of the *Air Quality App* generated plot of hourly pollutant concentration in Violet Grove, Alberta during winter 2020.

specifically designed to demonstrate the utility of simply "looking" at a plot within a specific context as the first step of effective data interpretation. Here is how the student who produced Figure 2A responded to this prompt:

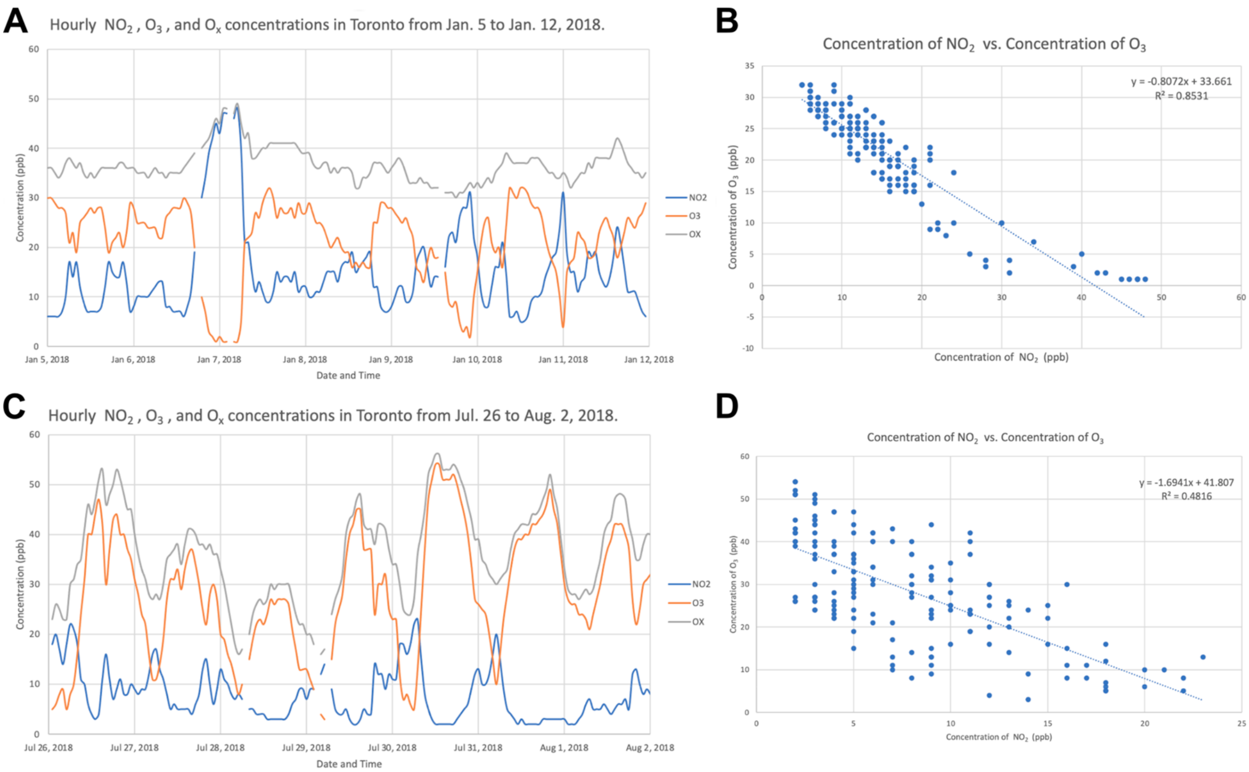
Yes, the dataset shows the diurnal trend but only for half of the time. As explained in the video, in the morning the O_3 is at a low level and the NO_2 is at a high level and the substance with the higher concentration oscillates throughout the day. After the light comes out in the morning, the O_3 is at a higher level and the NO_2 is at a lower level, following the trend shown in the video—the NO_2 reacting with light and O_2 to create O_3 and then when the light goes away at night, the O_3 reacts with NO to regenerate NO_2. After February 3rd though, the diurnal trend stops. What happens instead is the concentration of NO_2 always stays lower than the level of O_3. The amount of NO_2 is still relative to the amount of O_3 when the concentration of O_3 increases, the concentration of NO_2 decreases, and vice versa.

In this response the student effectively synthesized the plot they produced into the chemical context of urban air quality discussed in the introduction to this activity. They were able to identify regions of the plot that agreed with the expected chemistry and those that deviated from expectation. In the first iteration of the virtual activity students were asked to create this time series before calculating O_x and removing the "−999" values. This was done purposely to teach them how to add a series to a plot (adding the O_x variable) and to reinforce the utility of using data visualizations to explore the data during the data cleaning process (i.e., "−999" values were obvious in any plot). The workflow was streamlined in the current iteration to make room for the accompanying in-person lab work.

