Data User Guide: Using the Budburst Database

All Budburst data is available for anyone to search, download, and use. Use the Budburst database to create data visualizations and conduct scientific investigations with authentic datasets from across the country. Follow the steps in this guide to get started!

**This guide is written primarily for educators who want to use phenology data from the national Budburst database in their classrooms. However, it includes information and tips for everyone interested in using any type of Budburst data.*

Table of Contents

About Budburst Data

a) Data Overview

Budburst users have been sharing the observations of plants since 2007 and have submitted over 92,0000 observations on plants from all 50 states (and even internationally)! Their observations are freely available for download from the Budburst database, which is searchable by state, dates, species, and other criteria.

Note that like much community science data, Budburst data is collected opportunistically and can therefore be 'patchy'. Budburst community scientists may focus their observations on some species and locations over others, and contributions may vary with time. (i.e., many data points might be available for certain species/years/locations and very little for others). This guide will help you explore the Budburst database so you can find a dataset that works for you.

b) Budburst Database Views

You can access the Budburst database from three different 'views'. Detailed guides on how to search/filter/download data from these different views are available in the 'Getting your Data' document.

1. Full view (full database): Anyone can view and search for data in the full Budburst database, which includes all data collected by individual Budburst users and Groups in all locations.

2. Group view: If you belong to or manage a Budburst Group, you can access and download your Group data separately from the rest of the database. Most classrooms create a Budburst Group to aggregate student data, and some choose to focus their analysis solely on these data rather than including data submitted by other users.

3. Personal view: Budburst members who have submitted data can access all personally submitted data separately from data collected by other users.

c) Budburst Observation Types

Budburst community scientists can participate in several different ongoing research projects, each of which has specific data requirements. Depending on their interests, Budburst users can submit 3 different types of observations. Users can also filter the database by observation type:

1. Phenology: Phenology observations provide data on the timing of seasonal changes in plants. This observation type is used by Budburst members participating in our longest-running research project 'Phenology and Climate' (beginning in 2007).

2. Pollinators: Pollinator observations provide data on the number of pollinator visits to a plant in a specified period. This observation type is used by Budburst members participating in our

'Pollinators and Climate' (beginning in 2020). Many data on pollinator visits to a plant can be collected quickly , making this observation type the go-to for those wanting to collect and work primarily with their own data (see Step 1 below).

3. Milkweed & Monarchs: These observations provide data on the number of monarch butterfly eggs/caterpillars found on milkweed plant species. They are used by Budburst members participating in our 'Milkweed & Monarchs' project (beginning in 2020).

Of the three observation types, phenology observations are submitted most frequently. They have been made on many different species in all 50 states. Most of the examples in this guide thus focus on phenology observations. Still, many pollinator and milkweed & monarch observations have been made for select locations and species.

Now that you have a general understanding of the Budburst database, follow the steps below to help you get started using the Budburst database for your learning purposes!

Step 1: Choosing a focus

Ask yourself the following questions to help determine the dataset you want to work with.

1. Whose data do I want to use?

It is helpful to decide early on whether you want your students to use their own data, data collected by other Budburst users, or both. For those who want to use only student-collected data, we suggest structuring class data collection around that goal. For example, you could:

(a) commit to monitoring phenology on your campus long term to build a robust dataset on select species in your location.

(b) set up a natural phenology experiment as a class (e.g. compare whether trees near hot pavement flower earlier than trees further away from the pavement) or

(c) focus on observation types that allow students to gather lots of data quickly. For example, if they have access to some flowering plants, students can rapidly create a robust dataset of their own by submitting pollinator observations weekly, daily, or even multiple times a day.

For those who want to use data collected by other users, it can be helpful to focus on phenology observations, as they are the most frequently submitted nationwide.

2. Which plant species (and locations and time periods) are of interest?

Budburst users can submit data on any plant species. These species often differ in phenology and the type of pollinators they attract. It is therefore useful to focus your analyses on a specific species or set of species

Not sure which species to focus on? You can:

a) Start by browsing official Budburst species**.** These are species that Budburst has targeted for data collection because they are of special interest to researchers and Budburst community scientists. You can search for and read about Budburst species present in your area using the Budburst plant database (budburst.org/plants).
b) Choose a species on our top 15 species list. These are the species most often observed by Budburst users, and therefore will have the most data associated with them.
c) Focus within a Plant Group. Budburst sorts plants into five distinct Plant Groups based on general phenology.

