OVERVIEW

This highly customizable activity guides students through organizing and documenting spreadsheet data during a class investigation (e.g., experiment or lab). It is designed to be used at the start of an investigation you may already be doing with your class. The activity focuses on decisions during data organization and collection that facilitate future graphing and analysis.

Complete the “Customizable Student Handout” to use with a class investigation of your choosing. It is an editable template with seven parts, which you should adjust based on your class’s needs:

- In Part 1, students begin the investigation by writing the research question and forming a hypothesis.
- In Part 2, they use the investigation procedure to design their own data sheet.
- In Part 3, students collect measurements and then enter their data into a spreadsheet.
- In Part 4, they compare their spreadsheets with those of their classmates and reflect on their data organization decisions.
- In Part 5, they learn about some common effective practices for organizing data.
- In Part 6, the class decides on the final organization of their data, so that everyone can combine their data into one data set.
- In Part 7, they document their final spreadsheets by adding metadata.

This educator document contains multiple resources for implementing this activity with students, including the following (use the links to go directly to each section):

- **background** on why data and spreadsheet skills are important for students
- **teaching tips** for deciding the structure of the activity
- **procedures** for completing and implementing each part of the “Customizable Student Handout”
- **assessment guidance** with sample answers for the questions in the customizable handout
- **appendix** with a reference copy of the customizable handout and educator notes

Additional information can be found on [this resource’s webpage](#), including suggested audience, estimated time, and curriculum connections. You can also download example materials customized for the ruler drop investigation ([Ortolano, Finn, and Ortolano 2017](#)):

- The “Ruler Drop Student Handout” document provides a filled-in example of the “Customizable Student Handout.” It can be further customized or used as is for the ruler drop investigation.
- The “Ruler Drop Educator Materials” document provides additional guidance for implementing the ruler drop investigation in class.

**KEY CONCEPTS**

- Defining a research question and hypothesis guides data collection decisions.
- Data in spreadsheets should be consistently organized (e.g., with clearly labeled columns and rows) to facilitate interpretation and analysis.
- Standard spreadsheet organization practices and metadata make data easier to share and understand.
STUDENT LEARNING TARGETS

- Practice data collection, data entry, and spreadsheet skills.
- Identify and implement effective practices for data organization and documentation.
- Describe the importance of spreadsheet practices that make data understandable and shareable.

PRIOR KNOWLEDGE

Students should be familiar with:

- generating research questions and hypotheses
- collecting and recording variables in an experiment
- entering data (numbers and text) and performing basic functions (calculating averages using a formula and sorting) in the selected spreadsheet tool (e.g., Google Sheets, Microsoft Excel)

BACKGROUND

Collecting, organizing, and documenting data is important for many lines of work, from science and public policy to business and computing. Spreadsheet programs are important tools for working with data, but poorly organized spreadsheets can make the data more difficult to use. When first using spreadsheets, students often organize data based on ease of human viewing, which can cause problems for later graphing and computer analyses (Murrell 2013).

This activity guides students through effective practices for preparing data in spreadsheets. This includes making a data organization plan at the start of a study, alongside the generation of research questions and hypotheses. It also includes making spreadsheets understandable for future use, both by the researcher and by others who might want to confirm the findings or conduct additional studies.

Understanding principles of data organization and curation is important for using authentic data in the biology classroom (Kjelvik and Schultheis 2019). Additionally, teaching students about the importance of open and reproducible research can improve scientific literacy, student engagement, and trust and confidence in science (Pownall et al. 2022).

MATERIALS

- copies of the “Customizable Student Handout” that you have completed for the class investigation
- sheets to record data for each student group (e.g., blank pieces of paper, preprinted data tables, pages in lab notebooks, online documents in Google Docs/Microsoft Word/Notepad, etc.)
- access to a spreadsheet program (e.g., Google Sheets, Microsoft Excel)
- any other materials for the class investigation

TEACHING TIPS

There are many ways to customize this activity for your classroom context. Tips for the main customization opportunities are provided below (select links to go directly to each section):

- Selecting an Investigation
- Selecting a Facilitation Approach
- Selecting a Spreadsheet Program
- Customizing the Student Handout
Selecting an Investigation

This activity can be used with any class or lab investigation where students collect and record data, including controlled experiments or descriptive data collection. Select whichever type of investigation is most appropriate for your course and students.

The investigation should include multiple variables; this will create more variety in how students organize their data, which will spark more discussion during the activity. Examples include:

- comparisons between two or more groups
- two or more measurements collected per observation

Examples of common investigations in life science classes that would work well with this activity include:

- **Ruler drop** ([Ortolano, Finn, and Ortolano 2017](#)): Students explore differences in individual reaction times by measuring how long it takes to catch a ruler after it drops.
  - A customized “Ruler Drop Student Handout” and “Ruler Drop Educator Materials” for implementing this investigation are provided on [this resource’s webpage](#).
- **Wisconsin Fast Plants** ([Tomkins and Williams 1990](#)): Students can do a variety of experiments with Fast Plants, scientific name *Brassica campestris (rapa)* L. These plants have a five-week seed-to-seed cycle and are relatively easy to grow.
- **Leaf stomata density** ([Grant and Vatnick 2004](#)): Students examine leaf stomata density using species readily available to them.
- **Plant transpiration** ([McCallum 2001](#)): Students examine how wind and temperature affect plant transpiration, using celery as a model.
- **Pulse rate in blood worms and *Daphnia*** ([Bohrer 2006; Corotto et al. 2010](#)): Students examine the effects of nicotine, caffeine, ethanol, and other substances on these organisms.
- **Surface tension on a penny** ([Cavallo and Dunphey 2002](#)): Students examine different liquids on a penny to explore concepts of cohesion, surface tension, and intermolecular forces of water.

Selecting a Facilitation Approach

The “Procedure” section outlines two main approaches for facilitating the activity: predefined or collaborative.

For the **predefined** approach, you decide, ahead of time, the data collection parameters and variables that students measure. The default “Customizable Student Handout” uses this approach.

- This approach may be preferable when class discussion time is limited, or if you want to ensure that students are investigating a specific research question or using preplanned methodology.
- You should provide the background, experimental design, and data collection procedure for the class and make sure that all necessary information is included in the handout.

For the **collaborative** approach, you explicitly involve students when deciding on the experimental design, variables to measure, and data collection procedure.

- Although this approach may take more time, it can increase student voice, choice, and ownership of the investigation. However, all students should still agree on a consistent procedure before collecting data.
- For an example using this approach, download the customized “Ruler Drop Student Handout” on [this resource’s webpage](#). This example highlights a collaborative approach for selecting the experimental design prior to students engaging in the investigation.
- When using the “Customizable Student Handout” for the collaborative approach, you can add brainstorming questions before Part 1 and lead a class discussion that encourages students to generate experimental
design ideas. The class should ultimately decide on a single investigation with well-defined parameters, variables, and data collection protocols. The questions provided in the customizable handout begin after students have reached this stage.

