# Data Practices

Use this resource to reflect on your practice or identify opportunities to integrate data practices in your classroom. This resource was developed to support middle school science teachers to integrate computational thinking into NGSS-aligned lessons. The content can be adapted for other content areas, grade band, and contexts.

# What are computational data practices?

- **Collecting data** with computational tool(s) and outputting data to prepare for analysis in order to inform a driving question.
- Analyzing data in order to identify relationships and make predictions.
- Evaluating data to identify bias in data collection and reporting
- Communicating data-based claims to a particular audience.

# Look fors:

### **Collecting data**

Designing an experiment with computational tools

Collecting data that can be quantified

### Analyzing data

- Anipulating data with data moves (Erickson et al., 2019)
- Describing relationships between variables
- Using data to make predictions

#### **Evaluating data**

- ldentifying bias in data collection and reporting
- Considering if/how data sources are comparable

### **Communicating data**

- Designing a visual representation of data
- Selecting design features to communicate to a particular audience

# Activities & Examples

Activity	You might use this when	Template	Example
Collecting, Analyzing & Evaluating Data	Students are exploring a driving question by designing an investigation or manipulating existing datasets.	Template: <u>Collecting,</u> <u>Analyzing &amp; Evaluating Data</u>	Example: <u>Collecting,</u> <u>Analyzing &amp; Evaluating Data</u>
Communicating Data	Students are communicating data-based claims to a particular audience	Template: Communicating Data	Example: Communicating Data





Data Practices

# Prompting questions:

Ask students to reflect on their process or progress with these prompting questions:

#### **Collecting data**

- U What do you wonder about this topic?
- Uhat question(s) will your dataset address?
- What might someone learn from using the data you collect?
- U What will you measure to inform your question?
- What tools can you use to collect this data? What are the affordances and limitations of using a computational tool?
- How frequently will you collect data? Are you observing this data on a certain day or during certain times?
- □ Is the data you are collecting able to be compiled easily (using the same units, avoiding open-ended responses to be able to group)?

Could there be bias in your data (e.g. human error, missing data, or questions that lead to certain responses)?

#### **Analyzing data**

- □ Identify cases and attributes in the raw dataset. How will you modify or use cases/attributes to answer your question?
- How did you manipulate or organize your data? How did this reveal relationships/patterns within the dataset?
- Uhy do you think the numbers changed? Is there a relationship between numbers/categories?
- Could we use this information to make predictions?
- Can you develop a rule or formula to describe how one variable is related to another variable?
- What new questions do you have about the dataset?
- Uwhat other factors (not in the graph) might influence relationships between variables?

#### **Evaluating data**

- If there is more than one data set being used, is the data comparable?
- U What data might be missing? How might that data help you better understand the data you have already?
- How was the data collected and by who?
- Uwhat type of organizations can I trust to have collected valid data on this topic?
- U How might the interest or bias of groups change the story or findings we can learn from this dataset?
- Could there be bias in the instruments used to collect the data (e.g. questions that lead to certain responses)?
- □ Is the data self-reported? Think about the implications of self-reported data. How might people's biases or perceptions change the story or findings we can learn from this dataset?
- Think about the implications of observed data. How was the data collected and by who? How might the process or tool used for data collection change the story or findings we can learn from this dataset?

#### **Communicating data**

- How might you structure your dataset to help you or someone else to answer a question related to your findings?
- How can you provide context or cues (e.g. titles, labels, colors) to help someone else understand your data?
- Why might this data be interesting/important to other people?
- Explain how your design choices help you describe your data to your audience.



# Data Practices: Collecting, Analyzing and Evaluating Data

Use the flowchart below to 1) identify a driving question, 2) collect or locate data related to the question, and 3) structure the dataset to be analyzed to inform the question.