Keep in mind that the amount of data associated with each species varies with location and time. If you are viewing the full Budburst database and want to limit your analysis to specific locations or time periods, you can use the search/filter functions to determine how much data is available for a species given your additional parameters. See the 'Getting your Data' document for additional support searching the database.

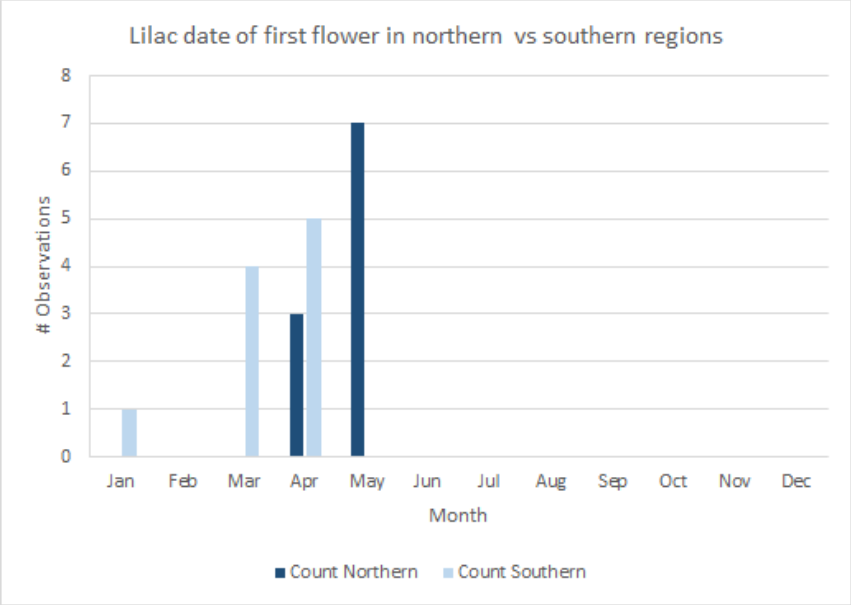
3. What is your research question?

Finally, think about what question(s) you can ask. Students often use Budburst phenology data to explore how the timing of specific phenophases varies with location, year, and species. Or they might compare the timing of phenophases in plants studied by their class to the timing of phenophases in plants observed by other Budburst users. They can also use Budburst phenology data to make predictions (e.g. They can try to determine what time of year they should go on a hike to see the best fall colors.). See the table (next page) for some examples of questions addressed using phenology observations downloaded from the full Budburst database.

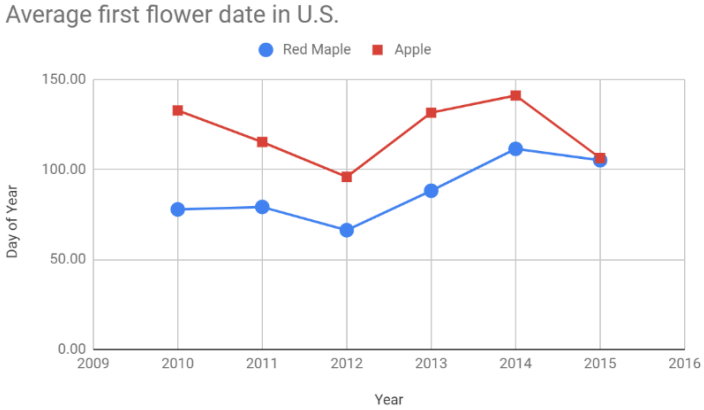
For examples of questions you can explore with other observation types (pollinators and milkweed & monarch), see this activity suggestion on the Budburst website.