- For this approach, you may customize the “Background” and “Procedure” sections of the handout to have more general information to guide class discussion rather than specific details. You may want to include information on the topic, the general parameters of the investigation, and constraints for students to consider (location, time, materials, technology, etc.).
- Refer to the Part 0 questions in the “Ruler Drop Student Handout” document for examples of how students can brainstorm potential research questions. These Part 0 questions may be helpful for using the collaborative approach with any system.

**Selecting a Spreadsheet Program**

You can do this activity with any spreadsheet program that allows students to enter and organize data. Common options include Google Sheets, Microsoft Excel, Numbers, and CODAP. Note that the “Customizable Student Handout” uses **Google Sheets** by default. Edit the instructions as needed if you are using a different program.

Select a program that you and your students are familiar with, or one that you want them to practice using. If you want students to use the same spreadsheet program for later graphing and statistical analyses, the program you choose may also impact the preferred data orientation students use to create the combined class data set in Part 6.

Make sure students are familiar with the basic functions of the program that they will need to use for the activity. In the default “Customizable Student Handout,” students are asked to:

- enter and edit data in cells
- add, copy, and rename tabs in a spreadsheet
- answer some questions about data sorting and the AVERAGE function

You may want to screenshot or demonstrate certain functions to the class. Students could also use the BioInteractive **“Spreadsheet Tutorial”** activities to learn or practice core spreadsheet skills in Microsoft Excel or Google Sheets.

Students can create their own spreadsheet file, or you can provide them with a starting template. For example, in Google Sheets:

- Students can create a spreadsheet file in their own Google account, then share or submit their file at the end of the activity.
- Alternatively, you can create blank or “template” spreadsheet files in a Google Drive folder for each student group in advance. This will allow you to monitor their progress during the activity.

You may also want to consider guidelines for making spreadsheet files accessible to people with disabilities (example for Microsoft Excel).

**Customizing the Student Handout**

The “Customizable Student Handout” is a template that contains example instructions and questions for students. Customize this handout based on your selected class investigation and the needs of your class.
Editable versions of the customizable handout are provided in the “Materials” box on this resource’s webpage in the Resource Google Folder (Link). The Google Drive folder is set as “View Only”; to save a copy of a document in this folder to your Google Drive, open that document, then select File → “Make a copy.”

General instructions for customizing this handout are as follows:

- The customizable handout contains “educator notes” on what you should fill in or modify. Delete all educator notes before distributing the handout to students.
  - In editable versions of the customizable handout, the educator notes are indicated using comment bubbles. Remove/resolve the comments as you complete your customizations.
  - The appendix of this document contains a reference copy of the customizable handout in which all the educator notes are labeled and colored within the text. This copy is not directly editable, but you can use it to confirm that you have viewed and addressed all the educator notes.
- You do not need to use everything in the customizable handout, just the parts/questions that are relevant to your class. Delete the parts/questions you do not use.
- You may want students to do various parts of the handout as group work, class discussions, or projects (presentations, posters, etc.). Modify the instructions in the handout accordingly.
- Once you are finished customizing, carefully review each part for completion and consistency before distributing it to students.

In particular, make sure to complete the following sections of the “Customizable Student Handout” for your selected investigation. More details for each section are provided in the “Procedure” section:

- **Background**: Include background information for the selected investigation.
- **Materials**: Include the materials that students will need to conduct the investigation.
- **Part 2**: Provide a procedure for the investigation. You will want to make sure all student groups are measuring and recording the same variables.
- **Part 5**: Optionally change the example data tables (Tables 1–5). By default, the examples are from the ruler drop experiment. To reduce cognitive load for students, you may want to replace these examples with representative data from the selected investigation or another study that is more relevant to your class.
- **Part 7**: Optionally change the example data tables (Tables 6–7). By default, the examples are from a study on finches. You may want to replace them with data from another study that is more relevant to your class.

For more ideas on how to customize the handout, download the example for the ruler drop investigation on this resource’s webpage. This example includes:

- **“Ruler Drop Student Handout”**: a ready-to-use student handout based on the “Customizable Student Handout” (can be further customized as needed)
- **“Ruler Drop Educator Materials”**: a guide for educators with more information on the ruler drop investigation and how it was translated into the customizable handout
PROCEDURE FOR CUSTOMIZATION AND IMPLEMENTATION

This section suggests procedures for completing the “Customizable Student Handout” and implementing the activity in class (use links to go directly to each section). Adjust the customizable handout as needed based on your activity.

- **Background Section**
- **Part 1: Writing the Research Question and Hypothesis**
- **Part 2: Designing the Methods and Setting Up Your Data Sheet**
  - Recording Data on a Data Sheet
- **Part 3: Collecting Data and Spreadsheet Entry**
- **Part 4: Comparing Spreadsheets**
- **Part 5: Organizing Data in Spreadsheets**
- **Part 6: Preparing to Combine Data**
  - Data Orientation
- **Part 7: Documenting Your Data**
- **Next Steps**

**Background Section**

Use the start of the activity to orient students to the investigation they will be doing. Students will need enough information to clearly define a research question, describe the relevance of their research, and generate a hypothesis in Part 1.

Students can generate questions after learning about a phenomenon that they will explore and explain through their investigation. You could introduce the phenomenon through a video clip, a lab demo, or anything else that sets the stage for students asking and generating research questions.

Based on the facilitation approach you choose, you can predefine the parameters for the investigation in the handout, or you can collaboratively design parts of the investigation via class discussion. For example, consider an investigation where students compare plants grown in different environments, such as sun or shade:

- **Predefined:** You could tell students to compare plants grown in sunny vs. shady environments.
- **Collaborative:** You could have a class discussion about what environmental differences are present outside your school, then have the class pick two environments to compare.

Similarly, you can predefine what variables should be measured in the handout, or have students discuss and collaboratively decide which variables to measure. Using the same example with plants:

- **Predefined:** You could tell students to collect data on leaf length.
- **Collaborative:** You could have a class discussion about which plant or leaf characteristics could be affected by the environment, then have the class pick which variable to measure.

Keep in mind that the data collected by each group will need to be combined into a class data set with a consistent organization. Thus, all groups must ultimately converge upon the same investigation and research question.

**PART 1: Writing the Research Question and Hypothesis**

Students write their research question and hypothesis and speculate on the importance and/or relevance of their investigation.
The tool *How Science Works* can be used to engage students with the overall scientific process. The activity “*Asking Scientific Questions*” can be used to help students develop research questions.

The handout suggests that students work in pairs, but you may modify group size based on your class’s needs. Smaller groups are preferred for this part of the activity, as groups are combined for discussion in Part 4.

As in the “*Background Section*,” you can redefine the investigation in the handout, including the parameters and variables that students will measure. Or you can decide on the investigation, parameters, and variables in a class discussion. These decisions will impact how students are generating a research question and hypothesis. Continuing with the plant leaf example:

- **Predefined:** You could tell students they will be collecting leaf length data on five leaves from a sunny environment and five leaves from a shady environment. Therefore, their research question should be directly related to this defined investigation.
- **Collaborative:** You could have a class discussion about which plant or leaf characteristics could be affected by the environment, then have students generate the class research question together.

