

Oregon Bee Project Data Pipeline Script

Summary

These scripts retrieve observation records from iNaturalist.org, reformat them, and create bee specimen labels from them.

There are two ways to run the scripts:

1. **Full Pipeline Mode** - the program runs the full process sequentially:
 1. Pulling data from iNaturalist.org
 2. Formatting the data
 3. Merging the data with an existing dataset and indexing it
 4. Creating labels from the data (optional)
2. **Labels Only Mode** - the program creates labels from a given formatted dataset
 1. This option exists because creating labels is time-consuming, so the user may wish to run it as a separate process.

These processes can be run by executing (double-clicking) Full_Process.bat or Labels_Process.bat, respectively (Full_Process.sh and Labels_Process.sh on MacOS). See the corresponding sections below for details.

Installation

These scripts depend on a several pieces of third-party software to function. For developers' convenience, there is a list in OBP-Script/config/dependencies.txt. This section will provide instructions on how to install each in order on a Windows computer.

Python

Firstly, the scripts require a version of Python 3 installed.

1. Open a Command Prompt terminal.
2. Type "python --version" and press enter.
 - If the command returns "Python 3.*.*", where * is any number, steps 3-5 are optional but recommended.



```
Command Prompt
Microsoft Windows [Version 10.0.22621.1992]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Myles Scholz>python --version
Python 3.11.4

C:\Users\Myles Scholz>
```

- If the command returns a "Python was not found" error or a version of Python less than 3, close the Command Prompt window and continue with steps 3-5.

```
Command Prompt
Microsoft Windows [Version 10.0.22621.1992]
(c) Microsoft Corporation. All rights reserved.

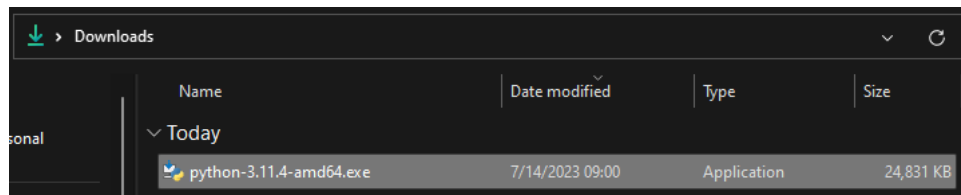
C:\Users\Myles Scholz>python --version
Python was not found; run without arguments to install from the Microsoft Store, or disable this shortcut from Settings
> Manage App Execution Aliases.

C:\Users\Myles Scholz>
```

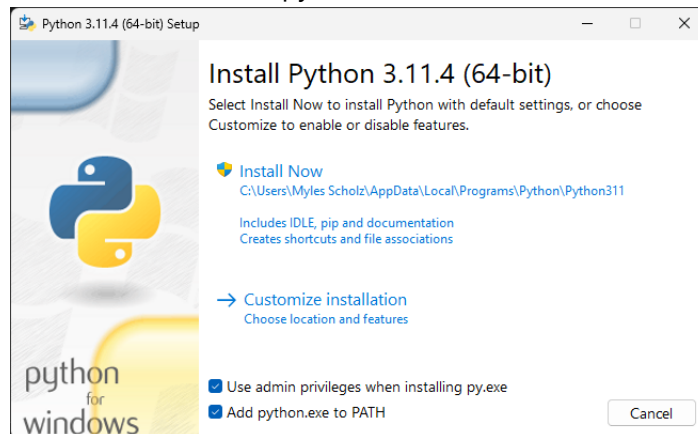
3. Go to <https://www.python.org/downloads/> and click "Download Python 3.*.*".



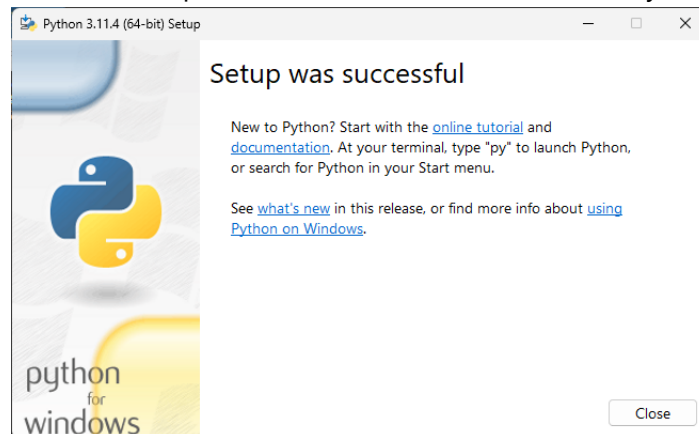
4. Go to the folder where the EXE file downloaded and execute (double-click) it.



- Follow the installation instructions.
- Check the box marked "Add python.exe to PATH variable" when prompted.



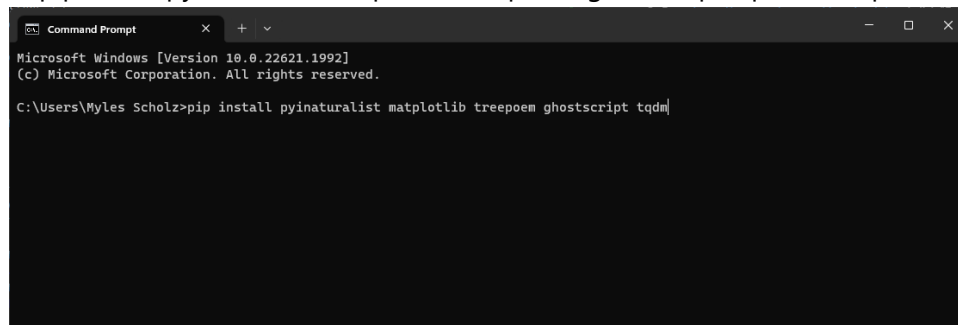
- When the installation is complete, continue to the next section ("Python Libraries").



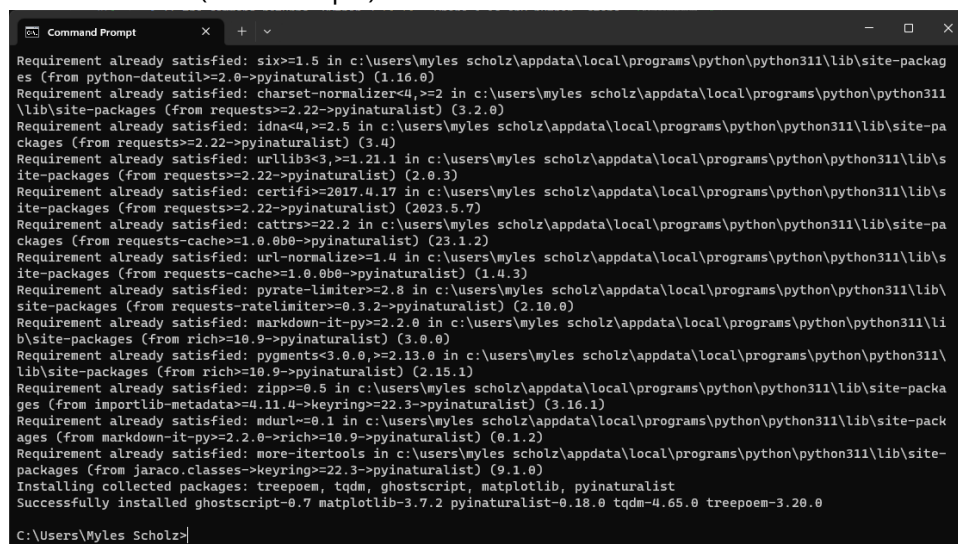
Python Libraries

Next, the scripts need some external Python libraries installed.

- Open a Command Prompt terminal.
- Type "pip install pyinaturalist matplotlib treepoem ghostscript tqdm" and press Enter.



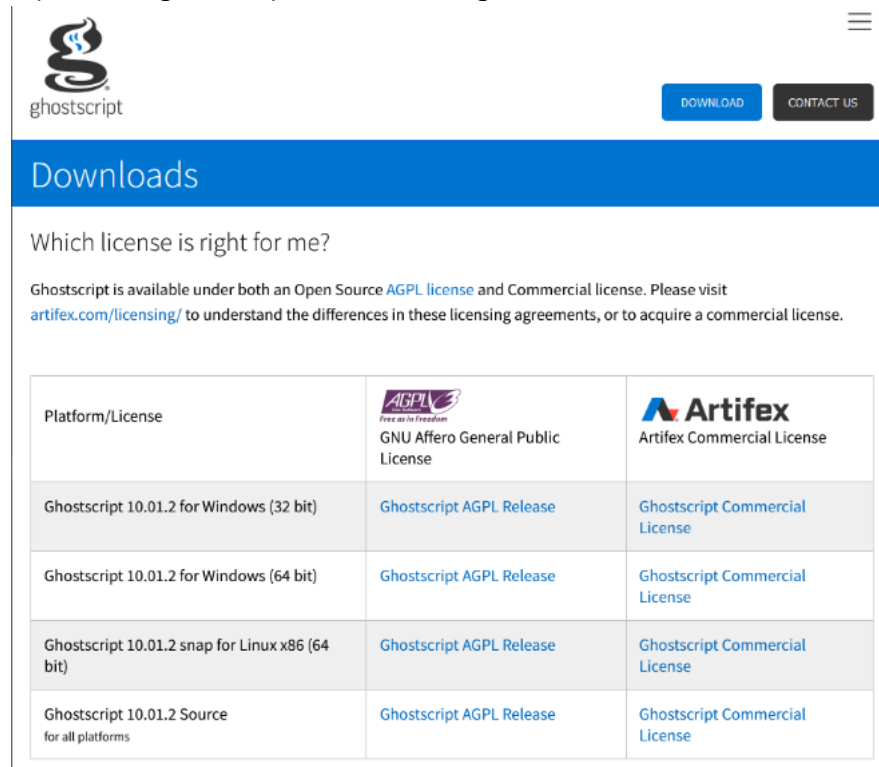
- When installation is complete (the cursor is flashing next to a line ending with ">"), continue to the next section ("Ghostscript").