Example Questions and Associated Visualizations of Budburst Data

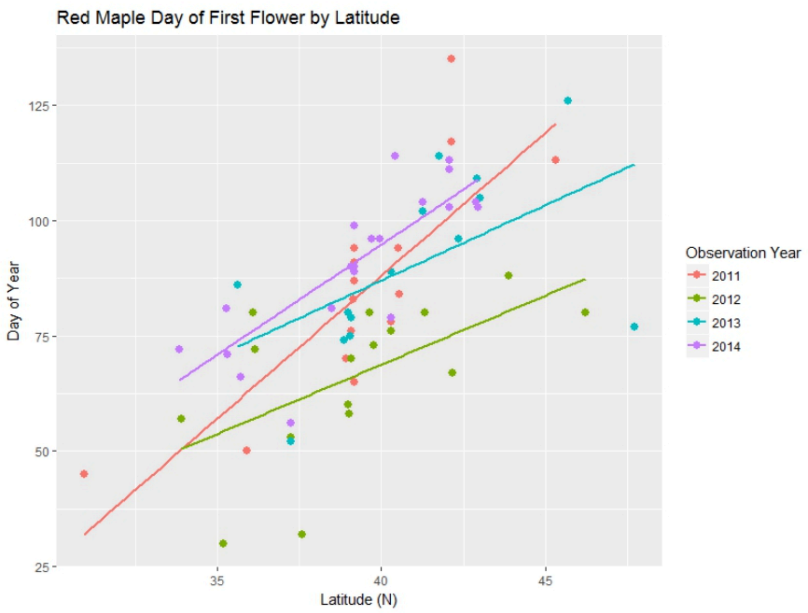
When can we expect to see lilacs flowering? Will my grandma in NC see lilac flowers before my family in New York?



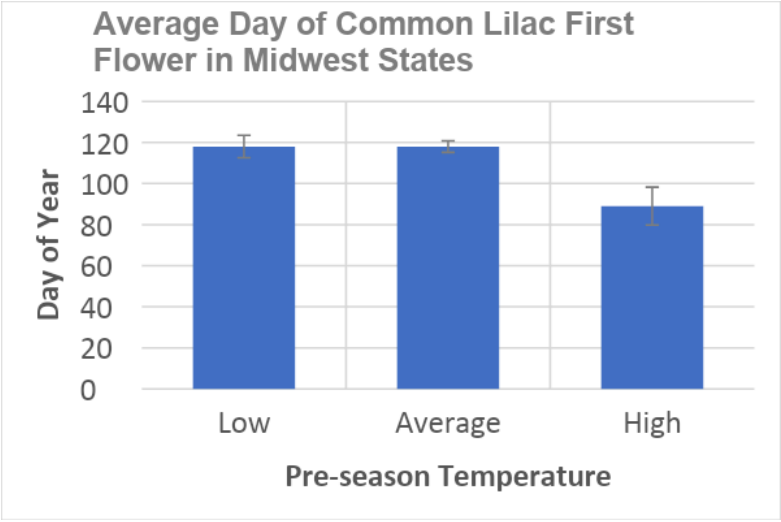
When do red maple and apple trees bloom each year? Do they bloom at the same time?



How does the timing of first flower in red maples vary across time and space?



Do common lilac trees bloom earlier in years with warmer winters?



Note that Budburst phenology observations do not include temperature data. However, national and local temperature data is easily available through websites such as NOAA's Climate at a Glance.

4. Do you have enough data to address your research question? If not, revise!

Because Budburst data is collected opportunistically, the data can be patchy. Allow yourself time to explore potential datasets to ensure that there is enough data to serve your purposes. If not, consider adjusting your research question. For example, a student in Illinois who observed flowering in a common lilac might form a question like: "What is the average time of first flower of common lilacs in Illinois and how does this vary across years?" This is a specific question that would require multiple data points on the timing of first flower in common lilac in Illinois across multiple years. The student will want to search the Budburst database and explore their downloaded data set (see steps 2-3) to ensure they can address their question before proceeding. If there is not enough data, they can change their focus or reframe their question to make it broader. For example, they could include data from other species in the same genus (e.g. Chinese lilacs), neighboring states, and additional flowering phenophases (e.g. early flowering). Their new question might be: 'What is the average time of flowering in lilac trees in Midwestern states and how does this vary across years?"

Note that many professional scientists working with opportunistically collected data must also go through this exploratory process while framing their question. It can be helpful to point this out to students who need to revise their questions!