If the scope of your class includes statistical analysis, and you would like your students to understand one-sided and two-sided hypotheses, consider adding the following question (or a variant thereof) after Question 4 in the default “*Customizable Student Handout*”:

---

There are two ways to write an alternative hypothesis. A **one-sided (directional) hypothesis** specifies whether the difference or effect will be positive or negative. A **two-sided (nondirectional) hypothesis** does not specify whether the difference or effect will be positive or negative.

- a. Write an alternative hypothesis for your research question, if you have not done so already.
- b. Is your alternative hypothesis one-sided or two-sided?
- c. Rewrite your alternative hypothesis below. If it was one-sided, rewrite it as two-sided. If it was two-sided, rewrite it as one-sided.

---

**PART 2: Designing the Methods and Setting Up Your Data Sheet**

Students plan how to collect their data. As in Part 1, there are multiple ways to decide on the investigation procedure.

- **Predefined:** You could provide detailed predetermined instructions for what students need to do in the handout. You may also want to demonstrate the data collection procedure before students do it themselves.
- **Collaborative:** You could provide more general guidance in the handout, then have students collectively develop the details of the data collection procedure via class discussion. Make sure the class agrees to use consistent rules for taking measurements (e.g., with the plant example, measuring plant height without touching the plant vs. grabbing the top leaf and stretching the plant).
  - The “Ruler Drop Student Handout” example on [this resource’s webpage](https://www.BioInteractive.org) has prompts to facilitate brainstorming and deciding an investigation procedure with students.

Question 5 in the “*Customizable Student Handout*” asks students which variables they will measure. Make sure all students in the class plan to measure the same variables. If some students measure different variables (e.g., leaf width instead of leaf length), they will be unable to combine their data into the class data set later.
As students are collecting data in the next part, circulate around the groups to ensure that they are all following the same data collection procedures, measuring the same variables, etc. Also make sure students recognize that:
● variables must be labeled on their data sheet
● values need to be recorded in a consistent manner
● their data sheet should be organized in an understandable way

Recording Data on a Data Sheet
The “Customizable Student Handout” instructs students to collect their data on a data sheet (e.g., blank paper or separate online document) first before entering it into a spreadsheet program. This extra step can be useful when:
● The way that students record data during the experiment differs from how the data needs to be organized for later analysis.
● You want students to keep records of data collection (e.g., lab notebooks) as a standard practice.
● You want students to have a backup of their data in case they make errors entering it into a spreadsheet.
● (For paper sheets) Students have limited access to digital devices or are working in a field/lab setting where it would be difficult to keep a digital device clean.

Depending on time constraints, classroom design, accessibility, and other considerations, directly entering data into a spreadsheet may be better for your students. If so, adjust the handout to omit the data sheet instructions.
● Before students enter their data, you may still want them to sketch their spreadsheet design to help them think more intentionally about how to set up the spreadsheet.

PART 3: Collecting Data and Spreadsheet Entry
Students conduct their investigation and record their data on their paper data sheet. After they have finished collecting data, they will enter it into a spreadsheet program of your choosing. (As a reminder, the default “Customizable Student Handout” uses Google Sheets. Edit the instructions as needed if you are using a different program.)

Encourage your students to consider whether they should organize their data differently in the spreadsheet compared to their paper datasheet. If you are doing this activity over multiple class periods, the end of Part 3 is a good pause point.

PART 4: Comparing Spreadsheets
Students compare their data sheets with those of another group. Combine student groups as appropriate. For example, student pairs from the previous parts of the activity can be combined into groups of four.

Emphasize that there are no “right” or “wrong” ways for students to have initially organized their data. The purpose of comparing with another group is to learn how others approached the task and to reflect on the elements that make a spreadsheet easy to understand. Highlight the opportunities for students to reflect on their choices.

Students can make minor organizational changes to their data at the end of Part 4, but they should not spend a lot of time reorganizing it. They will decide as a class how to organize everyone’s data later in the activity.
PART 5: Organizing Data in Spreadsheets

Students explore examples of common data organization issues that limit the use of important spreadsheet functions (calculations, formulas, sorting, etc.). The “Customizable Student Handout” uses examples from a ruler drop experiment that compared reaction time between students’ dominant and nondominant hands, but you can substitute data from your class’s own investigation or another relevant example.

Tables 1–5 should show the same data organized in different ways:

- **Table 1** should show the data organized according to standard practices and in long-form orientation.
- **Table 2** should include a list of each trial measurement in the same cell (separated by commas) instead of one measurement per cell.
- **Table 3** should include units with the numbers in each cell.
- **Table 4** should include inconsistent labels for the same group, so that sorting alphabetically would misorder the data.
- **Table 5** should show the data in wide-form orientation.

You can modify or remove examples based on the spreadsheet functions that are relevant for your class. If desired, you can also more broadly discuss effective practices for spreadsheet organization and/or include more examples not already in the activity.

- For additional examples, refer to Broman and Woo (2018) and the Data Carpentry module “Data Organization in Spreadsheets for Ecologists.”

Students can return to working in pairs or continue to work in their combined groups from Part 4. Alternatively, you can organize students into jigsaw groups, where each student group is assigned one of the four examples (Questions 16, 17, 18, or 21); after each group works through their assigned example, students recombine into mixed groups to share their insights.

Students can work through the data table examples directly in the handout. Or you could have them examine the examples in a spreadsheet program, so that they can experience the issues with the spreadsheet functions directly.

- You can ask students to attempt the tasks described in the handout (calculating means, sorting, etc.) in a spreadsheet program. This may take more time and require students to be more familiar with using spreadsheets.
  - For comparing wide-form and long-form data in Tables 1 and 5, students could compare the ease of calculating averages for each hand treatment.
- You can find spreadsheet files for the ruler drop data tables in the “Materials” box on this resource’s webpage.

Students can make minor organizational changes to their data at the end of Part 5. But, again, they should not spend a lot of time reorganizing it at this point, as they will be making final data organization decisions after a class discussion.

PART 6: Preparing to Combine Data

Students should now have ideas for how to effectively organize data in a spreadsheet. They will use what they’ve learned to decide the organization for a class data set, which combines the data from all the groups.
Engage the class in a discussion about how to organize the class data set, following the effective practices that they previously explored. If the data needs to be organized in a certain way for the graphing or analysis tools that your class is using, guide the class discussion accordingly. Make sure to cover the following decisions:

- how to organize the columns of the spreadsheet, including what labels/headers to use
- what units to use for the measurements. (Students who used different units can convert their data by creating a new column and using spreadsheet formulas; show them how to do this if they are unfamiliar.)
- whether to orient the data in long form or wide form (refer to the “Data Orientation” section below for details)

Students should document the column organization, labels, measurement units, etc., that the class decides should be used. Each group should then convert their data to the class organization.

If your investigation includes human subjects and/or the collection of personal data, like students’ names, then anonymizing personal data may also be necessary. You can add the following question after Question 25 in the “Customizable Student Handout” to encourage students to explore the importance of anonymizing data:

Investigations with human subjects often **anonymize**, or remove identifying details from, data collected from people. For example, if a data set includes people’s names, you can replace each name with another ID, such as a random number or letter.

a. Why do you think researchers anonymize data collected from humans?

b. Does any information in the class data set need to be anonymized? If so, decide as a class how to anonymize this information, then record your decisions below.