Ghostscript

Finally, some of the Python libraries (treepoem and ghostscript) require a version of the software Ghostscript installed.

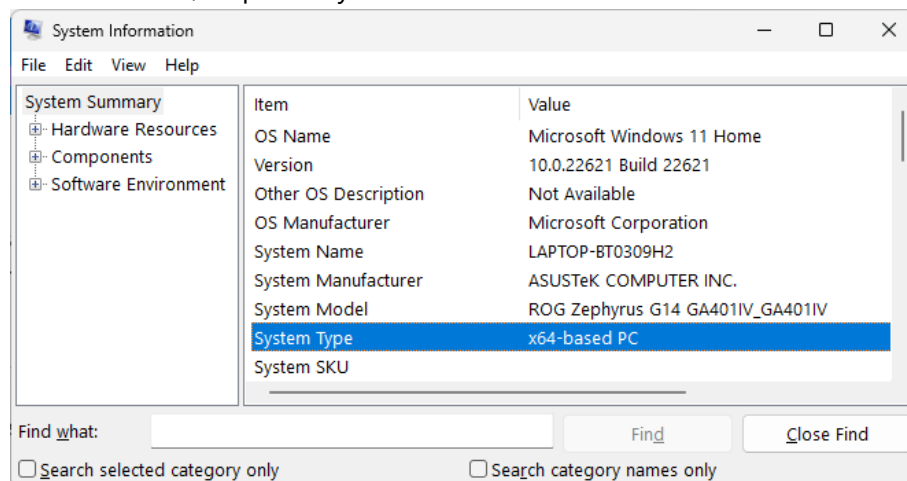
1. Go to <https://www.ghostscript.com/releases/gsdnld.html>.



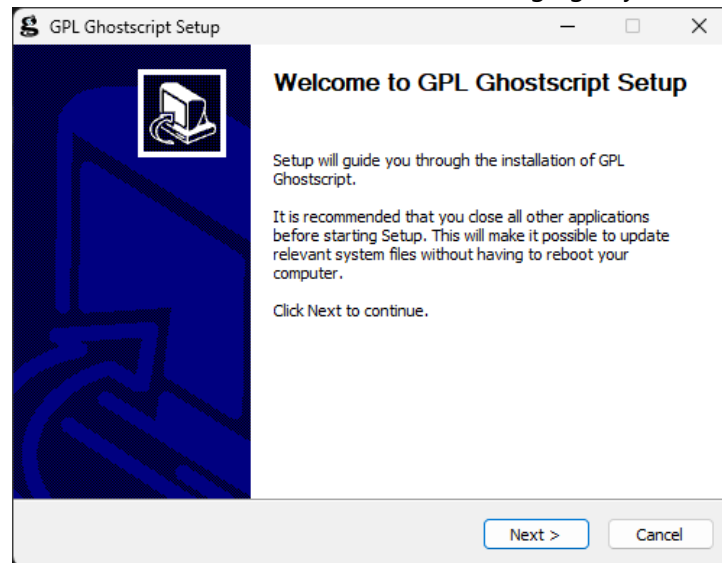
The screenshot shows the Ghostscript website's 'Downloads' section. At the top, there's a navigation bar with the Ghostscript logo, a 'DOWNLOAD' button, and a 'CONTACT US' button. Below the navigation bar, a blue header reads 'Downloads'. The main content area asks 'Which license is right for me?' and provides information about the Open Source AGPL license and the Commercial license, with a link to artifex.com/licensing/. A table follows, listing download links for various Ghostscript versions and platforms, categorized by license type.

Platform/License	AGPL Free as in Freedom GNU Affero General Public License	Artifex Artifex Commercial License
Ghostscript 10.01.2 for Windows (32 bit)	Ghostscript AGPL Release	Ghostscript Commercial License
Ghostscript 10.01.2 for Windows (64 bit)	Ghostscript AGPL Release	Ghostscript Commercial License
Ghostscript 10.01.2 snap for Linux x86 (64 bit)	Ghostscript AGPL Release	Ghostscript Commercial License
Ghostscript 10.01.2 Source for all platforms	Ghostscript AGPL Release	Ghostscript Commercial License

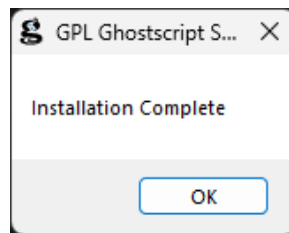
2. Click the link "Ghostscript AGPL release" next to either "Ghostscript 10.01.2 for Windows (64 bit)" or "Ghostscript 10.01.2 for Windows (32 bit)", depending on whether the computer is a 64-bit or 32-bit architecture.
 - o To check the computer's architecture, open the System Information application. The "System Type" field will contain either x64 or x32, corresponding to 64-bit and 32-bit architectures, respectively.



3. Go to the folder where the EXE file downloaded and execute (double-click) it.
 - This will require administrator privileges on the computer.
 - Follow the installation instructions without changing any of the options.



4. Wait until installation finishes.



After completing the above steps, the scripts should be able to run on the computer.

Full Pipeline Mode

This script (Full_Process.bat/Full_Process.sh) executes the full data pipeline in four steps:

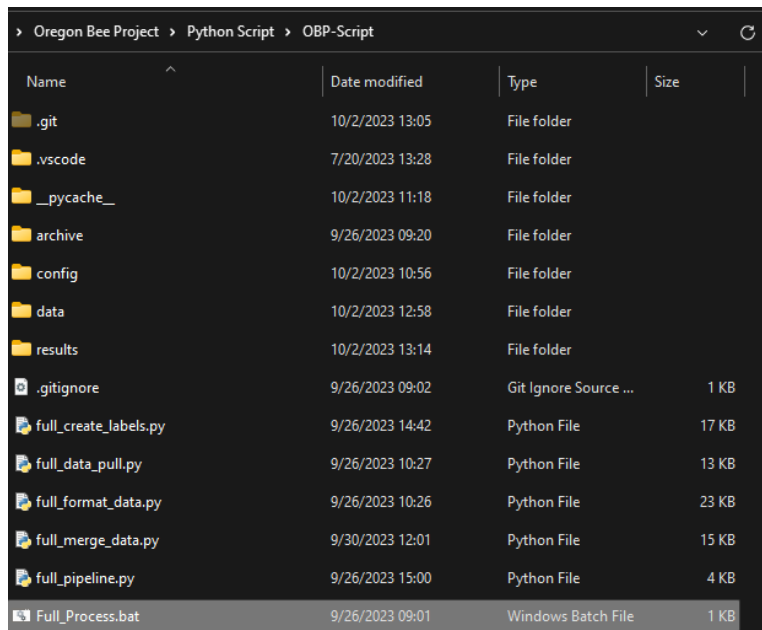
1. Pulling data
2. Formatting data
3. Merging data
4. Creating labels

The first three steps always run in Full Pipeline Mode and cannot be paused. The user has the option to run the fourth (label creation) step or end the process after the first three steps. No information will be lost if the user ends the process before creating labels.

Running the Process


0. Check that the script is configured properly. Each step has a configuration file in OBP-Script/config/. See the respective section below for details on how to configure them.

1. Execute (double-click) Full_Process.bat (Full_Process.sh on MacOS).



Name	Date modified	Type	Size
.git	10/2/2023 13:05	File folder	
.vscode	7/20/2023 13:28	File folder	
__pycache__	10/2/2023 11:18	File folder	
archive	9/26/2023 09:20	File folder	
config	10/2/2023 10:56	File folder	
data	10/2/2023 12:58	File folder	
results	10/2/2023 13:14	File folder	
.gitignore	9/26/2023 09:02	Git Ignore Source ...	1 KB
full_create_labels.py	9/26/2023 14:42	Python File	17 KB
full_data_pull.py	9/26/2023 10:27	Python File	13 KB
full_format_data.py	9/26/2023 10:26	Python File	23 KB
full_merge_data.py	9/30/2023 12:01	Python File	15 KB
full_pipeline.py	9/26/2023 15:00	Python File	4 KB
Full_Process.bat	9/26/2023 09:01	Windows Batch File	1 KB

2. A terminal will open and the program will run. When prompted, enter the requested information. See each step's section below for details on answering the prompts.



```
C:\WINDOWS\system32\cmd. X + v
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon
Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python full_pipeline.py
Pulling Data...
Year to Query: |
```

3. The locations of the output files for each step are detailed in the corresponding section below.

Step 1: Pulling Data from iNaturalist.org

The pipeline begins by querying iNaturalist.org for observation data from a given year and from a given list of iNaturalist projects. When a user runs the script, they will be prompted to type a year to query. The list of iNaturalist projects to query is stored in OBP-Script/config/sources.csv (see "Data Pulling Configuration" below).

