Data Detective:

Investigating the Human Health Effects of Air Pollution in Washington State

STEM Global Teacher Workshop

IHME | healthdata.org

Revised July 2020 with Remote Learning Adaptations

LESSON OVERVIEW

Time: 3 50 minute class periods, if taught during synchronous in-personal instruction. Adaptations for remote learning may change this time estimate.

Subject & Grade Level(s)

HS Earth and Space Sciences (Grades 9-12). This lesson incorporates mathematics and computational thinking concepts and practices.

Brief Overview

This lesson explores past, current, and future trends of air pollution in Washington State and provides opportunities for students to better understand the phenomenon through intensive interaction and manipulation of data using spreadsheet software. Through this lesson, students will develop an understanding of foundational data science principles and recognize techniques for manipulating and analyzing data. In particular, students will develop skills in vetting data quality and generating basic descriptive statistics, including calculating mean, median, min, and max. Students will also gain skills in interpreting trends and patterns in data and making informed and evidence-based conclusions. In addition, students will gain an understanding of how air pollution effects human health and the global epidemiology of outcomes attributed to air pollution.

This lesson was developed by the Institute for Health Metrics and Evaluation (IHME), an institute affiliated with the University of Washington focused on health metrics sciences. As such, this lesson attempts to introduce students to fundamental data sciences practices that are the work of scientists across fields, including global health.

Remote Learning Adaptations: Integrated into this lesson plan are suggestions for adapting the activities for remote teaching and learning settings. For remote instruction, this lesson assumes students have computer and internet access at home.

(Image credit: National Geographic).

STUDENT UNDERSTANDINGS

Anchoring Phenomenon

Exposure to outdoor (ambient) air pollution is a major risk factor for disease for people around the world. There is a wide array of health effects which are believed to be associated with air pollution exposure. Among them are respiratory diseases (including asthma and changes in lung function), cardiovascular diseases, adverse pregnancy outcomes (such as preterm birth), and even death. While air pollution is a global phenomenon, it has very local impacts that can profoundly affect communities in unique and profound ways. In Washington State for instance, increasing temperatures, rapid urbanization, increased manufacturing, and natural disasters can affect the region's air quality which, in turn, can contribute to poorer health outcomes for residents. In order to alleviate/mitigate the consequences of air pollution in Washington State, we must assess and understand levels and trends of exposure to air pollution to better identify solutions and interventions.

Driving Questions

- How can data be used to understand changing levels of air pollution exposure in the State of Washington?
- What are health effects from outdoor air pollution exposure to Washington residents?
- How can data be used to make predictions and recommendations about air pollution and human health?

Learning Objectives/Success Criteria

Students will be able to…

- Recognize and explain how air pollution affects human health.
- Evaluate and examine the quality of data including its completeness, outliers, consistency, and accuracy.
- Generate basic descriptive statistics using a set of data including mean, median, min, and max.
- Interpret trends and patterns of health outcomes related to air pollution by using various forms of data visualization.

Mathematical and Computational Thinking

This lesson was designed to introduce students to data science. It engages students in fundamental data science practices, including manipulating, analyzing, and visualizing data. For each of these steps, students are encouraged to engage in sensemaking around why it is important and what meaningful work it does with the data. In this lesson, students will use a spreadsheet as a computational tool for finding patterns and trends. The underlying assumption for this lesson is that students have not yet used spreadsheets, therefore scaffolding is provided in the student handouts. Students with prior expertise with spreadsheets could be challenged to engage with the data in more complex ways. See the *Suggested Lesson Extensions* section for ideas. The following linked teacher resource may be helpful for framing your approach to computational thinking in the science classroom:

- STEM Teaching Tool #56: *Engaging Students in Computational Thinking During Science Investigations*

NEXT GENERATION SCIENCE STANDARDS

This lesson builds toward the following bundle of high school level Performance Expectations (PEs). Standards marked with an asterisk (*) are concepts or practices aligned to this lesson, but not included in the PE bundle. The lesson materials are written to support the high school level PEs; therefore adaptations should be made to the lesson to make it appropriate for middle school students (e.g., scientific vocabulary, scaffolds for data analysis, etc.). Hyperlinks direct to relevant sections of the Next Generation Science Standards and *A Framework for K-12 Science Education*.