As described in the previous section, students also produced a correlation plot of O_3 and NO_2 by plotting the variables against one another, then adding a linear regression line-of-best fit with its associated equation and R^2 value displayed. An example of a student-generated correlation plot can be found in Figure 2B. This introduces another way to explore the diurnal trends between these variables and also demonstrates how data analysis often proceeds from qualitative to quantitative metrics. Following the creation of this plot, students were asked the following question: *Are NO_2 and O_3 positively correlated, negatively correlated, or not well-correlated? Does the direction of the correlation (positive or negative or lack of correlation) in the data set make sense given the background chemistry you know about these pollutants from the introduction?* Here is how the student who produced Figure 2B responded to this prompt:

NO_2 and O_3 are negatively correlated, as illustrated by the negative slope of the trendline. This means that as the concentration of one pollutant increases, the other decreases. To me, this makes sense given the background chemistry I know about these pollutants because when the concentration of NO_2 is low, the concentration of O_3 is high, and vice versa. For instance, as explained by the introductory video, in the morning when rush hour begins and the sun starts coming out, the concentration of NO_2 starts off high and then it starts decreasing into the day. On the other hand, O_3 starts off low and starts increasing into the day as NO_2 produces O_3 in the presence of light and molecular oxygen. Then at night after sunset, O_3 concentration decreases, and NO_2 increases. The negative correlation is seen throughout, and thus, this trendline makes sense to me.

https://doi.org/10.1021/acs.jchemed.3c00333
J. Chem. Educ. 2023, 100, 3410−3418



E

Pollutant	Winter (Jan. 5 to Jan. 12, 2018)			Summer (Jul. 26 to Aug. 2, 2018)		
	Mean	SD	RSD	Mean	SD	RSD
NO_2	15.1	8.9	59.0%	7.9	5.1	65.0%
O_3	21.4	7.8	36.0%	28.4	12.5	44.0%
O_x	36.6	3.5	9.5%	36.3	36.3	27.0%

Figure 3. Collection of submitted plots from a single student for the Fall 2021 iteration of the *Air Quality Activity*. (A) and (B) are time series and correlation plot, respectively, of student's winter data set. (C) and (D) are time series and correlation plot, respectively, of student's summer data set. (E) Table of summary statistics calculated by student for their assigned data sets, all units are ppb.

In this example the student can clearly enunciate how the chemical relationship between O_3 and NO_2 can explain the observed trend in their data (learning goal 3).

In addition to applying their chemical knowledge to the modeled data, it is important to understand its limits. As the linear regression is produced from a mathematical function, it can extend into negative values, however, the chemical phenomena it describes (e.g., concentration) cannot. To assess whether students could parse this difference, we asked the follow up question: *Is it possible to confidently extrapolate the linear regression of the correlation plot (meaning extend your line) beyond the data plotted? Can the lines extend past the axes (into negative values)? Why or why not? Explain your reasoning.* Here is how the student who produced Figure 2B responded to this prompt:

> *It is not possible to confidently extrapolate the linear regression of the correlation plot beyond the data plotted. In other words, the lines cannot extend past the axes into negative values. This is because concentration values cannot be negative—it doesn't make sense. The lowest concentration value possible is 0.*

The *Air Quality Activity* has always included an analysis of a second data set, so students have the opportunity to apply the knowledge about air quality they developed to a second scenario. The current iteration of the activity uses the *Exploring Air Quality* web-app to allow students to explore air quality across Canada. In their reports students are asked to select a rural community and answer the question: *How do the observed concentrations of O_3 and NO_2 indicate to you this is a rural location?* A student in Fall 2022 submitted Figure 2C, which shows the air quality in Violet Grove, Alberta, in February 2020. Here is how this student addressed the report sheet prompt:

> *While the O_3 levels are still high in this area, it can be seen that the levels of NO_2 are incredibly low—never surpassing the O_3 level at any given time. This is because NO, and by extension, NO_2, are direct products of vehicles. In rural areas with small populations, there is not a large amount of day-to-day commute or travel, not to mention less cars in general compared to a town. The O_3 levels are also high likely due to the lack of NO it cannot be degraded. Therefore, the low NO_2 levels and high O_3 levels indicate that this area has small amounts of vehicle activity, and it can be inferred it is a rural area.*

This student provided a very good analysis of the difference in air quality between an urban and rural location using two visualizations and the basic understanding of air quality provided in the introduction.

https://doi.org/10.1021/acs.jchemed.3c00333
J. Chem. Educ. 2023, 100, 3410−3418

As mentioned in the previous section, the virtual iteration of the *Air Quality Activity* had students analyzing a second data set collected in the summer (July or August) in Toronto and included calculations of mean, standard deviation, and relative standard deviation values for O_3, NO_2, and O_x in both seasons. A collection of these four plots and statistics calculations are shown in Figure 3.