As of September 2023, the script pulls from the following projects:

- Oregon Bee Atlas (OBA)
- Master Melittologist (MM)
- Washington Bee Atlas (WaBA)

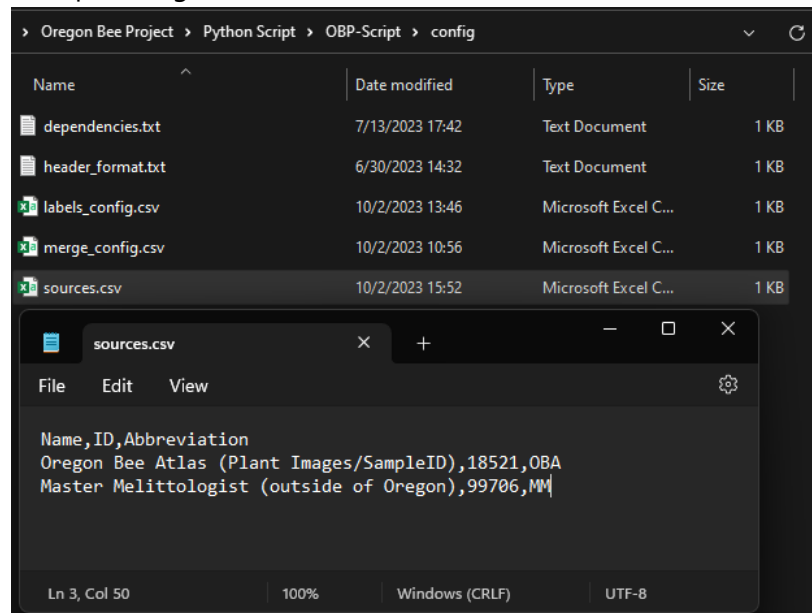
Data Pulling Configuration

The script pulls data from a list of iNaturalist projects specified in OBP-Script/config/sources.csv. The file is in a CSV format with three columns. They are, in order:

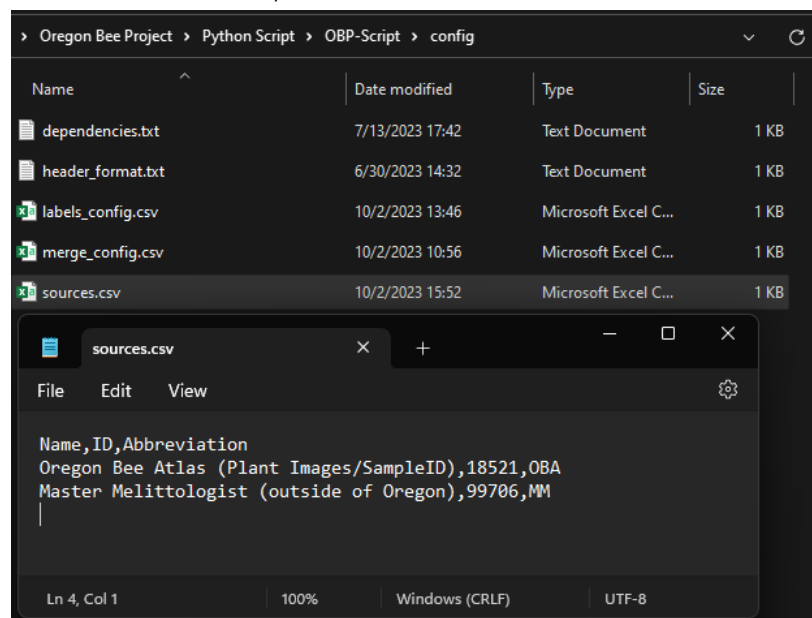
1. Name: the name of the source as it will appear in the terminal.
2. ID: the iNaturalist project ID; this is essential for pulling data from the correct sources.
3. Abbreviation: the abbreviation of the source's name; this must be unique.

To add a source, do the following:

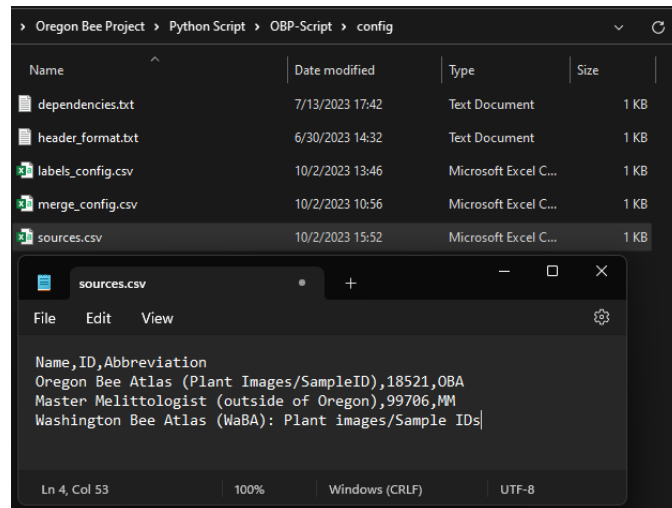
1. Open OBP-Script/config/sources.csv in a text editor or Excel.



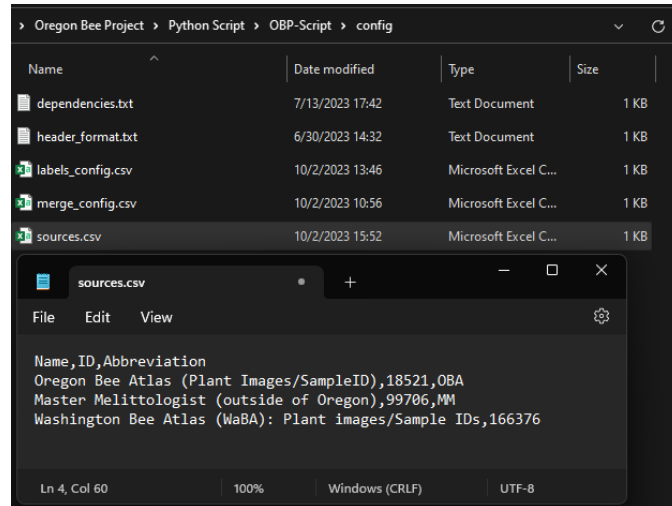
2. Add a line at the end of the file, or select the next blank cell in the A column if using Excel.



3. Type a name for the source. The name cannot contain commas, but all other keyboard characters are allowed.

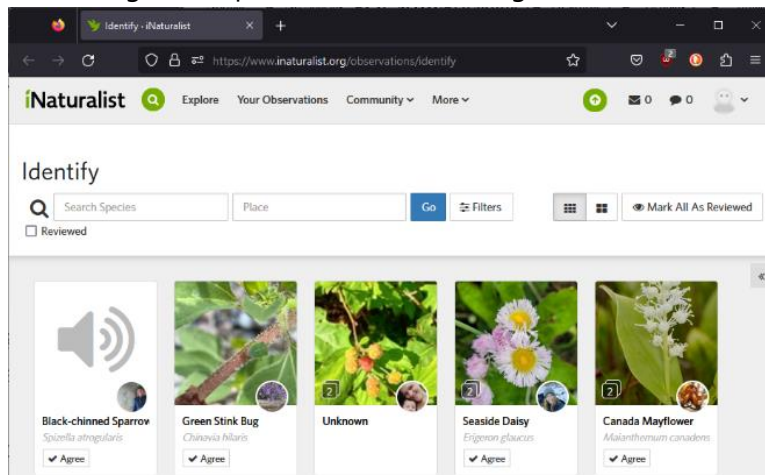


4. Type a comma (or move to B column), and then type the iNaturalist project ID.



The iNaturalist project ID can be found by doing the following:

1. In a browser, go to <https://www.inaturalist.org/observations/identify>.



2. Click the "Filters" button.

The screenshot shows the iNaturalist 'Identify' page. At the top, there's a navigation bar with 'iNaturalist' logo, 'Explore', 'Your Observations', 'Community', and 'More'. Below this is a search bar with 'Search Species' and 'Place' fields, a 'Go' button, and a 'Filters' button which is highlighted with a blue box. To the right of the search bar are icons for 'Mark All As Reviewed'. Below the search bar, there are several filter sections: 'Quality Grade (Select At Least One)' with checkboxes for 'Casual', 'Needs ID' (checked), and 'Research Grade'; 'Show' with checkboxes for 'Captive', 'Threatened', 'Introduced', 'Popular', 'Has Sounds', 'Has Photos', and 'Your Observations'; 'Description / Tags' with a text input field containing 'blue, butterfly, etc.'; 'Categories' with a grid of icons; 'Rank' with 'High' and 'Low' dropdowns; 'Sort By' with 'Date Added' and 'Descending' dropdowns; 'Date Observed' with radio buttons for 'Any' (selected), 'Exact Date', 'Range', and 'Months'; 'Photo Licensing' with a dropdown set to 'All'; and 'Reviewed' with radio buttons for 'Any', 'Yes', and 'No' (selected). At the bottom of the filter section are 'Update Search' and 'Reset Search Filters' buttons, and 'Atom' and 'Download' links.