Performance Expectations		
HS-ESS3-3: Create a computational simulation to illustrate the relationships among management of natural resources, the sustainability of human populations, and biodiversity. **HS-ESS3-6:** Use a computational representation to illustrate the relationship among Earth systems and how those relationships are being modified due to human activity.		
Disciplinary Core Idea(s)	**Crosscutting Concepts (CCCs)**	**Science and Engineering Practices (SEPs)**
ESS3.C Human Impacts on Earth Systems	Systems and systems models *Patterns *Stability and change	Using mathematics and computational thinking Influence of science, engineering, technology, and applications of science *Analyzing and interpreting data

Connections to Common Core State Standards in Mathematics include the following mathematical practices:

- MP.2 Reason abstractly and quantitatively
- MP.4 Model with mathematics

TEACHER PREPARATION

Materials

Material	Description/Source	Quantity
Classroom/ Teacher Computer	Computer with internet access, projector, and speakers (for live, in-class instruction).	1
Student Computers	Computers need to have access to spreadsheet software: Google sheets or Microsoft Excel. For remote instruction, this lesson assumes students have computer and internet access at home.	1/student or pair of students
Student Handouts	Make copies of the student handouts, one of each handout for each group of students -OR- make digital copies of handouts available to students via your learning management system. • **Student Handout 1:** *Instructions: Data Science in Global Health* (1/student or group) • **Student Handout 2:** *Time Series Plot for Washington State Data* (1/group) • **Student Handout 3 A-I:** *City Comparison Plots* (1/group to match city assignment) • **Student Handout 4:** *Washington State Map of PM2.5 Mean Values* (1/group)	1 for each group/pair of students
Teacher Slide Deck	Air Pollution slide deck (elements of slide deck below) • Part I: Introduction to Air Pollution (Slides #1-6) • Part II: Data Science and Air Pollution in Washington State (Slides #7-12) • Part III: Air Pollution and Human Health (Slides #13-26) • Part IV: How Air Pollution is Measured (Slides #27-34) • Part V: Data Quality Assessment (Slides #35-40) • Part VI: Basic Descriptive Statistics (Slides #41-50) • Part VII: City Comparisons (Slides #51-62)	1
Video: Data Analysis Example	**Data Driven: Data Analysis Example Video** Short video that demonstrates the data analysis process using Pullman as the example dataset. Teacher should watch this to help them understand the process and to decide if they want students to watch this prior to working on their own dataset. If so, load the video to your learning management system.	1
Data: Instructor version	https://docs.google.com/spreadsheets/d/11sJUkHGd2DdOZnsnYDb3J_eFhlP9bKT7wskeji-4B8Y/edit#gid=0 **(Includes an answer key)**	1
Data: Student version	https://docs.google.com/spreadsheets/d/15HWqvHQmZrgmB5KRkqj5SgG9utqS3LS1UbKhg6V-qeg/edit#gid=0	1

Notes to Teacher for Preparing to Teach this Lesson

- Computers are required for this lesson. Students must also have access to the internet and Google Sheets (preference) or Microsoft Excel. All directions on **Student Handout 1** are included for Google Sheets. If students will be using Excel instead, you may need to adapt some of the instructions for filtering and sorting data.
- Make copies of the student handouts, as indicated in the materials table above.

    - **Remote instruction:** Make digital copies of handouts available to students via your learning management system.
- Review the slide deck and speaker's notes.
    - **Remote instruction:** Consider recording a video of you presenting the slides (i.e., by using Peardeck for Google Slides or ScreenCast O Matic) for students to view from home.
- Watch the Data Analysis Example Video to help you understand the data analysis procedure. You may also want to go through the data analysis process, as demonstrated in the video, by using the Pullman data provided in the data spreadsheet.
- Decide if you want students to view the video before beginning data analysis.
    - **Remote instruction:** Load the video on your learning management system.
- An important component of engaging in data practices is having students pause during their computational work to engage in sensemaking together. This teacher resource on science talk and the embedded Student Talk Flow Chart provides ideas of how to structure student-to-student talk and teacher-student talk in equitable ways.
    - **STEM Teaching Tool #35:** *How Can I Foster Curiosity and Learning in my Classroom? Through Talk!*
    - **Remote instruction:** Consider how you might be able to use breakout rooms to allow for small group talk during live, synchronous remote instruction. Also consider using digital bulletin boards (i.e., Padlet, Google Jamboard app, or FlipGrid) as a way for students to communicate, share, and increase interaction.

ADAPTATIONS FOR REMOTE INSTRUCTION

This lesson plan has been updated to include adaptations for remote instruction settings. As written, the lesson is intended for in-person, live, classroom-based instruction. Suggestions are summarized in the table below for hybrid and fully remote teaching and learning settings. **Hybrid instruction** assumes several days each week of live, in-person, classroom-based instruction paired with several days of asynchronous, home-based, remote learning. **Asynchronous remote instruction** assumes no in-person, class-based instruction with all learning taking place in students' own homes. Suggestions focus on asynchronous learning. Teachers who can provide synchronous, live video conference meetings with students may sample from the original lesson plan or the hybrid setting suggestions. The adaptations described below should be considered in addition to the directions in the Procedure section of the lesson plan.