Students may describe privacy concerns related to the collection of personal information from human participants in a research study. Answers could also address the potential for researcher bias if identifying information about participants is known when analyzing the data.

**Data Orientation**

A major decision that the class must make is how to orient the data set. For data sets with multiple groups or treatments, there are two main orientations:

- **Long form**: Also called “tidy data” ([Wickham 2014](#)). Each variable forms a column, each observation forms a row, and each cell is a single measurement. In other words, each row contains all the information associated with a set of measurements for each observation (and only one observation). An example of long-form data from the ruler drop investigation is provided in Table 1 in the “Customizable Student Handout.”

- **Wide form**: Some variables may be “spread out” across multiple columns, where each column represents a different value of that variable (e.g., a different group or treatment). An example of wide-form data from the ruler drop investigation is provided in Table 5 in the “Customizable Student Handout.” In this case, the “Hand” variable is no longer a single column, but instead reported in both the “Dominant Hand” and “Nondominant Hand” columns.

Orientation can also affect the default arrangement of the variables for different graph types. Different orientations may be favored by different programs or tools.

- Long form is required by many statistical programs and tools, including BioInteractive’s [Data Explorer](#). It is the typical orientation for most large and/or publicly shared data sets.
● Wide form may be easier to interpret when viewing a small table of data. It can be useful for some analyses (e.g., performing summary calculations) depending on the spreadsheet programs or in paired study designs (when repeated measurements are collected on the same individual).

PART 7: Documenting Your Data

Students learn about documenting data so that it is understandable to others, including those who may want to reproduce their findings or conduct additional analyses. Clear and accurate experimental design, methodology, data recording, and data analysis enable future researchers to replicate a research study and/or apply its approach to a new population. (There may be circumstances that prevent complete reproducibility; for example, if the exact population in which the study was conducted no longer exists.)

Students should document each column of their data sets using additional descriptions, called “metadata.” As an example, the default “Customizable Student Handout” shows metadata for a data set on finches (from the “Finches in the Galápagos” data set also in Data Explorer).

● If the finch example is not as relevant to your class, you can replace it with another example. For instance, the “Ruler Drop Student Handout” example replaces the finch data with data from a physiological performance study on Karoo bush rats, to better align with the reaction time data collected by students.

Next Steps

At the end of the activity, students should submit their filled-out handouts and finalized spreadsheets according to your directions. You can assess their answers, spreadsheet revisions, metadata, etc., as appropriate for your class.

To create the class data set, you can either:

● Combine all the submitted group data sets into one data set yourself.
● Create a new, central Google Sheet that each group directly adds their data to. In this case:
  o Consider prepopulating the sheet with labels or an example of the final spreadsheet organization as a reminder to students.
  o You may need to check for consistency and make some corrections after all students have entered their data.
  o To prevent accidental changes, you may want to “lock” the spreadsheet once it’s done. Students can copy or download the spreadsheet for individual analyses.

Students can use the class data set to explore whether their hypotheses from Part 1 are supported via summary calculations, graphing, and/or statistical analyses. You may direct them to use additional resources and tools, such as the following:

● The web tool Data Explorer provides a quick and easy way to visualize and analyze data. It includes a variety of options for generating plots and performing statistical tests. If you plan to upload class data, please review the guidelines under “Import your own data” on the landing page of the tool.
● Provide students with scaffolded support for selecting appropriate graph types and statistical tests to visualize their data and test their hypotheses. The “Educator Materials” and “Selection Guides” for Data Explorer can be used for guidance.
● The “Spreadsheet Tutorials” can be used to help students learn about calculations or graphing in Microsoft Excel or Google Sheets.
ASSESSMENT GUIDANCE

The questions and sample answers in this section are based on the default “Customizable Student Handout.” Modify them as needed based on your customizations. Many student answers will vary based on the selected investigation. The “Ruler Drop Educator Materials” document provides examples of questions and answers customized for a specific investigation.

PART 1: Writing Your Research Question and Hypothesis

1. Write the research question that your class will be investigating.
   The specific research question will vary based on the investigation you selected and your facilitation approach. It should reflect either the question provided by you (predefined approach) or decided on by the class (collaborative approach). All research questions should be testable and specifically ask about the effects of an independent variable(s) on a dependent variable(s).

2. Describe at least one motivation for doing this investigation. For example, how might it be relevant to you, the public, or the scientific community?
   Students may discuss how the investigation is relevant to their own interests, community, or general human/environmental well-being based on their personal experience, prior knowledge, or other sources of information. Be open to a range of reasonable responses.

3. Write a hypothesis for the research question.
   The hypothesis should state the results that the student expects from the investigation.

4. Is your hypothesis a null hypothesis, meaning you do not expect a difference between groups or an effect based on the independent variable? Or is it an alternative hypothesis, meaning you do expect a difference or effect?
   Student answers will vary based on their response to Question 3.

PART 2: Designing the Methods and Setting Up Your Data Sheet

5. List the variables that your class will measure in this investigation.
   Student answers will vary depending on the investigation. Make sure that all students are planning to measure the same variables before they start collecting data.

6. Label each variable you listed as either an independent or dependent variable.
   Students should label the following as independent variables: treatments, conditions that they manipulate, or predictors they plan to compare. Students should label the following as dependent variables: variables that they measure or record as responses to conditions. Depending on the investigation, you could also modify this question to include control variables.

7. Consider your hypothesis (Question 3) and the variables you’ll be measuring (Question 5). Describe what result you could obtain from these variables that would support your hypothesis.
   Depending on the investigation, students may consider how data would differ between groups or how one variable may change in response to another. They should be able to describe expected trends in the data related to their hypothesis.

PART 3: Collecting Data and Spreadsheet Entry

8. What columns does your spreadsheet have?

9. What does each row in your spreadsheet represent?
   Columns and rows will vary depending on the design of the investigation. Students may have observations in rows or columns based on groups or replicates.
10. Why did you decide to arrange the data in your spreadsheet in this way?

*Students should be able to justify their spreadsheet organization. You should expect that spreadsheet organization will vary across student groups and emphasize that there is no “right” way to organize a spreadsheet at this point. (Later in the activity, students will modify their spreadsheets after iterative experiences with their peers, including self-reflection and exercises designed to help them discover and correct misconceptions.)*

11. What other information, if any, is needed to make your spreadsheet understandable to someone unfamiliar with the investigation?

*Students may identify variables that need additional labels or other information.*

PART 4: Comparing Spreadsheets

12. Consider the layout and variables in your spreadsheet and the other group’s.
   a. What are some similarities between your spreadsheets?
   b. What are some differences?

   *Similarities or differences may include the organization of the data (the placement of variables in rows and columns), the information recorded in each cell, which units are used, and the labels selected.*

13. What features of the other group’s spreadsheet help you understand the data they collected?

   *Students may identify structures or labeling used by the other group that helped explain the variables that were measured and what data were collected.*

14. Consider the units used for your measurements.
   a. What units did you use?
   b. What units did the other group use?
   c. Were the units originally labeled on your spreadsheets, and if so, where did each group place their unit labels?