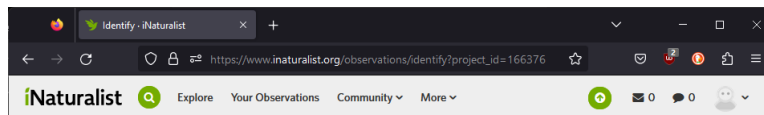
3. Click the "More Filters" button.

This screenshot shows the same iNaturalist 'Identify' page, but with the 'More Filters' button highlighted with a blue box. The 'More Filters' button is located below the 'Description / Tags' field. The filter sections are expanded to show additional options: 'Person' with a dropdown for 'Username or User ID'; 'Project' with a text input field for 'Name or URL slug, e.g. my-project'; 'Place' with a dropdown for 'Place'; 'With Annotation' with a dropdown for 'None'; 'Without Annotation' with a dropdown for 'None'; 'Account Creation' with a dropdown for 'Any'; and 'Date Added' with radio buttons for 'Any' (selected), 'Exact Date', 'Range', and 'Months'. The 'Update Search' and 'Reset Search Filters' buttons, as well as the 'Atom' and 'Download' links, remain at the bottom.

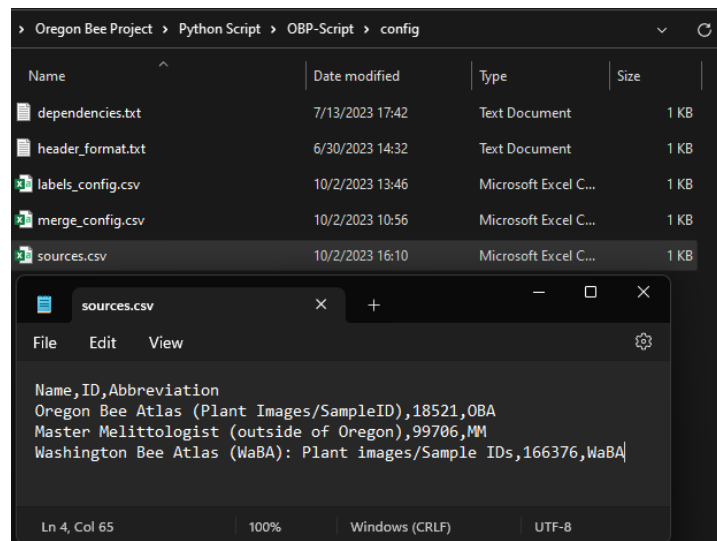
4. Type the name of the project in the "Project" field.
 - Be sure to select the project from the results drop-down menu. The full project name should appear in green.

The screenshot shows the iNaturalist 'Identify' page. The 'Project' field is highlighted with a green box and contains the text 'Washington Bee Atlas (WaBA): Plant Images/Sample IDs'. The 'Update Search' button is visible at the bottom of the filters panel.

5. The project ID will appear as a string of about five numbers at the end of the browser's URL bar.



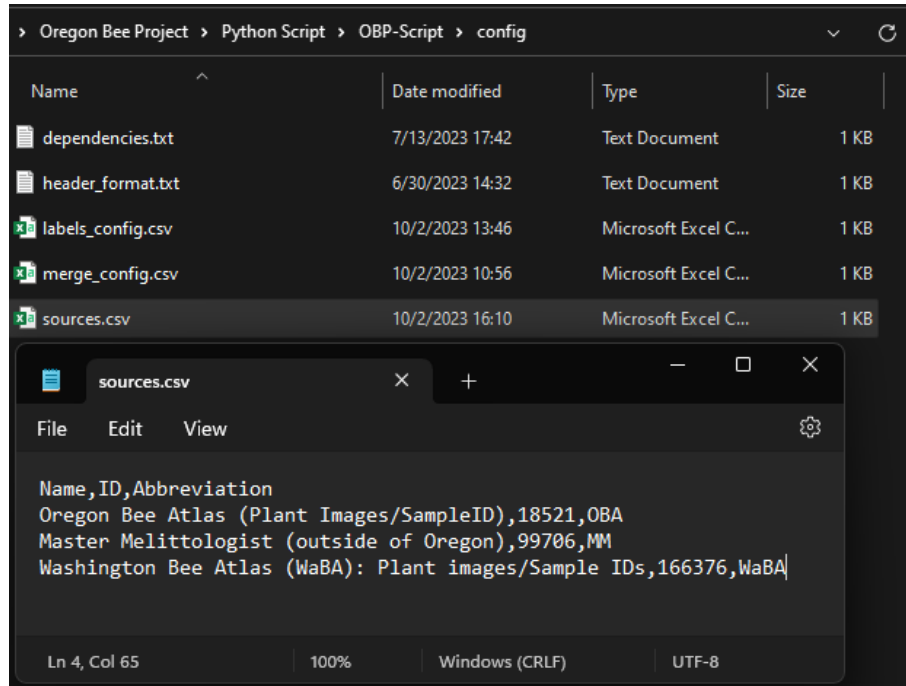
5. Type a comma (or move to the C column), and then type a unique abbreviation for the source.



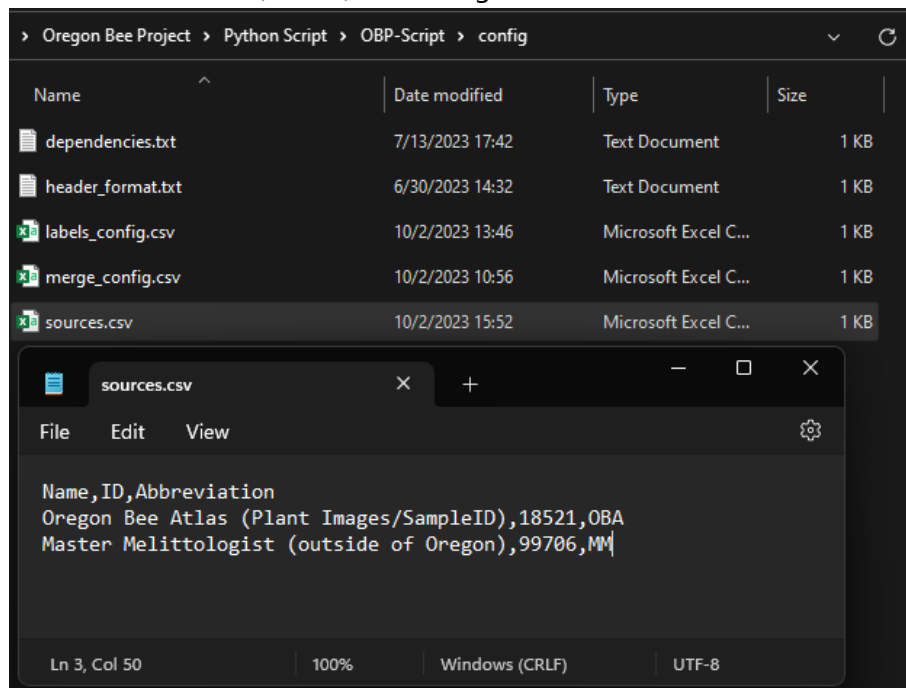
6. Save and close sources.csv.

To remove a source, do the following:

1. Open OBP-Script/config/sources.csv in a text editor or Excel.



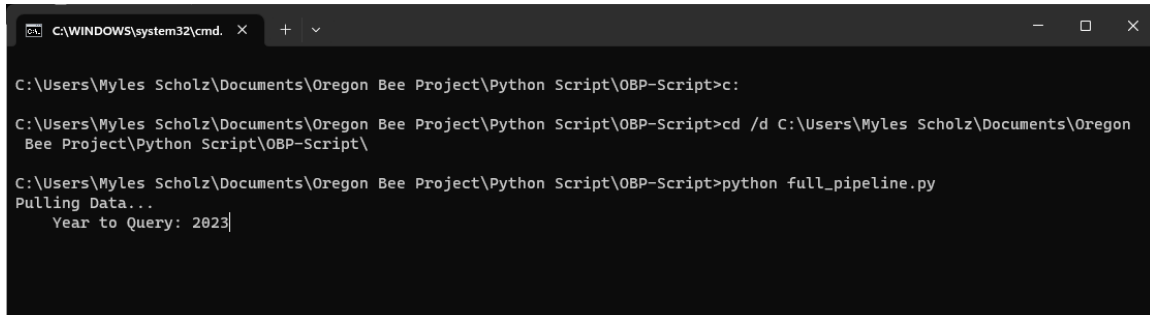
2. Select and delete the line (or row) containing the source to be removed.