Learning Activity	Adaptations for Hybrid Instruction	Adaptations for Asynchronous Remote Instruction
Teacher Prep for Lesson	Decide what students will do during in-class days and what they will do from home during remote days. Make digital copies of handouts available to students via your learning management system. Load the Data Analysis Example video on your learning management system so students can view from home. Focus in-person time in the classroom on group work, group discussion, and interaction between students.	Consider recording a video of you presenting the slides (i.e., by using Peardeck for Google Slides or ScreenCast O Matic) for students to view from home. Make digital copies of handouts available to students via your learning management system. Load the Data Analysis Example video on your learning management system so students can view from home. Consider using digital bulletin boards (i.e., Padlet, Google Jamboard app, or FlipGrid) as a way for students to communicate, share, and increase interaction.
Introduction: Air Pollution and Data Science (1.1 – 3.1)	If you will present the slide deck in class, follow the instructions as written in the Procedure. If students will review the slide deck on their own during a remote instruction day,	*Optional:* As an optional way to open the lesson which leverages that students are engaged in learning from home, students could be asked to engage in a self-documentation activity around sources of ambient air pollution in their own community and lives. Students could take photos using a cell phone and/or add text to a collaborative digital space using technology such as Padlet, Google Jamboard app, FlipGrid or Google Slides.

	follow the instructions in the column to the right. Discussion questions could be reviewed together during the next in-class session.	This resource provides an overview of how to engage in this kind of culturally-responsive launch to the lesson. ○ **STEM Teaching Tool #31:** *How to Launch STEM Investigations That Build on Student and Community Interests and Expertise.* (Slides 1-34) Have students review the slide deck and speaker's notes or view a video of you presenting the slide deck (preferable). See Procedure section as written for more information about the mini-presentations in the slide deck. At the end of each mini-presentation, engage students in sensemaking. They could be asked to respond in writing to the following questions to submit to the teacher. Another option to promote student interaction is to use a digital bulletin board for students to post responses to the prompts and comment on each other's responses. • **Beginning of Intro to Air Pollution and Data Science:** *What do you know about air pollution? Where have you observed air pollution in their community?* • **End of Intro to Air Pollution and Data Science:** *How can using data science help us better understand air pollution and its effects on human health?* • **Beginning of How Air Pollution Affects Human Health:** *What do you know about how air pollution affects human health? Can you think of any specific instance/event where air pollution impacted human health?* • **End of How Air Pollution Affects Human Health:** *What makes air quality good? What makes it bad?* • **End of The Science of Measuring Air Pollution:** *What are some different sources of air pollution (natural/human/rural/urban)? How do scientists track/monitor air pollution?*

Exercise Prep (4.1 – 4.3)	If students will be able to work in groups on data analysis during an in-classroom day, follow the instructions as written in the Procedure. Divide students into groups of 2 or 3. You will need one computer per student or group for classroom work. If students will need to work on data analysis from home during a remote instruction day, follow the instructions in the column to the right.	Decide if students will be able to work remotely in pairs, or if they will need to work individually on the data analysis project. If working in pairs, consider what technologies are available to your school/district that would allow students to virtually meet-up to complete the assignment together. **Option A: Working Remotely in Pairs** Follow the directions in the Exercise Prep section of the Procedure. Do not assign anyone to the Pullman dataset. **Option B: Working Remotely as Individuals** Follow the directions in the Exercise Prep section of the Procedure except do not break students into groups. Assign each individual student to a data set. Do not assign anyone to the Pullman dataset. Not counting Pullman, there are 18 available datasets. As needed, you may assign multiple students to a dataset. It will be important to look at the group assignments in the Procedure and Slide #52 in the slide deck, which show how students will later pair up to compare their data from two cities. Make sure that you assign cities so that you will be able to form these pairings later in the lesson.
Exercise Part 1: Assessing Data Quality (5.1 – 5.6)	If students will be able to work in groups on data analysis during an in-classroom day, follow the instructions as written in the Procedure. This is encouraged if possible because it allows for collaborative group work and also makes the instructor available for just-in-time troubleshooting assistance. If students will need to work on data analysis from home during a remote instruction day, follow the instructions in the column to the right.	Have students review the slide deck **Part V: Data Quality Assessment** (Slides #35-40), which focuses on data assessment, what is data quality, and why it is so important in data science. Students will need access to a digital version of **Student Handout 1:** *Instructions: Data Science in Global Health.* In Part I of the handout, students can be asked to complete their work on the digital version of the handout to submit to the instructor, or to submit answers to the instructor in some other format through the classroom learning management system. Ask students to review the raw dataset in the Google Sheet for the (a) accuracy, (b) completeness, (c) representation, (d) recency.

