**CURE in Traits and Species Distributions**

**Dataset Generation and Data Management Tutorial**

**Part I. Preparing the Datasheet**

1. Create a folder for your project, ideally on google drive or another filesharing platform, and give it a unique and informative name (e.g., ‘Traits\_Distribution\_CURE’). Using a filesharing platform will make it easier for your group to work collaboratively on data generation, analysis, and writing.
2. Create a spreadsheet and give it a name that includes the words ‘Master Data’ (e.g., ‘Traits\_Distribution\_Master\_Data’). This will be your central dataset that includes all of the data that you will generate during the course of your project. You may create files that are a subset of this dataset and give them a different name.
3. Name tab 1 ‘Metadata’ by double clicking the tab name to make it editable. You will use this tab to record data about the data.

Graphical user interface, application, table, Excel

Description automatically generated

You can begin by recording a few pieces of information that you should already have, including group members, your provisional project title, and the source of herbarium specimen images.

Graphical user interface, application, table, Excel

Description automatically generated

You will continue to add notes to your metadata tab over the course of the project, but this is enough information for now.

1. Use the plus sign button on the bottom left of the screen to add a new tab and rename it ‘summarized species data.’ This is the spreadsheet where you will end up aggregating most of your data and will then use for downstream analyses.

Graphical user interface, application

Description automatically generated with medium confidence

1. Begin building the scaffold of your main datasheet. Title the first column ‘Species name’ and list all of the names of species you are planning on including in your study. The rest of the columns then need to be given a title that reflects the data you are planning on collecting for each species. At minimum the following column titles are suggested. You may want to add other column titles if you are planning on including additional trait data.

Graphical user interface, application, Word

Description automatically generated

Column titles:

* Species name
* Total images examined
* Total specimens with isidia present
* Total specimens with soredia present
* Total specimens with lobules present
* Total specimens with asexual lichenized propagules present
* Total specimens with apothecia present
* Percent specimens with isidia present
* Percent specimens with soredia present
* Percent specimens with lobules present
* Percent specimens with asexual lichenized propagules present
* Percent specimens with apothecia present
* Distribution size (km2)

1. Make a new tab with the name of each that you plan on gathering data for.

**Part II. Generating Data**

1. Navigate to the C.V. Starr Virtual Herbarium website (<https://sweetgum.nybg.org/science/vh/>), then click ‘Advanced Search.’ Fill in the Taxonomy section with Genus and Species. Then, scroll down and select ‘Records with Images’ and click ‘Submit. If you are further restricting your data collection to specific dates or geographic units then fill in the appropriate fields.