3. Save and close sources.csv.

Data Pulling Prompts

At the beginning of this step, the program will prompt the user for a year with which to query iNaturalist.org. It accepts four-digit years less than or equal to the current year. Type a number of this format and hit Enter to continue. There are no other prompts in this step.



```
C:\WINDOWS\system32\cmd. X + v
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon
Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python full_pipeline.py
Pulling Data...
Year to Query: 2023|
```

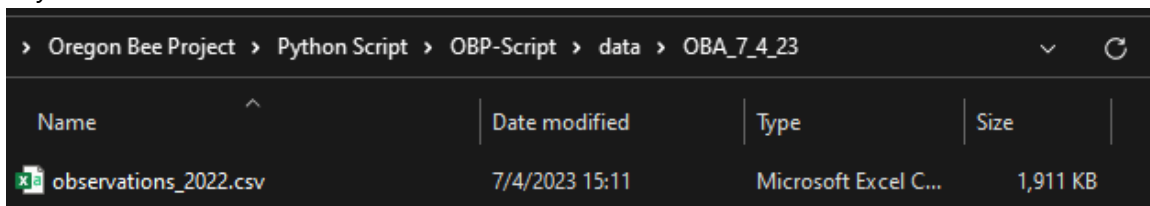
Data Pulling Output


The data pulling step outputs a minimally formatted CSV file in a source-specific folder under OBP-Script/data/.

Each source-specific folder will be named according to the format Abbr_M_D_YY, where Abbr is an abbreviation of the source and M_D_YY is the date when the program ran. For example, the results of fetching data from the Oregon Bee Atlas on July 4th, 2023 would appear in OBP-Script/data/OBA_7_4_23/.

The resulting CSV file will be named according to the format observations_YYYY.csv, where YYYY is the year that was queried. For example, querying Oregon Bee Atlas data from 2022 would produce a file named observations_2022.csv.

If the program is run a second time on the same day with the same query year, the output file will be entirely overwritten with new data.



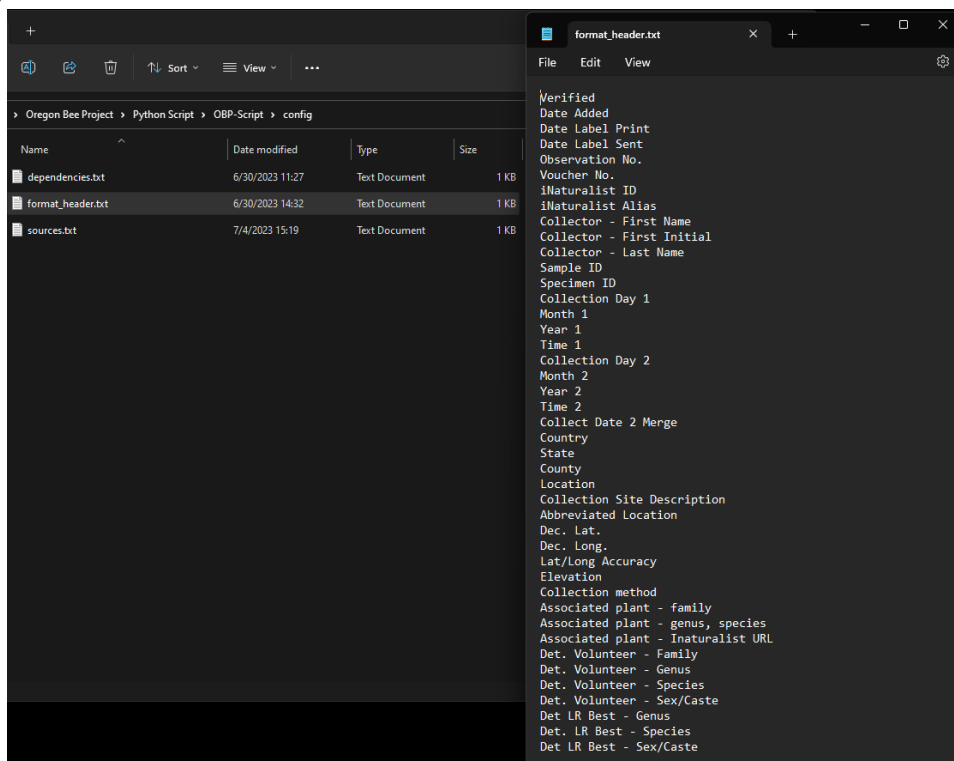
> Oregon Bee Project > Python Script > OBP-Script > data > OBA_7_4_23				
Name	Date modified	Type	Size	
 observations_2022.csv	7/4/2023 15:11	Microsoft Excel C...	1,911 KB	

Step 2: Formatting Data

The second step of the pipeline is to format the data that was pulled from iNaturalist.org previously. This step is entirely fixed without modifying the code, so it does not need user input or configuration. Nonetheless, this step's configuration file, header_format.txt, is explained below.

Data Formatting Configuration

The program formats the data into a CSV file with column names specified in OBP-Script/config/header_format.txt. The name of each column appears in order on each line of the file. For the merging step to work, these column names must match those of the input dataset for that step exactly.



Data Formatting Prompts

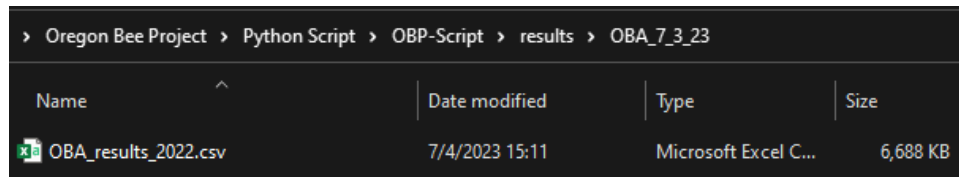
There are no user prompts for this step.

Data Formatting Output

The data formatting step outputs a CSV file in a source-specific folder under OBP-Script/results/. Each source-specific folder will be named according to the format Abbr_M_D_YY, where Abbr is an abbreviation of the source and M_D_YY is the date when the program ran. For example, the results of formatting data from the Oregon Bee Atlas on July 3rd, 2023 would appear in OBP-Script/results/OBA_7_3_23/.

The resulting CSV file will be named according to the format Abbr_results_YYYY.csv, where Abbr is an abbreviation of the source and YYYY is the year that was queried. For example, formatting Oregon Bee Atlas data from 2022 would produce a file named OBA_results_2022.csv.

If the program is run a second time on the same day with the same query year, the output file will be entirely overwritten with new data.



Name	Date modified	Type	Size
OBA_results_2022.csv	7/4/2023 15:11	Microsoft Excel C...	6,688 KB

Step 3: Merging and Indexing Formatted Data

In this step, the program will combine formatted data from the previous step with an existing dataset of the same format. It will detect duplicate entries and sort and index the output dataset. The input and output file paths for this step are defined in OBP-Script/config/merge_config.csv (see "Data Merging Configuration" below). These files must be CSV files, and the input file must have the exact header specified in OBP-Script/config/header_format.txt.

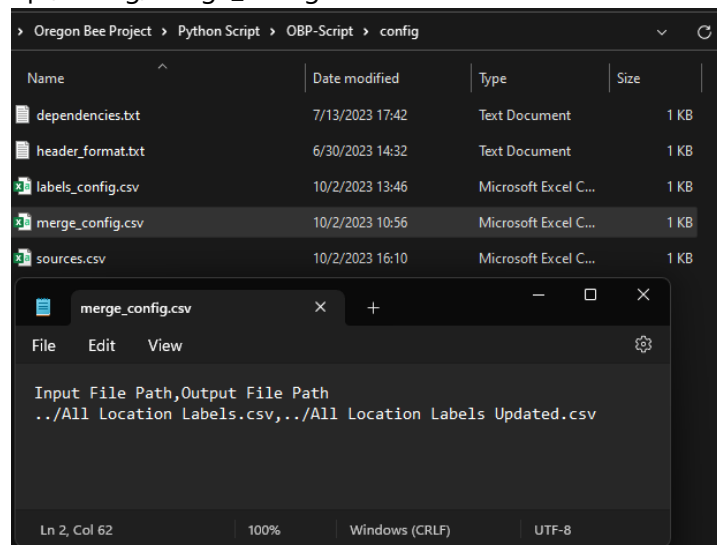
Data Merging Configuration

The input and output file paths for the data merging step are specified in OBP-Script/config/merge_config.csv. This file is in a CSV format with two columns:

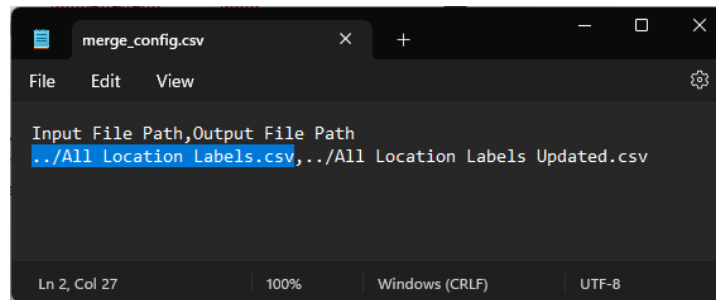
1. Input File Path: the relative or absolute file path of a formatted dataset to merge new data into. This must be a CSV file and must have the exact column names listed in OBP-Script/config/header_format.txt.
2. Output File Path: a relative or absolute file path where the resulting merged dataset will be saved. This must be a CSV file and can be the same as the Input File Path, but the original dataset will be overwritten.