   *Student answers will vary depending on the procedures provided by you or decided during class discussions. For distance measurements, for example, students may have used units of inches, centimeters, or millimeters.*

15. Data sets can combine data from multiple sources. If you were to combine your data with the other group's, how might you change the organization or structure of your spreadsheet?

   *Student answers will vary. Note that students are only being asked to reflect on potential changes at this point. They should not make any major changes to their spreadsheets until later in the activity.*

PART 5: Organizing Data in Spreadsheets

16. Table 2 shows how the first student group organized the data.

   **Table 2. First example of how students organized the data in Table 1.**

<table>
<thead>
<tr>
<th>Column A</th>
<th>Column B</th>
<th>Column C (cm)</th>
<th>Column D (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>Hand</td>
<td>Distance</td>
<td>Time</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td>Dominant</td>
<td>30.7, 17.7</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>Nondominant</td>
<td>35.8, 34.4</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>Dominant</td>
<td>11.0, 14.2</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>Nondominant</td>
<td>47.1, 15.9</td>
</tr>
</tbody>
</table>
The students tried to calculate the mean (average) distance for each type of hand using the AVERAGE function in a spreadsheet program. This function needs each cell to contain an individual numerical value.

a. The AVERAGE function didn’t work with Table 2. Why do you think this is?

*Both the “Distance” and “Time” columns have multiple measurements in each cell. The AVERAGE function doesn’t work because it needs each cell to contain an individual number.*

b. Describe the data organization practice used by Table 1, but not by Table 2, that allows the AVERAGE function to work with these data.

*The data organization practice is listing only one measurement/number in each cell. Table 1 does this by adding a “Trial” column that allows the numbers in the “Distance” and “Time” measurements to be separated into single cells in individual rows.*

17. Table 3 shows how the second student group organized the data.

**Table 3.** Second example of how students organized the data in Table 1.

<table>
<thead>
<tr>
<th>Column A</th>
<th>Column B</th>
<th>Column C</th>
<th>Column D</th>
<th>Column E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Student</td>
<td>Hand</td>
<td>Trial</td>
<td>Distance</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>Dominant</td>
<td>1</td>
<td>30.7cm</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>Dominant</td>
<td>2</td>
<td>17.7cm</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>Nondominant</td>
<td>1</td>
<td>35.8cm</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>Nondominant</td>
<td>2</td>
<td>34.4cm</td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>Dominant</td>
<td>1</td>
<td>11.0cm</td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>Dominant</td>
<td>2</td>
<td>14.2cm</td>
</tr>
<tr>
<td>8</td>
<td>B</td>
<td>Nondominant</td>
<td>1</td>
<td>47.1cm</td>
</tr>
<tr>
<td>9</td>
<td>B</td>
<td>Nondominant</td>
<td>2</td>
<td>15.9cm</td>
</tr>
</tbody>
</table>

The students tried to calculate the mean (average) for all the time measurements using the AVERAGE function. Again, this function needs each cell to contain an individual numerical value.

a. The AVERAGE function didn’t work with Table 3. Why do you think this is?

*Both the “Distance” and “Time” columns have the units written in text after each number. The AVERAGE function won’t work because it needs cells with just numbers. (In general, text in cells may prevent a spreadsheet program from recognizing numerical data in the cells.)*

b. Describe the data organization practice used by Table 1, but not by Table 3, that allows the AVERAGE function to work with these data.

*The data organization practice is removing text after the numbers in cells. Table 1 does this by removing the units in the cells and writing the units in the column headings instead.*

18. Table 4 shows how the third student group organized the data.
Table 4. Third example of how students organized the data in Table 1.

<table>
<thead>
<tr>
<th>Column A</th>
<th>Column B</th>
<th>Column C</th>
<th>Column D</th>
<th>Column E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Student</td>
<td>Hand</td>
<td>Trial</td>
<td>Distance (cm)</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>Dominant</td>
<td>1</td>
<td>30.7</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>Dominant</td>
<td>2</td>
<td>17.7</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>Nondominant</td>
<td>1</td>
<td>35.8</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>Nondominant</td>
<td>2</td>
<td>34.4</td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>SecondHand: Dom</td>
<td>1</td>
<td>11.0</td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>SecondHand: Dom</td>
<td>2</td>
<td>14.2</td>
</tr>
<tr>
<td>8</td>
<td>B</td>
<td>FirstHand: Nondom</td>
<td>1</td>
<td>47.1</td>
</tr>
<tr>
<td>9</td>
<td>B</td>
<td>FirstHand: Nondom</td>
<td>2</td>
<td>15.9</td>
</tr>
</tbody>
</table>

The students wanted to group the data based on “Hand” type (Column B) for additional calculations. They wanted all the rows of dominant hand data grouped together, followed by all the rows of nondominant hand data grouped together.

To reorder the data, they used the SORT function in a spreadsheet program, which sorts rows by the values in a specific column (in this case, the “Hand” column). The function sorts from smallest to largest values for numerical variables or alphabetically (A to Z) for text variables.

a. The SORT function didn’t group the data in Table 4 the way the students wanted. Why do you think this is?

_The treatments in the “Hand” column are named inconsistently. For example, the dominant hand treatment is called both “Dominant Hand” and “SecondHand: Dom.” When sorted alphabetically, the treatments will not be grouped together because some labels have different starting letters._

b. Describe the data organization practice used by Table 1, but not by Table 4, that allows the SORT function to group these data the way the students wanted.

_The data organization practice is naming groups/treatments consistently. Table 1 does this by using the same names (either “Dominant” or “Nondominant”) for each “Hand” treatment._

19. Now that you’ve identified several beneficial data organization practices, what other changes (if any) would you make to your own spreadsheet and why?

_Student answers will vary based on the initial layout of their data._

20. Based on Table 1 (long form) and Table 5 (wide form), what is the main difference between these two orientations?

_The long-form example (Table 1) places each trial in its own row, with separate columns for the measurements of each treatment. So, the “Dominant” and “Nondominant” treatments each have their own “Distance” and “Time” columns._

_The wide-form example (Table 5) places each observation in its own row, with the treatment being listed in its own column: the “Hand” column (recorded as either “Dominant” or “Nondominant”). Thus, all the “Distance” measurements are in a single column, as are all the “Time” measurements._

21. The orientation of the data may affect certain spreadsheet functions, such as calculating new variables or performing summary calculations. Consider the following examples based on the data in Tables 1 and 5.
a. A student wants to convert the time measurements from milliseconds (ms) to seconds (s) in a new column in the spreadsheet. Would this task be easier in Table 1 (long form) or Table 5 (wide form), and why?

It would probably be easier in Table 5. If the student uses Table 1, they would need to create two new columns: one to convert the “Dominant Time” column and another to convert the “Nondominant Time” column. If the student uses Table 5, they only need to create one new column to convert the “Time” column.

(In general, wide-form orientation is easier for calculations that need to be performed for each observation.)

b. A student needs to calculate the average time for each “Hand” treatment in a new row in the spreadsheet. Would this task be easier in Table 1 (long form) or Table 5 (wide form), and why?