Graphical user interface

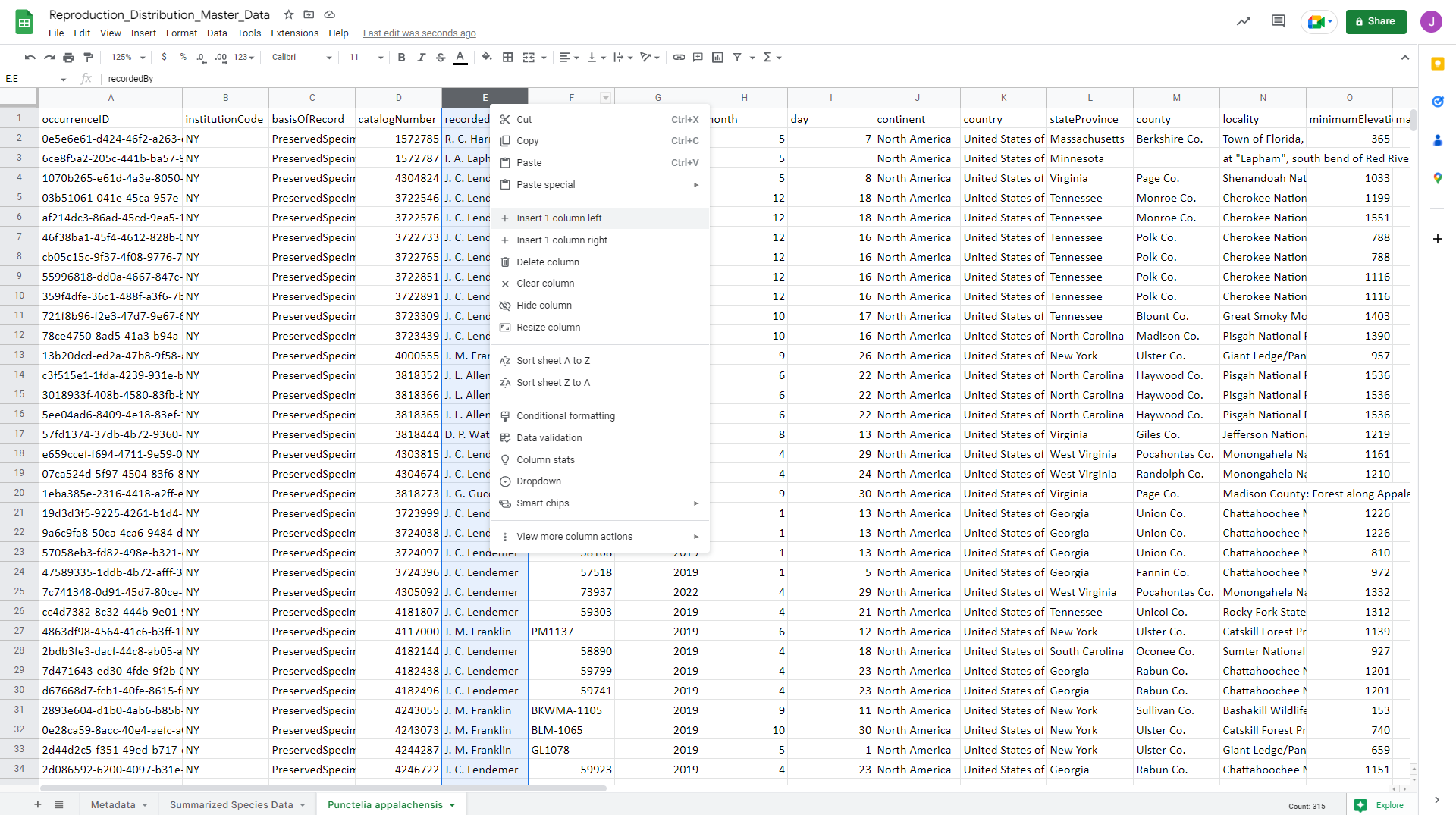
Description automatically generated with medium confidence

1. Once your search results are returned, click the ‘Download CSV’ button. Once the download is complete, open the file. Once the file is open, copy all of the content and paste it into the appropriate species sheet in your Master Data file.

Graphical user interface

Description automatically generated

1. Right click on the top of column E and select ‘Insert 1 column left.’ Repeat this process three more times to create four blank columns between ‘catalogNumber’ and ‘recordedBy.’



1. Title the new columns according to each reproductive structure that may be present. For most species this may be one column for a specific lichenized asexual propagule (isidia, soredia, or lobules) and one column for apothecia. Some species may produce more than one asexual lichenized propagule, in which case you will need extra columns. The new column should be title ‘Observer’ where you can record the name of the group member that examined the image of that particular collection. The final new column should be titled ‘Comments,’ which you can use to make any relevant notes as you go through the data generation process. Filling these columns with distinct colors can make it easier to keep track of which columns you are working with.