To set the input and output file paths, do the following:

1. Open OBP-Script/config/merge_config.csv in a text editor or Excel.

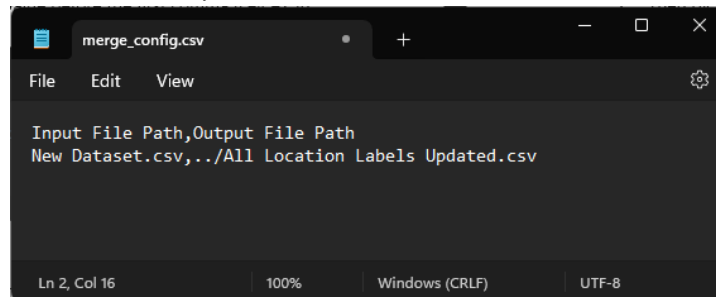


2. On the first line below the column names, select the value before the first comma (cell A2 in Excel).



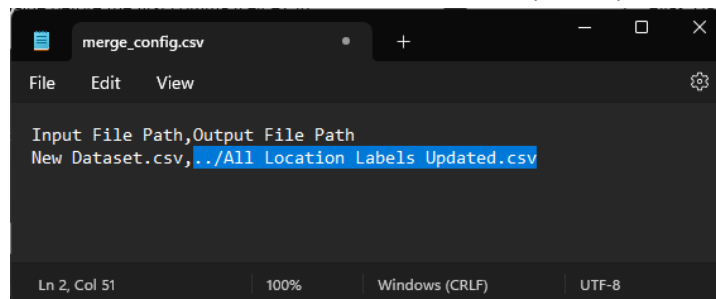
```
merge_config.csv
File Edit View
Input File Path,Output File Path
./All Location Labels.csv,../All Location Labels Updated.csv
Ln 2, Col 27 100% Windows (CRLF) UTF-8
```

3. Type the new value for the Input File Path.



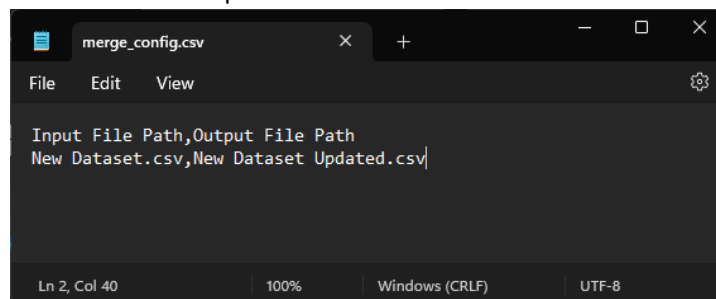
```
merge_config.csv
File Edit View
Input File Path,Output File Path
New Dataset.csv,../All Location Labels Updated.csv
Ln 2, Col 16 100% Windows (CRLF) UTF-8
```

4. On the same line, select the value after the first comma (cell B2).



```
merge_config.csv
File Edit View
Input File Path,Output File Path
New Dataset.csv,../All Location Labels Updated.csv
Ln 2, Col 51 100% Windows (CRLF) UTF-8
```

5. Type the new value for the Output File Path.



```
merge_config.csv
File Edit View
Input File Path,Output File Path
New Dataset.csv,New Dataset Updated.csv
Ln 2, Col 40 100% Windows (CRLF) UTF-8
```

6. Save and close merge_config.csv.

See "Help and Tips - File Paths" for information about file paths.

Data Merging Prompts

There are no user prompts for this step.

Data Merging Output

The data merging step outputs the resulting merged dataset to the file path specified in its configuration. The program checks the data for duplicate entries and adds indices to new data.

It checks for duplicate entries by comparing each entry in the appending data to each entry in the base dataset. If an entry in the appending data matches in any of the following ways (checked in order), it will not be added to the merged file.

1. Observation No.: If the "Observation No." values are not empty, two entries match if their "Observation No." fields match.
2. URL, Sample ID, and Specimen ID: If the "Associated plant - Inaturalist URL" values are not empty, two entries match if their "Associated plant - Inaturalist URL", "Sample ID", and "Specimen ID" fields all match.
3. Alias, Date, Sample ID, and Specimen ID: In all other cases, two entries match if their "iNaturalist Alias", "Collection Day 1", "Month 1", "Year 1", "Sample ID", and "Specimen ID" fields all match.

Entries with empty "Dec. Lat." or "Dec. Long." fields will also not be added to the merged file.

The program will index (assign a unique number to) the data in the "Observation No." field using the format YY#####, where YY is the two-digit abbreviation of the year when the script ran and ##### are five sequentially assigned digits.

The script assigns these numbers by sorting the data by

1. "Observation No."
2. Then by, "Collector - Last Name"
3. Then by, "Collector - First Name"
4. Then by, "Month 1"
5. Then by, "Collection Day 1"
6. Then by, "Sample ID"
7. Then by, "Specimen ID"

in ascending order, with blank values being put at the end. If all "Observation No." fields are blank in the merged data, the indexing will start at YY00000. Otherwise, the indexing will start at the previous largest "Observation No." plus one.

Step 4: Creating Labels from Formatted Data

This step will create a PDF of labels with formatted data from the previous steps or from a CSV file, depending on how it was executed. If this step is reached in Full Pipeline Mode, the user will have the opportunity to check the data before the program begins creating labels. If this step is reached in Labels Only Mode, the user will need to provide a file path of a formatted CSV dataset as input. See "Label Creation Prompts" below for details.

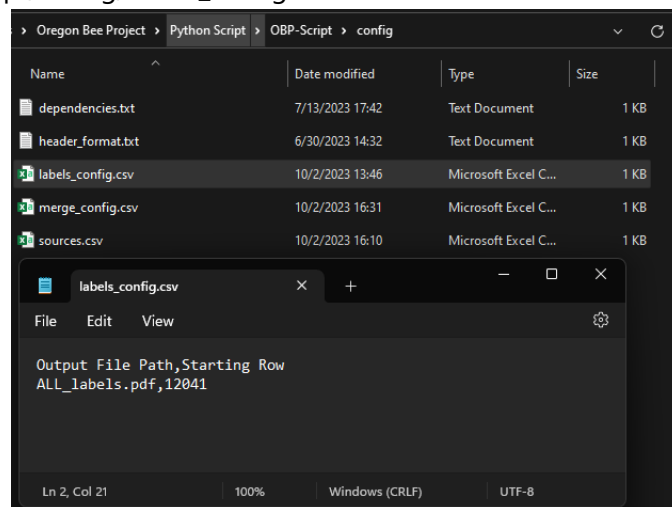
Label Creation Configuration

The label creation step will output the labels to a PDF file path specified in OBP-Script/config/labels_config.csv. The configuration file also specifies the default row number in the input dataset from which to create labels. This value is set by the merging process and is optional (see "Label Creation Prompts" below), so editing it is not recommended.

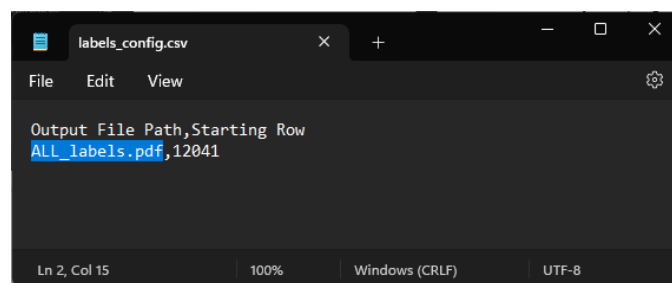
1. Output File Path: a relative or absolute file path where the labels will be output. This must be a PDF file.
2. Starting Row: the zero-indexed row number of the input dataset from which to create labels. It is not recommended for the user to change this manually.

To set the Output File Path, do the following:

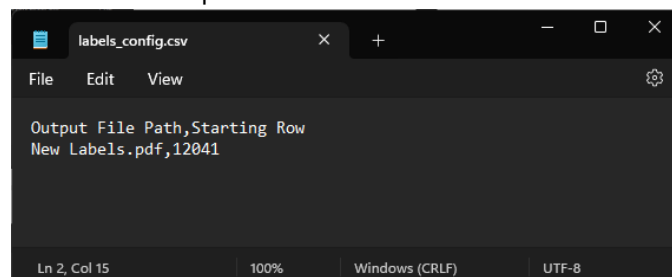
1. Open OBP-Script/config/labels_config.csv in a text editor or Excel.



2. On the first line after the column names, select the value before the first comma (cell A2 in Excel).



3. Type the new value for the Output File Path.



4. Save and close labels_config.csv.

Label Creation Prompts

The prompts for the label creation step begin differently, depending on whether the program is in Full Pipeline Mode or Labels Only Mode.