		Students will need access to a digital version of **Student Handout 2:** *Time Series Plot for Washington State Data* with the Washington State air pollution plot graphs. Students will examine the graphs and identify (a) accuracy, (b) completeness, (c) representation, and (d) recency. Consider if students should submit their observations of the plot graphs to the instructor. Direct students to **Slide #40:** *Washington State Air Pollution Data Time Series* to identify the data points that are questionable and why (information provided in the notes section of the slide deck).
Exercise Part 2: Generating Descriptive Statistics (6.1 – 6.3)	If students will be able to work in groups on data analysis (encouraged) during an in-classroom day, follow the instructions as written in the Procedure. If students will need to work on data analysis from home during a remote instruction day, follow the instructions in the column to the right.	Have students review the slide deck mini-presentation on basic descriptive statistics, **Part VI: Basic Descriptive Statistics** (Slides #41-50). This will introduce the concepts of mean, median, min, max, and variance. Individual students, from home, will use the dataset for their assigned city to find the mean, median, min, max, and variance in the data for their city. Instructions on the data analysis task are provided in Part II of **Student Handout 1:** *Instructions: Data Science in Global Health.* Students will document the findings for each metric on their handout. If working in pairs, consider what technologies are available to your school/district that would allow students to virtually meet-up to complete the assignment together and discuss their results. Also consider how students will get help if they encounter problems while engaged in data analysis from home. Consider if drop-in office hours via videoconference are a possibility on the day(s) students are likely to work on Part II of **Student Handout 1**. ***Optional:*** For assessment purposes, each student could be asked to submit their city's filtered/sorted/analyzed dataset to the instructor.

Exercise Part 3: Comparing Metrics Across Cities (7.1 – 7.4)	Whether students engaged in data analysis in class or at home, having students present their group comparison data in class is a great option for interaction and collaborative sensemaking. In this case, follow the instruction as written in the Procedure.	Consider what technologies are available to your school/district that would allow students to virtually meet-up to share and discuss their results. Another option could be using FlipGrid or Google slides to allow for each group (Groups A-I) to share and compare their results. For example, a Google slide could be created for students who make up Group A to input their results for Kent and Scotch Basin for comparison. Alternatively, FlipGrid allows for students to upload short, informal videos (could be filmed on a cell phone) of themselves presenting their data. Each group (A-I) could also integrate the city comparison plots using their group slide from the slide deck Part VII: City Comparisons (Slides #51-62). Students will need access to a digital version of Student Handout 3 (specific to their group). Depending on how students share their comparison plots and data in a virtual environment, ask all students to compare and contrast the results across all of the group's data. Ask students to respond to the following questions and submit to the instructor. ● *Which city had the highest maximum? Which city had the lowest minimum?* ● *What city had the highest mean? What city had the lowest mean?* ● *What city had the highest median? What city had the lowest median?* ● *What is the variance across the cities?* ● ***Note** that answers to these questions are found in the instructor version of the Goggle Sheet on the sheet titled "stats key".*
Exercise Part 4: Comparing WA and Global Results (8.1 – 8.3)	Part 4 of the lesson should work well for students to complete at home on a remote learning day. If so, follow the instructions in the column to the right.	Engage students in sensemaking focused on the first three parts of the exercise and to prepare them for comparing Washington State results with that from other countries. Students could be asked to respond in writing to the following questions to submit to the teacher. Another option to promote student interaction is to use a Padlet, Google Slides, or FlipGrid for students to post responses to the prompts and comment on each other's responses.

	If students will be able to work on Part 4 during an in-classroom day, follow the instructions as written in the Procedure.	• *How have you been impacted by air pollution?* • *Have you been impacted by any of the events found in the data? How did the event impact you (please explain why and how)?* • *What is the general trend for air pollution in Washington State (is it decreasing, increasing, staying stable)?* • *Why is this the case and what could be contributing to these trends/patterns?* • *What has the impact/effect of wildfires had on air pollution in Washington State?* • *How could wildfires affect human health in the Washington State?* Students will need access to a digital version of **Student Handout 4:** *Washington State Map of PM2.5 Mean Values.* Adapt the activity by having students work individually from home, rather than together in small groups in the classroom, to complete Part IV of **Student Handout 1:** *Instructions: Data Science in Global Health.* The handout will prompt students to access the State of Global Air webpage and to identify three countries: one that has lower, one that has higher, and one that has equivalent values to the maximum value across Washington State. The handout will also guide students in exploring the data plot view to compare time series. Finally, students will record their observations and summarize their ideas for how the data could be used to improve air quality.
Discussion and Wrap Up (9.1 – 9.2)	If the discussion will be happening during a live, in-class day, engage the students in a whole class discussion to wrap-up what they have learned about air pollution and human health, both locally, across our region, and globally. Ask students what we can do with this data and results to improve air quality/pollution in the state (from their responses to Part IV on	One option for remotely wrapping up the lesson is to ask students to look through the posts (i.e., via Padlet, Google Jamboard app, FlipGrid, etc.) made by other students about what they found on the State of Global Air webpage. A digital Exit Ticket could be assigned at this point asking students to: • Summarize what they have learned about air pollution and human health, both locally, across our region, and globally.