**How much data is 'enough'? There is no correct answer to this question. In general, more data is better. E.g. If you have 100 data points, you can be more confident in your claims than if you have only 10 data points. In the end, how much data you need depends on your goals (e.g. how confident you need to be in your claims, and how many comparisons you want to make) and the data itself. You might think you have enough data but change your mind after you do some visualizations. If your data is 'messy', with high variability and/or many potential outliers, a larger dataset will likely increase your confidence in your claims. You can quickly get a sense of how much data is available to address your question by searching the Budburst database (see step 2 below).*

Step 2: Getting the data

Once you've decided on a focus, create a Budburst account and/or log in. (Note: you can *view* the data without an account but you will need an account to *download* data). Navigate to the desired database view (see 'Budburst Database Views'). Apply the appropriate searches/filters and download the data in a CSV or Excel file. See the 'Getting your Data' document for step-by-step instructions.

Step 3: Reviewing & Preparing the Data

1. Understanding the raw data

The raw data (aka the downloaded data file) includes more data columns than you will need for your analysis, and some of these columns are only of interest to those actively managing the Budburst database. The table below describes the columns in phenology data downloads that are of interest to most users. See the 'Downloaded Budburst Data Files' document for descriptions of other data columns, including those associated with other observation types.

Table: Phenology Data Columns *Each row includes (1) the name of the column (2) a description of the data found in the column and (3) an example of data found in that column. The example data is from a phenology observation on a common lilac tree in New Hampshire. The plant was in the middle flowering phenophase at the time of observation (5/17/2020).*

Column Name	Description	Example
report_id	Each submitted Budburst observation form is given a unique report id. *The report id can be helpful when cleaning your data (See. Step #5 below).*	112824
common_name	The common name of the plant species observed	Common lilac
latitude	GPS coordinate for the latitudinal location of the plant being observed. *This is often of interest because the timing of seasonal shifts in weather – and subsequent changes in plant phenology - varies with latitude.*	43.0025000000000000
administrative_Area_level_1	U.S. state (abbreviated) in which the plant was observed (Note that other locational data, including zip code, county, and city, are available in additional columns)	NH
observation_date	The date the observation was submitted	5/17/2020
phenophases_plant_structure	The plant part being observed.More specifically, the name of the 'phenophase wheel' on the observation form this data point is associated with. (E.g. fruits, cones, autumn leaf color change)	Flowers
phenophases_title	The phenophase selected by the observer to describe the observed plant part (e.g. middle flowering, first ripe fruit)	Middle flowering
phenophase_id	This column combines information in the two above columns in an abbreviated form. *It is useful for filtering the data quickly by phenophase because it contains information about both the plant part and the selected phenophase.*	FL-MIDDLE
notes	This column includes any notes the observer included about the observation.	These bushes seem to be flowering a

		few days later than in the last few years.

2. Organizing the data file

We recommend making your data file easier to scroll through by:

a. Deleting any unwanted columns from your spreadsheet
b. Adjusting the row height to standard size (15 for excel and 21 for google sheets). See here for instructions on adjusting row height.

3. Reviewing & filtering phenophase data (phenology data only)

a. Many people working with phenology observations simplify their analyses by focusing on the timing of one specific phenophase (e.g., first flower or middle fruiting). You can use the count function and/or data filter option in excel or google sheets to determine which phenophases you have an adequate amount of data.
b. Keep in mind that the data for some phenophases will likely result in more interesting analyses than others. For example, you might start with 'first' or 'budburst' phenophases (e.g., first flower, leaf budburst). These usually occur over short periods and signify important shifts in the annual life cycle of a plant. On the other hand, data for phenophases like 'all leaves unfolded' in deciduous trees are often less precise. Trees are in this phenophase for prolonged periods, and we often cannot know from the data precisely when the phenophase first began.

4. Calculating day of year (phenology data only)

We recommend adding a 'day of year' column to your spreadsheet, where you can convert observation dates to the day of the year (For example, if an observation was made on January 2nd of 2015, the day of year will be 2 because Jan 2 is two days from the start of the year). Using numbers to represent day of year instead of dates will allow you to easily calculate and graph summary statistics, like averages. There are several different ways you can calculate the day of the year. Here is one way to do it in excel (Use the following equation, with A1 being the cell containing the observation date: **=A1-DATE(YEAR(A1),1,1)+1**)