It would probably be easier in Table 1. The student could simply add a new row at the bottom of Table 1 to calculate an average for each column, including the average time for each “Hand” treatment. Calculating the average time for each treatment in Table 5 would be more difficult, since the student would have to sort the data by “Hand” treatment and specify specific ranges of data for calculating the averages.

(In general, long-form orientation is easier for doing summary calculations, such as averages, per treatment.)

22. Return to the spreadsheet that you made earlier in the activity. Did you originally enter your data in long form or wide form?

Student answers will vary based on the initial layout of their data. The “Data Orientation” section of this document provides more considerations that you may want to discuss with your class.

PART 6: Preparing to Combine Data

23. What columns will the class data set have? List the headers/labels for each column.

24. What units will the class data set use for the measurements? Which, if any, of your units must you convert to match?

25. Which orientation (long form or wide form) will the class data set use?

Answers for Q23 through Q25 will depend on the data organization you decide. Make sure all students organize their data in the same way before combining into the class data set.

PART 7: Documenting Your Data

26. Why is it important for scientific findings to be reproducible, meaning that similar results are found by different people? Give a specific example where it’s important to reproduce the result of a study.

Reproducing the results of a study helps validate those results. This could be important for confirming high-stakes results, such as the outcome of a medical study that will affect the development of medicines and policies. It can also help check whether similar results still apply under different conditions or in different populations.

27. Based on just Table 6:

a. What can you tell about this study so far?

Students may infer that the study included at least 10 different birds, some of which did or did not
survive an unspecified event. Students may also recognize that the study measured several physical characteristics of the birds (body mass, tarsus length, and beak depth).

b. What seems confusing or requires more information to understand?

   Students may point out that there is no information about the units of the measurements, when/how the data were collected, and what the birds did or did not survive. Some terms like “tarsus length” may also be unfamiliar and require clarification.

28. Based on Table 7:

   a. What did you learn from Table 7 that was not clear from Table 6 alone?

      Students may say that the metadata defined certain terms, clarified how measurements were taken and which units were used, and provided more background on the study (e.g., dates of data collection and who collected the data).

   b. Why do you think metadata is important to include?

      Metadata is important for anyone who is trying to understand or replicate a study. It allows the people who did the study to explain what they did in more detail without making the data set labels too long or complicated.

REFERENCES


CREDITS
Written by Kristine Grayson, University of Richmond, VA; Kaitlin Bonner, St. John Fisher University, NY; Esther Shyu, HHMI
Edited by Valerie May, Woodstock Academy, CT; Paul Beardsley, California State Polytechnic University-Pomona, CA; Ann Brokaw, Rocky River High School, OH; Jason Crean, HHMI
Setting Up Your Data for Success

APPENDIX: Customizable Student Handout with Notes
This appendix provides a copy of the “Customizable Student Handout” with all the educator notes labeled within the text as follows:
- Red notes labeled “Educator to Complete” indicate parts you are required to fill in.
- Blue notes labeled “Educator Information” indicate additional notes or optional adjustments.

This copy of the handout is not directly editable and is meant to be used as a reference. For more information, including where to find editable versions of the customizable handout, refer to the “Customizing the Student Handout” section.

INTRODUCTION
Organizing data in a clear, understandable way is important for many lines of work — from science and public policy to business and computing. In this activity, you will decide how to collect and organize data for a class investigation. This process will help you better understand how scientists and other professionals make similar decisions.

During this activity, you will:
1. Write a research question and hypothesis.
2. Identify the variables to measure, then decide how to organize the data you collect.
3. Compare the data you collected and organized into a spreadsheet with that of your classmates.
4. Learn more about effective practices for organizing data.
5. Decide the final organization of your data as a class. This will allow everyone to combine their data into one data set for later analysis.
6. Document your data so that it can be understood and used by others.

BACKGROUND
Educator to Complete: Briefly describe the investigation that students will be conducting, including the context or phenomenon that motivates investigation. (Alternatively, provide this information during an in-class introduction or discussion.) For more guidance on customizing this section, refer to the “Background Section” procedure in the “Educator Materials.”

MATERIALS
- sheet to record data [Educator Information: Specify blank paper, handout, lab notebook, online document, etc., if needed.]
- access to Google Sheets [Educator Information: Change program if needed.]
- [Educator Information: Include any other materials that students will need for the investigation.]

PART 1: Writing the Research Question and Hypothesis
Based on what you know about the investigation, answer the following questions. Work in pairs or as directed by your instructor.
1. Write the research question that your class will investigate. [Educator Information: This assumes that students have already converged on a main research question. Refer to the “Selecting a Facilitation Approach” section of the “Educator Materials” for guidance on either predefining the investigation or collaboratively designing it with students.]

2. Describe at least one motivation for doing this investigation. For example, how might the investigation be relevant to you, the public, or the scientific community?

3. Write a hypothesis for the research question.

4. Is your hypothesis a null hypothesis, meaning you do not expect a difference between groups or an effect based on the independent variable? Or is it an alternative hypothesis, meaning you do expect a difference or effect?

PART 2: Designing the Methods and Setting Up Your Data Sheet

To collect data for this investigation, your class will use the following procedure:

Educator to Complete: Describe the procedure for the investigation. You can provide detailed instructions or more general guidance that will be supplemented with class discussion. For the latter, you may have students write parts of the data collection procedure together.

You’ll need to record your data on a data sheet, which should be clearly labeled with the goal of being understandable to others. Decide how you will organize your data sheet to record the data in your investigation, then create your data sheet. [Educator Information: Adjust these instructions if your students are directly entering their data into a spreadsheet program. For more information, refer to the “Part 2” procedure in the “Educator Materials.”]

5. List the variables that your class will measure in this investigation.

6. Label each variable you listed as either an independent or dependent variable. [Educator Information: Adjust the terminology for describing variables as needed.]

7. Consider your hypothesis (Question 3) and the variables you’ll be measuring (Question 5). Describe what result you could obtain from these variables that would support your hypothesis.

PART 3: Collecting Data and Spreadsheet Entry

Collect your data following the procedure in Part 2. Record your results on your data sheet.

After collecting your measurements, you will decide how to arrange and label your data in a spreadsheet program. Create or access a Google Sheet, as directed by your instructor: [Educator Information: These instructions are based on Google Sheets. Adjust as needed based on the program and files you want students to use.]

- Right-click or select the arrow on the “Sheet1” tab in the bottom-left corner, then rename the tab as “Original Data Collection.”
- Enter the data from your paper data sheet into the columns and rows of the spreadsheet.
- You can organize the data in the spreadsheet differently from how you organized it on paper. Just make sure the data in the spreadsheet are clearly labeled and understandable.

Answer the following questions based on your spreadsheet.

8. What columns does your spreadsheet have?
9. What does each row in your spreadsheet represent?

10. Why did you decide to arrange the data in your spreadsheet in this way?

11. What other information, if any, is needed to make your spreadsheet understandable to someone unfamiliar with the investigation?