	Student Handout 1). Student groups can report out and teacher will write ideas on the board. An Exit Ticket could be assigned at this point asking students to summarize what they learned from exploring the State of Global Air webpage. In addition, ask students to submit their completed **Student Handout 1** for grading/credit. If the wrap-up will occur during a remote instruction day, follow the instructions in the column to the right.	• Summarize what they learned from exploring the State of Global Air webpage and reviewing other students' posts. Did they see any patterns or trends? In addition, ask students to submit their completed **Student Handout 1** for grading/credit through your classroom learning management system. For accountability purposes, you could also ask students to submit their spreadsheet that shows the data analysis that they completed for the lesson.

INSTRUCTIONAL PROCEDURE (assuming synchronous, live, in-class instruction)

Topic	Teaching Activities	Student Activities
Introduction to Air Pollution and Data Science	1.1 Teacher asks students what they know about air pollution or where they have observed air pollution in their community. 1.2 Mini-presentation on air pollution using slide deck Part I: Introduction to Air Pollution (Slides #1-6). Define air pollution and describe why we are talking about it. 1.3 Mini-presentation on data science using the slide deck Part II: Data Science and Air Pollution in Washington State (Slides #7-12). What data science is and how using it can help us better understand air pollution and its effects on human health.	1.1 Students provide answers to this question 1.2 Students listen and ask questions 1.3 Students listen and ask questions
How Air Pollution Affects Human Health (defining the problem of air pollution)	2.1 Teacher asks students what they know about how air pollution affects human health. Can the students think of any specific instance/event they know of where air pollution impacted human health? 2.2 Mini-presentation on how air pollution affects human health (biological symptoms) using slide deck Part III: Air Pollution and Human Health (Slides #13-26). ● Play video about PM 2.5 (1:18 minutes) ● Ask students if they have any questions/thoughts. ● Include a discussion about what makes air quality bad and good (present examples from data or cities/countries).	2.1 Students answer the question by sharing their understanding of air pollution and human health 2.2 Students listen and ask questions
The Science of Measuring Air Pollution	3.1 Mini-presentation on how air pollution is measured using slide deck Part IV: How Air Pollution is Measured (Slides #27-34). Include a discussion about the different sources of air pollution (natural/human/rural/urban) and how scientists track/monitor air pollution (description of different instruments, measures).	3.1 Students will listen and ask questions
Exercise Prep	4.1 Teacher asks students to divide into groups of 2 or 3 students. Teacher can count off by numbering students 1 through 3 if necessary. Each student group will need their own computer. If there are limited computers in the classroom, then groups can be larger than 2 to 3 students. Limit the # of groups to 19 as there are only 19 cities in the monitoring dataset. ***Note:** If you choose to show students the Data Analysis Example video, limit the number of groups to 18 and do not include Pullman data (as the answers are provided in the video).*	4.1 Once students have their numbers, they can divide themselves into groups

	4.2 Teacher instructs students to log-on to computers and go to Google Sheet with data: Student version of Google Sheet data: https://docs.google.com/spreadsheets/d/15HWqvHQmZrgmB5KRkqj5SgG9utqS3LS1UbKhg6V-qeg/edit#gid=0 Instructor version of Google Sheet data: https://docs.google.com/spreadsheets/d/11sJUkHGd2DdOZnsnYDb3J_eFhlP9bKT7wskeji-4B8Y/edit#gid=0 Teacher will assign a city to each group. Cities can be assigned at random or students can choose based on interest. A state map could be shown to students to help them choose their city of interest. There is a total of 19 cities (18 if you do not include Pullman). **Note:** If you will assigning less than 19 cities, it will be important to look at the group assignments above and Slide #52 in the slide deck, which show how student groups will later pair up to compare two cities. Make sure that you assign cities so that you will be able to form these pairings later in the lesson. Given the uneven number of cities, Pullman does not have a comparison city, therefore, if you do not need 19 cities for your class size, this one can be skipped. 1. *Anacortes (Group E)* 2. *Bellingham (Group H)* 3. *Bremerton (Group C)* 4. *Bremerton-Silverdale (Group C)* 5. *Darrington (Group G)* 6. *Forks (Group D)* 7. *Kent (Group A)* 8. *Marysville (Group F)* 9. *Methow (Group G)* 10. *Mount Vernon-Anacortes (Group B)* 11. *Mountlake Terrace (Group D)* 12. *Pullman* 13. *Scotch Basin (Group A)*	4.2 Students will open/log-on to computers/laptops and go to the URL provided by the teachers