Students were asked a series of questions that challenged them to synthesize the plots and statistics and situate them within the chemistry taking place in both seasons. To probe the students' ability to relate the calculated statistics to the chemistry taking place they were asked: *Which of the three species calculated (O_3, NO_2, or O_x) had the smallest relative standard deviation in your winter data set? Does this observation make chemical sense? Explain your reasoning*. Here is how the student who generate the data in Figure 3 responded to this prompt:

> *O_x had the smallest relative standard deviation in my winter dataset (around 9%). This observation makes chemical sense. When the concentration of O_3 is high, NO_2 is low and vice versa but for the concentration of O_x, it would be relatively constant (and have the smallest relative standard deviation). The concentration of odd oxygen is the concentration of O_3 and NO_2 added together (cancelling out the lows and highs of the two gases) so the concentration of odd oxygen does not make sudden dips in the graph like O_3 and NO_2.*

This student clearly demonstrated an understanding of the simple statistics, standard deviation, relative standard deviation, and can apply that understanding to the air quality reactions described in this activity. To start students thinking about how these statistics could be used to explore differences in the chemistry taking place in the winter and summer they were asked: *Concentrations of O_x tend to be more variable in the summer than the winter. Use the plots and statistics you have generated thus far to comment on whether this trend was apparent in your data?* Here is how the student who generated the data in Figure 3 responded to this prompt:

> *This trend was apparent in my data. Both the standard deviation and relative standard deviation of the O_x concentrations in the summer were greater than in the winter, meaning that there was more variability in the summer. Also, looking at the graph for summer, the concentrations of odd oxygen show higher peaks compared to the graph for winter.*

Here the student again applies the calculated statistics successfully and is also able to relate the question to the two time series plots they made for the winter and summer data sets. Other students referred to the R^2 values in their correlation plots as additional evidence of the difference in variance between the two seasons.

To culminate this analysis of the two seasons, students were asked: *The relationship between NO_2 and O_3 described by reactions (2) and (4) in the introduction is an oversimplification of the complicated chemistry taking place in the atmosphere. The more chemistry taking place in the atmosphere overall, the more NO_2 and O_3 will deviated from this simplified relationship. Compare the data you generated for both the winter and summer data sets and comment on which season sees more additional chemistry. Does this make chemical sense?* Here is how the student who generated the data in Figure 3 responded to this prompt:

> *The summer season sees more additional chemistry as the concentrations of nitrogen dioxide and ozone deviate more from the simplified inverse relationship. This makes chemical sense as mentioned above, trees are a major source of VOCs in the summer, leading to higher concentrations (and the longer maintenance of high concentrations even at night) of ozone than predicted by the simplified relationship.*

This question of why there is more overall chemistry happening in the atmosphere in the summer is a complicated one related to meteorology, the concentration of many species including NO and VOCs, and the amount of sunlight traveling through the atmosphere.[17] Since the role of sunlight and VOCs in urban air quality were highlighted in the introductory material, we expected that students would mention one or more of these aspects in their answer. In the sample response provided above, the student does an excellent job making a connection between vegetation, VOC concentrations, and the seasons. The "as mentioned above" included in the student's response is a reference to the previous question in the report sheet which asked students to calculate a summertime concentration of isoprene in the air from provided values. This question was intentionally added to prompt students to think about the relationship between VOC concentrations and seasons.