PART 4: Comparing Spreadsheets

You’ll now compare your spreadsheet with that of another group of students, or as directed by your instructor, to learn how each group organized their data. Note that there are no “wrong” ways to have organized the spreadsheet. The goal is to learn how different groups approached the same task and to reflect on what makes a spreadsheet easier to understand.

After comparing spreadsheets, answer the following questions.

12. Consider the layout and variables in your spreadsheet and the other group’s.
   a. What are some similarities between your spreadsheets?
   b. What are some differences?

13. What features of the other group’s spreadsheet help you understand the data they collected?

14. Consider the units used for your measurements.
   a. What units did you use?
   b. What units did the other group use?
   c. Were the units originally labeled on your spreadsheets, and if so, where did each group place their unit labels?

15. Data sets can combine data from multiple sources. If you were to combine your data with the other group’s, how might you change the organization or structure of your spreadsheet?

Do not make any changes to your originally collected data. Use the instructions below to create a copy of your first spreadsheet tab: [Educator Information: These instructions are based on Google Sheets. Adjust as needed based on the program and files you want students to use.]

- Right-click or select the down arrow on your “Original Data Collection” tab at the bottom of the sheet, then select “Duplicate” to create a copy.
- Right-click on the new spreadsheet tab currently titled “Copy of Original Data Collection” and rename the tab as “Revised Data Collection.”

As you learn about the details of spreadsheet organization, you can make small changes in your “Revised Data Collection” tab. Do not make large changes to the organization or structure of your spreadsheet just yet, as we will be making final data organization decisions as a class later.

PART 5: Organizing Data in Spreadsheets

Although there are many ways to collect and record data, data sets are easier to understand when they are organized consistently. So, scientists and other professionals use standard practices for organizing data in spreadsheets. These practices make data easier to understand and share, as well as easier to summarize, visualize, and analyze using spreadsheet programs. You’ll explore how data organization affects spreadsheet program functions in the examples below.
The examples are from an experiment called the “ruler drop” test. A student dropped a ruler and another student tried to catch it with each hand (dominant or nondominant) multiple times. They recorded the distance where they caught the ruler and the amount of time they took to catch it. Table 1 shows a portion of the data from this experiment. [Educator Information: You can replace the examples with data from your selected investigation or another investigation more relevant to your class. Modify the descriptions and Tables 1–5 if so.]

**Table 1.** Data from the ruler drop test organized using standard data organization practices.

<table>
<thead>
<tr>
<th>Column A</th>
<th>Column B</th>
<th>Column C</th>
<th>Column D</th>
<th>Column E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Student</td>
<td>Hand</td>
<td>Trial</td>
<td>Distance (cm)</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>Dominant</td>
<td>1</td>
<td>30.7</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>Dominant</td>
<td>2</td>
<td>17.7</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>Nondominant</td>
<td>1</td>
<td>35.8</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>Nondominant</td>
<td>2</td>
<td>34.4</td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>Dominant</td>
<td>1</td>
<td>11.0</td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>Dominant</td>
<td>2</td>
<td>14.2</td>
</tr>
<tr>
<td>8</td>
<td>B</td>
<td>Nondominant</td>
<td>1</td>
<td>47.1</td>
</tr>
<tr>
<td>9</td>
<td>B</td>
<td>Nondominant</td>
<td>2</td>
<td>15.9</td>
</tr>
</tbody>
</table>

Before deciding on the data organization shown in Table 1, the student groups doing the experiment had originally organized these data in different ways. However, they discovered that some of the ways their data were originally organized limited important spreadsheet functions, such as performing additional calculations.

16. Table 2 shows how the first student group organized the data.

**Table 2.** First example of how students organized the data in Table 1.

<table>
<thead>
<tr>
<th>Column A</th>
<th>Column B</th>
<th>Column C</th>
<th>Column D</th>
<th>Column E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Student</td>
<td>Hand</td>
<td>Distance (cm)</td>
<td>Time (ms)</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>Dominant</td>
<td>30.7, 17.7</td>
<td>250.2, 190.0</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>Nondominant</td>
<td>35.8, 34.4</td>
<td>270.2, 264.8</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>Dominant</td>
<td>11.0, 14.2</td>
<td>149.8, 170.1</td>
</tr>
<tr>
<td>5</td>
<td>B</td>
<td>Nondominant</td>
<td>47.1, 15.9</td>
<td>309.9, 180.0</td>
</tr>
</tbody>
</table>

The students tried to calculate the mean (average) distance for each type of hand using the AVERAGE function in a spreadsheet program. This function needs each cell to contain an individual numerical value.

a. The AVERAGE function didn’t work with Table 2. Why do you think this is?

b. Describe the data organization practice used by Table 1, but not by Table 2, that allows the AVERAGE function to work with these data.

17. Table 3 shows how the second student group organized the data.

**Table 3.** Second example of how students organized the data in Table 1.

<table>
<thead>
<tr>
<th>Column A</th>
<th>Column B</th>
<th>Column C</th>
<th>Column D</th>
<th>Column E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Student</td>
<td>Hand</td>
<td>Trial</td>
<td>Distance</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>Dominant</td>
<td>1</td>
<td>30.7cm</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>Dominant</td>
<td>2</td>
<td>17.7cm</td>
</tr>
</tbody>
</table>
The students tried to calculate the mean (average) for all the time measurements using the AVERAGE function. Again, this function needs each cell to contain an individual numerical value.

a. The AVERAGE function didn’t work with Table 3. Why do you think this is?

b. Describe the data organization practice used by Table 1, but not by Table 3, that allows the AVERAGE function to work with these data.

18. Table 4 shows how the third student group organized the data.

<table>
<thead>
<tr>
<th>Column A</th>
<th>Column B</th>
<th>Column C</th>
<th>Column D</th>
<th>Column E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Student</td>
<td>Hand</td>
<td>Trial</td>
<td>Distance (cm)</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>Dominant</td>
<td>1</td>
<td>30.7</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>Dominant</td>
<td>2</td>
<td>17.7</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>Nondominant</td>
<td>1</td>
<td>35.8</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>Nondominant</td>
<td>2</td>
<td>34.4</td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>SecondHand: Dom</td>
<td>1</td>
<td>11.0</td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>SecondHand: Dom</td>
<td>2</td>
<td>14.2</td>
</tr>
<tr>
<td>8</td>
<td>B</td>
<td>FirstHand: Nondom</td>
<td>1</td>
<td>47.1</td>
</tr>
<tr>
<td>9</td>
<td>B</td>
<td>FirstHand: Nondom</td>
<td>2</td>
<td>15.9</td>
</tr>
</tbody>
</table>

The students wanted to group the data based on “Hand” type (Column B) for additional calculations. They wanted all the rows of dominant hand data grouped together, followed by all the rows of nondominant hand data grouped together.

To reorder the data, they used the SORT function in a spreadsheet program, which sorts rows by the values in a specific column (in this case, the “Hand” column). The function sorts from smallest to largest values for numerical variables or alphabetically (A to Z) for text variables.

a. The SORT function didn’t group the data in Table 4 the way the students wanted. Why do you think this is?

b. Describe the data organization practice used by Table 1, but not by Table 4, that allows the SORT function to group these data the way the students wanted.