	14. *Seattle (Group B)* 15. *Seattle-Tacoma-Bellevue (Group H)* 16. *Spokane (Group I)* 17. *Tacoma (Group E)* 18. *Walla Walla (Group F)* 19. *Yakima (Group I)* 4.3 Instructor asks students to create their own version of the data in a new Google Sheet. Instructions for creating a new database: • *Click the File tab* • *Click on the New tab* • *Select spreadsheet* • *Go to air pollution student sheet* • *Press Control + A to select all cells in the data set* • *Press Control + C to copy the data set* • *Press File > New > Spreadsheet* • *Press Control + V to paste data into a new sheet*	4.3 Student-groups will create their own version of the Google Sheet in order to make changes during the analysis.
Exercise Part 1: Assessing Data Quality	5.1 Mini-presentation on data assessment (what is data quality) and why it is so important in data science using the slide deck Part V: Data Quality Assessment (Slides #35-40). Teacher will define the criteria which the air pollution data will be assessed: (a) accuracy, (b) completeness, (c) representation, (d) recency (description and examples included in slides and on student handout). 5.2 Teacher will distribute copies of **Handout 1:** *Instructions* to students (1 copy per group of students). Instructor will ask students to spend 5 minutes reviewing the raw dataset in the Google Sheet for the criteria mentioned above: (a) accuracy, (b) completeness, (c) representation, (d) recency. Students can write down findings, insights, and discoveries on a loose-leaf sheet of paper. 5.3 Teacher will then distribute **Handout 2** with the Washington State air pollution plots graphs to each group. Students will spend 10 to 15 minutes looking at the graphs and identifying (a) accuracy, (b) completeness, (c) representation, and (d) recency.	5.1 Students will listen and ask questions. Students will answer 3+ questions. 5.2 Students will assess and analyze data and complete the handout. 5.3 Students will analyze graph and plots and write down their findings in the handout.

	5.4 Teacher will bring students back together and ask them report out to the entire class about their findings. ● Teacher will project Slide #40: Washington State Air Pollution Data Time Series onto the screen. Teacher will identify the data points that are questionable and describe why (teachers will have notes in the notes section of the slide deck—reproduced below--about the data points that have potential issues). *Answers to the data assessment exercise (included on Slide #40)* ● ***Completeness**: Does missingness appear to be random? (in the case of Marysville, data is missing each summer aka not random).* ● ***Accuracy**: Are there any invalid values (aka 0, negatives, extreme values – a number that is very high or low). One artifact that students should notice are that data from Kent in 2015 were reported in the wrong units and need to be multiplied by 100. Students should also notice very high values in 2016 for the Pullman data point. Given this strange time trend and that this is a student run station, we will need to remove it. Extreme values were seen in Methow Valley and Scotch Basin in 2014/2015. Students should be guided to observe that these reflect very large forest fires during that time period and are not invalid data points.* ● ***Recency**: Students should note that several monitors do not have data after 2014 and this should be taken into account in later analysis.* ● ***Representativeness**: Some data points (i.e., Forks, Methow Valley, Scotch Basin, Walla Walla) are located away from major population centers and as such, may not tell us much about health impacts.* 5.6 Teacher will ask students to remove the outliers in their data.	5.4 Students will report out about their findings and answer instructor questions. 5.6 Students will listen and ask questions
Exercise Part 2: Generating Descriptive Statistics	6.1 Mini-presentation on basic descriptive statistics using slide deck Part VI: Basic Descriptive Statistics (Slides #41-50). Teacher introduces the concepts of mean, median, min, max, and variance. Teacher will ask students if they understand what “average” means and how it is used to describe things (give examples).	6.1 Students will listen and ask questions 6.3 Students will generate mean, median, min, max, and variance.