Student Feedback

For several iterations of the *Air Quality Activity*, both virtual and in-person, we issued questions specific to the activity as a component of the broader end-of-course CHM135 surveys. The majority of students responded positively to the new learning experience. In the Fall 2022 term 991 students responded to our survey (Figure 4A), 75.0% of these students felt that the *Air Quality Activity* made subsequent Excel calculations or data-analysis in the class easier; 71.6% felt that the *Air Quality Activity* increased their confidence analyzing data; and 82.9% felt the activity increased their overall confidence using Excel. These results mirror earlier surveys of the virtual-only course (Figure S1). Moreover, 97.0% of students report returning to the *Excel4Chem* webbook to help them complete laboratories (Figure 4B), with the majority (64.6%) accessing it for three or more of the five laboratories in our curriculum. A large majority (86.4%) found the *Excel4Chem* webbook helpful, and more than half (54.2%) indicated it was a significant help for completing the lab component of the class (Figure 4C). Altogether, our survey results indicate that we have successfully met learning goal 4 and have provided students with an effective workflow and Excel skills to apply to subsequent CHM135 laboratories. The activity also seems to have increased students' confidence when conducting data analysis and working with Microsoft Excel. Additional details from these surveys, including student comments, can be found in the Supporting Information.

■ CONCLUSIONS

The *Air Quality Activity* is an introductory activity that provides incoming University students with an adaptable data analysis workflow and practical experience using Microsoft Excel. To support this activity, we leveraged the R computer language for the automated, and scalable, generation of unique data from real world atmospheric measurements from the NAPS data repository.[11] Alongside introductory atmospheric chemistry material and the *Excel4Chem* webbook, we