5. Cleaning the Data

The data you've downloaded has been entered by community scientists who have taken great care to make sure the data were recorded and entered correctly. However, typos or errors may have occurred. It is always a good idea to look through your dataset and/or visualize your data to

see if anything seems odd. A data point *should not* be thrown away just because it is odd; however, odd data are a clue to investigate more. Follow the below steps to get started cleaning your data set:

a. Look for oddities by graphing or skimming the data. For phenology data, look at the dates associated with specific phenophases. For example, you might be suspicious if an observer in New York recorded seeing first flower in a common lilac in January, as that is far earlier than we would expect to see lilac trees flowering in northern states.

b. Investigate further by:

   i. Checking the 'notes' column to see if the observer described or asked questions about what they saw

   ii. Looking up the observation in the database to view any images the observer submitted. You can do this by looking up the number found in the 'Report ID' column in your downloaded file (on the same row as your suspicious data point). Open a web browser and type in the URL: budburst.org/data/### (### being the report id). Hit enter and you'll see the online submission associated with the data point.

   For example, in the Phenology Data Columns Table (pg. 7), the report ID is listed as 112824. To view that observation, type in the URL: budburst.org/data/112824.

c. If there is sufficient evidence to believe there was an error, you may choose not to include that data point in your analysis.

Teaching Note: If you choose not to delete the suspicious data point(s) and a student asks about the odd dates, this could be an opportunity to talk about the advantage of certain statistical approaches. For example, you may decide to calculate a median day of year so that less weight is given to individual values that might not be correct. Or you may decide to analyze the data with and without the suspicious data points to determine whether your conclusions would change.

Step 4: Graph and Analyze

Congrats! You're now ready to graph and analyze your data. See 'Extra Resources', part c. below for guidance on analyzing and visualizing data in the classroom.

Extra Resources:

a. Classroom activities & lesson plans with phenology datasets

i. *(K-12 and College)* Students can learn about how climate change affects phenology and analyze real-world data with the 'Springing Forward' activity, developed by Data Nuggets. (Also available in Spanish.)
ii. *(High School and College)* Incorporate Project Biodiversify's slides into a lecture/assignment to highlight the research and personal stories of a diverse group of plant scientists. See slides featuring plant phenology research and graphs by Dr. Tanisha Williams and Dr. Erin Wolkovich (Dr. Wolkovich used Budburst data for her research!)
iii. *(Middle School)* Explore Budburst phenology data with BSCS Learning's Invitations to Inquiry. Spend ~3 class periods analyzing Budburst data on the timing of leaf color change in Fall (see the Fall Colors lesson) or the timing of lilac blooms in Spring (see the Lilac Festival lesson).
iv. *(Middle School and High School)* With this lesson plan from My NASA Data, students analyze historic phenology data from Washington, D.C.'s famous cherry blossom trees, as well as atmospheric near-surface temperatures, as evidence for explaining the phenomena of earlier peak blooms.
v. *(College)* Check out the 2021 paper by Lichti et al. describing how students participating in Budburst analyzed phenology and weather data to answer the research question "How does temperature affect the phenophases of your trees"?

b. Other publicly available datasets

i. Weather data:

Adding weather data to your analysis of Budburst data can provide extra insight into how abiotic factors, such as temperature and precipitation, affect plant phenology. There are many online repositories of weather data, including the NCEI's (National Centers for Environmental Information) 'Climate at a Glance' website. (ncei.noaa.gov/access/monitoring/climate-at-a-glance/). Users can request various types of weather data for specific cities, states, regions, and time periods.

ii. Phenology data
    1. Data Nuggets has a list of websites that offer freely available online data, including phenology datasets. https://datanuggets.org/resources/data/
    2. Browse the National Phenology Network's phenology datasets for historical phenology datasets. Your students can investigate whether a plant species' annual cycle responds to variability in climate by comparing their observations with historical data.

c. Resources & tools for using data in the classroom

i. Dataspire (dataspire.org)
    1. About: 'We inspire and empower educators and students to engage confidently with data and science in their daily lives.'
    2. Resources Page: dataspire.org/data-resources

ii. Data Nuggets (datanuggets.org)
    1. About: 'Activities that bring real scientific data into the classroom, guiding students through the entire process of science while building their quantitative abilities.'
    2. Resources Page: datanuggets.org/resources/
    3. Searchable Activities: datanuggets.org/view-in-searchable-table/