	6.3 Teachers instructs students to find the mean, median, min, max, and variance in the data for their city. Instructions on how to do this will be available on Handout #1. Students will document the findings for each metric on their handout.	Students will write metrics on the handout.
Exercise Part 3: Comparing Metrics Across Cities	7.1 Teacher will ask students to join groups with their other city group, as shown in the list below (forming groups of 4 to 5 students). Teacher will ask students to compare and contrast findings across the two cities. Students will need **Handout 3** specific to their group. Teacher will instruct students to document the findings on Part III of their **Handout 1**. Teacher will ask students to discuss their findings and evaluate the reasons for any differences and/or similarities. City comparison groups are listed below: ● *Group A: Kent vs. Scotch Basin* ● *Group B: Mount Vernon-Anacortes vs. Seattle* ● *Group C: Bremerton vs. Bremerton-Silverdale* ● *Group D: Forks vs. Mountlake Terrace* ● *Group E: Anacortes vs. Tacoma* ● *Group F: Marysville vs. Walla Walla* ● *Group G: Darrington vs. Methow* ● *Group H: Bellingham vs. Seattle-Tacoma-Bellevue* ● *Group I: Spokane vs. Yakima* **Note:** Due to the odd number of cities, Pullman does not have a comparison group. If assigned, the Pullman group can join any other group for this activity. 7.2 Teacher will bring class back together and ask students to present their findings by city comparison groups to the class. Teachers will project the city comparison plots as a backdrop found in the slide deck for each city comparison group. Use the slide deck Part VII: City Comparisons (Slides #51-62). 7.3 After all student groups present, teacher will ask the entire class to compare and contrast results. Students can generate their own questions to guide comparisons. In addition, teacher will ask the following questions: ● *Which city had the highest maximum? Which city had the lowest minimum?* ● *What city had the highest mean? What city had the lowest mean?* ● *What city had the highest median? What city had the lowest median?* ● *What is the variance across the cities?*	7.1 Students will review the data/findings from the two cities and document them on their handout. Students will discuss and critically examine the findings. 7.2 Students groups will come to the front of the class and report their findings 7.3 Students will discuss results and answer teacher questions

	• ***Note that answers to these questions are found in the instructor version of the Goggle Sheet on the sheet titled "stats key". • For each question, the teacher will challenge students to answer why they think those are the findings. Teacher will ask students what the general trends and patterns are for each metric. 7.4 Students will turn in their handouts to the teacher for grading.	7.4 Students will submit handouts to instructor
Exercise Part 4: Comparing Washington State Results with Global Results	8.1 Teacher will ask students probing questions: • *How have you been impacted by air pollution?* • *Have you been impacted by any of the events found in the data? How did the event impact you (please explain why and how)?* • *What is the general trend for air pollution in Washington State (is it decreasing, increasing, staying stable)?* • *Why is this the case and what could be contributing to these trends/patterns?* • *What has the impact/effect of wildfires had on air pollution in Washington State?* • *How could wildfires affect human health in the Washington State?* 8.2 Instructor will ask students to access the State of Global Air webpage and tell students to identify a country that has lower, higher, and equivalent values: http://www.stateofglobalair.org/air#PM. 8.3 Teacher will instruct students to go to the explore the data plot view to compare time series: http://www.stateofglobalair.org/data/#/air/plot • Students will need **Handout 4.** • Teacher will ask students to compare results for each country selected against the maximum value in Washington State (or each city).	8.1 Students will listen, ask questions and discuss 8.2 Students will go to the State of Global Air webpage and explore the map view 8.3 students will to the State of Global Air webpage and explore the plot view. Students will discuss the findings
Discussion & Wrap-up	9.1 Students will share if they found any interesting results from their exploration on the State of Global Air webpage. An Exit Ticket could be assigned at this point asking students to summarize what they learned from exploring the State of Global Air webpage. 9.2 Teacher will ask students what we can do with this data and results to improve air quality/pollution in the state. Teacher asks students to generate ideas in their groups.	9.1 Students will report out about what they found and discuss further 9.2 Students will ask each other questions and debate/discuss their ideas.

	Students will report out and teacher will write ideas on the board. Teacher will ask students to critique the ideas being shared.	

Student Assessment Opportunities

- Students' thoughtful participation during class discussions, group share-outs, and exit ticket, as well as responses on **Handout 1**, can all be used for assessment purposes.

Student Handouts & Teacher Resources

- Student Handouts:
  - **Student Handout 1:** *Instructions: Data Science in Global Health* (1/student or group)
  - **Student Handout 2:** *Time Series Plot for Washington State Data* (1/group)
  - **Student Handout 3 A-I:** *City Comparison Plots* (1/group to match city assignment)
  - **Student Handout 4:** *Washington State Map of PM2.5 Mean Values* (1/group)

- Teacher Resources:
  - Data Driven: Data Analysis Example video
  - Air Pollution slide deck (elements of slide deck below)
    - Part I: Introduction to Air Pollution (Slides #1-6)
    - Part II: Data Science and Air Pollution in Washington State (Slides #7-12)
    - Part III: Air Pollution and Human Health (Slides #13-26)
    - Part IV: How Air Pollution is Measured (Slides #27-34)
    - Part V: Data Quality Assessment (Slides #35-40)
    - Part VI: Basic Descriptive Statistics (Slides #41-50)
    - Part VII: City Comparisons (Slides #51-62)

Suggested Lesson Extensions

- For the Introduction, students could be asked to engage in a self-documentation activity around sources of ambient air pollution in their own community and lives. This resource provides an overview of how to engage in this kind of culturally-responsive launch to the lesson.
  - **STEM Teaching Tool #31:** How to launch STEM investigations that build on student and community interests and expertise.
- This lesson could be extended to focus additional time to understanding the health impacts of air pollution on different body systems.