CC (i) Attribution 4.0 International (CC BY 4.0)



# Part 2: Using Data Moves to Analyze Data

Use data moves (Erickson et al., 2019) to answer the questions about the dataset. Examples of data moves are:

- **Filtering** is selecting only a subset of the data that is available to explore. For example, you may view only a portion of the data because only some of it is relevant to the question you are exploring.
- Grouping is used to make comparisons between different subgroups of a data set.
- **Summarizing** is computing a value (e.g. mean, median, mode) to summarize a group.
- **Calculating** is creating a new attribute based on values of one or more existing attributes. For example, if an attribute is distance in kilometers, you may want to convert that value to distance in miles.
- **Merging/joining** is combining more than one dataset together. You may add more cases, or add more attributes to existing cases.
- Making a hierarchy is grouping data using multiple levels.

In order to complete data moves, you should become familiar with a program that allows you to manipulate data. Google Sheets, Microsoft Excel, CODAP (Concord Consortium) are all great resources for using data moves to analyze your data.

Question:			
Data Move(s) Dive deeper into the data set by combining moves	Sketch or describe how you organized or illustrated the data:	What does this tell you about the relationships/patterns within your dataset?	What new questions do you have?





Explore the sources that reported the data. Use the flowchart below to consider error and/or bias that could be present.





# Data Practices: Collecting, Analyzing and Evaluating Data



Part 1: Collecting and Structuring Data to answer a driving question

Use the flowchart below to 1) identify a driving question, 2) collect or locate data related to the question, and 3) structure the dataset to be analyzed to inform the question.





CC (1) Attribution 4.0 International (CC BY 4.0)

# In this activity, a student explored existing data about the novel coronavirus (COVID-19). The student followed the algorithm above to:

## 1. Identify a driving question:

The student began by brainstorming questions they had about the coronavirus, such as:

- Where in the world has the highest number of cases?
- How many cases have been diagnosed in the US?

## 2. Locate data related to the question:

The student identified a dataset related to their COVID-19 questions from <u>Our World in Data</u>. The raw data was downloaded from the Our World in Data website. Then, it was uploaded into a Google Sheet (see <u>Raw Data Sheet 1</u>).

### 3. Structure the dataset to be analyzed to inform the question

The data needed to be structured in a way that allowed it to be analyzed to answer the driving questions. First, the raw data was formatted into columns/rows. (See <u>Raw Data Formatted Sheet 2</u>.) Then, the rows and columns were flipped to make it easier to filter data for individual countries. (See <u>Select Countries</u> <u>Formatted Sheet 4</u>.)



# Part 2: Analyzing Data using data moves

Then, the student used data moves (Erickson et al., 2019) to answer the questions they had about the dataset. Examples of data moves are:

- **Filtering** is selecting only a subset of the data that is available to explore. For example, you may view only a portion of the data because only some of it is relevant to the question you are exploring.
- Grouping is used to make comparisons between different subgroups of a data set.
- Summarizing is computing a value (e.g. mean, median, mode) to summarize a group.
- **Calculating** is creating a new attribute based on values of one or more existing attributes. For example, if an attribute is distance in kilometers, you may want to convert that value to distance in miles.
- **Merging/joining** is combining more than one dataset together. You may add more cases, or add more attributes to existing cases.
- Making a hierarchy is grouping data using multiple levels.

In order to complete data moves, you should become familiar with a program that allows you to manipulate data. Google Sheets, Microsoft Excel, CODAP (Concord Consortium) are all great resources for using data moves to analyze your data.



# Question: How many cases have been confirmed in the US between Jan and Feb?

Data Move(s) Dive deeper into the data set by combining moves	Sketch or describe how you organized or illustrated the data:	What does this tell you about the relationships/patterns within your dataset?	What new questions do you have?
Filtering and summarizing	<ul> <li>Filter by country to view only confirmed cases from the US</li> <li>Find total number of cases at the end of February</li> <li>Find total number of cases at the end of January</li> </ul>	There have been approximately 60 cases diagnosed in the past month.	How long will the cases increase?
	Country T 2020-01-31 United States	<b>₹ 2020-02-29 ₹</b> 6 66	
	New (Total # of confirmed cases in .		