In Full Pipeline Mode, after the merging step is complete, the user will be prompted to check the dataset and respond affirmatively to proceed with label creation. It is recommended to look over the data in the output file of the merging step (located at the Output File Path in OBP-Script/config/merge_config.csv). When ready, type Y or y and Enter to continue. If the user wishes to end the program, they may type any other character (or no characters) and Enter to do so.

```
C:\WINDOWS\system32\cmd. X + v
Formatting Data...
Formatting 'Oregon Bee Atlas (Plant Images/SampleID)' data...
Observations: 100% | 5835/5835 [00:05<00:00, 989.94it/s]
Formatting 'Master Melittologist (outside of Oregon)' data...
Observations: 100% | 668/668 [00:00<00:00, 895.64it/s]
Formatting 'Washington Bee Atlas (WaBA): Plant images/Sample IDs' data...
Observations: 100% | 749/749 [00:00<00:00, 954.53it/s]

Writing 'Oregon Bee Atlas (Plant Images/SampleID)' observations to 'results\OBA_10_2_23\OBA_results_2023.csv'...
Writing 'Master Melittologist (outside of Oregon)' observations to 'results\MM_10_2_23\MM_results_2023.csv'...
Writing 'Washington Bee Atlas (WaBA): Plant images/Sample IDs' observations to 'results\WaBA_10_2_23\WaBA_results_20
23.csv'...
Formatting Data => Done

Merging Data...
Loading dataset...
Merging 'Oregon Bee Atlas (Plant Images/SampleID)' data with dataset...
Entries: 100% | 28012/28012 [24:02<00:00, 19.43it/s]
Merging 'Master Melittologist (outside of Oregon)' data with dataset...
Entries: 100% | 1706/1706 [02:24<00:00, 11.82it/s]
Merging 'Washington Bee Atlas (WaBA): Plant images/Sample IDs' data with dataset...
Entries: 100% | 2989/2989 [05:33<00:00, 8.97it/s]
Writing merged data to '..\All Location Labels Updated.csv'...
Merging Data => Done

WARNING: Creating labels takes a long time. Please check that the data in '..\All Location Labels Updated.csv' is prop
rly formatted to avoid costly errors. Make sure to save any changes before continuing.

Enter 'Y' to continue with label creation or any other key to quit: Y
```

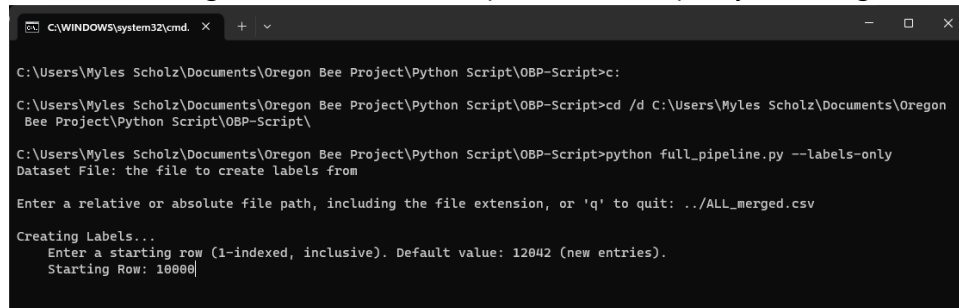
In Labels Only Mode, when the user runs the script, they will be prompted to enter a file path for an input dataset. The file path may be relative or absolute and may be wrapped in quotation marks (see "Help and Tips - File Paths" below). The input dataset should be a formatted CSV file with the exact header specified in OBP-Script/config/header_format.txt.

```
C:\WINDOWS\system32\cmd. X + v
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon
Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python full_pipeline.py --labels-only
Dataset File: the file to create labels from

Enter a relative or absolute file path, including the file extension, or 'q' to quit: ../ALL_merged.csv
```

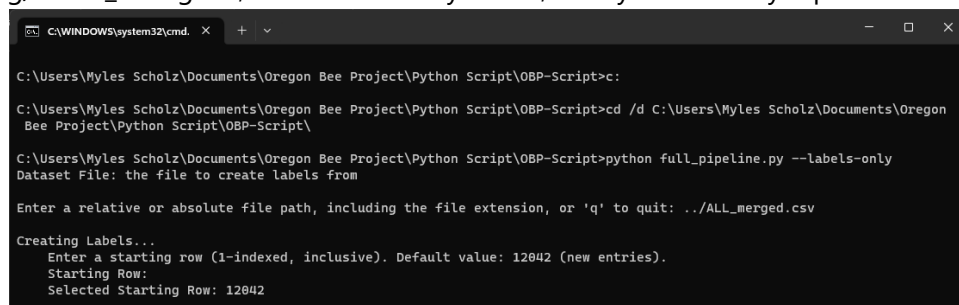
After the above prompts, Full Pipeline Mode and Labels Only Mode have the same prompts.

The program will next prompt the user for a starting row. This is an integer value representing the first row of the input dataset from which labels will be made. This allows the labels to be made from only part of the input dataset to avoid redundancy and save time. The user may type a number between 0 and the total length of the dataset and press Enter to specify a starting row.



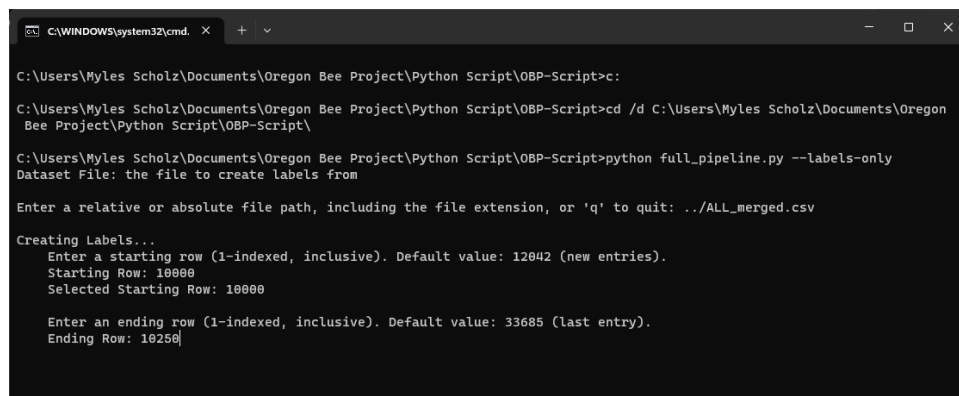
```
C:\WINDOWS\system32\cmd. X + v
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python full_pipeline.py --labels-only
Dataset File: the file to create labels from
Enter a relative or absolute file path, including the file extension, or 'q' to quit: ..\ALL_merged.csv
Creating Labels...
Enter a starting row (1-indexed, inclusive). Default value: 12042 (new entries).
Starting Row: 10000
```

If the user presses Enter without typing a number, the default value will be used. The default value for the starting row is the first new row. It is set by the data merging step and stored in OBP-Script/config/labels_config.csv, so in Labels Only Mode, it may not actually represent new data.



```
C:\WINDOWS\system32\cmd. X + v
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python full_pipeline.py --labels-only
Dataset File: the file to create labels from
Enter a relative or absolute file path, including the file extension, or 'q' to quit: ..\ALL_merged.csv
Creating Labels...
Enter a starting row (1-indexed, inclusive). Default value: 12042 (new entries).
Starting Row:
Selected Starting Row: 12042
```

Lastly, the program will prompt the user for an ending row. This is an integer value that represents the last row of the input dataset from which labels will be made. The user may type a number between the starting row and the total length of the dataset and press Enter to specify an ending row.



```
C:\WINDOWS\system32\cmd. X + v
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python full_pipeline.py --labels-only
Dataset File: the file to create labels from
Enter a relative or absolute file path, including the file extension, or 'q' to quit: ..\ALL_merged.csv
Creating Labels...
Enter a starting row (1-indexed, inclusive). Default value: 12042 (new entries).
Starting Row: 10000
Selected Starting Row: 10000
Enter an ending row (1-indexed, inclusive). Default value: 33685 (last entry).
Ending Row: 10250
```

If the user presses Enter without typing a number, the default value will be used. The default value for the ending row is the last row of the input dataset.

```
C:\WINDOWS\system32\cmd. x + v
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon
Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python full_pipeline.py --labels-only
Dataset File: the file to create labels from
Enter a relative or absolute file path, including the file extension, or 'q' to quit: ../ALL_merged.csv
Creating Labels...
Enter a starting row (1-indexed, inclusive). Default value: 12042 (new entries).
Starting Row:
Selected Starting Row: 12042
Enter an ending row (1-indexed, inclusive). Default value: 33685 (last entry).
Ending Row:
Selected Ending Row: 33685
Page 1/27
Labels: 0%|
| 0/250 [00:00<?, ?it/s]|
```

Label Creation Output

The label creation step produces a PDF file at the file path specified in its configuration. The resulting PDF will include one label for each entry of the input data within the given starting and ending row range. The pages have the following layout:

US Letter size paper (8.5" x 11")

Portrait orientation

0.25" horizontal margins

0.5" vertical margins

25 rows of labels

10 columns of labels

Equal horizontal and vertical spacing

0.666" label width

0.311" label height

All layout values are defined as constants in the first lines of full_create_labels.py.

Help and Tips

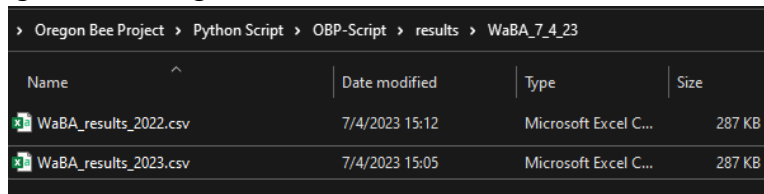
- **File Paths**

A file path is a string of characters representing the unique location of a file on a computer. It contains a sequence of slash-separated folders followed by a file name and its extension. This program accepts absolute and relative file paths and will display given file paths relative to the directory in which the program is running.