Notes on Adaptations and Inclusivity

- **Scientific & Computational Vocabulary:** Students may need some support in understanding the terminology embedded in this lesson. This may be particularly true for emerging bilingual students and students with lower reading levels. The following list captures some of the scientific, mathematical, and computational terms used in the lesson materials.
  - **Mean:** The mean or average that is used to derive the central tendency of the data in question. It is determined by adding all the data points in a population and then dividing the total by the number of points.
  - **Median**: A simple measure of central tendency. To find the median, we arrange the observations in order from smallest to largest value. If there is an odd number of

observations, the median is the middle value. If there is an even number of observations, the median is the average of the two middle values.
  - **Minimum**: The smallest observation (number) in a sample of data.
  - **Maximum**: The largest observation (number) in a sample of data.
  - **Data assessment**: The process of scientifically and statistically evaluating data in order to determine whether they meet the quality required for projects or business processes and are of the right type and quantity to be able to actually support their intended use.
  - **Accuracy:** The extent to which the result of a measurement, calculation, or specification conforms to the correct value or a standard.
  - **Completeness:** The extent to which there are any gaps in the data from what was expected to be collected, and what was actually collected.
  - **Recency:** The extent to which the data contains observations from recent time periods.
  - **Representativeness:** The extent to which the data provides insight into what is happening with the broader population of interest.
  - **PM2.5:** Refers to atmospheric particulate matter (PM) that have a diameter of less than 2.5 micrometers, which is about 3% the diameter of a human hair.

- **Inclusivity for All Learners:** Consider how the lesson activities may need to be adapted to be accessible for all learners. For example, what accommodations may a student with a visual or mobility impairment need to engage in the computational work? How might you elicit, build connections with, and leverage students' everyday expertise with data practices and spreadsheets? Is there a way that students with expertise with spreadsheets could help support students who do not yet have this experience? How might you group students with diverse expertise and learning needs into teams so that they can support each other?

- **Video:** The video that is embedded in the slide deck has a close captioning option. Press the CC icon on the bottom right of the video screen to access the closed captioning.

TEACHER RESOURCES

Background Information on Air Pollution in Global Health

WHO Ambient Air Pollution: Pollutants
https://www.who.int/airpollution/ambient/pollutants/en/

Our World in Data: Air Pollution
Hannah Ritchie and Max Roser, October 2017
https://ourworldindata.org/air-pollution

State of Global Air 2019 Report
IHME, 4/3/19
http://www.healthdata.org/news-release/state-global-air-2019-report

Health Impacts of Air Pollution

European Environment Agency, 4/22/16
https://www.eea.europa.eu/signals/signals-2013/infographics/health-impacts-of-air-pollution/view

Air Pollution, Climate, and Health
Breathe Life; Climate & Clean Air Coalition; WHO
https://www.who.int/sustainable-development/AirPollution_Climate_Health_Factsheet.pdf

The Weight of Numbers: Air Pollution and PM2.5
UnDark Magazine & the Pulitzer Center on Crisis Reporting
https://undark.org/breathtaking/

Career Links

The following careers are related to data sciences and global health.

- **Biostatistician:** https://www.careersinpublichealth.net/careers/biostatisticians/
- **Data analyst:** https://careerlink.com/quick/data-science-jobs
- **IHME Post-Bachelor Fellow:** http://www.healthdata.org/about/team/pbf

The career of a Data Analyst is featured on the STEM Global Pathways to Global Health Careers poster and accompanying fact sheets (see the Information Technology section):

- **Poster: Pathways to Global Health Careers:** https://www.wghalliance.org/resource/global-health-career-poster/
- **Fact Sheet: Pathways to Global Health Careers:** https://www.wghalliance.org/resource/global-health-career-fact-sheets/

Credit: *This activity was originally developed by the Institute for Health Metrics and Evaluation (IHME), a global health organization located in Seattle, WA and adapted for a STEM Global Teacher Workshop in April 2019. Authors include: Austin Carter, Doctoral Candidate and Researcher, IHME; Joseph Frostad, Doctoral Candidate and Researcher, IHME; Sean Lassiter, Senior Education Program Manager. Revisions in July 2020 by Sarah Wozniak and Justin Lo of the IHME. Lesson plan development, editing, and adaptations for remote instruction provided by Kristen Bergsman of Laughing Crow Curriculum LLC.*