https://doi.org/10.1021/acs.jchemed.3c00333
J. Chem. Educ. 2023, 100, 3410−3418

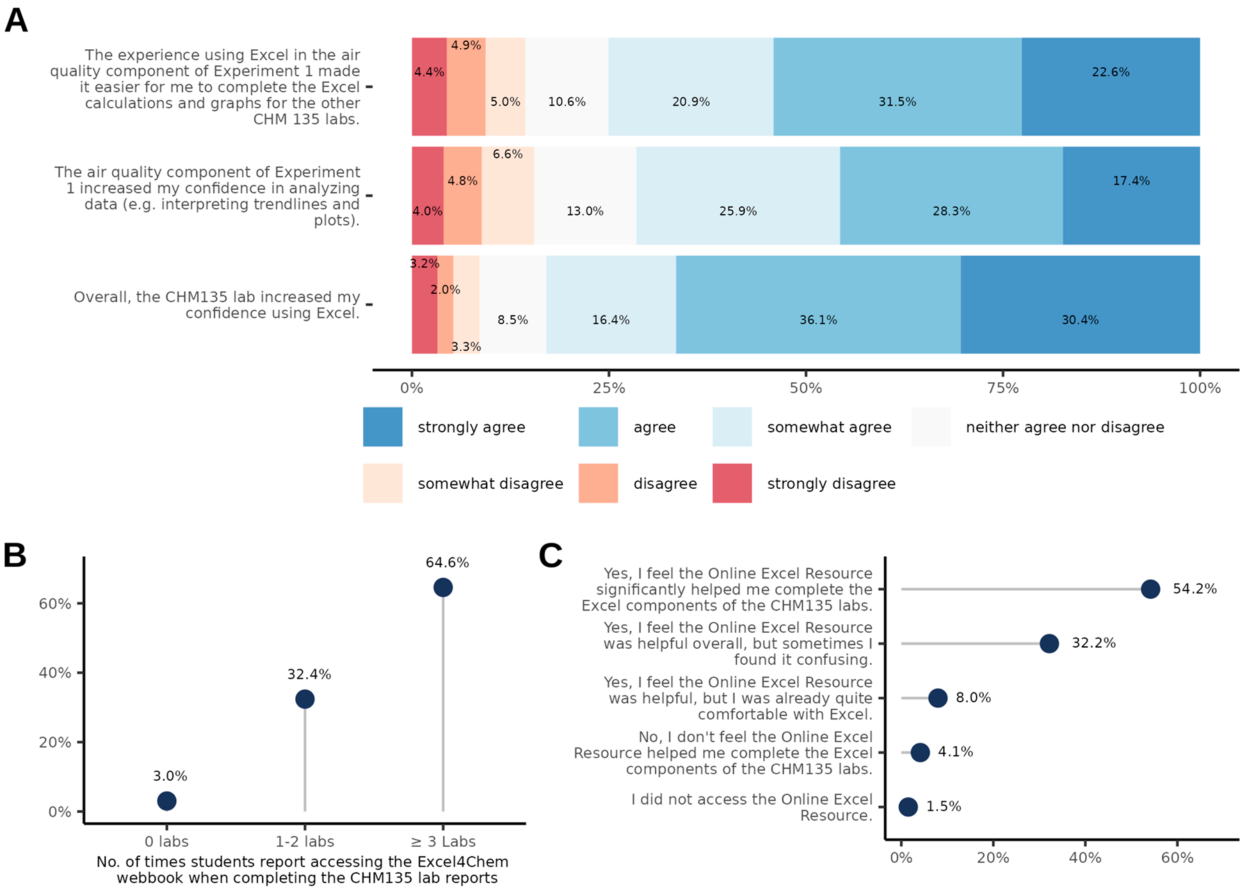


Figure 4. Survey results of student feedback for the in-person Fall 2022 *Air Quality Activity*. (A) The majority of students report that the *Air Quality Activity* was beneficial to their data analysis abilities. (B) Students report referencing the *Excel4Chem* webbook multiple times throughout the CHM135 Laboratories. (C) The majority of students found that the *Excel4Chem* webbook was helpful during the CHM135 laboratories. The total number of survey respondents was 991 students.

developed an online interactive web-app, allowing students to readily explore the entire NAPS data set to complement their individual analyses. This was a particularly important aspect of the work as the functionality of the app greatly reduced the "friction" of expanding the activity (i.e., rural vs urban questions, and the virtual to in-person transition), and allowed students to explore the larger NAPS data set without burdening them with lengthy data preparation or analysis. As the data is from real and complex environments, students often find data that does not support their hypothesis or that contradicts the prelab information. This contrasts traditional undergraduate laboratories which are often tailored so that students produce predictable results, which can inadvertently teach students to fit their data to the theory, in opposition to actual research where the data drives the theories. Students were excited by this activity as they were able to make real connections and real discoveries with real data. By including the activity at the outset of our lab curriculum, students were better equipped to tackle subsequent data analysis challenges both inside and outside the first-year chemistry laboratory. We are presently expanding our departmental wide foray in data-analysis pedagogy by incorporating R into upper year environmental chemistry courses. While still a work in progress, our open-source in-house textbook, *R for Environmental Chemistry*, can be found here: https://uoftchem-teaching.github.io/R4EnvChem/.

■ ASSOCIATED CONTENT

Data Availability Statement

The *Excel for General Chemistry*, the latest version of our Excel tip sheet, can be found freely online at: https://uoftchem-teaching.github.io/excel4chem/. The source code, air quality data, additional technical details on how the web-app works, and insights into hosting the *Air Quality App* can be found freely at: https://github.com/UofTChem-Teaching/AirQualityApp.

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.3c00333.

> Additional survey results and student comments; details on the production of data sets and implementation of the Shiny App; copies of the in-person Winter 2023 data analysis instructions and report sheet questions; copies of the virtual-only data analysis instruction, report sheet questions; a copy of the original *Excel Tip Sheet*; and a copy of a representative 7-day data set as it is assigned to students (PDF) (DOCX) (CSV)

https://doi.org/10.1021/acs.jchemed.3c00333
J. Chem. Educ. 2023, 100, 3410−3418

■ AUTHOR INFORMATION

Corresponding Author

Jessica C. D'eon − *Department of Chemistry, University of Toronto, Toronto, Ontario M5S3H6, Canada;* orcid.org/0000-0001-7448-8828; Email: jessica.deon@utoronto.ca

Author

David Ross Hall − *Department of Chemistry, University of Toronto, Toronto, Ontario M5S3H6, Canada*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jchemed.3c00333

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank all the CHM135 students who participated in this activity and specifically Abeeha Ahmad, Ainsley Base, Tabris Cao, and Sunny Yun for the use of their plots and short answer responses in this manuscript. We would also like to thank the summer 2020 TAs who facilitated the first iteration of the *Air Quality Activity*, and who generously took the time to communicate their, and students', experiences back to us. They are Illias Halimi, Jerry Augustune, Jessica Besaw, Kelly-Marie Yokuda, and Ying Shan Ma. Further thanks to Andrew Folkerson for ongoing support in integrating data science into the broader CHM135 lab. Also, thanks to Drs. Mima Staikova, C. Scott Browning, Andrew P. Dicks, and Kristine Quinlan for helping us tailor the experimental components and discussions for the CHM 135 audience. Thanks to David Liu for ideas and feedback on how to scaffold the Excel content from the Air Quality Activity into subsequent CHM 135 laboratories. Lastly, thank you to the Department of Chemistry's *Chemistry Teaching Fellowship Program* whose initial funding made this project possible and the University of Toronto's *Learning & Education Advancement Fund* that provided support for the development of the webbook and web-app.