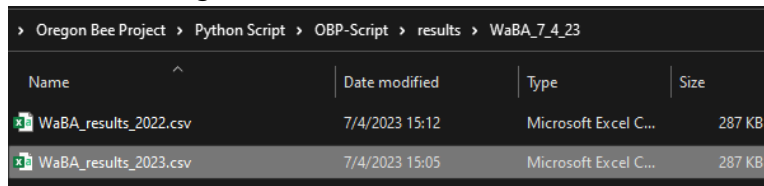
This script accepts file paths pasted directly from File Explorer. To copy a file path in File Explorer, do the following:

On Windows 11 and some versions of Windows 10,

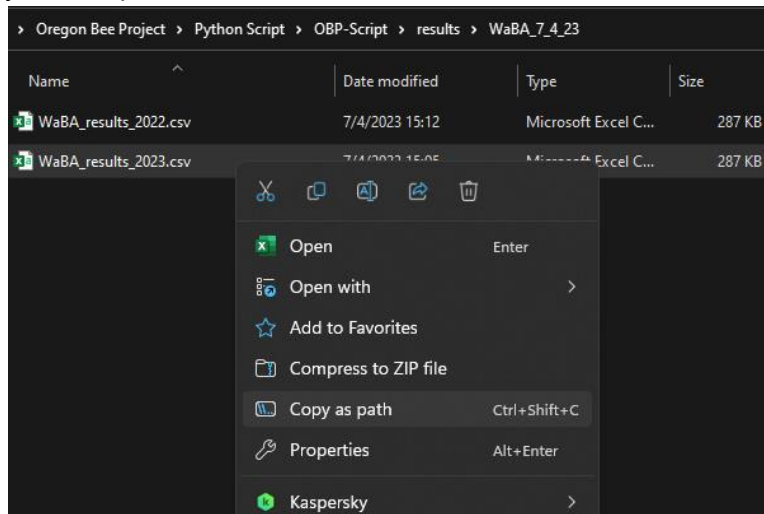
1. Navigate to the target file.



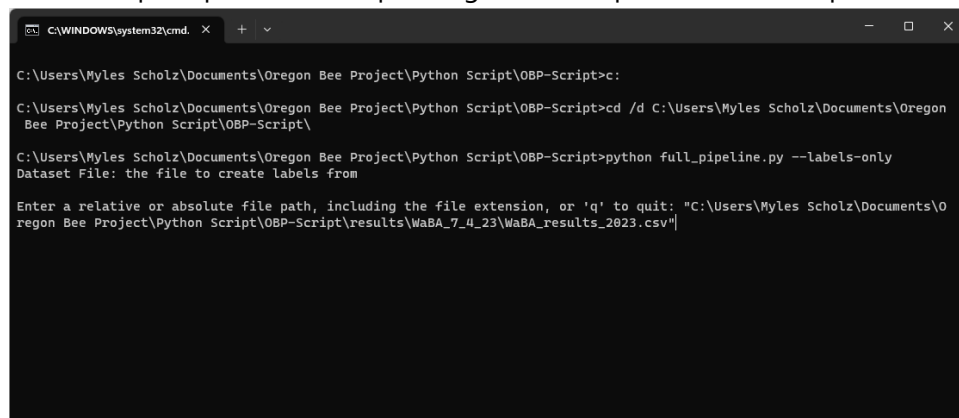
2. Left click on the target file to select it.



3. Right click on the target file and click "Copy as Path", or press Ctrl + Shift + C, to copy the file path.



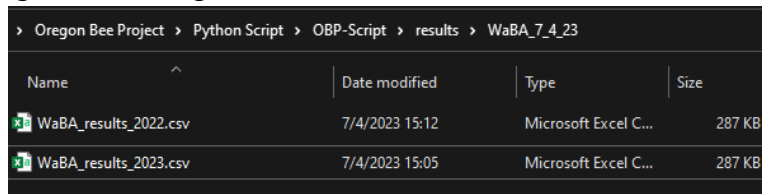
4. Click the terminal window in which the script is running.
5. When prompted for a file path, right click, or press Ctrl + V, to paste the file path.



6. Press Enter to submit the file path.

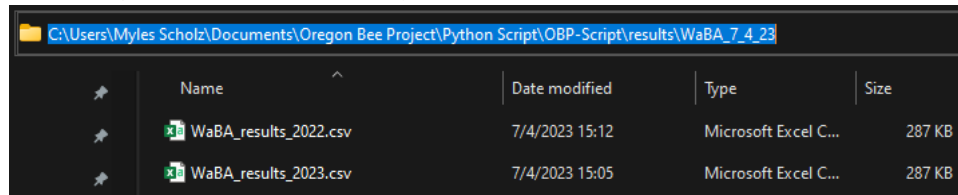
On Windows 10,

7. Navigate to the target file's folder.



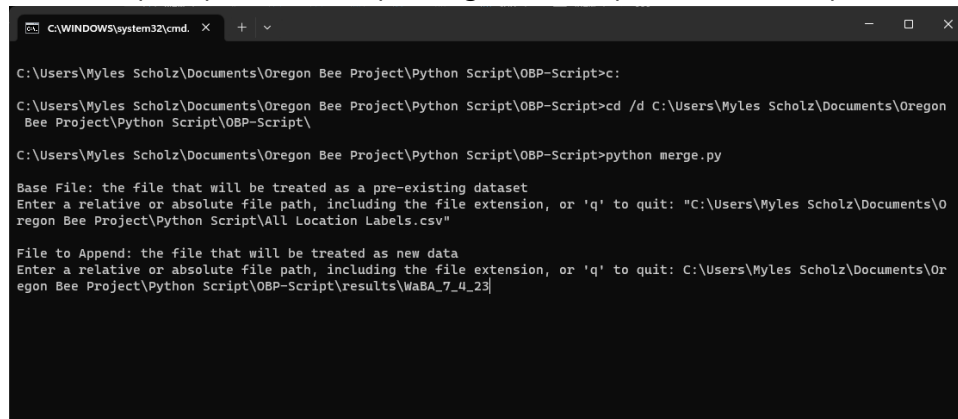
Name	Date modified	Type	Size
WaBA_results_2022.csv	7/4/2023 15:12	Microsoft Excel C...	287 KB
WaBA_results_2023.csv	7/4/2023 15:05	Microsoft Excel C...	287 KB

8. Click on the blank space in the file path bar near the top of the window. This should change the file path's display to text, which should be selected.



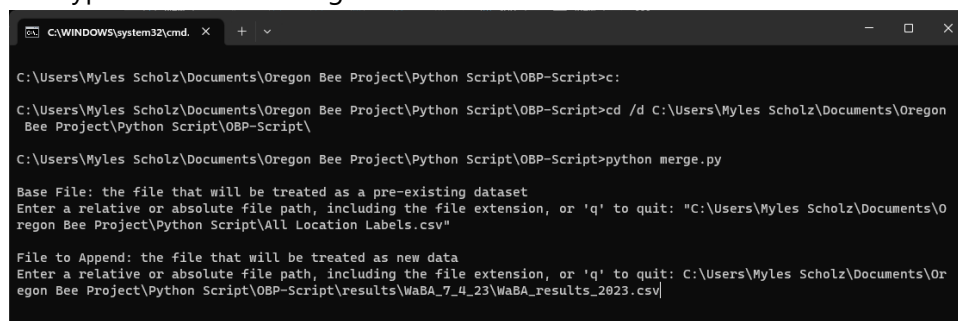
Name	Date modified	Type	Size
WaBA_results_2022.csv	7/4/2023 15:12	Microsoft Excel C...	287 KB
WaBA_results_2023.csv	7/4/2023 15:05	Microsoft Excel C...	287 KB

9. Press Ctrl + C, or right click and click "Copy", to copy the file path. This will only copy up to target file's parent folder.
10. Click the terminal window in which the script is running.
11. When prompted for a file path, right click, or press Ctrl + V, to paste the folder path.



```
C:\WINDOWS\system32\cmd. X + v
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon
Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python merge.py
Base File: the file that will be treated as a pre-existing dataset
Enter a relative or absolute file path, including the file extension, or 'q' to quit: "C:\Users\Myles Scholz\Documents\O
regon Bee Project\Python Script\All Location Labels.csv"
```

12. Type slash and the target file's name and extension.



```
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon
Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python merge.py
Base File: the file that will be treated as a pre-existing dataset
Enter a relative or absolute file path, including the file extension, or 'q' to quit: "C:\Users\Myles Scholz\Documents\O
regon Bee Project\Python Script\All Location Labels.csv"
File to Append: the file that will be treated as new data
Enter a relative or absolute file path, including the file extension, or 'q' to quit: C:\Users\Myles Scholz\Documents\Or
egon Bee Project\Python Script\OBP-Script\results\WaBA_7_4_23\WaBA_results_2023.csv\
```

13. Press Enter to submit the file path.