# Question: What country has the highest number of confirmed cases?

Data Move(s) Dive deeper into the data set by combining moves	Sketch or describe how you organized or illustrated the data:	What does this tell you about the relationships/patterns within your dataset?	What new questions do you have?
Sorting	Sorted from most recent date by highest to lowest Country 2020-05-03 T World 3,388,665 United States 1,133,069 Spain 216,582 Italy 209,328 United Kingdom 182,260	The United States has the highest number of cases, followed by Spain, Italy, the United Kingdom, and Germany.	What continent has the highest number of cases? (See analysis below)



Question:	What	continent	has th	e hig	hest nu	mber of	f confirmed	cases?

Data Move(s) Ske Dive deeper into the data set by combining data moves. dat	etch or describe w you organize llustrated the a:	What does this tell you abou the relationships/patterns within your dataset?	t What new questions do you have?
Sorting I d CC hi of CC CC W CC CC W CC CC CC CC	lentified the 25 pountries with the ghest total number f confirmed cases reated a new polumn to show hich continent ea pountry is on litered each colum y continent and ac otal confirmed case total c	North America has the most cases diagnosed followed by Europe. There are fewer cases in South American and Asia.	What measures would be effective to isolate the disease to a specific location or continent?





Explore the sources that reported the data. Use the flowchart below to consider error and/or bias that could be present.



This student found that <u>Our World in Data</u> is sourced from the European CDC. They stopped relying on data from the WHO in mid-March 2020 after finding <u>multiple reporting errors</u>. However, the total number of COVID-19 cases is likely higher than the number of confirmed cases reported on the infographic. This is because some individuals/countries have more access to testing than others.



# Data Practices: Communicating Data



Part 1: Determine the Audience

Before creating a visualization of data, consider your audience. Respond to the prompts below to consider why this data might be interesting or important to other people:





CC (\*) Attribution 4.0 International (CC BY 4.0)



There are many ways to visualize data (beyond just graphs!). Take a look at some different examples from <u>Information is Beautiful</u> and <u>Visual Capitalist</u>. Consider how you could design a visualization to help you explain data to your audience.

### Select two (or more) of the following choices:

- Combine/aggregate data
- Use shapes and/or colors to illustrate aggregates of data
- Eliminate information that isn't important
- Illustrate relationships
- Insert pictures
- Include multiple languages
- Another idea I have \_\_\_\_\_\_

Describe, sketch, or paste in your visualization:

Annotate your visualization to explain how your design choices help you describe your data to your audience:



Show your visualization to a partner. Is their understanding of the data the same as you intended? Are there modifications you can make to the design to clarify the message to your audience?



# Data Practices: Communicating Data



Part 1: Determine the Audience

Before creating a visualization of data, consider your audience. Respond to the prompts below to consider why this data might be interesting or important to other people:









There are many ways to visualize data (beyond just graphs!). Take a look at some different examples from <u>Information is Beautiful</u> and <u>Visual Capitalist</u>. Consider how you could design a visualization to help you explain data to your audience.

### Select two (or more) of the following choices:

- Combine/aggregate data
- Use shapes and/or colors to illustrate aggregates of data
- Eliminate information that isn't important
- Illustrate relationships
- Insert pictures
- Include multiple languages
- Another idea I have \_

### Describe, sketch, or paste in your visualization:





Annotate your visualization to explain how your design choices help you describe your data to your audience:



I designed this infographic for a person who has family living abroad. I included data for the five countries with the highest number of cases to help inform this person on where the virus has spread in the world. This may help the user determine if his/her family is at risk.



I designed this infographic for a person who is interested in traveling. I created a timeline from travel updates provided from the Center for Disease Control and Prevention (CDC). The purpose of this timeline is to give the user an idea of how travel restrictions have changed over time. This may help the user determine whether they still want to travel.





Show your visualization to a partner. Is their understanding of the data the same as you intended? Are there modifications you can make to the design to clarify the message to your audience?