■ REFERENCES

(1) Dietmüller, S.; Garny, H.; Eichinger, R.; Ball, W. T. Analysis of Recent Lower-Stratospheric Ozone Trends in Chemistry Climate Models. *Atmospheric Chemistry and Physics* **2021**, *21* (9), 6811−6837.

(2) de Almeida, A. F.; Moreira, R.; Rodrigues, T. Synthetic Organic Chemistry Driven by Artificial Intelligence. *Nat. Rev. Chem.* **2019**, *3* (10), 589−604.

(3) Blei, D. M.; Smyth, P. Science and Data Science. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (33), 8689−8692.

(4) Peters, E.; Vastfjall, D.; Slovic, P.; Mertz, C. K.; Mazzocco, K.; Dickert, S. Numeracy and Decision Making. *Psychological Science* **2006**, *17* (5), 407−413.

(5) Pinker, S.; Feedle, R. *A Theory of Graph Comprehension*; Psychology Press, 1990; pp 73−126.

(6) Cairo, A. *How Charts Lie: Getting Smarter about Visual Information*, 1st ed.; W. W. Norton & Company: New York, 2019.

(7) Glazer, N. Challenges with Graph Interpretation: A Review of the Literature. *Studies in Science Education* **2011**, *47* (2), 183−210.

(8) Rubin, S. J.; Abrams, B. Teaching Fundamental Skills in Microsoft Excel to First-Year Students in Quantitative Analysis. *J. Chem. Educ.* **2015**, *92* (11), 1840−1845.

(9) Johnson, A. M.; Jacovina, M. E.; Russell, D. G.; Soto, C. M. *Challenges and Solutions When Using Technologies in the Classroom*; Routledge, 2016.

(10) Potratz, J. P. Making Enzyme Kinetics Dynamic via Simulation Software. *J. Chem. Educ.* **2018**, *95* (3), 482−486.

(11) *National Air Pollution Surveillance Program*; Environment and Climate Change Canada, 2010.

(12) *Guidance Document on Achievement Determination: Canadian Ambient Air Quality Standards for Fine Particulate Matter and Ozone*; Candian Council of Ministers of the Environment: Ottawa, 2012.

(13) *Ontario's Ambient Air Quality Criteria*; Human Toxicology; Air Standards Section, Technical Assessment; Standards Development Branch, Ontario Ministry of the Environment, Conservation; Parks: Toronto, ON, Canada, 2020.

(14) D'Eon, J. C. The Chemistry of Air Quality. https://youtu.be/nkMUplFG6f0 (accessed on July 17, 2023).

(15) Atkinson, R. Atmospheric Chemistry of VOCs and NOx. *Atmos. Environ.* **2000**, *34* (12), 2063−2101.

(16) Kley, D.; Geiss, H.; Mohnen, V. A. Tropospheric Ozone at Elevated Sites and Precursor Emissions in the United States and Europe. *Atmos. Environ.* **1994**, *28* (1), 149−158.

(17) Seinfeld, J. H.; Pandis, S. N. Section 6.2: Basic Photochemical Cycle of NO, NO, and O. In *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*; John Wiley & Sons: Hoboken, NJ, 2016; pp 179−181.

(18) Johnstone, A. H. Why Is Science Difficult to Learn? Things Are Seldom What They Seem. *Journal of Computer Assisted Learning* **1991**, *7* (2), 75−83.

(19) Hall, D. R.; Liu, D.; D'Eon, J. C. Excel for General Chemistry, https://uoftchem-teaching.github.io/excel4chem/index.html (accessed on June 12, 2023).

(20) Hall, D. R.; D'Eon, J. C. Exploring Air Quality Data, https://uoft-chem.shinyapps.io/Air_Quality_App/ (accessed on July 27, 2023).

(21) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.

(22) Chang, W.; Cheng, J.; Allaire, J. J.; Xie, Y.; McPherson, J. *Shiny: Web Application Framework for r*; 2020, https://www.r-project.org.

(23) Ooms, J. *Sodium: A Modern and Easy-to-Use Crypto Library*, 2022, https://cran.r-project.org/web/packages/sodium/index.html.

(24) Wickham, H.; Grolemund, G. *R for Data Science*; Beaugureau, M., Loukides, M., Eds.; O'Reilly Media Inc.: Sebastopol, CA, 2017.

https://doi.org/10.1021/acs.jchemed.3c00333
J. Chem. Educ. 2023, 100, 3410−3418