Graphical user interface, application

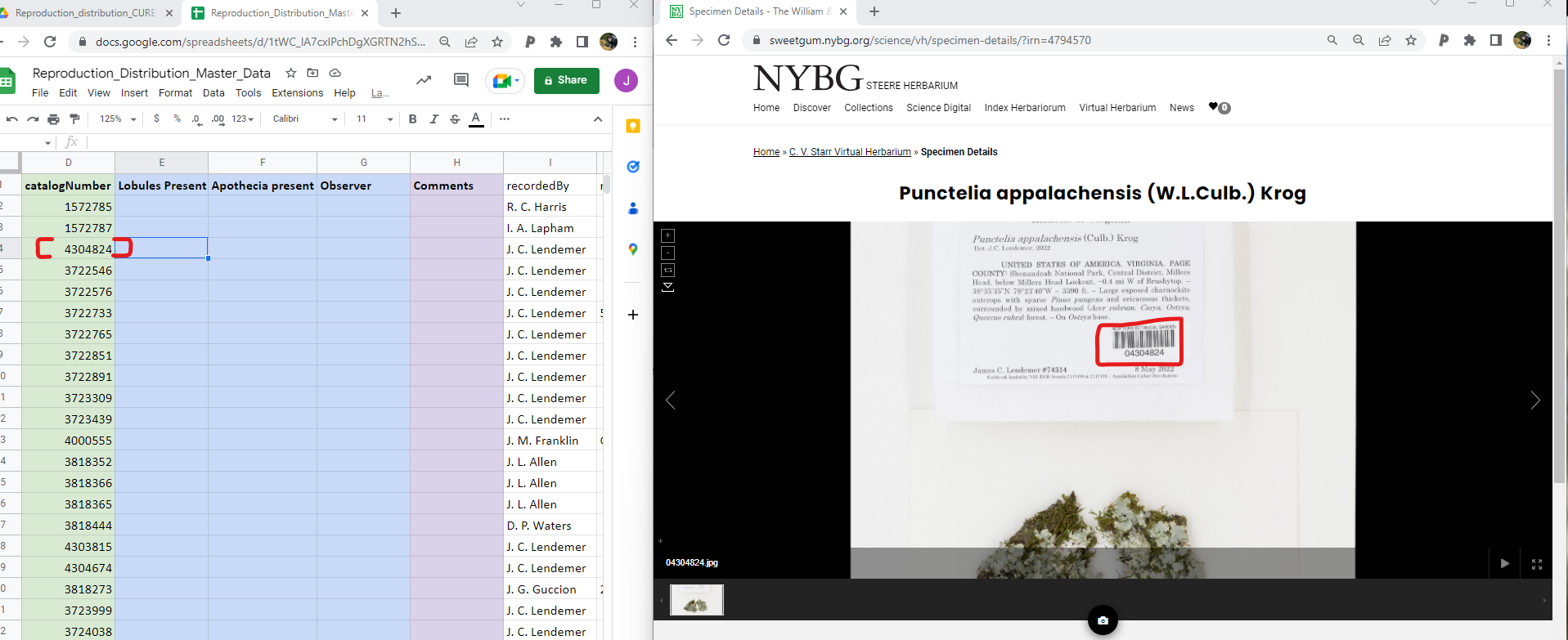
Description automatically generated

1. Next, arrange the windows on your screen so you can easily toggle between your datasheet and the results of your specimen search. Then, click on an image of a specimen. DO NOT click images of labels. Those will not show you pictures of lichens and you will not be able to score traits from them.

Graphical user interface, text, application

Description automatically generated

1. Once you have a specimen image loaded find the row with the catalogueNumber value matching the barcode number like the example below.



1. Zoom in on the image of the lichen and scroll over the whole specimen searching for reproductive structures. Once you have determined which structures are present and which are absent record those data in your spreadsheet. To mark a structure present use ‘1’ and to mark a structure absent use ‘0.’ Fill in the observer column and the comments column.

Graphical user interface, application, table, Excel

Description automatically generated

1. Once you finish recording data for a specimen, scroll to the bottom of the screen and click ‘View More.’ You should then see a list of the specimens and the current specimen you are looking at will be highlighted in black. Select the next specimen.

Graphical user interface

Description automatically generated

1. Repeat this process until you complete generating your dataset. Only a subset of specimens are shown on each page and you can use the arrows at the bottom of the screen to navigate to the next set of specimen images. Create new tabs for each species that you are including in your study. You may want to divide species among your groupmates so you avoid accidently re-gathering the same data, or overwriting each other’s data.
2. If there are specimens with photos of labels rather than specimens you delete those rows from your dataset.
3. Once you finish generating your basic data take a moment to celebrate your hard work!

**Part III. Aggregating data to summarized species data tab**

1. Once you finish generating data for a given species summarize it by using the following equations.
   1. To calculate the total number of specimens observed click the appropriate cell, then type **=COUNTIF( ,** then click the tab for the species in question, click and drag to select all of the cells in any of the columns where you recorded characters for your dataset. Then, click in the equation bar at the top of the screen and type **, “<>”)** after the range of cells that you selected. Then hit enter. Your file should then look as follows.

Graphical user interface, application, table, Excel

Description automatically generated

* 1. To calculate the number of each of the specimens with each of the respective structures you use a variation on the same equation. For instance, in my example I will calculate the number of specimens with lobules using the following equation:

=COUNTIF(‘Punctelia appalachensis’!E2:E38, “1”)

This tells the google sheet to count each cell where the number ‘1’ was recorded for the column where lobule data was recorded.

Repeat this process for each character.

* 1. Any characters are absent from all individuals in a species, simply type ‘0’.

1. To calculate percentage of specimens with a given structure use the following equation.

=(Cell with total specimens with a character present/Cell with total images examined)\*100

Graphical user interface, application

Description automatically generated