19. Now that you’ve identified several beneficial data organization practices, what other changes (if any) would you make to your own spreadsheet and why?
In addition to the practices you’ve identified, you should also consider the orientation, or the overall arrangement, of the data set. There are two main types of data orientation: wide form and long form. Table 1 shows the data organized in long form, and Table 5 shows the same data organized in wide form.

Table 5. The data in Table 1 organized in wide-form orientation.

<table>
<thead>
<tr>
<th></th>
<th>Column A</th>
<th>Column B</th>
<th>Column C</th>
<th>Column D</th>
<th>Column E</th>
<th>Column F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Student</td>
<td>Trial</td>
<td>Dominant</td>
<td>Dominant</td>
<td>Nondominant</td>
<td>Nondominant</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Distance (cm)</td>
<td>Time (ms)</td>
<td>Distance (cm)</td>
<td>Time (ms)</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>1</td>
<td>30.7</td>
<td>250.2</td>
<td>35.8</td>
<td>270.2</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>2</td>
<td>17.7</td>
<td>190.0</td>
<td>34.4</td>
<td>264.8</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>1</td>
<td>11.0</td>
<td>149.8</td>
<td>47.1</td>
<td>309.9</td>
</tr>
<tr>
<td>5</td>
<td>B</td>
<td>2</td>
<td>14.2</td>
<td>170.1</td>
<td>15.9</td>
<td>180.0</td>
</tr>
</tbody>
</table>

20. Based on Table 1 (long form) and Table 5 (wide form), what is the main difference between these two orientations?

21. The orientation of the data may affect certain spreadsheet functions, such as calculating new variables or performing summary calculations. Consider the following examples based on the data in Tables 1 and 5.

   a. A student wants to convert the time measurements from milliseconds (ms) to seconds (s) in a new column in the spreadsheet. Would this task be easier in Table 1 (long form) or Table 5 (wide form), and why?

   b. A student wants to calculate the average time for each “Hand” treatment in a new row in the spreadsheet. Would this task be easier in Table 1 (long form) or Table 5 (wide form), and why?

22. Return to the spreadsheet that you made earlier in the activity. Did you originally enter your data in long form or wide form?

   If you’d like, you can make small organizational changes in your “Revised Data Collection” tab. But again, do not make any large changes to the organization or structure of your spreadsheet yet. Your instructor will discuss which orientation to use for the final organization of your class data set.

PART 6: Preparing to Combine Data

You’ll now prepare to combine data from everyone in the class into one data set, which you can use to explore your research questions later. Following your instructor’s guidance, have a discussion with your class about how to organize the class data set. Once the class has decided which organization to use, record your decisions by answering the questions below.

23. What columns will the class data set have? List the headers/labels for each column.

24. What units will the class data set use for the measurements? Which, if any, of your units must you convert to match?

25. Which orientation (long form or wide form) will the class data set use?

Now make a copy of your “Revised Data Collection” tab (using the same procedure as at the end of Part 4) and rename the new tab “Class Data Collection.” Edit the data in this tab to match the class data set organization, accounting for all the decisions you recorded above.
PART 7: Documenting Your Data

In addition to organizing data in a clear and usable way, you should also document data sets by adding notes and explanations. Documenting data makes it easier to remember details that you might otherwise forget later. It also makes the data more understandable to others, including people who want to replicate your findings or build upon your work.

26. Why is it important for scientific findings to be reproducible, meaning that similar results are found by different people? Give a specific example where it’s important to reproduce the result of a study.

When scientists publish their results, they often share their data publicly in spreadsheet files. These files must contain enough information for others to understand their work. You’ll explore what information is useful to provide through the example below.

Table 6 shows some data from a study on birds called Galápagos finches. [Educator Information: You can replace the examples with data from another investigation more relevant to your class. Modify the descriptions and Tables 6–7 if so.]

Table 6. Data from a study on Galápagos finches.

<table>
<thead>
<tr>
<th>Bird ID</th>
<th>Status</th>
<th>Body mass</th>
<th>Tarsus length</th>
<th>Beak depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>560</td>
<td>survivor</td>
<td>14</td>
<td>19.1</td>
<td>8.5</td>
</tr>
<tr>
<td>561</td>
<td>nonsurvivor</td>
<td>16.5</td>
<td>20</td>
<td>10.2</td>
</tr>
<tr>
<td>564</td>
<td>nonsurvivor</td>
<td>14</td>
<td>18.8</td>
<td>9.3</td>
</tr>
<tr>
<td>572</td>
<td>survivor</td>
<td>18</td>
<td>20.2</td>
<td>10.3</td>
</tr>
<tr>
<td>605</td>
<td>nonsurvivor</td>
<td>15.5</td>
<td>19.9</td>
<td>10.2</td>
</tr>
<tr>
<td>609</td>
<td>nonsurvivor</td>
<td>16.5</td>
<td>19.6</td>
<td>10.5</td>
</tr>
<tr>
<td>610</td>
<td>nonsurvivor</td>
<td>14</td>
<td>18.8</td>
<td>9</td>
</tr>
<tr>
<td>611</td>
<td>nonsurvivor</td>
<td>16</td>
<td>18.9</td>
<td>9.8</td>
</tr>
<tr>
<td>616</td>
<td>survivor</td>
<td>19</td>
<td>20</td>
<td>9.6</td>
</tr>
<tr>
<td>618</td>
<td>survivor</td>
<td>17.5</td>
<td>20.7</td>
<td>9.9</td>
</tr>
</tbody>
</table>

27. Based on just Table 6:
   a. What can you tell about this study so far?
   b. What seems confusing or requires more information to understand?

When scientists share data sets publicly, they often include additional information called metadata to make the data easier to understand. Table 7 shows some metadata for the data set in Table 6.

Table 7. Metadata for the Galápagos finch data.

<table>
<thead>
<tr>
<th>Column Header</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bird ID</td>
<td>A unique identifier for each finch in the study, from the number on a metal leg band</td>
</tr>
</tbody>
</table>
Status | Recorded either as “nonsurvivor” or “survivor.” “Nonsurvivors” died during a drought in 1977. “Survivors” lived through the drought.
---|---
Body mass | The total mass of the finch in grams
Tarsus length | The length of the finch’s tarsus bone, which is the section of the leg from the ankle to the knee in millimeters
Beak depth | Measured from the top to the bottom of the beak at its base in millimeters

28. Based on Table 7:
   a. What did you learn from Table 7 that was not clear from Table 6 alone?
   b. Why do you think metadata is important to include?

You will now add metadata to your own data set. First, add a new sheet to your spreadsheet file (either click the “+” at the bottom left of the screen or Insert → Sheet), then right-click or use the arrow to rename this sheet as “Metadata.” [Educator Information: These instructions are based on Google Sheets. Adjust as needed based on the program and files you want students to use.]

- In Column A, copy-paste each column header from the “Class Data Collection” tab.
- In Column B, write a description for each column, like the examples in Table 7. Provide any relevant information on the variable in the column, including units.

Congratulations! You have created a well-organized and fully documented data set. Submit your data set and this handout as directed by your instructor, who will compile the class data set for further analysis. [Educator Information: Customize submission instructions as needed.]